



UNIVERSIDAD DE CÓRDOBA

## **Tesis Doctoral**

“Empleo de herramientas genómicas para la búsqueda de genes candidatos relacionados con la resistencia a *Ascochyta fabae* y *Orobanche crenata* en habas (*Vicia faba* L.)”

### **Doctoranda**

Sara Ocaña Moral

### **Directores**

Ana María Torres Romero

Eva María Madrid Herrero

*Área de Genómica y Biotecnología de cultivos. Instituto de investigación y  
Formación Agraria, Pesquera y Alimentaria (IFAPA),  
Centro Alameda del Obispo de Córdoba.*

Córdoba, Abril 2017.

---

TITULO: *Empleo de herramientas genómicas para la búsqueda de genes candidatos relacionados con la resistencia a Ascochyta fabae y Orobanche crenata en habas (Vicia faba L.)*

AUTOR: *Sara Ocaña Moral*

---

© Edita: UCOPress. 2017  
Campus de Rabanales  
Ctra. Nacional IV, Km. 396 A  
14071 Córdoba

[www.uco.es/publicaciones](http://www.uco.es/publicaciones)  
[publicaciones@uco.es](mailto:publicaciones@uco.es)

---





**TÍTULO DE LA TESIS:** “Empleo de herramientas genómicas para la búsqueda de genes candidatos relacionados con la resistencia a *Ascochyta fabae* y *Orobanche crenata* en habas (*Vicia faba* L.)”.

**DOCTORANDO/A:** Sara Ocaña Moral.

**INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS**  
(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma)

El trabajo titulado “Empleo de herramientas genómicas para la búsqueda de genes candidatos relacionados con la resistencia a *Ascochyta fabae* y *Orobanche crenata* en habas (*Vicia faba* L.)” se considera finalizado y reúne los requisitos para su exposición y defensa como tesis Doctoral.

La tesis se ha desarrollado dentro del Área de Genómica y Biotecnología del IFAPA, centro Alameda del Obispo de Córdoba de acuerdo a los objetivos propuestos. Las metodologías empleadas (genómica estructural, comparada y funcional), han permitido saturar los mapas disponibles, detectar nuevos marcadores asociados a QTLs y proponer genes candidatos relacionados con los caracteres analizados. Durante el desarrollo de su tesis, la doctoranda ha demostrado poseer una gran iniciativa y habilidad para manejar nuevas técnicas genómicas además de conocimientos bioinformáticos que adquirió tras una estancia en la Plataforma Andaluza de Bioinformática (PAB) de la Universidad de Málaga y posteriormente aplicó al estudio del transcriptoma de las habas.

La calidad y novedad de sus resultados está avalada por dos recientes publicaciones en revistas indexadas en el Journal Citation Report (JCR) dentro del primer cuartil, que constituyen los capítulos II y III del manuscrito, y por su colaboración en una tercera publicación de índole bioinformático (Anexo II). Los

---

---

resultados del capítulo IV se resumen en un trabajo enviado recientemente a la revista Theoretical and Applied Genetics y que en la actualidad se encuentra en fase de revisión. Dichas publicaciones junto a otras actividades científicas relacionadas con la línea de investigación desarrollada se reseñan a continuación:

**- Artículos en revistas indexadas en el Journal Citation Report (JCR):**

Ocaña S, Seoane P, Bautista R, Palomino C, Claros GM, Torres AM, Madrid E (2015). Large-Scale Transcriptome Analysis in Faba Bean (*Vicia faba* L.) under *Ascochyta fabae* Infection. PLoS ONE 10(8): 1-17; e0135143. doi:10.1371/journal.pone.0135143.

Satovic S, Avila CM, Cruz-Izquierdo S, Díaz-Ruiz R, García-Ruiz GM, Palomino C, Gutiérrez-Leiva N, Vitale S, Ocaña-Moral S, Gutiérrez MV, Cubero JI and Torres AM (2013). A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). BMC Genomics, 14:932 (<http://www.biomedcentral.com/1471-2164/14/932>).

Seoane P, Ocaña S, Carmona R, Bautista R, Madrid E, Torres A, Claros G. (2016). AutoFlow, a versatile workflow engine illustrated by assembling an optimised de novo transcriptome for a non-model species, such as Faba Bean (*Vicia faba*). Current Bioinformatics 11(4): 440-450.

Ocaña-Moral S, Gutiérrez N, Torres AM, Madrid E. Saturation mapping of regions determining resistance to *Ascochyta* blight and broomrape in faba bean using transcriptome-based SNP genotyping. Theoretical and Applied Genetics (enviada el 27/03/2017).

**- Congresos:**

Ávila C, Satovic Z, Madrid E, Ocaña S, Ruiz MD, Cubero JI, Torres AM (2012). Genomics research for faba bean (*Vicia faba* L.) breeding applications. Ponencia invitada en el VI International Conference on Legume Genetics and Genomics (VI ICLGG). Hyderabad, India. Octubre 2012.

Ocaña-Moral S, Madrid E, Millán T, García P, Pérez de la Vega M, Torres AM (2013). Strategies to identify candidate genes controlling *Ascochyta* blight in legume crops. Poster en el First Legume Society Conference (LSC1): A Legume Odyssey. Novi-Sad, Serbia. 9-11 mayo, 2013.

Ávila CM, Ruiz-Rodríguez MD, Ocaña-Moral S, Cubero JI and Torres AM (2013). Molecular approaches for the identification and validation of QTLs for *Orobanche* spp. resistance

---

---

in faba bean. Ponencia invitada en el Workshop and Mini-Symposium: Building a new research alliance to reclaim Faba bean production area abandoned to *Orobanche*. Rabat, Marruecos, 7-9 octubre, 2013.

Seoane P, Ocaña S, Carmona R, Bautista R, Guerrero-Fernández D, Madrid E, Torres AM, Gonzalo-Claros M (2014). Using AutoFlow, a framework to resolve workflows, to build a de novo plant transcriptome. Poster en el Plant genomics congress: utilizing Next generation sequencing as a tool for progresing plant research. ICRISAT. London, UK, 12-13 mayo, 2014.

Ocaña S, Seoane P, Claros MG, Bautista R, Torres AM, Madrid E (2014). Large-scale transcriptome analysis in faba bean (*Vicia faba* L.). Poster en el VI International Food Legume Research Conference & VII International Conference on Legume Genetics and Genomics (IFLRC VI & ICLGG VII). Saskatoon, Saskatchewan, Canada. Julio 2014.

Ocaña S, Seoane P, Gonzalo-Claros, Bautista R, Madrid E, Torres AM (2015). Análisis de la expresión diferencial de cDNAs y mapeo de genes candidatos para saturar QTLs de Resistencia a *Ascochyta fabae* en habas (*Vicia faba* L.). Poster en el XL Congreso de la Sociedad Española de Genética. Córdoba, 16-18 Septiembre 2015.

Ocaña S, Seoane P, Gonzalo-Claros, Bautista R, Torres AM, Madrid E (2016). Transcriptome analysis and mapping to identify candidate genes controlling *Ascochyta fabae* resistance in faba bean (*Vicia faba* L.). Ponencia invitada en Proceedings of the IV International *Ascochyta* Workshop (ASCOCHYTA 2016). Troia, Portugal. 10-11 Octubre, 2016.

**- Otras actividades:**

Participación en los seminarios de formación realizados por los doctorandos del Instituto de Investigación y Formación Agraria y Pesquera (IFAPA), Centro Alameda del Obispo con la presentación “Nuevas herramientas de apoyo en la mejora genética”. 5 Junio 2014 (Córdoba).

El gran número SNPs e InDels identificados en este trabajo ofrece una posibilidad rentable de desarrollar marcadores funcionales que ayuden a saturar los mapas de esta especie. Además, los transcritos expresados como respuesta a estímulos son un recurso valioso para identificar genes implicados en la respuesta a estreses. Es previsible que los resultados de esta investigación contribuyan a la identificación de genes clave para acelerar los futuros programas de mejora en este cultivo.

---

---

Por todo ello, se autoriza la presentación de la Tesis Doctoral.

Córdoba, 28 de Abril de 2017.

Firma del/de los directores/es



Fdo.: Dña. Ana Mª Torres Romero



Fdo.: Dña. Eva Mª Madrid Herrero

---

## Financiación

La presente Tesis Doctoral ha sido desarrollada gracias a la beca predoctoral otorgada a la doctoranda por el subprograma del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) de Formación de Personal Investigador (FPI-INIA) del Ministerio de Economía y Competitividad de España en resolución de la convocatoria del BOE 5 de febrero de 2011.

El trabajo de investigación se ha realizado gracias a la financiación de los siguientes proyectos cofinanciados con fondos FEDER de la Unión Europea (EU).

- INIA (RTA2010-00059).
- INIA (RTA2013-00025) cofinanciado por la EU a través de los Fondos FEDER 2014-2020 "Programa Operativo de Crecimiento Inteligente".
- INNPACTO (IPT-2011-1259-010000) del Ministerio de Economía y Competitividad.
- European Community's Seventh Framework Programme (FP7/ 2007-2013), proyecto LEGATO (nº FP7-613551).



UNIÓN EUROPEA  
Fondo Europeo de  
Desarrollo Regional



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE ECONOMÍA  
Y COMPETITIVIDAD



INIA  
Instituto Nacional de Investigación  
y Tecnología Agraria y Alimentaria



JUNTA DE ANDALUCÍA

Instituto de Investigación y Formación Agraria y Pesquera  
CONSEJERÍA DE AGRICULTURA, PESCA Y DESARROLLO RURAL

---





---

## Agradecimientos

Quiero mostrar mi más sincero agradecimiento a todas las personas que en mayor o menor medida han contribuido con su ayuda a la elaboración de este trabajo y aunque es muy difícil resumir lo vivido durante todos estos años, lo más complicado es escribir en pocas líneas los agradecimientos a tanta gente.

En primer lugar, agradecer a mis directoras de tesis: Ana M<sup>a</sup> Torres Romero y Eva M<sup>a</sup> Madrid Herrero, vuestra dedicación y apoyo constante en mi formación científica, por todo lo que me habéis enseñado durante estos años, por compartir conmigo conocimientos e inquietudes y por vuestra cercanía en momentos buenos y difíciles. Por estar siempre disponibles a mis constantes preguntas y no cerrar las puertas a nada, por vuestros consejos y apoyo en todo momento. Ana agradecerte la calma que me has transmitido durante todos estos años y a ti Eva el torbellino que llevas dentro que hace que te ilusiones y cargues las pilas en un solo instante.

En segundo lugar quiero agradecer a las dos personas que me brindaron la gran oportunidad de formar parte del mundo científico. Agradecer a Ana y José Rubio, la confianza depositada en mí desde el primer momento en aquella primera entrevista sin conocerme de nada.

Gracias a todos y cada uno de los compañeros con los que he compartido tiempo en este grupo (Carmen M<sup>a</sup> Ávila, Belén Román, Sergio Atienza, Josefina Sillero y demás). En especial a Natalia Gutiérrez, por su gran apoyo y por estar siempre dispuesta a ayudar y solucionar los muchos problemas que van surgiendo en el día a día tanto profesionalmente como personalmente, siempre recordaré nuestras charlas en el despacho.

A Mercedes Moreno, por estar siempre disponible para facilitarme el tema administrativo y recibirme todas las mañanas con una gran sonrisa.

A Gloria, por su gran ayuda y apoyo en el inicio de esta andanza y por supuesto a Juan Prieto, por su ayuda incondicional en todo momento. Gracias por tus consejos tan sabios y por estar dispuesto a facilitar el día a día a tus niñas en todo lo que necesitan.

---

---

A mis niñas Cordobesas: Dolo, Lola Madrid, Gema, Eva, Rocío, Vero, por todas esas charlas diarias con las que arreglábamos el mundo. Gracias por compartir mis penas y alegrías, por aguantarme día a día, por toda vuestra ayuda y comprensión. Vuestra amistad es uno de los recuerdos más bonitos que me llevo.

También agradecer toda la ayuda recibida durante mi estancia en la Plataforma Andaluza de Bioinformática de Málaga. Gracias a Rocío, Pedro y Gonzalo, por introducirme en el mundo de la Bioinformática, por todos los conocimientos adquiridos y por mostrarme la existencia de un nuevo mundo donde los ordenadores son una herramienta esencial para investigar, analizar e interpretar el significado biológico de una gran variedad de datos.

A Teresa Millán y Juan Gil, como miembros del Departamento de Genética de la Universidad de Córdoba, que siempre están dispuestos a dedicarte un ratito y aclararte dudas.

Y por último a mi familia, quienes siempre han confiado en mí y me han animado constantemente durante todos estos años, porque sin entender lo que estaba haciendo me habéis apoyado en todo momento. Gracias a mis Padres, por dejarme crecer y enseñarme desde muy pequeña que todo en esta vida se consigue con esfuerzo, ilusión y dedicación. Y a mi marido, mil gracias por tu cariño y paciencia. Gracias por animarme a completar esta etapa y acompañarme en todo momento a pesar de la distancia y los momentos difíciles.

Gracias a cada uno de vosotros, por haber hecho que me sienta tan querida y acogida desde el primer momento que llegue a vuestras vidas.

---

---

## Resumen

El objetivo principal de este trabajo ha sido desarrollar y explotar nuevas herramientas genómicas en habas para desvelar a nivel molecular posibles mecanismos de defensa frente al hongo patógeno *Ascochyta fabae* y la planta parásita *Orobanche crenata*, causantes de grandes pérdidas en su producción. Se ha pretendido ampliar la información genética sobre dichos mecanismos de resistencia e identificar posibles genes candidatos donde desarrollar marcadores diagnóstico que permitan extrapolar los resultados de QTL estables entre poblaciones y facilitar la selección precoz de materiales en sus programas de mejora.

Para conseguir estos fines, se han empleado distintas herramientas genómicas (estructural, comparada y funcional). Así, en el capítulo II, se han utilizado la genómica estructural y la macrosintenia existente entre leguminosas, para saturar dos de los mapas genéticos disponibles por el grupo con marcadores relacionados con la resistencia a ambos patógenos, descritos en otras leguminosas. Dicha saturación, ha permitido refinar la posición de los QTLs existentes e identificar marcadores estrechamente ligados que podrían facilitar el proceso de selección. Además, el genotipado de estos nuevos marcadores ha ayudado al desarrollo de un mapa consenso para este cultivo.

En el capítulo III, utilizando herramientas de genómica funcional, se ha tratado de profundizar en las rutas implicadas en la interacción planta-patógeno e identificar potenciales genes de resistencia que pudieran ser útiles en la mejora asistida por marcadores. Para ello se ha caracterizado el transcriptoma de dos líneas parentales (29H y Vf136) sometidas a la infección por *A. fabae* y obtenido sus perfiles de expresión génica. Aparte de proporcionar nueva información genómica, el estudio han revelado una amplia variedad de posibles mecanismos y rutas implicadas. El gran número SNPs e InDels identificados en los transcritos ofrece, además, una posibilidad rentable de saturar los mapas disponibles y de identificar genes responsables de la resistencia. Para finalizar, en el capítulo IV, se ha genotipado un conjunto de SNPs expresados diferencialmente entre ambos parentales, relacionados con mecanismos de resistencia. Ello ha permitido situar nuevos marcadores más robustos y transferibles (SNPs y ESTs) dentro de los intervalos de confianza de los QTLs de resistencia ya descritos, facilitando su comparación con otras poblaciones de habas y con otras leguminosas.

---



---

## Summary

The main objective of this work has been to develop and exploit new genomic tools in faba bean in order to reveal at molecular level possible defense mechanisms against the pathogenic fungus *Ascochyta fabae* and the parasitic plant *Orobancha crenata*, causing severe yield losses in this crop. The approach aimed to expand the genetic information on the resistance mechanisms involved and to identify possible candidate genes useful to develop diagnostic markers for validation of stable QTL among populations and for the efficient selection of superior materials in breeding programs.

To achieve these ends, different genomic tools (structural, comparative and functional) have been applied. Thus, in Chapter II, both the structural genomics and the synteny among legumes have been exploited to saturate two of the genetic maps available by the group using markers related with the resistance to both pathogens, described in other legumes. The saturation allowed to refine the position of the underlying QTLs and to identify closely linked markers that could facilitate the selection process. In addition, the new markers were used as a bridge to develop a consensus map in this crop.

In Chapter III, functional genomics tools have been applied to gain insight into the pathways involved in the plant-pathogen interactions and to identify potential resistance genes to be further used in marker assisted selection. The faba bean transcriptome from leaves of two lines (29H and Vf136) subjected to *A. fabae* infection were characterized in order to obtain its gene expression profiles. Apart from providing new genomic information, the study has revealed a wide variety of possible mechanisms and pathways involved. The large number of SNPs and InDels identified in the transcripts also offers a cost-effective way to saturate available maps and to identify genes responsible for resistance. Finally, in chapter IV, a set of differentially expressed SNPs between both parental lines, related with resistance mechanisms has been genotyped. The approach has allowed placing more robust and transferable markers (SNPs and ESTs) within the confidence intervals of the resistance QTLs already described, facilitating their comparison among faba bean and legume species.

---



---

**Nota:** A fin de establecer una coherencia formal a lo largo del presente documento, se han uniformado las referencias y se han editado los trabajos originales, eliminando de los mismos el apartado de referencias, el cual ha sido agrupado al final del documento.

---





---

## Índice de contenidos

<b>Capítulo I. INTRODUCCIÓN GENERAL</b> .....	1
I.1. El cultivo de habas .....	2
I.1.1. Origen y características .....	2
I.1.2. Importancia económica del cultivo .....	4
I.1.3. Limitaciones del cultivo de habas .....	6
I.2. El papel de la Mejora en el cultivo de habas.....	7
I.2.1. <i>Ascochyta fabae</i> .....	7
I.2.2. <i>Orobanche crenata</i> .....	8
I.2.3. Mecanismos moleculares de resistencia a patógenos en plantas .....	11
I.3. Selección asistida por marcadores (MAS) .....	14
I.4. Empleo de herramientas genómicas en la Mejora de las leguminosas.....	16
I.4.1. Genómica estructural .....	17
I.4.1.1. Mapa genético de habas .....	18
I.4.2. Genómica comparada y sintenia .....	22
I.4.2.1. <i>Medicago truncatula</i> como especie modelo en leguminosas.....	22
I.4.3. Genómica funcional .....	26
I.4.3.1. Tecnologías de secuenciación de nueva generación y bioinformática .....	28
I.4.3.2. Aplicaciones de la NGS en la genética y mejora vegetal .....	30
I.5. Hipótesis y Objetivos.....	32
<b>Capítulo II. Saturación de los mapas de habas disponibles mediante herramientas de Genómica estructural y comparativa (sintenia)</b> .....	3
II.1. Introducción .....	35
II.2. Materiales y métodos .....	39
II.2.1. Material vegetal .....	39
II.2.2. Análisis de marcadores moleculares .....	39
II.2.2.1. Extracción de ADN genómico .....	39
II.2.2.2. Análisis de ESTs (Expressed Sequence Tags) y UniTags .....	40
II.2.3. Construcción del mapa de ligamiento en la población RIL y análisis de QTLs.....	49
II.3. Resultados y Discusión .....	51
II.3.1. Marcadores polimórficos en <i>V. faba</i> .....	51
II.3.2. Nuevos mapas de ligamiento y análisis de QTLs implicados en la resistencia a <i>O. crenata</i> , <i>O. foetida</i> y <i>A. fabae</i> en habas.....	56
II.3.2.1. Población 29H x Vf136.....	56
II.3.2.2. Población Vf6 x Vf136 .....	61
II.4. Resumen y Conclusiones .....	68

---

---

<b>Capítulo III. Large-Scale Transcriptome Analysis in Faba Bean (<i>Vicia faba</i> L.) under <i>Ascochyta fabae</i> Infection</b> .....	73
III.1. Abstract .....	75
III.2. Introduction.....	77
III.3. Material and methods .....	79
III.3.1 Plant material and <i>A. fabae</i> inoculation .....	79
III.3.2. RNA extraction, library preparation and DNA sequencing .....	79
III.3.3. Sequence pre-processing, de novo assembly and annotation.....	80
III.3.4. Variant analyses.....	81
III.3.5. Differential gene expression and functional enrichment.....	81
III.3.6. Expression profiles via RT-qPCR.....	81
III.4. Results .....	82
III.4.1. A new faba bean transcriptome .....	82
III.4.2. Functional classification of the <i>V. faba</i> transcriptome .....	84
III.4.3. Sequence similarity of <i>V. faba</i> transcripts with other plants.....	86
III.4.4. Discovery of SNPs and InDels variants .....	87
III.4.5. Transcripts involved in blight resistance .....	87
III.4.6. Validation of Transcriptome by RT-QPCR .....	88
III.5. Discussion .....	89
III.6. Conclusions.....	94
III.7. Acknowledgement.....	94
<b>Capítulo IV. Saturation mapping of regions determining resistance to <i>Ascochyta</i> blight and broomrape in faba bean using transcriptome-based SNP genotyping</b> .....	95
IV.1. Abstract .....	97
IV.2. Introduction .....	99
IV.3. Material and methods .....	101
IV.3.1. Plant material .....	101
IV.3.2. SNP selection and genotyping.....	101
IV.3.3. Linkage map construction and QTL analysis .....	102
IV.3.4. Field trials and Resistance scoring .....	103
IV. 4. Results.....	104
I IV.4.1. SNP selection and genotyping.....	104
IV.4.2. Linkage analysis.....	105
IV.4.3. QTL analysis for <i>Ascochyta fabae</i> .....	106
IV.4.4. QTL analysis for <i>Orobanche sp</i> .....	108
IV.5. Discussion .....	111
IV.5.1.Genetic linkage mapping and synteny analysis .....	111
IV.5.2. QTL detection and candidate genes identification.....	112
IV.6. Acknowledgements .....	117
<b>Capítulo V. Discusión general</b> .....	119
<b>Capítulo VI. Conclusiones</b> .....	129
<b>Capítulo VII. Bibliografía</b> .....	133

---

---

**Anexos: Otras publicaciones y comunicaciones del autor ..... 163**

- Anexo I.** Satovic S, Avila CM, Cruz-Izquierdo S, Díaz-Ruíz R, García-Ruíz GM, Palomino C, Gutiérrez-Leiva N, Vitale S, Ocaña-Moral S, Gutiérrez MV, Cubero JI and Torres AM (2013). A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). *BMC Genomics*, 14:932 ..... 165
- Anexo II.** Seoane P, Ocaña S, Carmona R, Bautista R, Madrid E, Torres A, Claros G. (2016). AutoFlow, a versatile workflow engine illustrated by assembling an optimised de novo transcriptome for a non-model species, such as Faba Bean (*Vicia faba*). *Current Bioinformatics* 11(4): 440-450..... 181
- Anexo III.** Comunicaciones en congresos y conferencias ..... 193



---

## Índice de figuras

### Capítulo I. Introducción general.

<b>Figura 1.</b> Principales características de una planta de <i>V. faba</i> : (a) planta completa mostrando distintas inflorescencias, (b) forma de la hoja, (c) vaina y fruto y (d) flores.....	3
<b>Figura 2.</b> Tipos de haba según el tamaño de la semilla (Cubero, 1974) .....	3
<b>Figura 3.</b> Cromosomas de <i>Vicia faba</i> (Chapman, 1983) .....	4
<b>Figura 4.</b> Zonas de producción mundial de habas (FAOSTAT, 2014).....	5
<b>Figura 5.</b> Serie histórica de la producción en toneladas (línea roja) y del área cosechada en hectáreas (línea azul) del cultivo de habas en España (FAOSTAT, 2014).....	6
<b>Figura 6.</b> Lesiones en hojas (a), tallo (b), vainas (c) y semillas (d) causadas por <i>Ascochyta fabae</i> .....	7
<b>Figura 7.</b> Plantas de <i>O. crenata</i> parasitando a <i>Vicia faba</i> .....	9
<b>Figura 8.</b> a) Activación de la inmunidad activada por PAMPs o PTI mediante el reconocimiento de PAMPs/MAMPs por receptores PRRs de la planta. b) Susceptibilidad desencadenada por un efector (ETS) producido por el patógeno que suprime PTI. c) Activación de la inmunidad desencadenada por efector (ETI) al ser reconocido específicamente por proteínas de resistencia de la planta (Fuente: Pieterse et al. 2009). .....	12
<b>Figura 9.</b> Esquema con los principales componentes que actúan en las rutas de señalización del ácido salicílico (SA), ácido jasmónico (JA) y etileno (ET). a) SA origina la monomerización de NPR1 en citosol y su transporte al núcleo para interactuar con factores de transcripción (TGAs/ WRKYs) y expresar genes que codifican proteínas relacionadas con la patogénesis (PRs). WRKY70 actúa como interruptor molecular entre la vía del SA y la de JA. b) En la ruta del JA, la proteína JAZ interactúa con factores de transcripción MYC2 y ERF1 responsables de la expresión de genes de defensa y de integrar las señales de las rutas del JA y el ET. c) El etileno activa una cascada de quinasas MAPK, cuya señal es transmitida a EIN2 que evita la degradación del factor de transcripción EIN3. Su acumulación activa a ERF1 que regula la expresión génica (Fuente: Pieterse et al. 2009). .....	13
<b>Figura 10.</b> Relaciones filogenéticas de la subfamilia Papilionoideae. Fuente: Zhu et al. 2005. A la derecha de las especies está indicado el número cromosómico (N) y el tamaño del genoma expresado en Mpb, MA = Millones de años .....	24

---

<b>Figura 11.</b> Mapa consenso simplificado de ocho leguminosas. La figura está basada en la publicaciones de Choi et al. (2004b) con modificaciones de Zhu et al. (2005). Mt, Medicago truncatula; Ms, Medicago sativa; Lj, Lotus japonicus; Ps, Pisum sativum; Ca, Cicer arinetum; Vr, Vigna radiata; Pv, Phaseolus vulgaris; Gm, Glycine max. S y L son el brazo corto y largo de cada cromosoma en M. truncatula. Los bloques sinténicos están dibujados a escala basándose en la distancia genética .....	25
<b>Figura 12.</b> Estudios y herramientas que engloba el uso de la genómica funcional .....	27
<b>Figure 13.</b> Esquema de los estudios de genómica estructural, comparativa y funcional aplicados durante el desarrollo de esta tesis para la saturación del mapa y la identificación de genes candidatos de resistencia a jopo y <i>Ascochyta fabae</i> en habas.....	31

**Capítulo II. Saturación de los mapas de habas disponibles mediante herramientas de Genómica estructural y comparativa (sintenia)**

<b>Figura 1.</b> Representación esquemática de bloques de sintenia entre cromosomas y segmentos cromosómicos de <i>M. truncatula</i> (Mt) y leguminosas grano tradicionales; Guisante ( <i>Pisum sativum</i> , Ps), habas ( <i>Vicia faba</i> , Vf), garbanzo ( <i>Cicer arietinum</i> , Ca) y lentejas ( <i>Lens culinaris</i> , Lc). Las barras que representan las regiones cromosómicas homólogas de Medicago y guisante se muestran con la misma intensidad o patrón de grises. Las flechas indican la orientación de los cromosomas (brazo corto - brazo largo) en el caso de Medicago. Los bloques en habas, garbanzo y lenteja están representados por barras en blanco. Las barras no reflejan los tamaños relativos de los segmentos cromosómicos o cromosómicos y los puntos de rotura de los cromosomas se indican de modo aproximado. Fuente: Kalo et al. 2011 con modificaciones. La figura integra datos de Choi et al. 2004, Kaló et al. 2004, Zhu et al. 2005, Nelson et al. 2006, Phan et al. 2006, Ellwood et al. 2008, Nayak et al. 2010, Cruz-Izquierdo et al. 2012 y Satovic et al. 2013 .....	36
--	----

<b>Figura 2.</b> Nuevo mapa de ligamiento y QTLs para resistencia a <i>A. fabae</i> (Af), <i>O. crenata</i> (Oc) y <i>O. foetida</i> (Of) en la población 29H x Vf136. En rojo se distinguen los 8 marcadores mapeados en este estudio. La línea punteada separa los GL que no han podido ser asignados a cromosomas (CR) específicos.....	59
--	----

<b>Figura 3.</b> Nuevo mapa de ligamiento y QTLs para la resistencia a <i>A. fabae</i> (Af), <i>O. crenata</i> (Oc) y <i>O. foetida</i> (Of) en la población RIL Vf6 x Vf136 .....	65
--	----

**Capítulo III. Large-Scale Transcriptome Analysis in Faba Bean (*Vicia faba* L.) under *Ascochyta fabae* Infection**

<b>Figura 1.</b> Venn diagram depicting the number of overlapping sequences and unique genes present among the three transcriptomes. KT: transcriptome developed by Kaur et al. (2012); PT: Illumina reads from the present study;	
--	--

---

ET: enriched transcriptome (Illumina reads and database sequences) ..... 85

**Figura 2.** Distribution of the GO categories assigned to the faba bean transcriptome. Unique transcripts were annotated in three categories: biological process, cellular components and molecular functions..... 86

**Figura 3.** Analysis of the kinetics of 4 transcripts analyzed by RT-qPCR in the resistant and susceptible genotype. (a) LEA-18; (b) Transcription factor NAI1; (c) Jasmonate O-methyltransferase; (d) F-box/LRR-repeat protein At3g59200 analyzed at 4, 8 and 12 hours after inoculation with *A. fabae*. Relative mRNA quantification was performed using ACT1 and CYP2, as reference genes for normalization ..... 89

#### **Capítulo IV. Saturation mapping of regions determining resistance to Ascochyta blight and broomrape in faba bean using transcriptome-based SNP genotyping**

**Figura 1.** Genetic map and summary of the *Orobanche crenata* and *Ascochyta fabae* resistance QTL in the faba bean RIL population derived from the cross 29H x Vf136. *O. crenata* QTLs were named using the initials Oc followed by a number assigned in previous works and the identifier for each location (Eg, Egypt and Cor, Córdoba). For Cordoba analyses, a season identifier was also included (06, season 2006/2007; 07, season 2007/2008; 08, season 2008/2009). In the case of *A. fabae*, the QTLs were named considering the experiment (field, isolate Co99 in growth chamber, isolate Lo98 in growth chamber) and trait (DSL, disease severity on leaves; DSS, disease severity on stems; DSP, disease severity on pods). QTL positions are represented as 2-LOD bars. .... 109

---





---

## Índice de Tablas

### Capítulo I. Introducción general.

<b>Tabla 1:</b> Principales QTLs para resistencia a <i>Ascochyta fabae</i> detectados en poblaciones F2 y RILs de habas en los cruzamientos Vf6 × Vf136 y 29H × Vf136....	20
<b>Tabla 2:</b> Principales QTLs para resistencia a <i>Orobanche crenata</i> y <i>Orobanche foetida</i> detectados en poblaciones F2 y RIL de habas en los cruzamientos Vf6 × Vf136 y 29H × Vf136.....	21

### Capítulo II. Saturación de los mapas de habas disponibles mediante herramientas de Genómica estructural y comparativa (sintenia)

<b>Tabla 1.</b> Mezcla de reacción de PCR para los diferentes marcadores .....	41
<b>Tabla 2.</b> Condiciones de amplificación de PCR para marcadores ESTs .....	42
<b>Tabla 3.</b> Condiciones de amplificación de PCR para marcadores UniTags de lenteja .....	42
<b>Tabla 4.</b> Condiciones de amplificación de PCR para marcadores UniTags de habas..	42
<b>Tabla 5.</b> Condiciones de amplificación de PCR para los marcadores de garbanzo ....	42
<b>Tabla 6.</b> Marcadores moleculares analizados en los cruzamientos Vf6 x Vf136 y 29H x Vf136 en poblaciones RILs en <i>V. faba</i> .....	44
<b>Tabla 7.</b> Relación de enzimas utilizadas para desarrollar marcadores CAPs. Secuencias diana y temperatura (°C) de corte de cada una de ellas .....	49
<b>Tabla 8.</b> Resultado de los 116 marcadores analizados en las poblaciones RILs Vf6 x Vf136 y 29H x Vf136 .....	52
<b>Tabla 9.</b> Listado de los 27 marcadores moleculares genotipados en las poblaciones Vf6 x Vf136 y 29H x Vf136 y metodologías para la detección del polimorfismo .....	54
<b>Tabla 10.</b> Posición en el mapa de los marcadores genotipados en la población 29H x Vf136.....	56
<b>Tabla 11.</b> Grupos de ligamiento (GL), asignación a cromosomas y distribución de marcadores en el mapa 29H x Vf136 .....	57
<b>Tabla 12.</b> QTLs detectados para resistencia a <i>A. fabae</i> ( <i>Af</i> ), <i>O. foetida</i> ( <i>Of</i> ) y <i>O. crenata</i> ( <i>Oc</i> ) en la población RIL 29H x Vf136 de habas por análisis de intervalos (IM). El LOD crítico fue determinado usando el análisis de permutaciones en todos los ambientes.....	60
<b>Tabla 13.</b> Marcadores moleculares genotipados y ligados en la población Vf6 x Vf136.....	62

---

---

<b>Tabla 14.</b> Grupos de ligamiento (GL), asignación a cromosomas, características y distribución en el mapa Vf6 x Vf136 .....	63
--	----

<b>Tabla 15.</b> QTLs detectados para resistencia a <i>A. fabae</i> (Af), <i>O. foetida</i> (Of) y <i>O. crenata</i> (Oc) en la población RIL Vf6 x Vf136 mediante análisis de intervalos (IM) .....	67
--	----

### **Capítulo III. Large-Scale Transcriptome Analysis in Faba Bean (*Vicia faba* L.) under *Ascochyta fabae* Infection**

<b>Table 1.</b> Primer pairs designed to validate RNAseq data by RT-qPCR and number of reads detected in each of the faba bean genotypes. ....	82
--	----

<b>Table 2.</b> Summary of the attributes assemblies obtained using three different datasets after the analysis using Full-LengtherNext. Transcriptomes of Kaur et al. [23] and version 1.0 were analysed with older releases of Full-LengtherNext, while the current transcriptome (version 1.1) was analysed with a new release of this software.....	84
---	----

### **Capítulo VI. Saturation mapping of regions determining resistance to *Ascochyta* blight and broomrape in faba bean using transcriptome-based SNP genotyping**

<b>Table 1.</b> Origin and characteristics of the SNPs assayed in the faba bean RIL population Vf29H x Vf136 .....	104
--	-----

<b>Table 2.</b> Linkage groups, chromosome (Chr.) assignment and marker distribution in the 29H x Vf136 map.....	105
--	-----

<b>Table 3.</b> Putative QTLs for <i>O. crenata</i> (Oc) and <i>A. fabae</i> (Af) detected in the faba bean cross 29H x Vf136 by Interval mapping (IM) analysis.....	107
--	-----

---

# Capítulo I

## Introducción General





## INTRODUCCION GENERAL

Las leguminosas representan una de las familias más numerosas e importantes en la agricultura, incluyendo más de 650 géneros y 18.000 especies que destacan por su gran valor nutritivo (Doyle, 2000). La familia abarca desde plantas herbáceas hasta grandes árboles (Doyle y Luckow 2003; Lewis et al. 2005) teniendo múltiples usos en agricultura (leguminosas grano, cultivos hortícolas o leguminosas forrajeras). Son consideradas cultivos mejorantes del terreno por su capacidad para fijar nitrógeno atmosférico mediante complejas interacciones con bacterias del género *Rhizobium* (hasta 100-120 Kg N/ha, en el caso concreto de las habas) jugando un papel importante en la rotación de cultivos (Ghaham y Vance 2003; Garg y Geetanjali 2009; Udvardi y Poole 2013).

Nutricionalmente, las leguminosas son una fuente principal de proteína vegetal (entre un 20 y un 40%) ricas además en carbohidratos, fibra, vitaminas y minerales esenciales para la dieta humana y la alimentación animal. Su composición en aminoácidos esenciales, ricos en lisina y pobres en los de tipo azufrado y triptófano (Wang et al. 2003), las hace complementarias a los cereales. En concreto, las habas, cuentan con un contenido de proteína en grano seco cercano al 30%, una alta proporción de lisina y una buena palatabilidad para su empleo como pienso en rumiantes y ganado de carne en general. Como planta hortícola, su destino es para consumo humano, aprovechándose vainas y granos en industrias transformadoras para enlatado y congelado (Nadal et al. 2004). Estudios recientes han demostrado que ciertos compuestos antinutricionales (Roy et al. 2010), que reducen su valor biológico tienen, sin embargo, efectos beneficiosos para la salud como son los compuestos bioactivos (Muzquiz et al. 2012) que reducen ciertos tipos de cáncer, activan mecanismos de defensa, tienen propiedades anti-inflamatorias, reducen la hipertensión, poseen efecto antioxidante o mejoran la obesidad. También se ha demostrado que dietas ricas en proteínas con bajo contenido en metionina, como lo son las proteínas de leguminosas, retardan el envejecimiento (McCarty et al. 2009) y que compuestos como la 3,4-dihidroxifenilalanina (L-DOPA) son un importante principio activo en la industria farmacéutica para combatir la enfermedad del Parkinson, la hipertensión, el fallo renal o la cirrosis (Rabey et al. 1993; Randhir et al.

2002). Otro de los muchos beneficios de las leguminosas es su aplicación como abono sideral o como cultivo trampa para uno de sus mayores patógenos, la planta parásita *Orobanche crenata* (Nadal et al. 2004).

El cultivo de las leguminosas, junto con el de los cereales, es uno de los más antiguos y ha estado presente en la agricultura desde sus orígenes. Uno de los principales centros de origen de muchas leguminosas (habas, garbanzos, guisantes, lentejas, yeros, etc.) es el Próximo Oriente, donde se practica su cultivo desde casi los comienzos de la agricultura (Abbo et al. 2003). En la agricultura mediterránea, y desde los tiempos de los romanos, la rotación cereal-leguminosa se ha venido realizando de una forma tradicional por los beneficios que aportan estas últimas, incrementando la fertilidad del suelo, mejorando su estructura y contribuyendo al saneamiento de otros cultivos al romper sus ciclos de enfermedades y plagas.

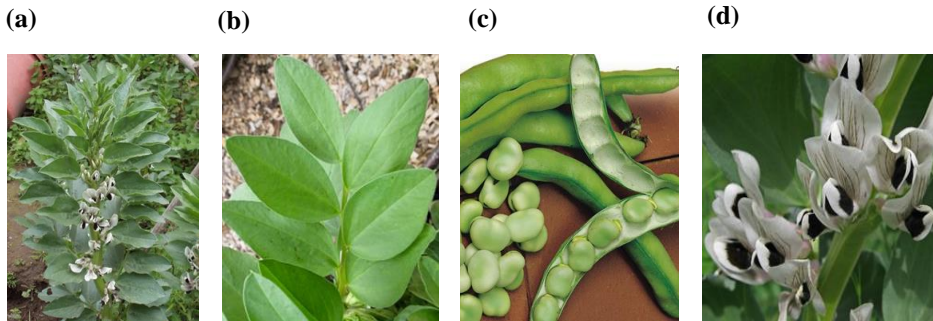
### ***1.1. El cultivo de habas.***

#### ***1.1.1. Origen y características.***

Las habas (*Vicia faba* L.) son plantas anuales, alogamas y diploides ( $2n=2x=12$ ). Pertenecen a la familia *Fabaceae*, tribu *Vicieae*, género *Vicia* y especie *V. faba* (Cronquist, 1981). Fueron unas de las primeras leguminosas domesticadas por el hombre al inicio de la agricultura. Tanto su origen exacto como su antecesor silvestre se desconocen, concluyendo algunos autores que la especie, probablemente, es originaria del Cercano Oriente (Cubero, 1974) o Asia Central (Ladizinski, 1975) desde donde se extendió al Mediterráneo y posteriormente al resto del mundo. Actualmente, se considera que la forma primitiva más cercana al extinguido progenitor silvestre es el grupo *paucijuga* (Cubero, 1974).

*V. faba* es una especie anual, erecta y glabra (Figura 1a). Presenta una fuerte raíz primaria muy desarrollada y numerosas raíces secundarias y terciarias responsables de formar nódulos que se asocian simbióticamente con *Rhizobium leguminosarum* para fijar nitrógeno atmosférico (Nutman, 1976; Beck y Duc 1991). Sus hojas son compuestas, alternas, sin zarcillos y formadas por folíolos con un borde recto o ligeramente ondulado de color verde-grisáceo (Figura 1b). Las estipulas dentadas están provistas de un nectario y muestran una mancha oscura. Las flores se agrupan en inflorescencias axilares de 1 hasta 12 flores y cada flor contiene un pétalo

grande llamado estandarte y dos pétalos de igual tamaño conocidos como alas. Los pétalos normalmente son blancos con una mancha negra-purpúrea en las alas (Figura 1d).



**Figura 1.** Principales características de una planta de *V. faba*: (a) planta completa mostrando distintas inflorescencias, (b) forma de la hoja, (c) vaina y fruto y (d) flores.

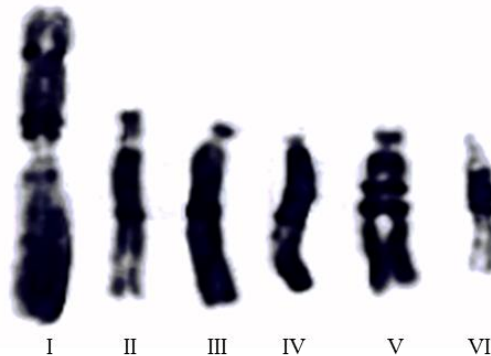
Su fruto es carnoso, de color verde-pardo; siendo el número de semillas por vaina variable (hasta 8-10 granos en algunos cultivares *major*) (Figura 1c). Sus semillas son ricas en proteínas (alrededor del 30 %), carbohidratos, fibra, vitaminas y minerales. Según el tamaño de la semilla son cuatro los tipos botánicos existentes: *paucijuga*, *minor*, *equina* y *major* (Cubero, 1974) (Figura 2). La semilla de *paucijuga* pesa entre 0.31-0.40 gr, la *minor* 0.41-0.60 gr, la *equina* 0.61-1.10 y la *major* 1.11-1.70 gr. Las semillas grandes (*major*) emergieron en los países Mediterráneos y China, las de tamaño medio (*equina*) en el Este y Norte de África, mientras que las semillas pequeñas (*minor*) son típicas del Norte de Europa (Duc, 1997).



**Figura 2.** Tipos de haba según el tamaño de la semilla (Cubero, 1974).



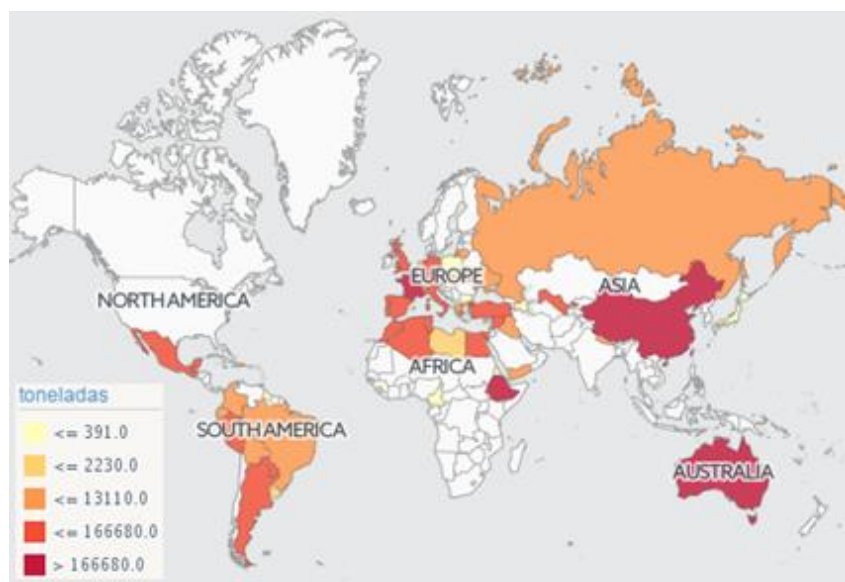
El genoma de *V. faba* es citogenéticamente uno de los mejor caracterizados entre las plantas. *V. faba* es una planta diploide con una dotación cromosómica de  $2n = 12$ , formada por seis grandes cromosomas (Figura. 3) de los cuales, cinco son acrocéntricos y uno metacéntrico y de mayor tamaño ( $15 \mu\text{m}$ ) (Chapman, 1983). El tamaño de los cromosomas y el contenido de ADN por núcleo en esta especie es casi el doble (13.000 Mbps) que en otras del genero *Vicia* (Chooi, 1971; Lawes et al. 1983; Raina y Ogihara 1995; Bennett et al. 1995). Es 25 veces mayor que el de la especie modelo *Medicago truncatula* y presenta una alta proporción de ADN repetitivo (Flavell et al. 1974) lo que ha retrasado los avances genéticos y el desarrollo de herramientas genómicas en esta especie.



**Figura 3.** Cromosomas de *Vicia faba* (Chapman, 1983).

### ***1.1.2. Importancia económica del cultivo.***

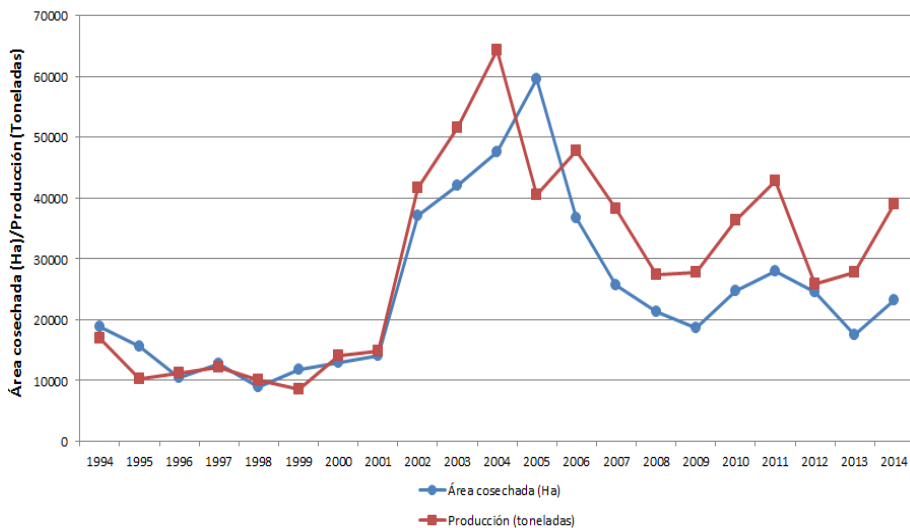
Tradicionalmente las habas se han cultivado en la cuenca Mediterránea, Oriente Medio, China y América Latina aunque actualmente está logrando especial relevancia en Australia, Europa y América del Norte (Figura 4). En 2014, la superficie mundial de habas cultivada superó los 2 millones de hectáreas (has), con una producción total de unos 4,14 millones de toneladas, lo que hace que sea la cuarta leguminosa grano más importante tras el garbanzo, el guisante y la lenteja (FAOSTAT, 2014). Los mayores productores mundiales son China, Etiopía, Australia y Francia. España se encuentra en decimoquinto lugar en cuanto a producción, lo que representa casi un 1% de la producción mundial total.



**Figura 4.** Zonas de producción mundial de habas (FAOSTAT, 2014).

En España, la superficie destinada al cultivo es de 23.155 ha, alcanzado una producción de 38.934 toneladas (FAOSTAT, 2014). Tanto la superficie cultivada como la producción han ido fluctuando en los últimos 20 años (Figura 5). Estas fluctuaciones se han debido principalmente al abandono de las rotaciones leguminosa-cereal al cambiar las prácticas agrícolas y a la disponibilidad de fertilizantes nitrogenados relativamente baratos que parecían ser más eficientes en mejorar los rendimientos. El aumento significativo de la superficie de cultivo durante el periodo 2001 al 2005, estuvo influenciado por las ayudas de la UE a las leguminosas (Reforma de Política Agraria Comunitaria de 2003), pero cayó posteriormente hasta niveles de 1994. Este hecho evidencia la dependencia de nuestro país de las ayudas externas a la producción y por tanto de las normas de la PAC que se apliquen en cada momento. Así, la agricultura intensiva española y europea actual, basada principalmente en la rotación oleaginosa-cereal, ha generado una total dependencia de las importaciones de proteínas vegetales. La producción de proteína vegetal de la UE ocupa sólo el 3% de sus tierras de cultivo, lo que cubre apenas el 30% de nuestra demanda de proteaginosas para piensos. Resulta absurdo, sin embargo, que debamos importar el 70 % de proteínas vegetales en su mayor parte de Estados Unidos, Méjico

o Argentina, con unos costes energéticos y económicos inasumibles en una sociedad moderna.



**Figura 5.** Serie histórica de la producción en toneladas (línea roja) y del área cosechada en hectáreas (línea azul) del cultivo de habas en España (FAOSTAT, 2014).

### 1.1.3. Limitaciones del cultivo de habas.

Tradicionalmente el cultivo de habas se ha caracterizado por la baja estabilidad de su rendimiento. Entre los principales factores que afectan a la producción del cultivo y a la calidad de la semilla, se encuentran estreses abióticos como la sequía, las heladas y las altas temperaturas y bióticos como el jopo (*Orobanche crenata*), una fanerógama parásita muy extendida en toda la cuenca Mediterránea y hongos patógenos como *Ascochyta fabae*, *Uromyces viciae-fabae* (causante de la roya) o *Botrytis fabae* que produce la denominada botritis o mancha chocolate. La frecuencia con la que ocurren estos estreses y su severidad es variable dependiendo de las condiciones climáticas y de la zona geográfica en la que se cultive. Por lo tanto, los principales objetivos de mejora en habas incluyen la mejora de la calidad de semillas y la resistencia o tolerancia a estreses bióticos y abióticos (Sal, 1983; Bond et al. 1994; Sillero et al. 2010).

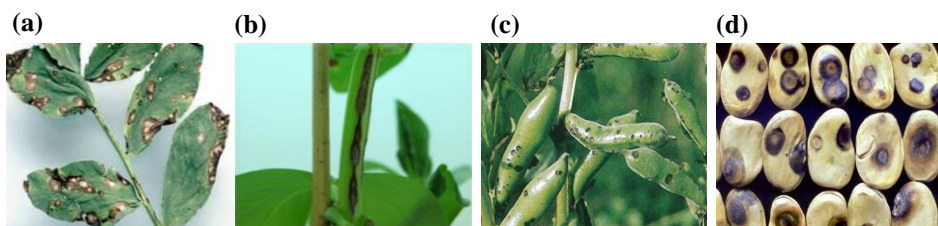
## ***1.2. El papel de la Mejora en el cultivo de habas.***

La mejora vegetal ha sido una de las tecnologías que se ha desarrollado con mayor éxito en la agricultura moderna, dando lugar a grandes cambios en la calidad, estructura y adaptación de los cultivos a su entorno, mejorando su potencial genético e incrementado su rendimiento. Aunque algunos esfuerzos realizados han tenido poco éxito, la mayoría de las modificaciones genéticas de los cultivos logradas mediante mejora clásica han sido muy positivas e importantes (Lee, 1995).

Como se ha mencionado anteriormente, al igual que para el resto de cultivos, el principal objetivo de la mejora en habas es incrementar y estabilizar su rendimiento mediante el desarrollo de cultivos adaptados a los estreses desarrollados en cada ambiente. Entre los factores limitantes en la producción y rendimiento de esta especie están la susceptibilidad al hongo patógeno *A. fabae* y a la planta parasita *Orobanche crenata*.

### ***1.2.1. Ascochyta fabae.***

Se trata de un hongo aéreo causante de una de las enfermedades fúngicas (ascoquitosis) más comunes del cultivo en todo el mundo (Gaunt, 1983). El hongo es responsable de pérdidas del rendimiento tanto de peso como de calidad de las semillas, entre el 30 y el 40%, pudiendo llegar incluso al 90% en cultivos muy susceptibles (Hanounik y Robertson 1989). El patógeno ataca a todas las partes aéreas de la planta causando necrosis y colapso tisular. Los síntomas consisten en lesiones circulares de color marrón oscuro que evolucionan a necrosis en hojas, tallos, vainas y semillas (Figura 6).



**Figura 6.** Lesiones en hojas (a), tallo (b), vainas (c) y semillas (d) causadas por *Ascochyta fabae*.

El patógeno es capaz de infectar la semilla, por lo que una de las medidas de control de la enfermedad es el uso de semillas libres de inóculo (Beaumont, 1950; Hewett, 1973; Kharbanda y Bernier 1979; Gaunt y Liew 1981; Kaiser, 1997). Sin embargo, la existencia de otras fuentes de inóculo, como restos de cultivo infectados, requiere medidas de control adicionales como son la rotación de cultivos o el uso de productos químicos (Jiménez-Díaz et al. 1998). Frente a ello el empleo de variedades resistentes, se considera la forma más efectiva para el control de la enfermedad a largo plazo.

Se han identificado varias fuentes de resistencia que se están utilizando en programas de mejora (Bond y Pope 1980; Tivoli et al. 1988; Hanounik y Robertson 1989; Rashid et al. 1991a, b; Ondrej, 1993; Sillero et al. 2001) pero a día de hoy aún se desconoce en detalle, su base genética.

### ***1.2.2. Orobanche crenata.***

Planta holoparásita (Figura 7) caracterizada por carecer de una raíz verdadera, no realizar fotosíntesis, tener haustorios especializados y nutrirse del xilema y floema de la planta parasitada (Young et al. 1999; Nickrent, 2002). Entre los cultivos que parasita se encuentran habas, garbanzos, lentejas, arvejas, guisantes, apio y zanahoria (Sauerborn, 1991; Eizenberg et al. 2001; Rubiales et al. 2003). Se distribuye extensamente a través de la cuenca mediterránea, especialmente en el norte y el este de África (Rubiales y Fernández-Aparicio 2012) ocasionando grandes pérdidas en estos cultivos (Cubero, 1983; Mesa-García y García-Torres 1984; Parker y Riches 1993). En habas, se estiman pérdidas en la producción entre 5 y 95%, dependiendo de la fecha de siembra y el nivel de infestación (Rubiales et al. 2006).

La infección comienza con el crecimiento de la radícula de la semilla germinada hacia la raíz de la planta hospedante, penetrando en ella por la epidermis y el córtex con ayuda de enzimas hidrolíticas hasta ubicarse en el sistema vascular y formar el haustorio. Seguidamente, el haustorio se diferencia en nódulos que crecen hasta desarrollar raíces simples que darán lugar a raíces secundarias y que serán las que absorban los nutrientes y el agua de la planta hospedadora. A partir de este momento, el jopo emerge del suelo y crece rápidamente hasta llegar al estado de floración y producción de semillas (López-Granados y García-Torres 2002). Sus

semillas son pequeñas (0.3 x 0.2 mm) y su número puede ascender a más de 200,000 por planta (Parker y Riches 1993) pudiendo tener una longevidad en el suelo de unos 14 años (López-Granados y García-Torres 1999). La germinación se ve afectada por el grado de humedad, la temperatura del suelo y la intensidad de la luz (Parker y Riches 1993; López-Granados y García-Torres 1993; Manschadi et al. 2001) y ocurre cuando la semilla se encuentra a menos de 1 cm de distancia de la raíz donde contacta con los exudados radiculares generados por la planta hospedadora (López-Granados y García-Torres 2002).



**Figura 7.** Plantas de *O. crenata* parasitando a *Vicia faba*.

Para hacer frente al ataque de jopo, se han investigado diversas medidas de control, destacando el control químico con glifosato (Parker y Riches 1993; Parker, 1994; López-Granados y García-Torres 1998), la solarización del suelo (Mauromicale et al. 2001), la rotación de cultivos (Al-Menoufi, 1994), el cambio de fecha de siembra (López-Granados y García-Torres 1998; Rubiales et al. 2003), la fertilización (Pieterse, 1991), la inundación del terreno (Zahran, 1982) y el control biológico (Linke et al. 1992). A pesar de todos estos estudios, la aplicación de estos métodos de control por separado mostró resultados no satisfactorios y se recomienda la integración de diferentes métodos. García-Torres y López-Granados (1991) integran la aplicación de herbicidas, el cambio de la fecha de siembra y la rotación de cultivos para reducir la infección de jopo en terrenos agrícolas. A partir de todas estas investigaciones se concluye que la mejora genética es el método más económico y

eficiente para el control de jopo (López-Granados y García-Torres 2002; Pérez-de-Luque et al. 2007).

La resistencia contra jopo es de naturaleza compleja, difícil de evaluar y de baja heredabilidad lo que convierte el proceso de mejora en una tarea muy difícil (Rubiales, 2003). En habas se ha observado que la resistencia puede estar correlacionada con el tamaño de la semilla, de forma que semillas de tamaño pequeño tienden a ser más resistentes a jopo que semillas más grandes (Cubero, 1973), aunque hasta la fecha, no se han identificado ninguna fuente de resistencia completa se han obtenido líneas que muestran resistencia incompleta (Pérez-de-Luque et al. 2010) como Vf172 (Cubero, 1973), con resistencia parcial pero sin características agronómicas deseables, la línea F402 (Nassib et al. 1982) con alto nivel de resistencia a jopo y con características agronómicas favorables, la línea Giza 402 resistente tanto en condiciones de invernadero como de campo y estable en diferentes ambientes (Valle del Nilo, La Costa de Siria y Andalucía al Sur de España) y la línea Vf1071 utilizada en Andalucía como fuente de resistencia para desarrollar líneas de interés en los programas de mejora (Cubero, 1991; Sillero et al. 1999). Actualmente la variedad Baraca, derivada de Vf1071, caracterizada por su buena adaptación, altos rendimientos y tolerancia al ataque de jopo es una de las más utilizadas comercialmente.

Además de *O. crenata*, otras especies de jopo son cada vez más importantes en ciertas áreas. *Orobanche foetida* Poir., distribuida ampliamente en la cuenca mediterránea y que inicialmente solo parasitaba huéspedes silvestres (Pujadas-Salva, 1999), ha atacado intensamente campos de habas y garbanzos en la región de Beja, Túnez (Kharrat et al. 1992), produciendo pérdidas de rendimiento entre 66 y 83% (Kharrat, 1999) y en Marruecos (Rubiales et al. 2005). Estos hechos hacen que se considere la importancia de esta especie y se evite su propagación por su cercanía geográfica a Andalucía donde aún parasita a especies silvestres.

Tanto para *A. fabae* como para *O. crenata* el uso de fuentes de resistencia en programas de mejora se puede ver condicionado por factores como la variabilidad patogénica del agente causal. En muchas ocasiones se puede observar una interacción diferencial entre los genotipos del agente causal de una enfermedad y los de la especie hospedadora. Ello marca la necesidad de conocer el espectro racial del patógeno en cada zona de cultivo a fin de elegir las fuentes de resistencia más idóneas para el

desarrollo de nuevas variedades resistentes. Por otra parte, la selección de genotipos interesantes dentro de cada programa, conlleva tediosos procesos de inoculación y evaluación, tareas que se agravan por la dificultad de establecer las condiciones óptimas para el desarrollo de la enfermedad dependiendo, además, de la disponibilidad del patógeno en cada momento. Por otra parte, si se está seleccionando frente a diversos patotipos del agente causal, la situación se verá agravada ya que habría que proceder a la inoculación y evaluación con cada uno de ellos.

### ***1.2.3. Mecanismos moleculares de resistencia a patógenos en plantas.***

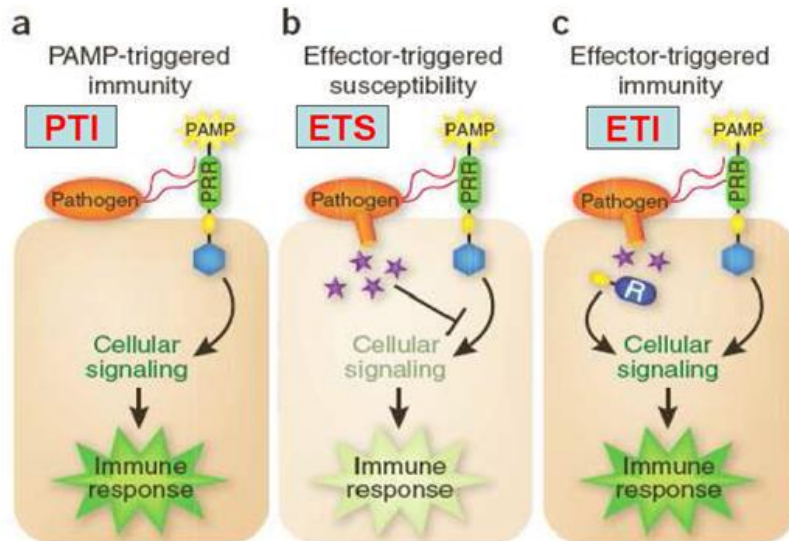
En la naturaleza, las plantas están expuestas al ataque de diversos patógenos (bacterias, virus, hongos, insectos) que perjudican su desarrollo y producción (Boyd et al. 2013). Para defenderse, las plantas han desarrollado mecanismos que les permiten reconocer, detener y contrarrestar la infección (Walbot, 1985; Durrant y Dong, 2004). El conocimiento molecular de estos mecanismos y la identificación de genes de resistencia es de gran interés en los programas de mejora.

Para acceder al interior de la planta, el microorganismo debe atravesar diferentes barreras físicas (superficie de hojas, raíces, pared celular) y químicas como fenoles, lignina, taninos, saponinas, antocianinas, flavonoides, lectinas, glucanasas, quitinasas (Kliebeinstein, 2004) que actúan como mecanismos de defensa frente a diversos patógenos (Heath, 2000). Modificaciones en la composición de la pared celular (lignificación o defectos en su composición) dificultan la penetración de patógenos (Bechinger et al. 1999, Vogel et al. 2002; 2004) y favorecen la resistencia en la planta.

Una vez en el interior de la planta, se activa la inmunidad iniciada por PAMPs (PTI, del inglés PAMP triggered immunity) (Jones y Dangl, 2006; Boller y Felix, 2009; Figura 8a) donde receptores (PRRs) de la superficie extracelular de la planta reconocen MAMPs (patrones moleculares asociados a microorganismos como quitina, ergosterol, flagelina o peptidoglucanos), ausentes en plantas (Boller y He, 2009) y DAMPs o patrones moleculares asociados a lesiones originadas por el patógeno al intentar atravesar la pared celular (De Lorenzo et al. 2011; Boyd et al. 2013). Este reconocimiento da lugar a una serie de reacciones intracelulares donde aumenta la concentración de  $Ca^{2+}$  y se activan cascadas de proteínas MAPK, que



producen cambios en la fosforilación de proteínas (Peck et al. 2001), activación génica (Zipfel et al. 2004, 2008), producción de especies reactivas del oxígeno, biosíntesis de etileno (Spanu et al. 1994), endocitosis de receptores (He et al. 2007; Kwon et al. 2008; Boller y Felix, 2009) y deposición de callosa (Gómez-Gómez et al. 1999) para reforzar la pared celular en los sitios de infección y evitar la propagación del patógeno.

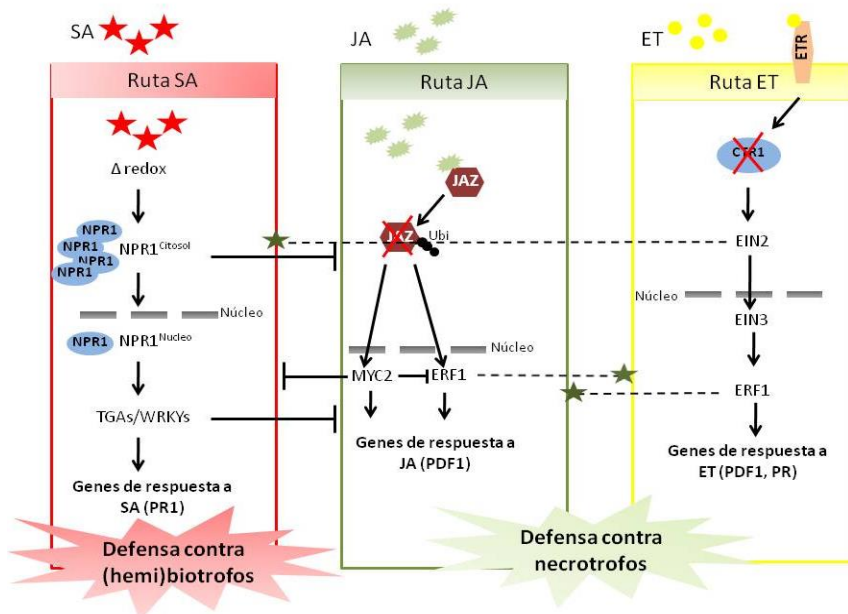


**Figura 8.** a) Activación de la inmunidad activada por PAMPs o PTI mediante el reconocimiento de PAMPs/MAMPs por receptores PRRs de la planta. b) Susceptibilidad desencadenada por un efector (ETS) producido por el patógeno que suprime PTI. c) Activación de la inmunidad desencadenada por efector (ETI) al ser reconocido específicamente por proteínas de resistencia de la planta (Fuente: Pieterse et al. 2009).

Algunos patógenos logran evadir o suprimir la PTI mediante la síntesis de efectores (Bent y Mackey, 2007) que modifican proteínas de la planta y mimetizan el efecto de sus hormonas (Jones y Dangl, 2006). Este mecanismo se conoce como susceptibilidad desencadenada por efector o ETS (Figura 8b). Frente a este mecanismo, la planta evoluciona y desarrolla la inmunidad activada por efector (ETI, Figura 8c), una respuesta más especializada, rápida y fuerte que PTI (Jones y Dangl, 2006; Chisholm et al. 2006) capaz de detectar efectores (fitotoxinas y factores proteicos) mediante proteínas de resistencia tipo NB-LRR de la planta. Entre las respuestas de defensa se produce un aumento del flujo de  $Ca^{2+}$ , la alcalinización del espacio extracelular, actividad proteínaquinasa, expresión de genes y síntesis de

especies reactivas de oxígeno (ROS), óxido nítrico (NO) y de hormonas (ácido salicílico (SA), ácido jasmónico (JA) y etileno (ET)), reforzamiento de la pared celular, lignificación, deposición de calosa, producción de agentes antimicrobianos y activación de la respuesta hipersensible (HR).

Entre las moléculas señalizadoras implicadas en las rutas de transducción de señales destacan el SA, JA y ET (Pieterse et al. 2009; Figura 9). La inducción de estas rutas depende del tipo de patógeno, de forma que en respuesta a patógenos necrotrofos (destruyen células huésped para alimentarse) se induce la ruta JA/ET y frente a biotrofos (obtienen nutrientes de células vivas) o hemibiotrofos (dependiendo del estado de su ciclo, son biotrofos o necrotrofos) la ruta del SA (Thaler y Bostock 2004).



**Figura 9.** Esquema de los principales componentes que actúan en las rutas de señalización del ácido salicílico (SA), ácido jasmónico (JA) y etileno (ET). a) SA origina la monomerización de NPR1 en citosol y su transporte al núcleo para interactuar con factores de transcripción (TGAs/ WRKYs) y expresar genes que codifican proteínas relacionadas con la patogénesis (PRs). WRKY70 actúa como interruptor molecular entre la vía del SA y la del JA. b) En la ruta del JA, la proteína JAZ interacciona con factores de transcripción MYC2 y ERF1 responsables de la expresión de genes de defensa y de integrar las señales de las rutas del JA y el ET. c) El etileno activa una cascada de quinasas MAPK, cuya señal es transmitida a EIN2 que evita la degradación del factor de transcripción EIN3. Su acumulación activa a ERF1 que regula la expresión génica. —|, indica inhibición de la ruta, ★ indica activación de la ruta. Fuente: Pieterse et al. 2009.

Aunque la comunicación cruzada entre las vías de señalización hormonal proporciona a la planta una gran capacidad reguladora que puede adaptar su respuesta de defensa a diferentes tipos de atacantes, los patógenos han evolucionado con potentes efectores que alteran la red de señales, manipulan las defensas y pueden llegar a suprimir la ETI (Alfano y Collmer, 2004; Kamoun, 2007). Si el patógeno suprime este tipo de inmunidad y la planta no es capaz de detectarlo se nutrirá, dispersará y dará lugar a la llamada susceptibilidad activada por efector (ETS) (Jones y Dangl, 2006).

### *1.3. Selección asistida por marcadores (MAS).*

El objetivo principal de la mejora clásica es obtener nuevas combinaciones genéticas a partir de la selección de los mejores fenotipos de una progenie obtenidos tras realizar uno o varios cruzamientos. Este tipo de mejora ha sido muy eficaz a lo largo de la historia, sin embargo, presenta algunos inconvenientes tales como la influencia del ambiente sobre el fenotipo, el alto coste de los procesos de fenotipado, el gran tiempo invertido en el proceso de selección y la dificultad de evaluación de ciertos caracteres como son los caracteres poligénicos o cuantitativos.

La mayor parte de los caracteres de interés agronómico, incluyendo la resistencia a enfermedades, son de naturaleza cuantitativa, es decir, muestran amplias distribuciones fenotípicas más o menos continuas. Estos caracteres están controlados por una importante cantidad de genes sujetos a la influencia del ambiente. Su disección genética es una tarea ardua y costosa, por lo que las estrategias de la mejora genética tradicional se centraron en aproximaciones estadísticas mediante el estudio de correlaciones entre parientes para estimar la heredabilidad, el parámetro que permite predecir la respuesta a la selección. A pesar de esta dificultad, desde la década de los 90, los marcadores moleculares han permitido una disección más detallada de la base genética de los caracteres cuantitativos o QTLs (Quantitative Trait Loci), pudiéndose detectar los efectos genéticos más notables de un carácter mediante el rastreo con mapas genéticos (Collard et al. 2005). Para ello se evalúan las asociaciones estadísticas entre los marcadores de un mapa y los fenotipos del carácter bajo estudio, teniendo en cuenta que los genes/marcadores localizados en la proximidad en los genomas tienden a transmitirse juntos. Empleando marcadores en

la selección del carácter se elimina el efecto ambiental, aumentando la repuesta a la selección de genes de importancia económica, pero de herencia compleja (Collard y Mackill 2008; Xu y Crouch 2008).

Así desde la década de los 90 (Lander y Botstein, 1989), los marcadores moleculares han demostrado ser eficientes para la detección de genes y QTLs de interés mediante la denominada selección asistida por marcadores o MAS (Marker Assisted Selection). La MAS fue desarrollada para evitar los inconvenientes de la mejora clásica, originando un cambio en el criterio de selección, que pasó de una selección de fenotipos a una selección de marcadores ligados a genes de interés. Entre sus principales ventajas destacan la sencillez, el ahorro de tiempo al seleccionar los genotipos durante los primeros estadios de desarrollo de la planta y la fiabilidad del proceso. Además no se encuentra afectada por condiciones ambientales, facilita la selección de varios caracteres de forma simultánea y puede ser utilizada tanto para caracteres cualitativos como cuantitativos (Francia et al. 2005). En el caso de la resistencia a patógenos puede sustituir los procesos de inoculación y evaluación en generaciones avanzadas del programa de mejora. El empleo de dichos marcadores, puede facilitar la selección simultánea frente a varias enfermedades o frente a varios patotipos de la misma enfermedad con un gran ahorro de tiempo y dinero.

Para que un marcador sea utilizado en MAS debe ser fiable, fácilmente intercambiable y preferiblemente codominante, permitiendo distinguir individuos homocigotos de heterocigotos en la progenie. Se pueden considerar buenos marcadores aquellos que se encuentran ligados a distancias menores de 5 cM del carácter que se desea seleccionar (Collard et al. 2005). El éxito de la MAS para caracteres agrónomicamente importantes como son la resistencia a enfermedades y los rasgos de calidad, depende de la identificación de los genes responsables o relacionados con dichos caracteres y que puedan ser utilizados para el desarrollo de marcadores diagnóstico (Hospital, 2009). Este tipo de marcadores presentan ventajas sobre los marcadores estrechamente ligados a un carácter, como son la ausencia de fenómenos de recombinación entre el marcador y el gen que controla el carácter, por lo que no se producen errores en el proceso de selección.

Diferentes tipos de marcadores moleculares han sido utilizados en MAS (Collard et al 2008): RFLPs (Restriction Fragment Length Polymorphisms), RAPDs (Random Amplified polymorphic DNA), AFLPs (Amplified Fragment Length Polymorphism), SSRs (Simple Sequence Repeats), SCARs (Sequence Characterized Amplified Regions), CAPS (Cleaned Amplified Polymorphic Sequences), STSs (Sequence-Tagged Sites), ESTs (Expressed Sequence Tags), SSCP (Single-Strand Conformation Polymorphism) y SNPs (Single Nucleotide Polymorphisms). Los SNPs son la más reciente generación de marcadores moleculares. Se basan en la identificación de la sustitución de un nucleótido por otro y representan dos alelos simples. En plantas su frecuencia es de uno cada 100-300 bp pares de bases (Edwards et al., 2007; Xu, 2010). Tradicionalmente la identificación de SNPs se ha realizado mediante digestión con enzimas de restricción seguida de electroforesis. En la actualidad, el análisis de SNPs se puede realizar a gran escala y a un coste muy competitivo mediante diversas técnicas de última generación (Affymetrix, Illumina Agilent, Sequenom, etc.). El desarrollo de plataformas de genotipado como por ejemplo la TaqMan/Kaspar (LGC/KBiosciences), o el sistema de MassArray-Sequenom hacen posible explorar la variabilidad genética de SNPs de forma eficiente y asequible, lo que posibilita la identificación de asociaciones para la detección de genes candidatos relacionados con caracteres de interés en mejora.

#### ***1.4. Empleo de herramientas genómicas en la Mejora de las leguminosas.***

Desde el redescubrimiento de las leyes de Mendel hasta la actualidad, avances en Genética Molecular como el descubrimiento de las enzimas de restricción o el desarrollo de técnicas como la PCR, han repercutido enormemente en la mejora genética vegetal. Recientemente, la irrupción de nuevas tecnologías de secuenciación, junto con el progresivo abaratamiento de los costes de genotipado están propiciando un cambio de escenario. A la publicación del primer genoma vegetal secuenciado (The *Arabidopsis thaliana* Genome Initiative 2000) le siguieron el de otras muchas especies modelo. Dentro de la familia de las leguminosas (*Fabaceae*), además de los genomas de cuatro especies modelo: *Medicago truncatula* (Young et al. 2011), *Lotus japonicus* (Sato et al. 2008), *Phaseolus vulgaris* (Schmutz et al. 2014) y *Cajanus cajan* (Varshney et al. 2012), ya se dispone de genomas más o menos completos de otras especies, como la soja (Schmutz et al. 2010) y el garbanzo (Jain et al. 2013;

Varshney et al. 2013) y están en marcha la secuenciación de otros. Los avances mencionados han permitido un mayor desarrollo genómico y tecnológico en herramientas moleculares y bioinformáticas que han revolucionado la mejora y aportado un mejor conocimiento de las especies estudiadas y de sus procesos biológicos (Echenique et al. 2004).

Para abordar distintos problemas biológicos la genómica se subdivide en distintas áreas de conocimiento como la genómica estructural, la genómica comparativa y la genómica funcional. Aunque exploten herramientas distintas todas tienen un objetivo común, generar información exacta para identificar genes de interés y poder entender mejor los procesos biológicos en los que están implicados. Si bien de forma más lenta que en otras especies vegetales, las leguminosas y en particular las habas, han ido incorporando herramientas para el rastreo genómico mediante la saturación de sus mapas genéticos (Satovic et al. 2013; Duc et al. 2015; Webb et al. 2016), la identificación de genes candidatos mediante estrategias de genómica comparada (Zhu et al 2005; Choi et al. 2004; Ávila et al. 2007; Gujaria-Verma et al. 2014) o análisis de expresión (Madrid et al. 2010; 2012) génica y más recientemente, la utilización de nuevas tecnologías de secuenciación masiva que están incrementando los recursos genómicos de forma exponencial (Kaur et al. 2012; Ocaña et al. 2015; Ray et al. 2015).

#### ***1.4.1. Genómica estructural.***

La genómica estructural se centra en la caracterización y localización de secuencias génicas. Uno de los primeros pasos para caracterizar un genoma suele ser la obtención de mapas genéticos (mapas de ligamiento) que indican la localización aproximada de genes de interés respecto a otros conocidos. Estos mapas se basan en la frecuencia de recombinación, cuyo valor dependerá de que los genes estén o no ligados, de forma que si la frecuencia de recombinación entre dos loci es mayor al 50% estarán localizados en cromosomas diferentes o muy separados en el mismo cromosoma; mientras que si la frecuencia de recombinación es menor del 50%, los loci están ligados, muy próximos en el mismo cromosoma y pertenecerán al mismo grupo de ligamiento.

#### ***1.4.1.1. Mapa genético de habas.***

Tradicionalmente los mapas genéticos han sido el método para detectar marcadores ligados a caracteres cualitativos y cuantitativos de importancia agronómica que puedan ser utilizados en programas de mejora para seleccionar genotipos de interés. Han sido fundamentales para el desarrollo de mapas físicos y para la secuenciación de genomas completos. Actualmente, y a pesar de las limitaciones biológicas de las habas, como son el gran tamaño de su genoma y la alta proporción de elementos de ADN repetitivo, se dispone de mapas genéticos que han permitido localizar genes y QTLs que controlan la resistencia a diferentes patógenos como el jopo (Román et al. 2002; Díaz-Ruiz et al. 2009; 2010; Gutiérrez et al. 2013), *A. fabae* (Román et al. 2003; Ávila et al. 2004; Díaz-Ruiz et al. 2009; Kaur et al. 2014; Atienza et al. 2016) y la roya (Ávila et al. 2003). También se dispone de mapas genéticos donde se localizan las regiones que controlan otras características agronómicas, como, la fecha de floración, el rendimiento y la adaptación a la sequía (Cruz-Izquierdo et al. 2012; Khazaei et al. 2014). Toda esta información ha dado origen al desarrollo de dos mapas consenso que facilitan el mapeo comparativo y la transferencia de información con el resto de leguminosas (Satovic et al. 2013; Webb et al. 2016). Sin embargo, se requiere mejorar la saturación de estos mapas para identificar marcadores estrechamente vinculados y determinar la posición exacta de QTLs de interés agronómico. Este ha sido uno de los objetivos desarrollados en esta tesis con el fin de identificar genes candidatos para el control de la resistencia a *A. fabae* y jopo en las poblaciones de habas Vf6 x Vf136 y más profundamente en la población 29H x Vf136.

A pesar de los avances recientes, ha sido muy difícil diseñar marcadores moleculares sobre genes responsables de caracteres de interés agronómico en habas. Sólo se han descrito marcadores asociados al gen TFL-1, responsable del crecimiento determinado en habas (Ávila et al. 2006; 2007) y al gen TTG-1 responsable de la ausencia de taninos en semillas (Webb et al. 2016), ambos caracteres de tipo monogénico. Aún no se tienen referencias de marcadores diagnóstico para caracteres poligénicos responsables de la resistencia a estreses bióticos como la resistencia a patógenos, debido a que las regiones genómicas que controlan estos caracteres todavía no están bien caracterizadas y requieren una mayor saturación.

En las tablas 1 y 2 pueden verse los diferentes QTLs identificados para el control de la resistencia a *A. fabae*, *O. crenata* y *O. foetida* en los dos cruzamientos (Vf6 x Vf136 y 29H x Vf136) de habas que van ocupar esta tesis. Algunos de estos QTLs han sido validados tanto en poblaciones F2 como en líneas avanzadas o RILs (Recombinant Inbred Lines). De los diferentes QTLs localizados para *A. fabae* (Tabla 1), solo han podido validarse en F2 y RILs, *Af1* y *Af2* en el cruzamiento Vf6 x Vf136 y *Af3* en 29H x Vf136. Además *Af1* (Román et al. 2003, Díaz-Ruiz et al. (2009a) identificado en la población Vf6 x Vf136 y *Af3* (Ávila et al. 2004, Atienza et al. 2016) identificado en la población 29H x Vf136 están asignados al cromosoma III y podrían corresponder a las mismas regiones genómicas. Lo mismo ocurre con *Af2* (Román et al. 2003, Díaz-Ruiz et al. (2009a) identificado en la población Vf6 x Vf136 y *Af4* (Ávila et al. 2004, Atienza et al. 2016) presente en la población 29H x Vf136, asignados ambos al cromosoma II.

Para la resistencia a *O. crenata* (Tabla 2), han podido validarse *Oc2* y *Oc3* en las poblaciones F2 y RILs del cruzamiento Vf6 x Vf136 mientras que en el cruzamiento 29H x Vf136 se han validado en diferentes ambientes *Oc7* y *Oc8*. En este caso, *Oc2*, identificado en la población Vf6 x Vf136 (Román et al. 2002; Díaz-Ruiz et al. 2009) y *Oc7* detectado en 29H x Vf136 (Gutiérrez et al. 2013) están asignados al cromosoma VI y podrían corresponder a las mismas regiones genómicas. La falta de marcadores comunes en ambas poblaciones dificulta estas comparaciones por lo que la saturación de estas zonas con marcadores comunes y los avances en sintenia con especies modelo están ayudando a aclarar estas hipótesis.

Las perspectivas de MAS para la resistencia a *Ascochyta* y jopo es identificar los genes responsables y desarrollar marcadores diagnóstico que permitan extrapolar los resultados de QTL estables de unas poblaciones a otras y faciliten la selección de materiales en los programas de mejora.



**Tabla 1:** Principales QTLs para resistencia a *Ascochyta fabae* detectados en poblaciones F2 y RILs de habas en los cruzamientos Vf6 × Vf136 y 29H xVf136.

QTL	Experimento	Cruzamiento (Generación)	CR / GL	Referencia
<i>Af1</i>	CO95_DSL	Vf6×Vf136 (F2)	III	Román et al. 2003
<i>Af2</i>	CO95_DSL		II	
<i>Af1</i>	CO95_DSL		III	
<i>Af1</i>	CO95_DSS	Vf6×Vf136 (RILs)	III	Díaz-Ruiz et al. 2009
<i>Af2</i>	CO95_DSL		II	
<i>Af2</i>	CO95_DSS		II	
<i>Af3</i>	CO99_DSL	29H xVf136 (F2)	III	Ávila et al. 2004
<i>Af3</i>	LO98_DSL		III	
<i>Af3</i>	CO99_DSS		III	
<i>Af3</i>	LO98_DSS		III	
<i>Af4</i>	CO99_DSL		II	
<i>Af4</i>	LO98_DSL		II	
<i>Af4</i>	CO99_DSS		II	
<i>Af4</i>	LO98_DSS		II	
<i>Af5</i>	CO99_DSL		Lg15	
<i>Af5</i>	CO99_DSS		Lg15	
<i>Af6</i>	LO98_DSL		Lg12	
<i>Af7</i>	LO98_DSS		Lg3	
<i>Af7</i>	LO98_DSL		Lg3	
<i>Af8</i>	LO98_DSS		Lg14	
-	<i>field_DSL</i>	29H xVf136 (RILs)	II	Atienza et al. 2016
-	CO99_DSL		II	
-	LO98_DSL		II	
<i>Af3</i>	<i>field_DSL</i>		III	
<i>Af3</i>	<i>field_DSP</i>		III	
<i>Af3</i>	<i>field_DSS</i>		III	
<i>Af3</i>	CO99_DSL		III	
<i>Af3</i>	LO98_DSL		III	
<i>Af3</i>	LO98_DSS		III	
-	<i>field_DSP</i>		VI	

Af: Resistencia a *Ascochyta fabae*; QTL: quantitative trait locus; CR: cromosoma; GL: grupo de ligamiento; DSL: Severidad de la infección de *Ascochyta fabae* en hojas; DSS: Severidad de la infección de *Ascochyta fabae* en tallo; DSP: Severidad de la infección de *Ascochyta fabae* en vaina; CO95, CO99 y LO98: Cepas de *A. fabae* obtenidas en Córdoba (CO) y Logroño (LO).

**Tabla 2:** Principales QTLs para resistencia a *Orobanche crenata* y *Orobanche foetida* detectados en poblaciones F2 y RIL de habas en los cruzamientos Vf6 × Vf136 y 29H xVf136.

QTL	Localidad y Año	Cruzamiento (Generación)	CR / GL	Referencia
<b><i>O.crenata</i></b>				
<i>Oc1</i>	Córdoba 1997/08	Vf6×Vf136 (F2)	LG03	Román et al. 2002
<i>Oc2</i>	Córdoba 1997/08		VI	
<i>Oc3</i>	Córdoba 1997/08		II	
<i>Oc2</i>	Córdoba 2003/04	Vf6×Vf136 (RILs)	VI	Díaz-Ruíz et al. 2010
<i>Oc2</i>	Mengíbar 2004/05		VI	
<i>Oc3</i>	Córdoba 2003/04		II	
<i>Oc3</i>	Córdoba 2004/05		II	
<i>Oc4</i>	Córdoba 2004/05		I	
<i>Oc5</i>	Mengíbar 2004/05		I	
<i>Oc6_C3</i>	Córdoba 2003/04	Vf6×Vf136 (RILs)	I	Díaz-Ruíz et al. 2009
<i>Oc14_C3</i>	Córdoba 2003/04		II	Satovic et al. 2013
<i>Oc15_M4</i>	Mengíbar 2004/05		II	
<i>Oc16_C4</i>	Córdoba 2004/05		V	
<i>Oc7</i>	Córdoba 2006/07	29H xVf136 (RILs)	VI	Gutiérrez et al. 2013
<i>Oc7</i>	Córdoba 2007/08		VI	
<i>Oc7</i>	Córdoba 2008/09		VI	
<i>Oc8</i>	Córdoba 2006/07		V	
<i>Oc8</i>	Córdoba 2007/08		V	
<i>Oc9</i>	Córdoba 2008/09		I	
<i>Oc10</i>	Kafr El-Sheikh 2007/08		III	
<i>Oc11</i>	Kafr El-Sheikh 2007/08		I	
<i>Oc12</i>	Kafr El-Sheikh2007/08		LG8	
<i>Oc13</i>	Kafr El-Sheikh 2007/08		I	
<b><i>O. foetida</i></b>				
<i>Of1</i>	Beja 2003/04	Vf6xVf136 (RILs)	I	Díaz-Ruíz et al. 2010
<i>Of2</i>	Beja 2004/05		III	
<i>Of3</i>	Beja 2006/07	29H xVf136 (RILs)	V	Gutiérrez et al. 2013
<i>Of4</i>	Beja 2006/07		I	
<i>Of5</i>	Beja 2006/07		LG29	

Oc: Resistencia a *Orobanche crenata*; Of: Resistencia a *Orobanche foetida*; QTL: quantitative trait locus; CR: cromosoma; GL: grupo de ligamiento.

#### ***1.4.2. Genómica comparada y sintenia.***

El mapeo comparativo y el análisis de secuencias genómicas confirma que la organización de los genes dentro de los genomas ha permanecido muy conservada a lo largo de la evolución. Parece que todos los seres vivos están relacionados, descienden de un ancestro común y utilizan grupos génicos parecidos para realizar las funciones celulares básicas, como los mecanismos de replicación, transcripción y traducción. La genómica comparativa hace uso de esta información, permitiendo establecer las posibles relaciones estructurales y funcionales entre una especie modelo y las especies cultivadas relacionadas, de forma que resulta posible predecir caracteres no observados en una especie sobre la base de los existentes en especies evolutivamente próximas (Weeden et al. 1992; Menancio-Hautea et al. 1993). De esta forma es posible descubrir nuevos genes, conocer las relaciones entre organismos y su ambiente, reducir el coste e incrementar la eficiencia en la mejora genética vegetal (Phan et al. 2006). Este campo utiliza una amplia gama de técnicas y recursos, incluyendo la construcción y utilización de bases de datos que contengan secuencias nucleotídicas y aminoacídicas, técnicas citogenéticas de cartografía génica como la hibridación in situ fluorescente (FISH) y métodos experimentales (mutagénesis).

En genómica comparada, sintenia es sinónimo de colinearidad y ambos hacen referencia al grado de conservación en el contenido de genes, el orden y la orientación de los mismos entre cromosomas de distintas especies o cromosomas no homólogos de una misma especie. El desarrollo de especies modelo han permitido utilizar la genómica comparada para estudiar enfermedades e interacciones entre genes en procesos básicos, transferir información a especies de interés agronómico y desarrollar estudios de sintenia a gran escala (Gepts et al. 2005). En leguminosas, la transferencia de conocimiento entre especies de interés agronómico está ampliamente documentada a partir del uso de marcadores moleculares basados en regiones conservadas entre especies (Choi et al. 2004b; Kaló et al. 2004; Zhu et al. 2005; Ellwood et al. 2008).

##### ***1.4.2.1 Medicago truncatula como especie modelo en leguminosas.***

La importancia de las especies modelo se centra en el origen común de genes conservados en diferentes especies a lo largo de la evolución. En biología, el uso de

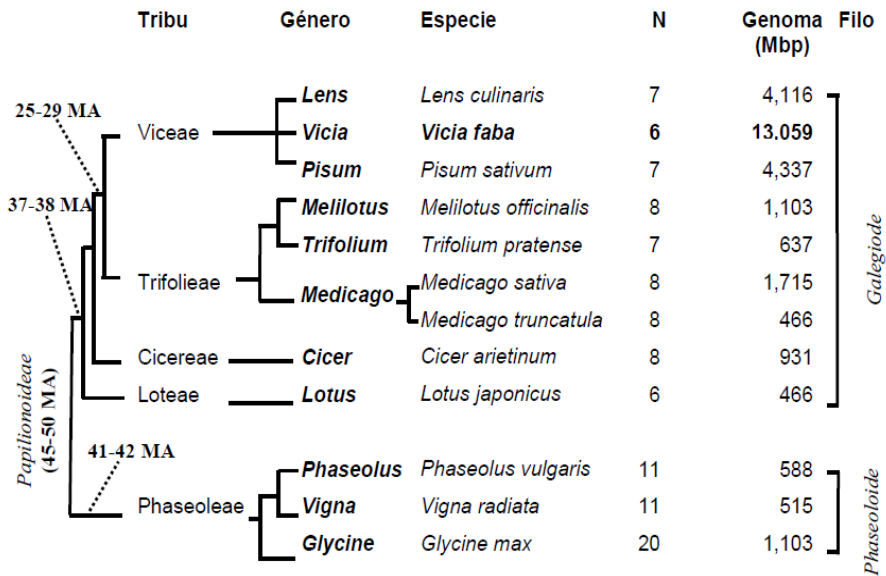
especies modelo es necesario para comprender mejor cualquier proceso biológico. Entre sus ventajas respecto a especies cultivadas, están su fácil conservación y manipulación en laboratorio, su ciclo de vida corto, su gran número de descendientes, la disponibilidad de variabilidad fenotípica y genotípica, la observación fácil de fenotipos y la disponibilidad de técnicas de estudio cada vez más desarrolladas. *M. truncatula* junto a *L. japonicus* fueron considerados en los 90, organismos modelo para leguminosas (Cook et al. 1997; Cook, 1999). La presente introducción se centrará en la especie modelo *M. truncatula*, al haber sido la utilizada en este trabajo por su mayor cercanía al género *Vicia*.

*Medicago* es una leguminosa que pertenece al filo *Galeoide*, tribu *Trifoliae* y que presenta todas las ventajas como especie modelo, citadas anteriormente. Se trata de una especie diploide ( $2n=16$ ) (Zhu et al. 2005), autógena, con un genoma pequeño (<550 Mb), que produce numerosas semillas y presenta un ciclo de generación menor a 3 meses, lo que la convierte en una herramienta muy útil tanto para estudios genéticos como genómicos (Cook, 1999, Young et al. 2005). Todas estas características la convirtieron en un modelo para estudios de genómica funcional y estructural, desarrollándose numerosas herramientas genómicas como mapas genéticos saturados (Thoquet et al. 2002; Choi et al. 2004), mapas físicos (Young et al. 2005; Mun et al. 2006), bibliotecas de cromosomas artificiales bacterianos (BAC), 286.176 ESTs (Expressed Sequence Tag), “chips” desarrollados por Afimetrix® o la secuenciación de su genoma por un Consorcio Internacional (Young et al. 2011).

La estrecha relación filogenética entre *M. truncatula* y otras leguminosas como *Pisum sativum*, *Lens culinaris*, *C. arietinum* o *V. faba* (Figura 11) y el desarrollo de herramientas genéticas y genómicas en esta planta convierten a *Medicago* en un organismo muy atractivo para aumentar el conocimiento sobre caracteres agronómicos importantes en especies próximas (Denarie, 2002).

Así, el uso de las herramientas genómicas y moleculares disponibles en la planta modelo, permiten no sólo la investigación de los procesos básicos importantes en leguminosas, sino también la transferencia de esa información a especies de interés agronómico, utilizando lo que se denomina *sintenia*, es decir, la colinealidad en la organización genómica de diferentes especies. Esta colinealidad permite identificar genes ortólogos (genes que están presentes en diferentes especies y que proceden de un gen ancestral común por especiación), lo que resulta esencial para la predicción

fiable de la función de genes de interés en especies donde el conocimiento a nivel genómico y transcriptómico es mucho más escaso y complicado de conseguir.



**Figura 10.** Relaciones filogenéticas de la subfamilia Papilionoideae. Fuente: Zhu et al. 2005. A la derecha de las especies está indicado el número cromosómico (N) y el tamaño del genoma expresado en Mpb, MA = Millones de años.

En leguminosas, el primer estudio exhaustivo de macrosintenia fue realizado por Choi et al. (2004 a y b) que utilizó a *M. truncatula* como especie modelo y aportó la primera representación del grado de sintenia existente entre leguminosas siendo bastante elevada dentro de los filos *Galegoide* y *Phaseoloide* y disminuyendo al comparar leguminosas de diferente filo, aunque persiste un cierto grado de conservación. Posteriormente Zhu et al. (2005) aumentaron el número de especies estudiadas y modificaron el mapa consenso existente (Figura 11).

Diferentes estudios han mostrado la colinealidad entre *M. truncatula* y otras especies como con *Arabidopsis* (Zhu et al. 2003) y las leguminosas: *M. sativa* (Choi et al. 2004a y b); *G. max* (Mudge et al. 2005); *L. angustifolium* (Nelson et al. 2006); *L. japonicus*, (Cannon et al. 2006); *L. culinaris* y *L. albus* (Phan et al. 2006, 2007). La aplicación de la sintenia para identificar genes en otras especies ha proporcionado ya buenos resultados. Yang et al. (2008) describieron mediante paseo cromosómico (“map-based cloning”) un gen de resistencia a anthracnosis en alfalfa utilizando



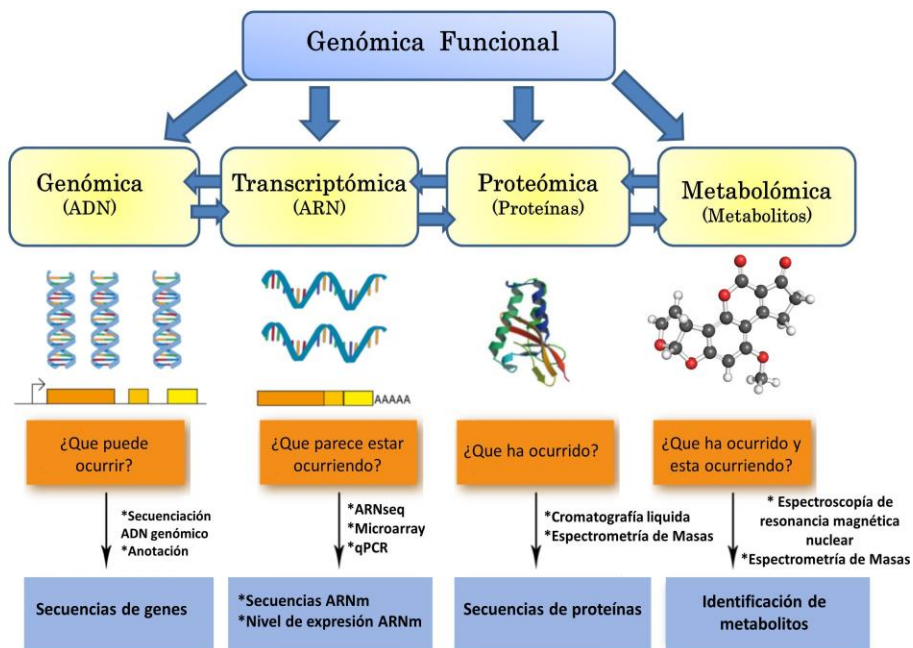
A pesar de las limitaciones citadas anteriormente en *V. faba*: genoma 25 veces mayor que la especie modelo *M. truncatula*, gran cantidad de ADN repetitivo y falta de marcadores robustos y transferibles en sus mapas de ligamiento, el desarrollo de técnicas moleculares y bioinformáticas junto a la información disponible de especies filogenéticamente cercanas, ha permitido el mapeo comparativo con *M. truncatula*, *P. sativum* y *L. culinaris* y la identificación de regiones homólogas responsables de caracteres agronómicos de interés (Gutiérrez, 2007; Ellwood et al. 2008; Cruz Izquierdo et al. 2012; Webb et al. 2016).

### ***1.4.3. Genómica Funcional.***

En las últimas décadas, se han desarrollado numerosos marcadores moleculares para los más importantes cultivos. Estos marcadores han servido para construir mapas genéticos e identificar genes o QTLs para futuras estrategias de MAS. Junto al tradicional mapeo de QTLs, recientemente se aplican nuevos enfoques como el TILLING (Targeting Induced Local Lesions in Genomes) que permite la identificación de una mutación en un gen concreto, el mapeo por asociación o la genómica funcional, entre otros (Varshney et al. 2013; Libault and Dickstein 2013; Pandey et al. 2016).

La genómica funcional estudia aspectos relativos a la función e interacciones entre genes y proteínas. Mientras los estudios clásicos sólo permitían estudiar la variación de la expresión gen a gen, la genómica funcional moderna ha revolucionado este proceso. Previamente, para identificar la función y propiedades de una proteína se tenía que aislar y sintetizar mediante la predicción de la secuencia de aminoácidos a partir de la secuencia de nucleótidos de un gen. Este proceso largo y costoso se ha optimizado enormemente con el desarrollo de tecnologías de alto rendimiento (high throughput) y métodos bioinformáticos que permiten analizar gran cantidad de genes simultáneamente e identificar su función a partir de su secuencia. Este tipo de análisis se conoce como análisis transcriptómico global y proporciona una percepción de las funciones biológicas del genoma, a través del análisis automatizado de su expresión. Para ello se combinan metodologías experimentales de secuenciación a gran escala con el análisis estadístico o computacional de los resultados (Hieter y Boguski, 1997).

Entre las aplicaciones de la genómica funcional destacan la caracterización de transcriptomas específicos de tejidos, el estudio de patrones de expresión génica asociados a procesos celulares y el estudio de las alteraciones en los perfiles de expresión en procesos patológicos (infecciones). Entre sus métodos de análisis cabe destacar: Northern Blot, PCR cuantitativa, differential display, microarrays, el desarrollo de librerías substractivas mediante SSH (suppression subtractive hybridization), SAGE (serial análisis de la expresión génica) y su versión mejorada SuperSAGE y la secuenciación masiva de librerías de cDNA o RNA-Seq.



**Figura12.** Estudios y herramientas que engloba el uso de la genómica funcional.

Todas estas técnicas han proporcionado enormes bases de datos de secuencias de proteínas (muchas de función desconocida) que han permitido evaluar y estudiar la célula o el organismo como un sistema para llegar a comprender mejor los procesos biológicos desarrollados en su interior (Mittler y Shulaev 2013).



***1.4.3.1. Tecnologías de secuenciación de nueva generación y bioinformática.***

En los últimos años han tenido lugar avances espectaculares en la velocidad, capacidades y abaratamiento de la tecnología de secuenciación de genomas. Frente a la tecnología convencional de secuenciación Sanger, la secuenciación masiva permite hoy en día obtener millones de secuencias de ADN a una velocidad sin precedentes y a un coste cada vez más reducido. Ello unido a la disponibilidad de análisis bioinformáticos a gran escala, está revolucionando los estudios de genómica funcional.

Con la secuenciación de nueva generación (next generation sequencing [NGS]), se ha logrado que el número de bases que se pueden secuenciar por el mismo precio haya crecido exponencialmente (Egan et al. 2012). Las nuevas plataformas pueden secuenciar millones de fragmentos de ADN de forma paralela a un precio por base mucho más barato y permiten detectar todos los tipos de variación genómica en un único experimento, incluyendo mutaciones puntuales, pequeñas inserciones y deleciones y también variantes estructurales tanto equilibradas (inversiones y traslocaciones) como desequilibradas (deleciones o duplicaciones).

Las tecnologías de NGS utilizadas mayoritariamente por la comunidad científica son: el 454 GSFlex de Roche, el Genome Analyzer o HiSeq de Illumina y el SOLiD de Life Technologies (Varhsney et al. 2009). Difieren en varios aspectos pero el esquema principal de trabajo es conceptualmente similar para todas ellas. El ADN se fragmenta y mediante ligación se le añaden secuencias adaptadoras a los extremos. Los fragmentos de ADN se amplifican clonalmente y se agrupan (clustering) para ser utilizados como entidades a secuenciar. La secuenciación se realiza entonces alternando ciclos de terminación reversible cíclica (cyclic reversible termination [CRT]) donde se utiliza terminadores reversibles para incorporar nucleótidos marcados con fluorescencia que a continuación son «fotografiados» en la toma de imágenes y posteriormente procesados. Las secuencias cortas producidas por el instrumento a partir de los extremos del ADN con los adaptadores se denominan lecturas o “reads”. Dichas secuencias son más cortas que las obtenidas mediante el tradicional método Sanger, un hecho que dificulta el ensamblaje y análisis de datos. Además de secuencia de ADN cortas, estas tecnologías pueden generar archivos de datos del tamaño de terabytes ( $10^{12}$  bytes) en cada lectura lo que incrementa

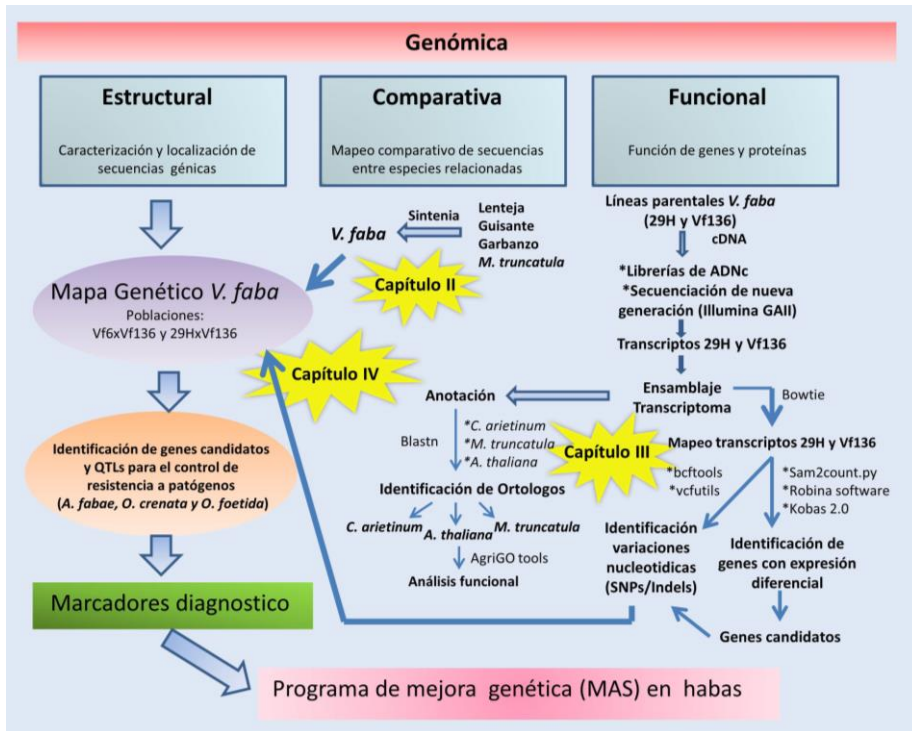
enormemente los requerimientos de recursos informáticos en los laboratorios de secuenciación (Exposito et al 2016).

Aunque ya se dispone de herramientas bioinformáticas y algoritmos, los esfuerzos hoy en día se dirigen principalmente a mejorar la precisión de los alineamientos de los datos de distintos laboratorios, por lo que hay necesidad de herramientas bioinformáticas o plataformas que faciliten el análisis de las secuencias de NGS de modo más eficiente, fiable y fácil.

Las ciencias informáticas están tomando una relevancia crucial en la NGS ya que sus capacidades son esenciales para manejar y analizar una cantidad de datos sin precedentes que un ordenador común no puede manejar. Para que un análisis de datos tenga sentido biológico se debe incorporar al sistema toda la información relevante que exista. Aunque para algunas plataformas ya existen programas para el manejo y análisis de datos, las tareas más complejas requieren de una persona con conocimientos en bioinformática que pueda integrar la información disponible en las distintas bases de datos y los conocimientos en ciencias computacionales. La creciente complejidad que representa encontrar información útil en un “laberinto” de datos ha generado el desarrollo de técnicas que integran la información dispersa, gestionan bases de datos, las seleccionan automáticamente, evalúan su calidad, y facilitan su accesibilidad para los investigadores. En este sentido la bioinformática, ofrece la capacidad de comparar y relacionar la información genética con una finalidad deductiva, ofreciendo respuestas que no parecen obvias a la vista de los resultados experimentales. Todas estas tecnologías vienen justificadas por la necesidad de tratar información masiva, con enfoques integrados como las requeridas en estudios de genómica funcional, o expresión multigénica como los tratados en este estudio.

#### ***1.4.3.2. Aplicaciones de la NGS en genética y mejora vegetal.***

Las tecnologías NGS permiten generar recursos genómicos con aplicaciones potenciales en genética y mejora incluyendo el desarrollo de marcadores moleculares, el mapeo de QTLs y la identificación de genes candidatos. En este trabajo, la secuenciación del transcriptoma de genotipos parentales contrastados utilizando NGS, ha proporcionado una nueva batería de transcritos donde se han identificado marcadores moleculares basados en SNPs (Figura 13), aumentando en gran medida el número de marcadores SNPs localizados directamente en genes. Entre otras aplicaciones, estos marcadores pueden ser utilizados para construir mapas genéticos, identificar QTLs y comprobar la introgresión de un carácter tras un cruzamiento. Marcadores asociados a QTLs que controlan un carácter de interés son útiles en la selección de la progenie portadora de alelos favorables a través de la selección asistida por marcadores (MAS) y el desarrollo de marcadores diagnóstico. Así, la secuenciación de los genes que se están expresando en un momento concreto y en dos genotipos que difieren para un carácter, puede ser utilizada para identificar genes candidatos implicados o asociados con el carácter. El mapeo de estos genes candidatos, junto con el fenotipado de las poblaciones segregantes (RILs) obtenidas a partir de los genotipos contrastantes, proporcionan QTLs de expresión (eQTLs) y los marcadores asociados a estos QTLs, por lo tanto, pueden llegar a ser marcadores perfectos para MAS. Otras aplicaciones importantes de la NGS se encuentran en la genética de asociación o genética de poblaciones, donde genomas o mezcla de productos de PCR de miles de genes candidatos pueden ser secuenciados en cientos de individuos. Los datos de secuencia así obtenidos, pueden ser utilizados para identificar SNPs o haplotipos entre genes o genomas para su uso en dichos campos. Las tecnologías NGS son también especialmente adecuadas para la resecuenciación de genomas, la secuenciación “de novo” de especies sin genoma de referencia, mapeo por asociación usando poblaciones naturales así como en genética de poblaciones y biología evolutiva (Varshney et al. 2009).



**Figura 13.** Esquema de los estudios de genómica estructural, comparativa y funcional aplicados durante el desarrollo de este trabajo de tesis para la saturación del mapa y la identificación de genes candidatos de resistencia a jopo y *Ascochyta fabae* en habas.

***1.5. Hipótesis y Objetivos.***

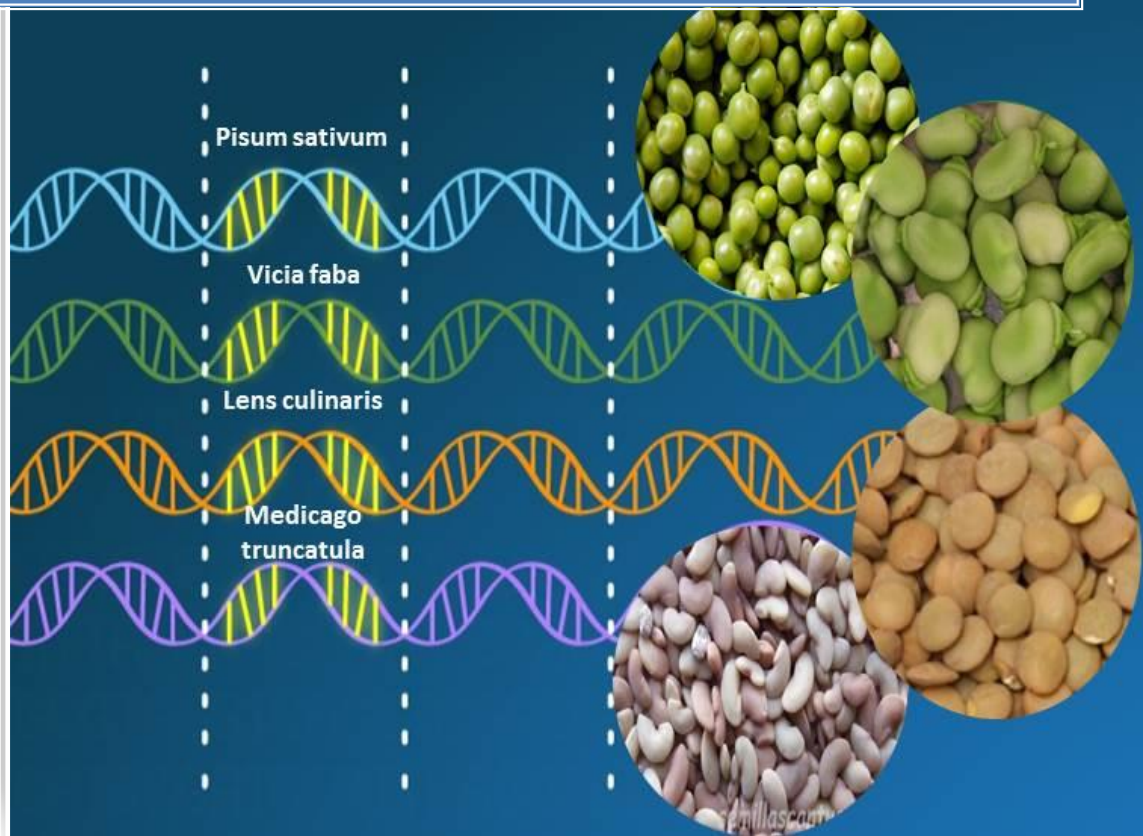
El desarrollo de marcadores moleculares basados en estudios de genómica estructural, comparada y funcional ha abierto una nueva etapa en los programas de mejora tradicionales. El objetivo general de esta tesis ha sido aplicar distintas herramientas genómicas para identificar genes candidatos y posibles rutas metabólicas implicadas en la resistencia a jopo y *Ascochyta* en habas y facilitar el desarrollo de marcadores para la selección asistida por marcadores o MAS.

Para ello se han abordado los siguientes objetivos concretos:

1. Empleo de herramientas de genómica estructural y comparada (sintenia) para mejorar los mapas genéticos disponibles, desarrollar un mapa consenso del cultivo (Anexo I), refinar la posición de QTLs de interés e identificar marcadores ligados, útiles en selección (Capítulo II).
2. Ampliar los datos genómicos del cultivo mediante la caracterización de un nuevo transcriptoma derivado de las líneas parentales 29H y Vf136 sometidas a la infección por el hongo *Ascochyta fabae* (Capítulo III).
3. Genotipado de SNPs presentes en transcritos mediante plataformas de genotipado masivo, mapeo y análisis de QTLs a fin de identificar candidatos que colocalicen o saturen QTLs de resistencia estables (Capítulo IV).

# Capítulo II

**Saturación de los mapas de habas disponibles mediante herramientas de genómica estructural y comparativa (sintenia).**





## II.1. INTRODUCCION.

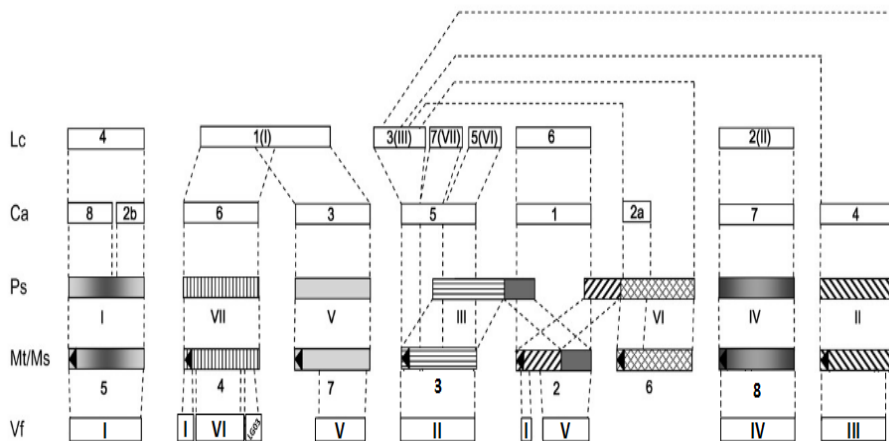
*Vicia faba* es una especie diploide ( $2n = 2x = 12$ ) con uno de los genomas más grandes entre las leguminosas (~ 13.000 Mb) y con más del 85% de ADN repetitivo (Flavell et al. 1974). El gran tamaño de su genoma y el desconocimiento de su secuencia dificultan la saturación de sus mapas genéticos y la identificación de marcadores ligados a caracteres de interés. En consecuencia, el desarrollo de nuevas variedades utilizando nuevas tecnologías como la mejora asistida por marcadores (MAS) no es posible y en la actualidad el desarrollo de nuevas variedades requiere un proceso más largo y complejo que en otros cultivos con más recursos genómicos.

La integración de distintas herramientas y metodologías como el análisis genético mendeliano, la elaboración de mapas genéticos saturados y la secuenciación de genomas, ha revelado la existencia de ordenaciones lineales de genes localizados en distintas regiones cromosómicas, que se encuentran conservadas entre especies más o menos próximas filogenéticamente (Zhu et al. 2005). Esta colinealidad se denomina sintenia y permite identificar genes ortólogos (genes que están presentes en diferentes especies y que proceden de un gen ancestral común por especiación), lo que resulta esencial para la predicción fiable de la función de genes de interés en especies donde el conocimiento a nivel genómico y transcriptómico es mucho más complicado. En este contexto se sitúa lo que se conoce como genómica comparativa que se centra en el estudio de las similitudes y diferencias en la estructura, organización y función de los genes entre genomas de distintas especies; de forma que resulta posible predecir caracteres no observados en una especie sobre la base de los existentes en especies evolutivamente próximas (Zhu et al. 2005). Asimismo, es posible utilizar la información surgida de la secuenciación de especies modelo para la identificación y caracterización de genes en otras especies de interés. Tanto la información posicional obtenida de los mapas genéticos como la información funcional pueden ser utilizadas para la mejora de especies de interés agrícola con menos avances genómicos.

Los primeros mapas comparativos se desarrollaron entre especies de la familia de las solanáceas. Bonierbale et al. (1998) demostraron que marcadores de cDNA estaban en gran medida alineados en 12 de los cromosomas de tomate y patata, mientras que en pimiento se observaba una mayor propensión a la reordenación. Los estudios realizados en la familia de las gramíneas son quizás los más desarrollados.



Empleando RFLPs (restriction fragment length polymorphism) desarrollados en una especie, se crearon mapas genéticos en especies cercanas y se desarrollaron mapas comparativos entre más de 7 especies de gramíneas (Devos y Gale, 1997). Ello demostró que el número y orden de los genes en estas especies está muy conservado. En leguminosas, las herramientas genómicas desarrolladas en la especie modelo *Medicago truncatula* han permitido no sólo la investigación de sus procesos básicos, sino la transferencia de esa información a especies de interés agronómico (Figura 1), utilizando lo que se denomina *sintenia*, es decir, la colinealidad en la organización genómica de diferentes especies (Choi et al. 2004 a; b; Nelson et al. 2006; Phan et al. 2006; 2007; Zhu et al. 2005). Algunos ejemplos incluyen la comparación de *M. truncatula* con alfalfa, guisantes, garbanzos, soja, frijol, mungo, lentejas, altramuces y habas (Choi et al. 2004, 2006; Kaló et al. 2004; Zhu et al. 2005; Nelson et al. 2006; Phan et al. 2006, 2007 b; Aubert et al. 2006; Ellwood et al. 2008; Nayak et al. 2010; Cruz-Izquierdo et al. 2012 y Satovic et al 2013).



**Figura 1.** Representación esquemática de bloques de sintenia entre cromosomas y segmentos cromosómicos de *M. truncatula* (Mt) y leguminosas grano tradicionales; Guisante (*Pisum sativum*, Ps), habas (*Vicia faba*, Vf), garbanzo (*Cicer arietinum*, Ca) y lentejas (*Lens culinaris*, Lc). Las barras que representan las regiones cromosómicas homólogas de *Medicago* y guisante se muestran con la misma intensidad o patrón de grises. Las flechas indican la orientación de los cromosomas (brazo corto - brazo largo) en el caso de *Medicago*. Los bloques en habas, garbanzo y lenteja están representados por barras en blanco. Las barras no reflejan los tamaños relativos de los segmentos cromosómicos o cromosómicos y los puntos de rotura de los cromosomas se indican de modo aproximado. Fuente: Kalo et al. 2011 con modificaciones. La figura integra datos de Choi et al. 2004, Kaló et al. 2004, Zhu et al. 2005, Nelson et al. 2006, Phan et al. 2006, Ellwood et al. 2008, Nayak et al. 2010, Cruz-Izquierdo et al. 2012 y Satovic et al. 2013.

Los primeros marcadores empleados en los mapas de habas, al igual que en otras especies de leguminosas, eran anónimos (ej. RAPDs o Random Amplified Polimorphic DNA) y difíciles de transferir entre laboratorios y/o mapas genéticos. Sin embargo, el empleo posterior de marcadores desarrollados en secuencias expresadas, como los ESTs (Expressed Sequenced Tags), permitió establecer un marco mundial de genómica comparada entre las leguminosas (Frugoli y Harris, 2001). Los ESTs son herramientas esenciales en la búsqueda de un gen de interés. El valor de dichas secuencias fue expuesto por primera vez por Adams et al. (1991) al argumentar que serían la forma más eficiente y rentable de etiquetar la mayoría de los genes de un organismo, mucho antes de que su secuencia completa estuviera disponible. Este tipo de marcador tiene el atractivo adicional de centrarse sólo en las partes más conservadas de los genomas y en secuencias que probablemente tienen una importancia funcional. Por lo tanto, los EST pueden ser utilizados como una fuente para la identificación de genes candidatos y QTLs implicados en el control de caracteres deseados. Dado que las estructuras genómicas están altamente conservadas, el mapeo de genes candidatos de especies modelo podría ayudar a identificar los genes responsables en otros cultivos (Pflieger et al. 2001; Gao et al. 2004).

Pese a las perspectivas descritas, los resultados de distintos estudios han suscitado cuestiones sobre la naturaleza predictiva de los mapas comparativos. Kilian et al. (1997) fueron los primeros en intentar clonar un gen en una especie basándose en la posición detallada y la información sobre su secuencia (microsintenia) en una región homóloga en otro género. Aunque los segmentos cromosómicos en cebada (*Hordeum vulgare*) y arroz (*Oryza sativa*) eran claramente homólogos en la región portadora de gen de resistencia a la roya del tallo en cebada (*Rpg1*), el gen responsable no pudo identificarse en el lugar previsto en arroz. Este estudio mostró las limitaciones de la aplicación de mapas comparativos para el aislamiento de genes, aunque es evidente su utilidad para saturar regiones homólogas en especies relacionadas.

Un creciente número de estudios en leguminosas empleando marcadores ESTs, ha demostrado la extensa sintenia existente entre especies modelo y cultivadas (Zhu et al. 2005; Choi et al. 2004, Kaló et al. 2004, Nelson et al. 2006, Phan et al. 2006, Ellwood et al. 2008, Nayak et al. 2010; Satovic et al. 2013). De hecho, la aplicación de la sintenia a la identificación de genes en especies no modelo está comenzando a dar resultados. Así, Yang et al. (2008) describieron mediante paseos

cromosómico (“map-based cloning”) un gen de resistencia a anthracnosis en alfalfa (*M. sativa*) utilizando información génica de *M. truncatula*. Wang et al. (2008) identificaron un gen candidato responsable de dos mutantes florares de guisante, usando información genómica de otra especie modelo, *L. japonicus*. Estos artículos ilustran cómo la información genética de las especies modelo resulta útil en la identificación y caracterización de genes en especies cultivadas.

Con este trabajo se pretende contribuir al desarrollo del mapa consenso en habas que pueda ser de utilidad para saturar regiones específicas e identificar genes responsables de la variabilidad de caracteres de interés mediante la estrategia de genes candidatos. Los objetivos concretos fueron:

1. Saturar los mapas desarrollados en las poblaciones RILs de dos cruzamientos (29H x Vf136 y Vf6 x Vf136) de habas que segregan para resistencia a *Ascochyta* y jopo con marcadores génicos o ESTs procedentes de diferentes leguminosas.
2. Actualizar y refinar la posición de QTLs que controlan la resistencia a ambos patógenos, identificados en estudios precedentes.
3. Construir un mapa de referencia para integrar la información genómica disponible por el grupo.

## II.2. MATERIALES Y MÉTODOS.

### II.2.1. Material vegetal.

Se emplearon dos poblaciones de líneas recombinantes (RILs) de habas, derivadas de los cruzamientos Vf6 x Vf136 y 29H x Vf136 formadas por 165 y 119 individuos, respectivamente. Los parentales femeninos Vf6 y 29H han sido descritos como resistentes a *Ascochyta fabae* en varios estudios (Tívoli et al. 1998; Maurin et al. 1992; Bond et al. 1994; Sillero et al. 2001) y susceptibles a jopo. El parental masculino Vf136 pertenece a la colección de germoplasma del Centro IFAPA Alameda del Obispo de Córdoba (España). Vf136 es resistente a jopo y susceptible a *Ascochyta* (Román et al. 2003; Díaz-Ruiz et al. 2009, 2010; Ávila et al. 2004; Gutiérrez et al. 2013) y procede de la selección realizada en Córdoba en la descendencia del cruzamiento de Vf1071 x Alameda (Cubero et al. 1992). Vf1071 deriva, a su vez, de la línea egipcia F402 fuente primaria de resistencia a este patógeno.

### II.2.2. Análisis de marcadores moleculares.

#### II.2.2.1. Extracción de ADN genómico.

La extracción de ADN genómico se realizó a partir de hojas tiernas tanto de parentales, como de individuos de las poblaciones RIL, siguiendo la metodología descrita por Torres et al. (1993), y que se detalla a continuación.

Cada muestra se colocó en un mortero con nitrógeno líquido y se trituró con un mazo en la campana extractora. Se añadió 1 ml de tampón de extracción (CTAB) modificado con mercaptoetanol (0.4%) y una pizca de bisulfito sódico. En tubos eppendorf de 1.5 ml se mezclaron 100 µl de cloroformo : octanol (24:1) y el contenido de los morteros. Las muestras se incubaron al baño maría a 65°C durante 30 minutos y se dejaron enfriar a temperatura ambiente durante 15 minutos. A cada tubo se le añadió 1 ml de cloroformo : octanol y se agitó hasta obtener una emulsión homogénea, la cual se centrifugó durante cinco minutos a 7000 revoluciones por minuto (rpm). El sobrenadante se depositó en tubos eppendorf, y se añadió 1 ml de etanol (95%). Las muestras se colocaron en un congelador a -20°C durante 10 minutos. En una nueva tanda de tubos se depositó 1 ml de etanol al 75% (0.2 M Na-Acetato), en los que se lavan los filamentos de ADN recogidos con una pipeta pasteur

a modo de gancho. Tras 5-10 minutos de lavado disolvimos el ADN en tampón TE (EDTA) pH 8. Finalmente, la digestión de ARN se realizó añadiendo 0.7 unidades de RNAsa y manteniendo las muestras a 37°C un mínimo de tres horas. La solución stock fue guardada en un congelador a -80°C.

En un espectrofotómetro (UV-Visible Spectrophotometer, Model Evolution 500) se determinó la concentración de ADN y se realizó la dilución del ADN stock hasta concentraciones de trabajo de 30 ng/μl (29H x Vf136) y 50 ng/μl (Vf6 x Vf136).

#### **II.2.2.2. Análisis de ESTs (Expressed Sequence Tags) y UniTags.**

Se analizaron 116 marcadores ESTs procedentes de guisante (*Pisum sativum*), *M. truncatula*, lenteja (*Lens culinaris*), garbanzo (*Cicer arietinum*) y habas (*V. faba*). El número y origen de cada uno de ellos se describe a continuación:

- 22 marcadores genotipados en la población Vf6 x Vf27 de habas (Ellwood et al. 2008; Cruz-Izquierdo et al. 2012) con el objetivo de disponer de marcadores comunes en los distintos fondos genéticos implicados en los cruzamientos.
- 12 marcadores relacionados con la resistencia a *A. fabae* en guisante (Prioul et al. 2007).
- 23 marcadores de *M. truncatula* (Choi et al. 2004) seleccionados por situarse en regiones homólogas a los cromosomas II, III, V y VI de habas portadores de QTLs importantes y estables que controlan la resistencia a *A. fabae* (II y III), jopo (II y VI) y la precocidad en floración (V).
- 18 UniTags diferencialmente expresados tras el análisis con SuperSage de la interacción *Ascochyta*-lenteja (García et al 2012; 2016).
- 24 UniTags diferencialmente expresados tras el análisis con SuperSage de la interacción *Ascochyta*-habas validados por qPCR (Madrid et al. 2013).
- 14 marcadores asociados al QTL3 y GL2, de garbanzo asociados con la resistencia a *Ascochyta rabiei*.
- 3 marcadores de garbanzo ligados al carácter fecha de floración.

Para cada cebador se probó su amplificación y posible polimorfismo entre los parentales (29H, Vf6 y Vf136) y cinco individuos de las dos poblaciones RIL (Vf6 x Vf136 y 29H x Vf136). En caso de mostrar polimorfismo el marcador se genotipó en el resto de la población.

La mezcla de reacción para la amplificación del ADN de los diferentes cebadores se detallada en la Tabla 1.

**Tabla 1.** Mezcla de reacción de PCR para los diferentes marcadores.

Mezcla de Reacción	$\mu\text{l}^a$	$\mu\text{l}^b$	$\mu\text{l}^c$
ADN	2	2	2
Buffer (10x)	2	1	1
MgCl <sub>2</sub> (25 mM)	0,8	0,4	0,5
dNTP (10 mM)	1,6	0,2	0,4
Cebador Fw (10 $\mu\text{M}$ )	0,6	0,2	0,4
Cebador Rev (10 $\mu\text{M}$ )	0,6	0,2	0,4
Taq polimerasa (5 u/ $\mu\text{l}$ )	0,2	0,04	0,05
Agua estéril	14,2	5,96	5,25
Volumen final	22	10	10

<sup>(a)</sup> ESTs Ellwood et al. 2008, Prioul et al. 2007, Choi et al. 2004.

<sup>(b)</sup> UniTags de lenteja.

<sup>(c)</sup> UniTags de habas (Madrid et al. 2013) y marcadores de garbanzo.

Las condiciones de amplificación se optimizaron probando distintos parámetros hasta obtener amplicones claros, variando la temperatura de hibridación en función del cebador utilizado (Tablas 2-5). La reacción de amplificación se realizó en termocicladores PCR Eppendorf AG® y BIO-RAD T100. Todos los marcadores se visualizaron en geles de agarosa al 2%, con una mezcla de NUSIEVE y SEAKEM en buffer 1xTBE. La electroforesis se inició a 80 V y transcurridos 5 minutos, los geles se cubrieron de tampón y se incrementó el voltaje a 150 V. La tinción del gel se realizó en 250 ml de agua destilada con 3  $\mu\text{l}$  de bromuro de etídio. Pasados 20 minutos de tinción, se observaron las bandas con luz ultravioleta y se fotografiaron los geles.

**Tabla 2.** Condiciones de amplificación de PCR para marcadores ESTs.

Fase	Temperatura (°C)	Tiempo (min)	Ciclos
Inicio	95	5	
Desnaturalización	95	1	} X 40
Hibridación*	50-60	1	
Extensión	72	2	
Extensión final	72	8	
Enfriamiento	4	∞	

\* La temperatura de hibridación se cambió de acuerdo al cebador utilizado.

**Tabla 3.** Condiciones de amplificación de PCR para marcadores UniTags de lenteja.

Fase	Temperatura (°C)	Tiempo (min)	Ciclos
Inicio	94	5'	
Desnaturalización	94	30''	} X 35
Hibridación*	48-55	30''	
Extensión	72	2'	
Extensión final	72	10	
Enfriamiento	4	α	

\* La temperatura de hibridación se cambió de acuerdo al cebador utilizado

**Tabla 4.** Condiciones de amplificación de PCR para marcadores UniTags de habas.

Fase	Temperatura (°C)	Tiempo (min)	Ciclos
Inicio	95	5'	
Desnaturalización	95	30''	} X 40
Hibridación	55	30''	
Extensión	72	30''	
Extensión final	72	10	
Enfriamiento	4	α	

**Tabla 5.** Condiciones de amplificación de PCR para los marcadores de garbanzo (*ein3*, *avr* y *agr*).

Fase	Temperatura (°C)	Tiempo (min)	Ciclos
Inicio	95	1'	
Desnaturalización	95	30''	} X 35
Hibridación*	48-62	30''	
Extensión	72	1'	
Extensión final	72	7'	
Enfriamiento	4	α	

\* La temperatura de hibridación se cambió de acuerdo al cebador utilizado

Las características de todos los marcadores analizados se detallan en la Tabla 6 (a-e). Dependiendo del tipo de polimorfismo observado entre los parentales, se emplearon distintos métodos para genotipar su correspondiente población. Cuando las parejas de cebadores revelaban diferencias de peso molecular (>10pb) el polimorfismo ALP o “Amplicon Length Polymorphism”, se genotipo directamente en geles de agarosa al 2%.

Los productos que resultaban monomórficos, se digirieron inicialmente con la enzima que estuviera descripta para dicho marcador en cada uno de sus estudios previos, y si aun así, el marcador resultaba monomórfico se digería con una batería de 10 enzimas de restricción que reconocen secuencias de 4 a 5 pb (Tabla 7). La digestión se realizó con 2 unidades (0,2 µl) de cada enzima en un volumen final de 25 µl. Los otros componentes fueron 12 µl de agua destilada estéril, 2,5 µl de buffer específico para cada enzima y 5-10 µl de producto amplificado de PCR. La incubación se realizó durante 12 horas en una estufa a 37°C. El polimorfismo se observó cuando la banda inicial origino dos bandas de menor peso molecular, obteniendo un marcador conocido como CAP (Cleaved Amplified Polymorphism).



**Tabla 6 (a).** Marcadores moleculares analizados en las poblaciones RILs de los cruzamientos Vf6 x Vf136 y 29H x Vf136 de *V. faba*.

Marcador	Referencia	Origen	Forward primer sequence	Reverse primer sequence
Anmts 37	Ellwood et al. 2008	<i>V. faba</i>	GATCCAGCATTCTGTGTCGCC	CATCCACTCCAAGTTACGTGGATC
cgP137FSNP	Ellwood et al. 2008	<i>V. faba</i>	TGTTAACCTCAGTTACTGCAACAGA	AGGATGCAATCAAGCATATATCTTGA
CP450	Ellwood et al. 2008	<i>V. faba</i>	AGTGTGAGATCAATGGTTATGTGATC	CATCATCACCTTTCAATATTTGTC
CPCB2	Ellwood et al. 2008	<i>V. faba</i>	AGAAAGAGTGAAGTCTGTGGATCTACATC	GGATGAACAGCCACACACCTAATGTAATC
Mtmt_Gen_00733_03_3	Ellwood et al. 2008	<i>V. faba</i>	TGGGAAGAAAAGTGCTGCTT	GCCTCGTTAATTGGTCTTCG
GLIP 99	Ellwood et al. 2008	<i>V. faba</i>	ACAAGCAACCATGTTGGACA	GAGCAGATCACCTCCTGGAC
Mtmt_Gen_00510_01_1	Ellwood et al. 2008	<i>V. faba</i>	GTCAAGCTTGGGTGAATGGT	TTGGATGCTTTCAACCATGA
Mtmt_Gen_01017_03_1	Ellwood et al. 2008	<i>V. faba</i>	ATTGACTGCTGTGGTGGTGA	CTTGCCCAATGGAAAGAAAA
Mtmt_Gen_01115_02_1	Ellwood et al. 2008	<i>V. faba</i>	ATGAGCAAAGAAAGCCAAA	TACCAGCACGTGGTTGATGT
Mtmt_Gen_01965_03_1	Ellwood et al. 2008	<i>V. faba</i>	TTGTTGCAAGCAATGGAGAG	AAGAAGCCAAGCAAGTTGA
GLPSNP	Ellwood et al. 2008	<i>V. faba</i>	GACTCAACACACTTGGTATATCTTTGGCTC	TTATCCACTTGAAAGGCTTTGGTGAGAAC
HYPTE3SNP	Ellwood et al. 2008	<i>V. faba</i>	TCGTCTCATGGTGAATCGTGATGGT	TTCTCCTTTAAACAAGCAAATTGGA
Lg 33	Ellwood et al. 2008	<i>V. faba</i>	GCTGATATGAGGCGCGAAAGTG	GCCCCGAGAGGGCAGAGG
Lg 34	Ellwood et al. 2008	<i>V. faba</i>	TGCCCTCTGCGGGGCAGG	TTCTCCTCAGCCTCCCGTCG
Lup 36	Ellwood et al. 2008	<i>V. faba</i>	CTGATATCCATTCACCTCAAG	AACACCTGGAAGTCCAACATA
PESR1	Ellwood et al. 2008	<i>V. faba</i>	ACAAGCCTAGAAAATCCATTC	CGATATGAGTGACTCAATTTGTGGTC
RBPC_0SNP	Ellwood et al. 2008	<i>V. faba</i>	CCCCATCATGATGAGTGTGGAGA	TTGAGAAATCGTTACCGGTGACAATGATG
REPSNP	Ellwood et al. 2008	<i>V. faba</i>	GCCAAATTCCTCGTCTACTTC	GCTAAAAGATTCATCTCCGGTATCCA
UNK 28	Ellwood et al. 2008	<i>V. faba</i>	GTAGATTCAATTCAAGTGCAGAAAGGCTTTC	TACCAACCCATCAATGTTCTTCTAACTTTG
PsPRP 4A	Prioul et al. 2007	<i>P. sativum</i>	ACATACAATAACTACAACCC	TCTCACTGTTACCTGAGCTCC
DRR 230 b	Prioul et al. 2007	<i>P. sativum</i>	ATGGAGAAGAAAATCAMTAGC	GTGCGCTTTGTTYYTRCAGTG
Hmm6	Prioul et al. 2007	<i>P. sativum</i>	GGAGATATGTTCAAGTCTGT	AGCAGTTCTTTCTAATGTACTC
IJB174	Prioul et al. 2007	<i>P. sativum</i>	GACGACTCTTGCTACTGC	ATGAAGTAGATGAGGTAT
IJB91	Prioul et al. 2007	<i>P. sativum</i>	GACGACTCTTGCTACTGC	TGAGGGAGACGGATCGGG
Peachi 21	Prioul et al. 2007	<i>P. sativum</i>	CTTTCCCAACTTCGCCAATAA	TAGTCGAGAATGAAATGGTCTGAGA

**Tabla 6 (b).** Marcadores moleculares analizados en las poblaciones RILs de los cruzamientos Vf6 x Vf136 y 29H x Vf136 de *V. faba*.

Marcador	Referencia	Origen	Forward primer sequence	Reverse primer sequence
PsDof 1	Prioul et al. 2007	<i>P. sativum</i>	AAGCCTCAGCCGGAACAAGC	TCCATTTCCAAGGAATAAACC
PsMnSOD	Prioul et al. 2007	<i>P. sativum</i>	GCTCTCGAACAGCTTCACGAT	CTTTGGTGGTTCACCACCTCC
RGA2.97	Prioul et al. 2007	<i>P. sativum</i>	TTAGGAACGAGGGAGCATAG	ATTCTTTAGCCATTTACGC
RGA2P	Prioul et al. 2007	<i>P. sativum</i>	GCTACTGCTTTGTATGCTA	TCTTTTGTCAAGTATTAGTAA
RGA3P	Prioul et al. 2007	<i>P. sativum</i>	AGACGACTCTTGCTACTG	GAGAGGAACTTGGTAAAC
RGA-G3A	Prioul et al. 2007	<i>P. sativum</i>	GTATGCTAGAATCTCTAATCA	TGTC AAGTATATGTAACCACTC
Pis_Gen_6_3_1	Cruz-Izquierdo et al. 2012	<i>P. sativum</i>	CTCCCTGCACCTTCTCTGTC	ATCAAAGTTGCTCCGACCG
psat_Est_00190_01_1	Cruz-Izquierdo et al. 2012	<i>P. sativum</i>	GGGAACCTCTTGATCCCAAT	CAGGTTGTTGAAGCTCTGA
mtmt_GEN_01114_05_1	Cruz-Izquierdo et al. 2012	<i>M. truncatula</i>	CTGGAGGGATGATGACTGGT	TTGTAGAGCCATTGTCACGC
CAK	Choi et al. 2004	<i>M. truncatula</i>	TTCAACCCCTCTGCGAACC	CATCTATAGCAATTGCTGTTGTCATCT
CysPr 1	Choi et al. 2004	<i>M. truncatula</i>	GAGAATTC AAAGAAGAAATTAAGACAAAGA	GAAGAATTCATGGGGAGCAAAGT
DNABP	Choi et al. 2004	<i>M. truncatula</i>	CCCTATGAGCTTGGGTTTGTCT	CTCATGGCATACTGTGTTTCAGC
DSI	Choi et al. 2004	<i>M. truncatula</i>	CCAAGACATCTTTGGTTTCATCC	ACTGCAGAATCACTTGCCGAGTT
ENOLB	Choi et al. 2004	<i>M. truncatula</i>	TTCCATCAAGGCCCGTCAGA	TTGCACCAACCCCATTCATT
EPS	Choi et al. 2004	<i>M. truncatula</i>	GCTGTTGTGGAAGGCAGTGG	ACGACATACGGAACAGAAATCAGT
EST 718	Choi et al. 2004	<i>M. truncatula</i>	CGGCGGCATGCTTAGTGG	TCTTGG AATGCCTTTGAATGAATA
EST 948	Choi et al. 2004	<i>M. truncatula</i>	GCAGGGGTTTCGCTCCAGTG	AACTTAATGAATGATTGGAAGGTTTAGCG
HYPTE_116	Choi et al. 2004	<i>M. truncatula</i>	AACACAGAGGTAGCGTTTGGTTTAT	TCGGGTCAAAGATCTCGTTCAA
KCOAT	Choi et al. 2004	<i>M. truncatula</i>	TGCTACTGCGGGTGGTAGATTTA	CTCCAGCACCATCACTCACCT
MPP	Choi et al. 2004	<i>M. truncatula</i>	TCCCCGAAACAATCCTCATCTG	GCAAATGTGTAGCCCCAAAAGTTA
PCT	Choi et al. 2004	<i>M. truncatula</i>	TTGGCAAAAACGATAAACCTGT	CACGGCACATCTGGAATAACTT
PRTS	Choi et al. 2004	<i>M. truncatula</i>	CATAGCTACTTGATCTGAAAACCTGACA	TGGTGA ACTTCACTACCATTACAACC
RNAH	Choi et al. 2004	<i>M. truncatula</i>	GCTTCCACCAGCTGATACACG	TTAGCCCTAGCAAGAATGTCACTG
Te001	Choi et al. 2004	<i>M. truncatula</i>	CGGCGCCGGAGATTACTG	AATCACAAACCCACCCAACATCTG

**Tabla 6 (c).** Marcadores moleculares analizados en las poblaciones RILs de los cruzamientos Vf6 x Vf136 y 29H x Vf136 de *V. faba*.

<b>Marcador</b>	<b>Referencia</b>	<b>Origen</b>	<b>Forward primer sequence</b>	<b>Reverse primer sequence</b>
TE-011	Choi et al. 2004	<i>M. truncatula</i>	GGAGAGAAACCGACTGAAGAAACA	CAAGAAGAAGCCCTAGTCCCTCCATT
Te013	Choi et al. 2004	<i>M. truncatula</i>	TCCGCATGTACGAGTTCAAGATAAG	AAATCACAAACCCACCCAACATC
TE-016	Choi et al. 2004	<i>M. truncatula</i>	TCCCCAGGCCTTACAAGATGATTAT	AAACACTCCCACGTCGCTACTAAG
TUP	Choi et al. 2004	<i>M. truncatula</i>	GAATGGGATGCTATGGGAAGTG	TGGATCAGTGGCACCATCTTTAT
UDPGD	Choi et al. 2004	<i>M. truncatula</i>	CAAAAGCGTTTCATCACTCATCTCT	ATCGTCAAGGCCAGGTTTCATAG
UNK16	Choi et al. 2004	<i>M. truncatula</i>	CCTTCCAATATCCCTCCACAT	GAAGAAAATGATGAAAAGCCAAAAG
UNK21	Choi et al. 2004	<i>M. truncatula</i>	TCGCCTCCATGTCCACCTC	CGGCCTTGCTAAATCAGTCAG
UNK3	Choi et al. 2004	<i>M. truncatula</i>	CACCGGAAATTCAACAGCAAC	GACCTAGGCAACACAACCTCCATTA
Ein3Fw2	Madrid et al. 2014	<i>C. arietinum</i>	TGCTCAACACCCTCTGTCTCTCA	TGGCCGTCTTCGGGAACCTCCA
Ein3Fw3	Madrid et al. 2014	<i>C. arietinum</i>	TTCAGTGCATAAAGAGGTCCGCC	TGGCCGTCTTCGGGAACCTCCA
Ca2_Avr9	Madrid et al. 2014	<i>C. arietinum</i>	GACTCGCCACTCACTGACTCATCC	CGCCGTCAACTGATCCACCGTC
Ca2_Ago	Madrid et al. 2014	<i>C. arietinum</i>	GGCGTCTGTGAAGGGAAGGG	CAATGGGCACGTAAGTTGGCCT
P00023	García et al. 2012; 2016.	<i>L. culinaris</i>	TGAAGCACCATCAGAAACCA	TGGCACAGATGTGTCCTCTT
P00042	García et al. 2012; 2016.	<i>L. culinaris</i>	TTGGCTCCACCCTATGATTC	CGTGGCGTAACCCATAAGTG
P00043	García et al. 2012; 2016.	<i>L. culinaris</i>	GGCTCGTCTTTCACTAACG	TCTGGAACCCATGAAACCTT
P00046	García et al. 2012; 2016.	<i>L. culinaris</i>	TATCTTCGTGCAACCGTTA	TGCCCCGCTCAAAATAGTTA
P00053	García et al. 2012; 2016.	<i>L. culinaris</i>	TCATGGTGAAGACTGGCATT	TTGTCGATATCTCCACCAACA
P00073	García et al. 2012; 2016.	<i>L. culinaris</i>	GTTAGTTGCTTCCGGCTCAG	TGCGGTAATCAAAATGGACAA
P00075	García et al. 2012; 2016.	<i>L. culinaris</i>	CCCTCTCCACCCTTCTAC	GTTGGATCAGGTTTGCCAGT
P00084	García et al. 2012; 2016.	<i>L. culinaris</i>	TTGAATCAGCTTCCGCTTTT	AAGCAGCAAAGACCCCTAGCA
P00087	García et al. 2012; 2016.	<i>L. culinaris</i>	AGCTCTTGATTTGGCTTCCA	AGCAGCAGGTGCAGTTTCTT
P00088	García et al. 2012; 2016.	<i>L. culinaris</i>	TTGGATGCTGCTAAGACTTTTC	TTTCTTTTACATGTTTCGCTATCAA
P00098	García et al. 2012; 2016.	<i>L. culinaris</i>	AAGATCGTGAAAGGGCTCAA	GCTTCCACCAGCATTATTCC
P00179	García et al. 2012; 2016.	<i>L. culinaris</i>	CGTTGAATTCGCAACTCGTA	GGTTTCGCAGTTCTTTCAGC
P00260	García et al. 2012; 2016.	<i>L. culinaris</i>	TATGGCTATGCTGCTCCTGA	GTTTGGCCATTTCGATTA

**Tabla 6 (d).** Marcadores moleculares analizados en las poblaciones RILs de los cruzamientos Vf6 x Vf136 y 29H x Vf136 de *V. faba*.

<b>Marcador</b>	<b>Referencia</b>	<b>Origen</b>	<b>Forward primer sequence</b>	<b>Reverse primer sequence</b>
P51982	García et al. 2012; 2016.	<i>L. culinaris</i>	ATAATGCCGGTAATGCTGCT	AGACGAGCATCAGCTGGAAC
P52010	García et al. 2012; 2016.	<i>L. culinaris</i>	CCTCTTGGACTGCTCGTCTC	CCATTGGATGCTCTTCATCA
P52014	García et al. 2012; 2016.	<i>L. culinaris</i>	CATGCAGCAAGGTTTGACAT	GCCTGAGCCGTCAAAGTT
P52052	García et al. 2012; 2016.	<i>L. culinaris</i>	CATGGCTAACATCCAGAAGAAA	CCACTTTTCTTGAGGCCAGA
P52059	García et al. 2012; 2016.	<i>L. culinaris</i>	ATGACAGCAACGTCGTTTGA	AAAACTTTGGTTGAACAAGAACG
ERTF	Madrid et al. 2013	<i>V. faba</i>	GAGTCCAACAAGGGTGGGGTCCA	ATGGATGCGCCATGGCTAAAATGTC
HSP	Madrid et al. 2013	<i>V. faba</i>	TGGAGTTCAAGCGAATGGCAATGGAG	AGAACACCATCTCTAACCAAGTGCCT
Leu-TF	Madrid et al. 2013	<i>V. faba</i>	AGCCTACAACCATGACACAACCTCCTT	GCGGCTGCTGATCAGGCCAT
P450	Madrid et al. 2013	<i>V. faba</i>	GCCTTTTGGTTTTGGAGGGAGAGCA	GAGGAAGTCCATGACTAGGTCTAAGGG
WRKY58	Madrid et al. 2013	<i>V. faba</i>	GAGGAGTGTCTCGCGGACGA	GCTGAAATCAGAAAACACCCAAAACC
b-ZIP	Madrid et al. 2013	<i>V. faba</i>	ATCCCCACCGCCGCTTCT	ACACCCACACGTGTAACCCCTCCA
Tag 1	Madrid et al. 2013	<i>V. faba</i>	AGGCTGCAAGTACCAGAGAACACTA	ACTCAGAAACTGTGTGTAGAGTGTGGC
bcl2	Madrid et al. 2013	<i>V. faba</i>	ACCCGCCCCGCAAGTCTATCA	AGGACAAATCAAAGCTAGTGCTAGTGG
SerK	Madrid et al. 2013	<i>V. faba</i>	TGTCACCCAATCCACTTGCCAATAGC	TGACGGAGCGTCATCAGTGTTT
AKR	Madrid et al. 2013	<i>V. faba</i>	CAGCTGAAGTCATGGCAGGCA	AACAAGTAACTGGTGTGTTAGCTGCAT
ATP-Dep	Madrid et al. 2013	<i>V. faba</i>	GCTTCATCCGCTATGTTCTCCCC	ATGATAGTCTACGGTATGAGGCACCTG
SerP	Madrid et al. 2013	<i>V. faba</i>	TGTTTCGCGGTTACACACTCC	TGACTGCACCTCCTCTTTTCAG
U-box	Madrid et al. 2013	<i>V. faba</i>	GAAGGCCAACAGTATTCTCGAGAGGC	TCTCACTACTAAGCACTACTAAGCA
ERTF-Rev2	Madrid et al. 2013	<i>V. faba</i>	GAGTCCAACAAGGGTGGGGTCCA	GCAGAAAGGGGATTCTCTGTGG
AKR-Fw2	Madrid et al. 2013	<i>V. faba</i>	CAGCTGAAGTCATGGCAGGCA	GCGGCGAACACTCTCAGGA
HSP-Fw2	Madrid et al. 2013	<i>V. faba</i>	TGGAGTTCAAGCGAATGGCAATGGAG	GGCAAAGCAGCATCAGTTGCCA
Bcl2_Rev2	Madrid et al. 2013	<i>V. faba</i>	ACCCGCCCCGCAAGTCTATCA	TGGGAGATGAGGCCAGGTGGA
bZip-Fw2	Madrid et al. 2013	<i>V. faba</i>	ATCCCCACCGCCGCTTCT	AGCCCCAAATTCTGGCCTCCG
qPCR1	Madrid et al. 2013	<i>V. faba</i>	TGGCTGGGATTGCAATATCAACAGGAC	TCGTATTGGAGCTTCGCAAGTTTCTCT
qPCR-3	Madrid et al. 2013	<i>V. faba</i>	GCTGACGCCATGGTCACTTGGG	AGCAGCAACCTCAGAAGCTTTGGT

**Tabla 6 (e).** Marcadores moleculares analizados en las poblaciones RILs de los cruzamientos Vf6 x Vf136 y 29H x Vf136 de *V. faba*.

<b>Marcador</b>	<b>Referencia</b>	<b>Origen</b>	<b>Forward primer sequence</b>	<b>Reverse primer sequence</b>
qPCR-4	Madrid et al. 2013	<i>V. faba</i>	GGGACTCCTAAGAGGCCTAATGGACC	TGTCTCTCAATTCAGGACCTGGCT
qPCR-6	Madrid et al. 2013	<i>V. faba</i>	GAGGCTTATGGCTGTGGCTGCT	CCAGCCCATGCAGACACCAATCC
qPCR-7	Madrid et al. 2013	<i>V. faba</i>	GGCCAAGATCATGGAGTGAACCTGC	AGGTCATTGTGATTGGGTTTGAGAAGC
qPCR-8/SSA_5	Madrid et al. 2013	<i>V. faba</i>	TGTTACAGCTTGTGTAAGGTGCATCGC	TCCTTCACATGTCTGGACGTCGCT
Ca2_ABF5	Madrid et al. 2014	<i>C. arietinum</i>	GGGAGCATGAATCTCGACGA	ACCAGTTCATTATGGTAAGCCTG
Ca2_Ribncl	Madrid et al. 2014	<i>C. arietinum</i>	GCTTCGATCCTCCGACCCCGA	CGCCAGGACCAGGGTGAGGT
Ca2_Ein3	Madrid et al. 2014	<i>C. arietinum</i>	GGTACACTTGGGGAAGGGGA	AAGGCATTTGGACTGCCTCA
Ca2_Rk	Madrid et al. 2014	<i>C. arietinum</i>	GCCGGAAGTCAAGCGGG	GGCCGTGCCGATGGAGAAGG
Ca2_ERTF	Madrid et al. 2014	<i>C. arietinum</i>	TCCACACCCCTTCTCCTCCA	CCCCCTCCACTTTGCAATTCTGAT
Ca2_ZnF	Madrid et al. 2014	<i>C. arietinum</i>	GTTTCAAGGTGAAGCCGTGC	CCAGACCGAGCAGTTCATC
Ca2_SerK	Madrid et al. 2014	<i>C. arietinum</i>	TGGCTCCGGAAGTCATTCAG	TTC AATTGCAGACACGCCAA
Ca2_Transf	Madrid et al. 2014	<i>C. arietinum</i>	GTGTTAAAGCCAGGAGCATGGT	GTGCAGGCTTCTGGAATTTCCCAT
Ca2_SAP1	Madrid et al. 2014	<i>C. arietinum</i>	GTGGAAGCATTGCTGGACCCC	GGAGACTGCTGTGCTCCATGACC
CaLF1	T Millán (No publicado)	<i>C. arietinum</i>	AAGACAAGGCGGCTATGAGA	GTGATTCTGGCCCAAGTGAT
CaLE1	T Millán (No publicado)	<i>C. arietinum</i>	AACCCGCTTATGTCGTTTTG	GGTCCCTTTTGGACATTTT
Ca2Calm	Madrid et al. 2014	<i>C. arietinum</i>	TGGCCGATCAGCTCACCGAT	GGTTCTGTCCAAGCGACCGCA
CaDET2	T Millán (No publicado)	<i>C. arietinum</i>	GTGATTCTGGCCCAAGTGAT	CCAAATGTGGCATCTGTTGT

**Tabla 7.** Relación de enzimas utilizadas para desarrollar marcadores CAPs. Secuencias diana y temperatura (°C) de corte de cada una de ellas.

Enzima de Restricción	Secuencia diana	Temperatura de Corte
Alu I	AG <sup>↓</sup> CT	37°C
Fsp B I (Mae I)	C <sup>↓</sup> TAG	37°C
Hind III	A <sup>↓</sup> AGCT <sup>↓</sup> T	37°C
Hinf I	G <sup>↓</sup> ANT <sup>↓</sup> C	37°C
Hpa II	C <sup>↓</sup> CG <sup>↓</sup> G	37°C
Mbo I	<sup>↓</sup> GATC <sup>↓</sup>	37°C
Rsa I	GT <sup>↓</sup> AC	37°C
Ssi I (Aci I)	C <sup>↓</sup> CG <sup>↓</sup> C	37°C
Tas I (TspEI)	<sup>↓</sup> AATT <sup>↓</sup>	37°C
Pfe I	G <sup>↓</sup> AWTC	37°C

### II.2.3. Construcción del mapa de ligamiento en la población RIL y análisis de QTLs.

Para todos los marcadores genotipados en las dos poblaciones RIL la banda heredada del parental materno se identificó con la letra A (alelo a) y la del paterno con la B (alelo b). El ajuste de los marcadores analizados a la segregación esperada en la población (1:1) se comprobó mediante la prueba  $\chi^2$  ( $p > 0.05$ ) con  $n-1$  grados de libertad, utilizando la fórmula:

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i}$$

- Dónde:  $\chi^2$  = test de ajuste  
*i* = clase genotípica *i*  
*n* = número total de las clases  
*o<sub>i</sub>* = valores observados  
*e<sub>i</sub>* = valores esperados

El mapa de ligamiento se elaboró añadiendo los nuevos marcadores a la matriz de datos disponible de estudios previos en ambas poblaciones (Díaz-Ruiz et al. 2009 a, b, 2010; Gutiérrez et al. 2013; Atienza et al. 2016; Satovic et al. 2013) utilizando el programa JoinMap® versión 4.1 (Van Ooijen, 2006) a un LOD de 4 y 5 en la población 29HxVf136 y Vf6xVf136 respectivamente. Las fracciones de recombinación se convirtieron en centiMorgans (cM) utilizando la función de mapeo

de Kosambi (1994). Los grupos de ligamiento se orientaron siempre que fue posible en base a los marcadores comunes de mapas de ligamiento publicados previamente (Díaz-Ruiz et al. 2009 a, b, 2010; Gutiérrez et al 2013; Atienza et al 2016; Satovic et al 2013).

Para llevar a cabo el análisis de QTLs se utilizaron evaluaciones previas realizadas tanto en campo como en condiciones controladas para *Orobanche* y *Ascochyta*. La población RIL F<sub>6</sub> Vf6xVf136 fue evaluada para resistencia a *O. crenata* y *O. foetida* por Díaz-Ruiz et al. (2010; 2009b) durante dos campañas (2002/2003 y 2003/2004) en campos infestado de Córdoba y Mengibar y para *O. foetida* en Beja y Loubna (Túnez). La evaluación de la resistencia a *A. fabae* se realizó a los 15 días de la inoculación, bajo condiciones controladas en cámara de crecimiento con el aislado CO95-01 originado en Córdoba (Díaz-Ruiz et al. 2009 a). La enfermedad se evaluó como porcentaje de área sintomática en hojas (DSL) y tallos (DSS).

La otra población RIL 29HxVf136 fue evaluada para la resistencia a *O. crenata* durante tres campañas (2006/2007, 2007/2008 y 2008/2009) en ensayos de campo en Córdoba y Kafr El-Sheikh (Egipto) durante 2007/2008 (Gutiérrez et al. 2013). Además, se evaluó la resistencia para *O. foetida* en Beja (Túnez) durante dos campañas (2006/2007 y 2007/2008). En esta población también se evaluó la resistencia a *A. fabae* en plántulas desarrolladas en cámara de cultivo e inoculadas con las cepas CO99 y LO98, recogidas en Córdoba y Logroño (España), respectivamente (Ávila et al. 2004; Atienza et al. 2016). La enfermedad se evaluó 15 días después de la inoculación en hojas (DSL) y tallos (DSS) basándose en el porcentaje de área sintomático. Posteriormente se evaluó la resistencia en la planta adulta en hojas (DSL), tallos (DSS) y vainas (DSP) en ensayos de campo durante la campaña 2005-06 en Córdoba y con un aislado local.

El análisis de QTL se realizó con el programa MapQTL® V5 (Van Ooijen 2006) empleando los datos fenotípicos disponibles para ambas poblaciones (Díaz-Ruiz et al. 2009 a y b; 2010; Gutiérrez et al. 2013 y Atienza et al. 2016) y descritos anteriormente. Para identificar QTL relacionados con la resistencia a estas dos enfermedades se utilizó el método de intervalos simple (IM) en todos los grupos de ligamiento. El valor de LOD para aceptar la existencia de un QTL fue estimado por el análisis de permutaciones (Churchill y Doerge, 1994) utilizando 1000 permutaciones, considerándose también QTL detectados en estudios previos cuya diferencia con el

LOD crítico fuera  $< 0.5$ . La posición del QTL fue estimada con la localización del máximo valor de LOD con un intervalo de confianza LOD-2 (Li 2011).

### **II.3. RESULTADOS Y DISCUSIÓN.**

#### **II.3.1. Marcadores polimórficos en *V. faba*.**

El objetivo de este trabajo, es saturar los mapas genéticos desarrollados en estudios previos en dos poblaciones RILs de habas (Vf6 x Vf136 y 29H x Vf136) con el mayor número de marcadores procedentes de diferentes leguminosas (*P. sativum*, *M. truncatula*, *L. culinaris*, *C. arietinum* y *V. faba*) a fin de refinar la posición de QTLs que controlan la resistencia a *Ascochyta* y jopo causantes de importantes pérdidas económicas en el cultivo.

Un total de 116 marcadores moleculares fueron seleccionados para ser genotipados en ambas poblaciones. Los resultados se detallan en la Tabla 8. De los 116 marcadores, sólo 27 (23%) mostraron buen polimorfismo y pudieron genotiparse en una o ambas poblaciones RILs. Once de estos marcadores mostraron polimorfismo directo tipo ALP (Amplified Length Polymorphism), uno mostró presencia/ausencia de banda y 15 fueron sometidos a digestiones enzimáticas con el fin de desarrollar marcadores tipo CAPs (Cleavage Amplified Polymorphism) (Tabla 9). El resto de marcadores no pudieron incluirse en el análisis por falta de amplificación, por ser marcadores monomórficos tras el corte con enzimas de restricción o por no mostrar un patrón de bandas repetible y fiable en las poblaciones.

De estos 27 marcadores, 12 se genotiparon en la población 29H x Vf136 y 22 en la Vf6 x Vf136. Siete fueron polimórficos en ambas poblaciones, cinco sólo en el cruzamiento 29H x Vf136 y quince en el cruzamiento Vf6 x Vf136 (Tabla 9). De los siete nuevos marcadores comunes solo cuatro se ligaron a ambos mapas. Dichos marcadores, junto a dos genotipados previamente (RGA3P y PisGen\_6\_3\_1) en la población 29H x Vf136 y que en este estudio lo han sido en la Vf6 x Vf136, fueron puntos de anclaje para comparar ambos mapas.



**Tabla 8.** Resultado de los 116 marcadores analizados en las poblaciones RILs Vf6 x Vf136 y 29H x Vf136.

Marcador	Vf6xVf136	29hxVf136	Marcador	Vf6xVf136	29hxVf136
cgP137FSNP	M	X <sup>(a)</sup>	P00098	X	X
Mtmt_Gen_01017_03_1	X <sup>(a)</sup>	M	P00179	M-I	M-I
Mtmt_Gen_01965_03_1	X <sup>(a)</sup>	M	P00260	X	M-ER
GLPSNP	M	M	P51982	M-ER	M-ER
REPSNP	X <sup>(a)</sup>	M	P52010	M-ER	M-ER
Mtmt_Gen_00510_01_1	M	X <sup>(a)</sup>	P52014	M-ER	M-ER
Mtmt_Gen_01115_02_1	M	X <sup>(a)</sup>	P52052	M-ER	M-ER
HYPTE3SNP	M	M	P52059	X	M-ER
RBPC_0SNP	M	X <sup>(a)</sup>	ERTF	M	M
UNK 28	X <sup>(a)</sup>	M-ER	HSP	M	M
CPCB2	X	X	Leu-TF	M	M
Lg 33	M-ER	M-ER	P450	M	M
PESR1	X	M-ER	WRKY58	M	M
Lup 36	M-ER	M-ER	bZIP	M	M
GLIP 99	M-ER	M-ER	psat_Est_00190_01_1	M-ER	X <sup>(b)</sup>
Lg 34	M	M	mtmt_GEN_0114_05_1	I	I
CP450	M-ER	M-ER	CAK	I	I
Mtmt_Gen_00733_03_3	X <sup>(a)</sup>	M-ER	CysPr 1	I	I
Anmts 37	X <sup>(a)</sup>	M-ER	DNABP	M-ER	M-ER
PsPRP 4A	I	I	DSI	M-ER	M-ER
IJB174	I	I	ENOLB	M-ER	M-ER
IJB91	I	I	EPS	I	I
RGA2.97	X <sup>(a)</sup>	X	EST 718	I	I
Hmm6	I	I	EST 948	M-ER	M-ER
PsDof 1	I	X <sup>(b)</sup>	HYPTE116	M-ER	X
DRR 230 b	I	X <sup>(b)</sup>	KCOAT	M-ER	M-ER
PsMnSOD	X <sup>(a)</sup>	X <sup>(a)</sup>	MPP	I	I
RGA-G3A	X	X	PCT	X <sup>(a)</sup>	M-ER
Peachi 21	I	I	PRTS	I	I
RGA2P	I	I	RNAH	M-ER	M-ER
RGA3P	X	X <sup>(b)</sup>	Te001	I	I
Pis_Gen_6_3_1	X <sup>(a)</sup>	X <sup>(b)</sup>	TE-011	I	I
UDPGD	M-I	M-I	Te013	M	M
UNK16	M-ER	M-ER	TE-016	M-ER	M-ER
UNK21	M-I	M-I	TUP	I	I
UNK3	M-I	M-I	Ca2_ERTF	M-I	M-I
Ein3Fw2	M-I	M-I	Ca2_ZnF	M-I	M-I
Ein3Fw3	M-I	M-I	Ca2_SerK	M-I	M-I
Avr	M-I	M-I	Ca2_Transf	X	X
AgR	M-I	M-I	Ca2Calm	M-I	M-I
P00023	X	M-I	Tag 1	M	M
P00042	X	M-ER	bcl2	M	M
P00043	M-ER	M-ER	SerK	M	M
P00046	M-I	M-I	AKR	M	M
P00053	M-I	M-I	ATP-Dep	M	M
P00073	M-ER	M-ER	SerP	M	M
P00075	X	M-ER	U-box	M	M

**Tabla 8 (continuación).** Resultado de los 116 marcadores analizados en las poblaciones RILs Vf6 x Vf136 y 29H x Vf136.

Marcador	Vf6xVf136	29HxVf136	Marcador	Vf6xVf136	29HxVf136
P00084	M	M	ERTF-Rev2	M	M
P00087	M-ER	M-ER	AKR-Fw2	M	M
P00088	M	M	HSP-Fw2	M	M
Bcl2_Rev2	M	M	Ca2_ABF5	M-ER	M-ER
bZip-Fw2	M	M	Ca2_Ribncl	MI	M-I
qPCR1	MI	MI	Ca2_Ein3	MI	M-I
qPCR-3	M	M	Ca2_Rk	M-ER	M-ER
qPCR-4	M	M	Ca2_SAP1	MI	M-I
qPCR-6	M	M	CaLF1	M	M
qPCR-7	M	M	CaLE1	M	M
SSAf_5	<b>X<sup>(a)</sup></b>	M-I	CaDET2	<b>X</b>	<b>X</b>

X Marcador genotipado. En negrita se destacan los realizados en este trabajo.  
X<sup>(a)</sup> Marcador genotipado en la población y publicado en un mapa consenso (Satovic et al. 2013).  
X<sup>(b)</sup> Marcador genotipado en estudios previos.  
M Marcador monomórfico (análisis en parentales y 5 individuos de la población).  
M-ER Marcador monomórfico tras cortar con una batería de 10 enzimas de restricción.  
I Marcador con amplificación imprecisa.  
M-I Marcador monomórfico y con amplificación imprecisa.

A pesar de los intentos para incrementar la especificidad y mejorar la amplificación del resto de marcadores analizados (modificación de la temperatura de hibridación y concentración de MgCl<sub>2</sub>, entre otros), no se mejoraron los resultados y los marcadores que no amplificaron o mostraron un polimorfismo ambiguo, fueron excluidos del análisis. Los resultados expuestos, muestran que el genotipado de marcadores desde especies relacionadas en *V. faba* es una tarea complicada y que el nivel de transferencia se ve limitado por la falta de especificidad en la secuencia de los cebadores empleados. A ello se le une que *V. faba* es una especie parcialmente alógama (35%) (Bond and Poulsen, 1983) y el polimorfismo inicial puede perderse en la descendencia por algún cruzamiento incontrolado. Por todos estos motivos la puesta a punto de la amplificación de nuevos cebadores y la obtención de amplificaciones de banda única para poder trabajar con las diferentes técnicas de búsqueda de polimorfismo se hacen más complejos en esta especie.

El escaso polimorfismo presente en el cultivo ya se constató en el proyecto europeo GLIP (Grain Legume Integrated Project- Food-CT-2004-506223) donde *V. faba* mostró uno de los índices de polimorfismo más bajos (12,5%) entre los cultivos analizados. Al manejarse únicamente cruzamientos intraespecíficos (por no existir ningún antecesor silvestre conocido), la posibilidad de encontrar polimorfismo

disminuye bastante frente al manejo de cruzamientos interespecíficos, como los empleados en otras especies relacionadas como garbanzo, lenteja, guisante etc.

**Tabla 9.** Lista de los 27 marcadores moleculares genotipados en las poblaciones Vf6 x Vf136 y 29H x Vf136 y metodologías para la detección del polimorfismo.

Marcador	Vf6xVf136	29HxVf136	Método de genotipado	Enzima de restricción
cgP137FSNP	M	<b>X</b> <sup>(a)</sup>	CAP	Hpa II
Mtmt_Gen_01017_03_1	<b>X</b> <sup>(a)</sup>	M	CAP	Mse I
Mtmt_Gen_01965_03_1	<b>X</b> <sup>(a)</sup>	M	CAP	Bsa A I
REPSNP	<b>X</b> <sup>(a)</sup>	M	CAP	Taq I
Mtmt_Gen_00510_01_1	M	<b>X</b> <sup>(a)</sup>	CAP	Alu I
Mtmt_Gen_01115_02_1	M	<b>X</b> <sup>(a)</sup>	CAP	Ase I
RBPC_0SNP	M	<b>X</b> <sup>(a)</sup>	CAP	Bsrl
UNK 28	<b>X</b> <sup>(a)</sup>	M-ER	CAP	Hpa II
CPCB2	<b>X</b>	<b>X</b>	ALP	ALP
PESR1	<b>X</b>	M-ER	CAP	Ssil
Mtmt_Gen_00733_03_3	<b>X</b>	M-ER	ALP	ALP
Anmts 37	<b>X</b>	M-ER	CAP	AvaII
RGA2.97	<b>X</b> <sup>(a)</sup>	<b>X</b>	P/A	P/A
PsMnSOD	<b>X</b> <sup>(a)</sup>	<b>X</b> <sup>(a)</sup>	CAP	MboII
RGA-G3A	<b>X</b>	<b>X</b>	CAP	RsaI
RGA3P	<b>X</b>	<b>X</b> <sup>(b)</sup>	ALP	ALP
Pis_Gen_6_3_1	<b>X</b> <sup>(a)</sup>	<b>X</b> <sup>(b)</sup>	ALP	ALP
HYPTE_116	M-ER	<b>X</b>	ALP	ALP
P00023	<b>X</b>	M-I	ALP	ALP
P00042	<b>X</b>	M-ER	CAP	Ssil I
P00075	<b>X</b>	M-ER	ALP	ALP
P00098	<b>X</b>	<b>X</b>	CAP	Alu I
P00260	<b>X</b>	M-ER	ALP	ALP
P52059	<b>X</b>	M-ER	ALP	ALP
SSAf_5	<b>X</b> <sup>(a)</sup>	M-I	ALP	ALP
Ca2_Transf	<b>X</b>	<b>X</b>	CAP	RsaI
CaDET2	<b>X</b>	<b>X</b>	ALP	ALP

X: Marcador genotipado. En negrita se destacan los realizados en este trabajo.

**X**<sup>(a)</sup>: Marcador genotipado en la población y publicado en un mapa consenso (Satovic et al. 2013).

**X**<sup>(b)</sup>: Marcador genotipado en estudios previos.

M: Marcador monomórfico (análisis en parentales y 5 individuos de la población).

M-ER: Marcador monomórfico tras cortar con una batería de 10 enzimas de restricción.

I: Marcador con amplificación imprecisa.

M-I: Monomórfico y con amplificación imprecisa.

CAP: Cleavage Amplified Polymorphism. Secuencias de restricción amplificadas y polimórficas.

ÁLP: Amplified Length Polymorphism. Polimorfismo en la longitud (peso molecular) de la banda.

P/A: Presencia y ausencia de banda.

De los 27 marcadores genotipados en una u otra población (Tabla 9), la transferencia ha sido más eficiente en el caso de marcadores que resultaron ya polimórficos en el cruce Vf6 x Vf27 de *V. faba* (Ellwood et al. 2008; Cruz-Izquierdo et al. 2012), pudiéndose genotipar 12 de los 22 (54,54%) marcadores analizados. Seguidamente pudimos transferir 5 de los 12 (41,7%) marcadores procedentes de guisante relacionados con la resistencia a *A. fabae* (Prioul et al. 2007), 6 de los 18 (33,3%) UniTags expresados diferencialmente tras el análisis con SuperSage de la interacción *Ascochyta*-lenteja (García et al 2012; 2016) y 2 de los 17 (11,8%) marcadores asociados con la resistencia a *A. fabae* en garbanzo y al carácter fecha de floración. Los marcadores que peor se han transferido, solo 1 de los 23 (4,35%), han sido los procedentes de *M. truncatula* y diseñados por Choi et al. (2004) así como los UniTags expresados diferencialmente tras el análisis con SuperSage de la interacción *Ascochyta*-habas (Madrid et al. 2013) con sólo 1 de los 24 (4,17%) analizados. Estos resultados están de acuerdo con las relaciones filogenéticas entre las diferentes leguminosas (Zhu et al. 2005). Así, las habas están en el mismo clade que guisantes y lentejas (*Viceae*), por lo que los marcadores desarrollados en estas especies se transferirán mejor, que los desarrollados en *Medicago* o garbanzo, especies pertenecientes a otro clade diferente (*Trifolieae* y *Cicereae*, respectivamente).

Además de saturar los mapas de cada población, 14 de los marcadores genotipados fueron incluidos en un mapa consenso del cultivo (Satovic et al. 2013; Anexo I). Once de ellos resultaron comunes y facilitaron la integración de los grupos de ligamiento en el nuevo mapa.

**II.3.2. Nuevos mapas de ligamiento y análisis de QTLs implicados en la resistencia a *O. crenata*, *O. foetida* y *A. fabae* en habas.**

**II.3.2.1. Población 29H x Vf136.**

Como se ha mencionado anteriormente, de los 116 marcadores analizados solo 12 se pudieron genotipar en esta población RIL. Dichos marcadores se añadieron a los 200 genotipados en estudios previos (Gutiérrez et al. 2013; Atienza et al. 2016) a fin de saturar el mapa y la posición de los QTLs de dicha población. Un 10.37% de los marcadores incluidos en el mapa no se ajustaron a la segregación 1:1 esperada. Como parámetros de mapeo, consideramos un LOD  $\geq 4$  y una fracción de recombinación  $\leq 0.25$ . Cinco de estos marcadores fueron incluidos en un nuevo mapa consenso realizado para este cultivo por Satovic et al. (2013).

De los 12 marcadores genotipados (Tabla 10), se han conseguido añadir a los mapas previos (Gutiérrez et al. 2013; Atienza et al. 2016) 8 (2 ALPs, 5 CAPs y 1P/A) distribuidos en los cromosomas I, II, VI y en dos grupos de ligamiento (GL) que no han podido ser asignados a cromosomas específicos (GL 26 y GL\*). Los 4 marcadores restantes (P00098, CPCB2, Ca2\_Transf y RBPC/0) no se ligaron a ninguno de los GL actuales, aunque serán incluidos en sucesivas actualizaciones de este mapa (Tabla 10, Figura 2). La identificación y orientación de cada GL se determinó por comparación con mapas previos (Gutiérrez et al. 2013; Satovic et al. 2013; Atienza et al. 2016).

**Tabla 10.** Posición en el mapa de los marcadores genotipados en la población 29H x Vf136.

<b>Marcador</b>	<b>LG</b>	<b>Marcador</b>	<b>LG</b>
cgP137FSNP	Cr Ia	PsMnSOD	Cr II
Mtmt_Gen_00510_01_1	Cr VI	RGA-G3A	LG*
Mtmt_Gen_01115_02_1	Cr VI	HYPTE_116	LG 26
RBPC_0SNP	No ligado	CaDET2	Cr Ia
CPCB2	No ligado	P00098	No ligado
RGA2.97	Cr VI	Ca2_Transf	No ligado

El nuevo mapa comprende 182 marcadores distribuidos en 26 GL cubriendo una longitud de 2125.39 cM con una distancia media entre marcadores de 10.67cM (Tabla 11, Figura 2). El mayor grupo de ligamiento corresponde al cromosoma VI

compuesto por 22 marcadores y 284.9 cM y el menor fue el LG 18 asociado al cromosoma II con 3 marcadores y 16.6 cM (Tabla 11).

**Tabla 11.** Grupos de ligamiento (GL), asignación a cromosomas y distribución de marcadores en el mapa 29H x Vf136.

GLs ID *	Cromosoma	Longitud (cM)	Número de marcadores mapeados	Distancia media entre marcadores (cM)
LG8	I	79.15	7	11.31
LG16	I	31.28	3	10.43
LG5+LG6	I	225.66	18	12.54
LG7+LG17	I	168.78	11	15.34
LG12	I	50.15	5	10.03
LG19	I	15.27	3	5.09
LG11	I	45.27	6	7.55
LG24	IB	18.19	2	9.10
LG14	IB	12.70	4	3.18
LG4	II	200.19	17	11.78
LG29	IIA	22.77	2	11.39
LG2	III	220.27	18	12.24
LG9	III	111.58	8	13.95
LG18	III	16.62	3	5.54
LG13	IIIA	27.09	4	6.77
LG3	V	210.66	15	14.04
LG23	V	23.15	2	11.58
LG1	VI	284.90	22	12.95
LG1	VI	35.78	6	5.96
LG*	-	73.31	4	18.33
LG26	-	39.16	3	13.05
LG15	-	22.63	3	7.54
LG10	-	82.92	7	11.85
LG20	-	32.18	3	10.73
LG22	-	41.96	3	13.99
LG21	-	33.76	3	11.26
<b>Total</b>		<b>2125.4</b>	<b>182</b>	<b>10.67</b>

GLs ID \*: Grupos de ligamiento identificados en este estudio (26) en correspondencia con los descritos por Gutiérrez et al. (2013).

Si comparamos el nuevo mapa con el publicado por Gutiérrez et al. (2013), la incorporación de nuevos marcadores ha permitido unir cuatro pequeños GL en dos

de mayor tamaño (LG17+LG7 y LG6+LG5) en el cromosoma I. Además, si comparamos con el mapa consenso (Satovic et al. 2013), el GL 9 asignado al cromosoma I por Gutiérrez et al. (2013) pasaría a formar parte del cromosoma III, el GL 8 se integraría en el cromosoma I, el GL 29 al cromosoma IIa y el GL13 adscrito al cromosoma II pasaría a formar parte del cromosoma IIIa en nuestro estudio. El mapa descrito cubre 2,125.4 cM frente a los 1,402.1 cM descritos por Gutiérrez et al. 2013.

Utilizando el nuevo mapa y las evaluaciones publicadas en Gutiérrez et al. (2013) para *O. foetida* y *O. crenata* y Atienza et al. (2016) para *A. fabae*, se ha realizado un nuevo análisis de QTLs (Tabla 12) con el objetivo de reubicar los QTLs en el nuevo mapa de la población. Dado el limitado número de marcadores añadidos al mapa, las modificaciones respecto a su posición no han sido muy apreciables.

Para *A. fabae*, se detectaron cuatro regiones relacionadas con la resistencia situadas en los cromosomas II, III, V y VI. En concordancia con los datos de Ávila et al. (2004) y Atienza et al. (2016), *Af2* se ubicó en el cromosoma II limitado por los marcadores OPR16\_1577/OPB13\_1133. Este QTL fue ratificado en los tres ensayos para DSL (F\_DSL, Co99\_DSL y Lo98\_DSL) a un LOD máximo de 3.4, 2.61 y 3.08 respectivamente (Figura. 2, Tabla 12). *Af2* explica un porcentaje de variación fenotípica ( $R^2$ ) entre el 13,1% (Co99\_DSL) y el 15,7% (Field\_DSL). El cromosoma II fue saturado con un sólo marcador (PsMnSOD) ubicado lejos del área de influencia del QTL. Un segundo QTL (*Af3*) fue detectado en el cromosoma III, para las evaluaciones Lo98\_DSL, Lo98\_DSS y F\_DSP, no detectándose en ninguna de las otras tres situaciones descritas por Atienza et al. (2016). Lo98\_DSL, Lo98\_DSS y F\_DSP fueron detectados a un LOD de 3.52, 2.41 y 2.61, respectivamente. Están limitados por los marcadores OPI01\_1985/OPD16\_1732 y explican entre el 11.1% (Lo98\_DSS) y el 15.7% (Lo98\_DSL) de la variación fenotípica. Por último, en el cromosoma VI se detectaron dos regiones. La primera región corresponde a la identificada por Atienza et al. (2016) correspondiente a F\_DSP1, F\_DSS y Lo98\_DSL con valores de LOD entre 2.6 (F\_DSS) y 3.59 (F\_DSP1). Dicha región está flanqueada por los marcadores mtmt\_GEN\_01130\_02\_1/OPI10\_1178, explicando entre 11.4% (F\_DSS) y el 17.1% (F\_DSP1) de la variación fenotípica. F\_DSP1 se sitúa en la misma posición que *Oc7* identificado y validado para la resistencia a *O. crenata* lo que sugiere la existencia de posibles mecanismos de defensa comunes para diferentes patógenos.





El intervalo de confianza para este QTL contiene los marcadores mtmt\_GEN\_01130\_02\_1 y RNAR que podrían estar involucrados en procesos de resistencia. Así, mtmt\_GEN\_01130\_02\_1 es un receptor LRR de quinasa (LRR-RKs) involucrado en la regulación de procesos relacionados con el desarrollo y la defensa, incluyendo la proliferación celular, la percepción hormonal, la respuesta de defensa específica y no específica del huésped, la respuesta a la herida y la simbiosis (Torii, 2004) y RNAR es una oxidoreductasa involucrada en la síntesis de ADN. En la segunda región, la inclusión del nuevo marcador RGA2\_97 ha permitido detectar por primera vez a F\_DSP2 y Lo98\_DSL explicando un 21.1% y 15.2% respectivamente de la variación fenotípica del carácter. Sería de gran interés la saturación del cromosoma VI con nuevos marcadores para unificar los dos grupos de ligamiento y precisar la posición de estos QTLs sobretodo de F\_DSP1 debido a su interés al colocalizar con Oc7 validado para la resistencia a *O. crenata*.

**Tabla 12.** QTLs detectados para resistencia a *A. fabae* (*Af*), *O. foetida* (*Of*) y *O. Crenata* (*Oc*) en la población RIL 29H x Vf136 de habas por análisis de intervalos (IM). El LOD crítico fue determinado usando el análisis de permutaciones en todos los ambientes.

QTL	Ambiente, aislado y/o órgano analizados	CR	Marcadores flanqueantes	Marcador más asociado	LOD	LOD crítico	R <sup>2</sup> (%)
<i>Oc7</i>	Cor06/07	VI	OPI10_1178/OPU09_831	OPA11_554	7.89	3	34
<i>Oc7</i>	Cor07/08	VI	OPA17_633/OPU09_831	OPA11_554	5.92	3.1	27.6
<i>Oc7</i>	Cor08/09	VI	OPI10_1178/OPU09_831	OPA11_554	6.91	3	30.8
<i>Oc8</i>	Cor07/08	V	OPAC06_1034/mtmt_GEN_00866_02_1	Vfg69	3.13	3.1	17.7
<i>Af2</i>	F_DSL	II	OPR16_1577/OPK16_1144	1433P	3.4	3.2	15.7
<i>Af2</i>	Co99_DSL	II	OPR16_1577/OPB13_1133	1433P	2.61	2.8	13.1
<i>Af2</i>	Lo98_DSL	II	OPR16_1577/OPB13_1133	1433P	3.08	2.9	14.1
<i>Af3</i>	Lo98_DSS	III	OPI01_1985/OPD16_1732	OPD20_322	2.41	2.9	11.1
<i>Af3</i>	Lo98_DSL	III	SECA/OPP17_944	OPD20_322	3.52	2.9	15.7
<i>Af3</i>	F_DSP	III	OPI01_1985/OPP17_944	OPD20_322	2.61	3	12.1
<i>Af</i>	F_DSP2	VI	HBP2/RGA2_97	mtmt_GEN_01109_01_1	3.66	3	21.1
<i>Af</i>	Lo98_DSL	VI	OPJ18_430/RGA3P	OPB03_1294	2.47	2.9	15.2
<i>Af</i>	F_DSP1	VI	mtmt_GEN_01130_02_1/OPI10_1178	RNAR	3.59	3	17.1
<i>Af</i>	F_DSS	VI	mtmt_GEN_01130_02_1/RNAR	mtmt_GEN_01130_02_1	2.6	2.9	11.4
<i>Af</i>	Lo98_DSL	VI	mtmt_GEN_01130_02_1/RNAR	mtmt_GEN_01130_02_1	2.89	2.9	11.8

NOTA: Cada QTL se nombró como en trabajos previos (Gutiérrez et al. 2013, Atienza et al. 2016). Para la resistencia a *O. crenata* y *O. foetida*, cada localidad se refiere como Co (Córdoba), Be (Beja) y Eg (Egipto) identificándose el año como 06 (2006/2007), 07 (2007/2008) y 08 (2008/2009). Para *A. fabae* se detallan los aislados utilizados (Co99 y Lo98) en evaluaciones en fase de plántula o en planta adulta y a campo abierto (F) empleando un aislado local. Además, se distingue entre la evaluación realizada en hoja (DSL), tallo (DSS) y vaina (DSP).

Para el control de la resistencia a *O. crenata*, se confirmaron dos QTLs estables previamente descritos por Gutiérrez et al. 2013. *Oc7* fue identificado en el cromosoma VI limitado por los marcadores OPA17\_633/OPU09\_831. Este QTL fue identificado durante tres campañas consecutivas en Córdoba (Córdoba 06/07, 07/08 y 08/09) explicando entre el 27,6% y el 34% de la variación fenotípica total (Tabla 12, Figura.2). *Oc7* se localiza en la misma posición que F\_DSP1 detectado para *A. fabae* lo que aumenta el interés de saturación de esta zona para refinar su posición.

De acuerdo con estudios previos, *Oc8* fue localizado en el cromosoma V (LOD = 3.13 y R<sup>2</sup> del 17.7%), pero sólo en la campaña 07/08 y no en 06/07 ambos detectados por Gutiérrez et al. 2013 (Figura 1, Tabla 12). Este QTL está limitado por los marcadores OPAC06\_1034/mtmt\_GEN\_00866\_02\_1 situados en el extremo del cromosoma V. Tampoco fueron detectados los QTLs *Oc9-Oc13* ni los QTLs relacionados con la resistencia a *O. foetida* (*Of3-Of5*) descritos por Gutiérrez et al. 2013.

Pese a existir una zona con cierta significación para *O. foetida* en el GL2 (chr. III) el valor de LOD no superó el umbral para aceptarlo. La construcción del mapa mediante el programa JoinMap v. 4.1 en lugar del MAPMAKER v. 3.0b (Lander et al. 1987) y el análisis de QTLs mediante el programa MapQTL® V5 (Van Ooijen 2006) en lugar del Windows QTL Cartographer v. 2.5 (Wang et al. 2007), empleados previamente, junto a la variación en algunos de los parámetros podría estar entre las causas de esta diferencia.

### **II.3.2.2. Población Vf6 x Vf136.**

De los 116 marcadores analizados en esta población, sólo 22 pudieron ser genotipados de manera fiable y añadirse a la base de datos disponible (Tabla 13, Figura 3). Diez de los nuevos marcadores fueron incluidos en un reciente mapa consenso del cultivo (Satovic et al. 2013). El nuevo análisis de ligamiento implicó un total de 416 marcadores de los cuales un 14,9% no se ajustaron a la segregación 1:1 esperada. Un valor de LOD de 5 y una fracción de recombinación máxima de 0.25 fueron considerados para formar los distintos GL. Finalmente, el nuevo mapa consistió en 371 marcadores asociados a 18 GL que cubren una longitud de 4,970.92 cM con una distancia media entre marcadores de 12.86 cM (Tabla 14, Figura 3). La identificación y orientación de cada GL se determinó por comparación con los mapas previamente publicados para esta población (Díaz-Ruiz et al. 2009; Satovic et al. 2013) siendo

excluidos GL con dos marcadores. En este caso, se han conseguido añadir al mapa previo los 22 marcadores (10 ALPs, 11 CAPs y 1P/A) genotipados en este estudio (Tabla 13, Figura 3). Los dos GL mayores corresponden a partes de los cromosoma IA (50 marcadores que cubren una longitud de 786.37 cM) y IIA (54 marcadores y 747.34 cM). El GL menor es el GL 6B con 4 marcadores y 42,12 cM (Tabla 14). Con la incorporación de nuevos marcadores y la nueva reorganización, el mapa cubre 4,970.92 cM frente a los 3,537 cM previos (Satovic et al. 2013).

**Tabla 13.** Marcadores moleculares genotipados y ligados en la población Vf6 x Vf136.

Marcador	Cromosoma/GL	Marcador	Cromosoma/GL
Mtmt_Gen_01017_03_1	III	RGA3P	VI-A
Mtmt_Gen_01965_03_1	I-A	Pis_Gen_6_3_1	V
REPSNP	III	P00023	II-A
UNK 28	GL2	P00042	I-A
CPCB2	GL2	P00075	I-A
PESR1	I-B	P00098	GL12
Mtmt_Gen_00733_03_3	V	P00260	I-A
Anmts 37	V	P52059	I-A
RGA2.97	VI-A	SSAf_5	VI-B
PsMnSOD	II	Ca2_Transf	III
RGA-G3A	VI-A	CaDET2	I-A

Para el nuevo análisis de QTL fueron consideradas las evaluaciones realizadas por Díaz-Ruiz et al. (2009 a, b y 2010) para *A. fabae*, *O. foetida* y *O. crenata*, respectivamente. Los resultados del análisis de QTLs por análisis de intervalos y permutaciones pueden verse detalladamente en la Tabla 15.

En el caso de *A. fabae*, se identificaron cinco regiones relacionadas con la resistencia (*Af1-Af5*). *Af1* y *Af2* fueron descritos previamente por Díaz-Ruiz et al. (2009a) y Satovic et al. (2013) en hojas (DSL) y tallos (DSS). Este trabajo no detectó a *Af1\_DSS* pero sí a *Af1\_DSL* (III) limitado por los marcadores mtmt\_GEN\_00012\_03\_1/OPAC02\_579. Este QTL explica un 12,5% de la variación fenotípica ( $R^2$ ), porcentaje mayor al descrito por Díaz-Ruiz et al. (2009a). Aunque este cromosoma ha sido saturado con tres nuevos marcadores ninguno co-localiza en esta región. *Af2*, adscrito al cromosoma II, fue de nuevo detectado tanto en hoja (*Af2\_DSL*) como en tallo (*Af2\_DSS*) siendo sus marcadores más asociados Mer04\_790 (DSL) y OPE17\_1326 (DSS), respectivamente. *Af2\_DSL* explica un

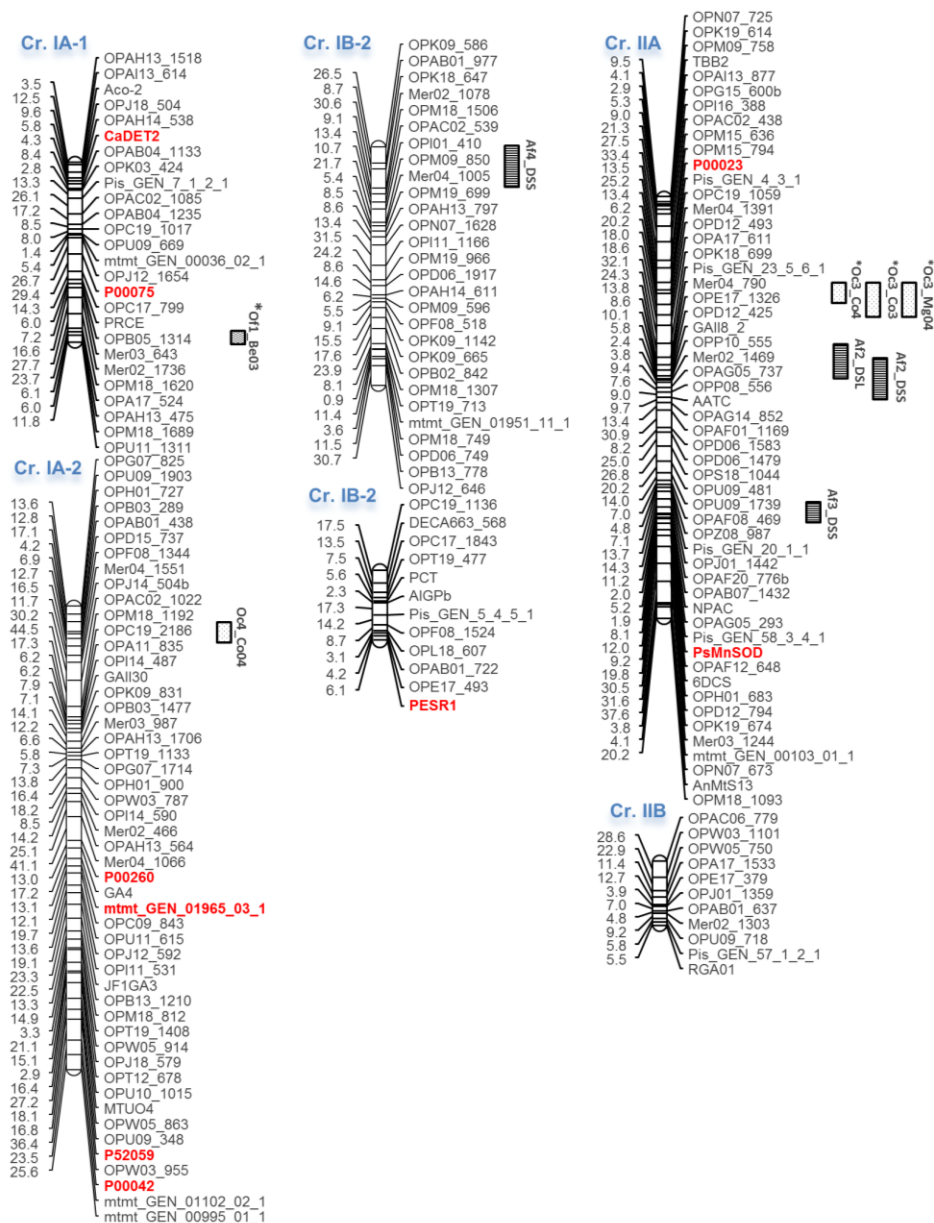
13.1% de la variación fenotípica en hoja, valor similar (12.54 %) al descrito por Satovic et al. 2013. Por el contrario, *Af2\_DSS* explicó un porcentaje de variación del 12% ligeramente superior al previo (9.85%). *Af2* está muy próximo a *Oc3* lo que sugiere que en dicha región del genoma podrían coexistir mecanismos comunes implicados en la resistencia a ambos patógenos. Por ello, la adición de nuevos marcadores a la región candidata facilitará la identificación de genes que desempeñen un papel funcional en mecanismos de defensa frente a distintos patógenos (hongos y plantas parásitas, en este caso) y el desarrollo de marcadores moleculares que mejoren la eficiencia de los programas de mejora asistida por marcadores.

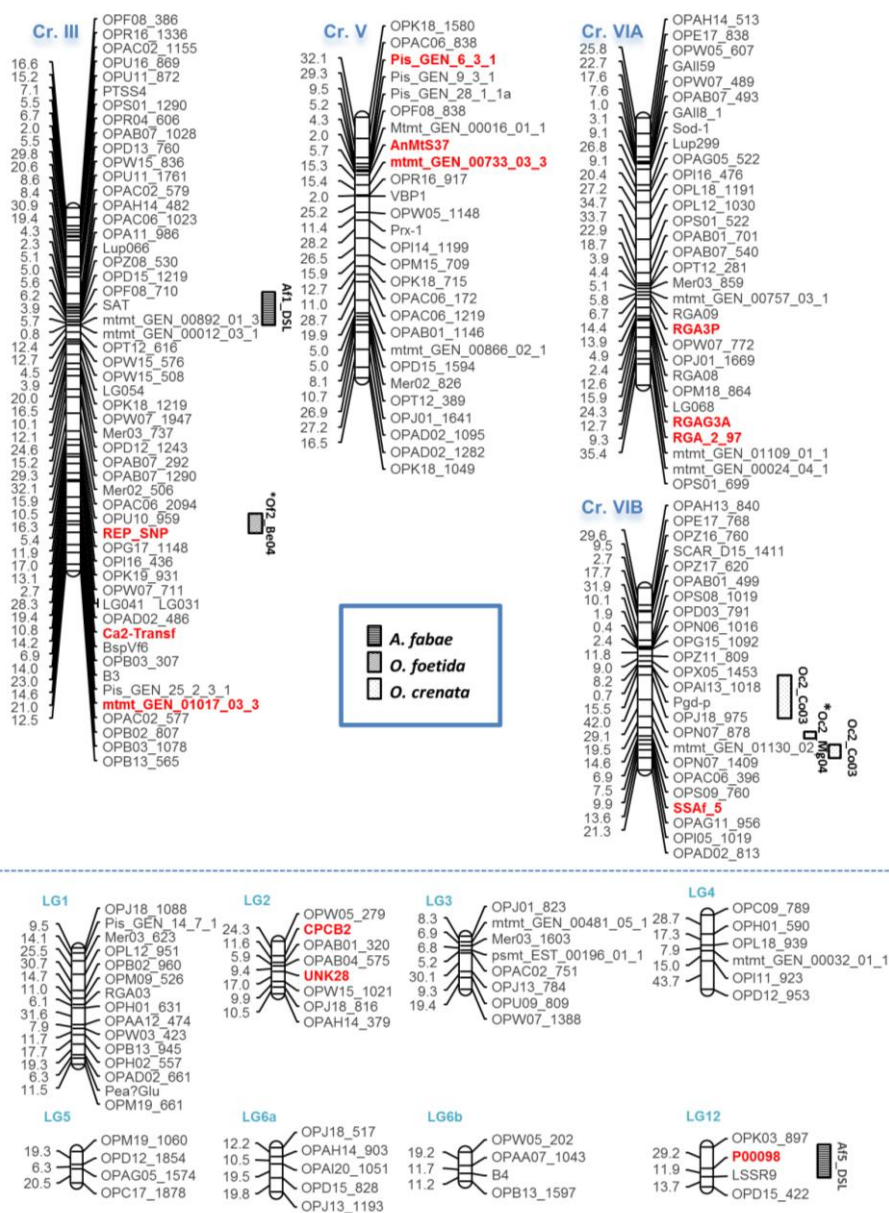
**Tabla 14.** Grupos de ligamiento (GL), asignación a cromosomas, características y distribución en el mapa Vf6 x Vf136.

Cromosoma /GLs*	Longitud (cM)	Numero de marcadores mapeados	Densidad media de marcadores (cM)
IA-1	302.28	26	11.63
IA-2	786.37	50	15.73
IB-1	379.53	28	13.55
IB-2	100.03	12	8.34
IIA	747.34	54	13.84
IIB	111.80	11	10.16
III	666.07	54	12.33
V	399.71	27	14.80
VIA	452.09	31	14.58
VIB	315.76	24	13.16
LG1	217.63	15	14.51
LG2	88.64	8	11.08
LG3	85.99	8	10.75
LG4	112.64	6	18.77
LG5	46.14	4	11.53
LG6a	62.03	5	12.41
LG6b	42.12	4	10.53
LG12	54.78	4	13.69
<b>Total</b>	<b>4970.92</b>	<b>371</b>	<b>12.86</b>

GLs ID \*: Grupos de ligamiento identificados en este estudio (18) que se corresponden con los descritos por Díaz-Ruiz et al. 2009.

**Figura 3.** Nuevo mapa de ligamiento y QTLs para la resistencia a *A. fabae* (*Af*), *O. crenata* (*Oc*) y *O. foetida* (*Of*) en la población RIL Vf6 x Vf136.





Nota: Cada QTL está identificado con el número que se le asigno en trabajos previos (Díaz-Ruiz et al. 2009/2010). Para la resistencia a *O. crenata* y *O. foetida*, cada localidad está identificada como Co (Córdoba), Be (Beja) y Mg (Mengibar) identificándose el año de campaña como 03 (2002/2003) y 04 (2003/2004). Para *A. fabae* las evaluaciones se identifican como DSL (Hoja) y DSS (Tallo). El umbral de LOD usado para identificar los QTLs fue obtenido usando el análisis de permutaciones. En rojo se señalan los 22 marcadores genotipados en este estudio, en azul los diferentes cromosomas y grupos de ligamiento que aún no han podido ser asignados a cromosomas específicos. QTLs con \* son QTLs detectados por Díaz-Ruiz et al. (2009b; 2010) a LOD 2 que no han superado el valor umbral de LOD en este estudio.

Además, se identificaron tres nuevos QTLs (*Af3\_DSS*, *Af4\_DSS* y *Af5\_DSL*). La incorporación de los nuevos marcadores PsMnSOD y P00098 en el cromosoma II y GL12, ha originado que se detecte por primera vez *Af3\_DSS* (OPAF20\_776b/PsMnSOD) y *Af5\_DSL* (LSSR9/OPK03\_897) explicando un 10.2% y un 15,6% de la variación fenotípica del carácter. El nuevo marcador PsMnSOD está asociado a la resistencia a *A. fabae* en guisante (Prioul et al. 2007) y codifica para la enzima superóxido dismutasa que se induce en procesos de defensa frente a patógenos en plantas. Por otro lado, P00098 (marcador más asociado a *Af5\_DSL*) deriva de un estudio transcriptómico asociado con la resistencia a *Ascochyta* en lenteja (García et al. 2012; 2016). Finalmente, *Af4\_DSS* identificado en el cromosoma IB está limitado por los marcadores OPK09\_586/Mer02\_1078 y explica un 12% de la variación fenotípica. Es necesaria la inclusión de marcadores más robustos (ESTs, SNPs, etc.) para identificar candidatos génicos para esta resistencia.

Para el control de la resistencia a *O. crenata* y de acuerdo con los datos de Díaz-Ruiz et al. (2010), se identificaron dos QTLs (*Oc2* y *Oc4*) validados previamente. *Oc2* se localizó en el cromosoma VIb en las evaluaciones realizadas en Córdoba durante la campaña 2002/2003, explicando un porcentaje de explicación (22.7%) bastante mayor que el descrito previamente (8%) por Díaz-Ruiz et al. (2010). En esta región hemos incorporado un nuevo marcador (SSAf\_5) asociado con la resistencia a *Ascochyta* en otra población de habas. Díaz-Ruiz et al. (2010) también identificó este QTL en Mengibar pero el presente estudio no lo detecta debido a que los criterios empleados para identificar QTLs se basan en el valor de LOD obtenido por permutaciones y no en la fijación de un valor de LOD fijo (LOD 2). Si consideramos los criterios utilizados por Díaz-Ruiz et al. (2010) sí que se detecta dicho QTL explicando un 7.8% de la variación fenotípica del carácter. Además, considerando estos criterios dicho QTL se detectaría también en Córdoba 2003/2004 con un LOD 2.84 y explicaría 13.8% de la variación fenotípica.

*Oc3* identificado por Díaz-Ruiz et al. 2010 en Córdoba 2002/2003 y 2003/2004 y Mengibar 2003/2004, no es detectado en nuestro estudio al no superar el LOD fijado por permutación. De nuevo, si empleamos los criterios previos (LOD 2) sí que se identifica dicho QTL en el cromosoma IIA limitado por los marcadores OPD12\_493/P00023. *Oc3* explicaría entre un 8.5% (Córdoba 02/03) y un 11.2% (Córdoba 03/04 y Mengibar 2003/2004) de la variación del carácter. Pis\_GEN\_4\_3\_1 se localiza junto a un nuevo marcador genotipado en este estudio (P00023) asociado

con la resistencia a *Ascochyta* en lenteja. *Oc4* se detecta en el cromosoma I.A-2 en Córdoba 2003/2004 (LOD de 4.45 y  $R^2=17.7\%$ ) teniendo como marcador más asociado OPB03\_289. Estos datos concuerdan con los de Díaz-Ruiz et al. (2010) donde *Oc4* explica un 17% de la variación del carácter. *Oc5*, detectado por Díaz-Ruiz et al. (2010) en el cromosoma IA en Mengíbar 2003/2004 no ha sido detectado en el presente estudio.

**Tabla 15.** QTLs detectados para resistencia a *A. fabae* (*Af*), *O. foetida* (*Of*) y *O. crenata* (*Oc*) en la población RIL Vf6 x Vf136 mediante análisis de intervalos (IM).

QTL	Ensayo	Cr.	Marcadores flanqueantes	Marcador más asociado	LOD	LOD crítico	R <sup>2</sup> (%)
<i>Af1</i>	DSL	III	mtmt_GEN_00012_03_1/ OPAC02_579	Lup066	3.88	3.2	12.5
<i>Af2</i>	DSS	II	Pis_GEN_23_5_6_1/ AATC	OPE17_1326		3.2	
<i>Af2</i>	DSL	II	OPK18_699/GAII8_2	Mer04_790	3.89	3.2	13.1
<i>Af3</i>	DSS	II	OPAF20_776/PsMnSOD	OPAB07_1432	3.29	3.2	10.2
<i>Af4</i>	DSS	IB2	OPK09_586/Mer02_1078	OPAB01_977	3.31	3.2	12
<i>Af5</i>	DSL	LG12	LSSR9/OPK03_897	P00098	3.89	3.2	15.6
<i>Oc2</i>	Cor03	VIB	OPI05_1019/ OPAG11_956	OPAG11_956	5.87	3.3	22.7
<i>Oc2</i>	Cor03	VIB	OPJ18_975/ mtmt_GEN_01130_02_1	OPN07_878	3.56	3.3	25.8
<i>Oc4</i>	Cor04	IA	OPH01_727/OPD15_737	OPB03_289	4.45	3.3	17.7
* <i>Oc2</i>	Mg04	VIB	OPAC06_396/OPS09_760	OPAC06_396	2.39	3.2	7.8
* <i>Oc3</i>	Cor04	IIA	P00023/Mer04_1391	Pis_GEN_4_3_1	2.47	3.3	11.2
* <i>Oc3</i>	Cor03	IIA	OPD12_493/P00023	Mer04_1391	2.47	3.3	8.5
* <i>Oc3</i>	Mg04	IIA	OPD12_493/P00023	Pis_GEN_4_3_1	2.47	3.3	11.2
* <i>Of1</i>	Bej03	IA	OPA17_524/OPU11_1311	OPAH13_475	2.01	3.2	6.7
* <i>Of2</i>	Bej04	III	mtmt_GEN_01017_03_3/ OPB03_307	Pis_GEN_25_2_3_1	2.09	3.2	7.2

Nota: Los QTLs con un \* son QTLs detectados por Díaz-Ruiz et al. (2009b; 2010) a LOD 2 que no han superado el valor umbral de LOD por análisis de permutaciones en este estudio para ser considerados como tales.

Finalmente, para *O. foetida*, *Of1* y *Of2* no fueron detectados al no pasar el umbral fijado para la detección de un QTL por el análisis de permutaciones. Con los criterios fijados por Díaz-Ruiz et al. (2009b), que considera la detección de un QTL a LOD 2 (Figura. 3, Tabla 15), *Of1* es detectado en Beja 2003 en el cromosoma IA (LOD 2.01 y  $R^2=6.7\%$ ) y *Of2* es detectado en Beja 2004 en el cromosoma III (LOD 2.09 y  $R^2=7.2\%$ ). Siguiendo estos criterios se habría incorporado un nuevo marcador (mt\_Gen\_1017\_03\_3) a la región limitada por *Of2*.



#### II.4. RESUMEN Y CONCLUSIONES.

Para identificar genes implicados en el control de caracteres de interés, la genómica translacional (desde plantas modelo bien caracterizadas a cultivos menos estudiados) proporciona una valiosa fuente de marcadores para saturar zonas concretas y de genes candidatos caracterizados funcionalmente. En este trabajo, aplicamos dicha estrategia para desarrollar los mapas genéticos de las habas. Centrándonos en una serie de regiones portadoras de QTLs de resistencia a jopo y *Ascochyta* y basándonos en relaciones sinténicas y en resultados de trabajos previos, identificamos posibles regiones homólogas en *M. truncatula*, guisante, garbanzo y lenteja. Así seleccionamos 116 marcadores en dos poblaciones recombinantes (29H x Vf136 y Vf6 x Vf136) de los que sólo pudimos genotipar 12 y 22, respectivamente. Cinco de los marcadores genotipados en el primer cruce y diez del segundo se emplearon como marcadores puente en un reciente mapa consenso de este cultivo (Satovic et al. 2013). La identificación y orientación de cada LG se determinó por comparación con los mapas previamente publicados para estas dos poblaciones (Díaz-Ruiz et al. 2009; Gutiérrez et al. 2013; Satovic et al. 2013; Atienza et al. 2016). Los nuevos mapas han pasado a cubrir 4,970.92 cM (Vf6 x Vf136) y 2,125.39 cM (29H x Vf136) siendo los más completos para estas dos poblaciones de habas, en la actualidad. Los resultados expuestos muestran que el genotipado de marcadores en *V. faba* es una tarea complicada y que el nivel de transferencia se ve limitado por la falta de especificidad en la secuencia de los cebadores empleados. Si bien la genómica translacional identifica regiones colineales para saturar zonas concretas, la secuencia entre especies no siempre está lo suficientemente conservada como para garantizar la transferencia. A ello se le suma que:

1. *V. faba* es una especie parcialmente alógama y el polimorfismo inicial detectado en líneas parentales puede perderse por algún cruzamiento incontrolado.
2. Manejamos unos de los genomas más grandes entre las leguminosas (~ 13.000 Mb) con más del 85% de ADN repetitivo.
3. Al no conocerse ningún antecesor silvestre la mejora en habas se basa únicamente en cruzamientos intraespecíficos por lo que la posibilidad de encontrar polimorfismo disminuye respecto al empleo de cruzamientos interespecíficos.

4. La evaluación de enfermedades con naturaleza poligénica añade complejidad a la hora de detectar los efectos de los QTLs y determinar su posición exacta en el genoma.
5. La puesta a punto de la amplificación de nuevos cebadores derivados de especies cercanas y la obtención de bandas únicas y específicas para la detección de polimorfismos y el genotipado, es compleja en esta especie.

Pese a estas limitaciones, estudios previos han descrito diferentes QTL estables para la resistencia a *A. fabae*, *O. crenata* y *O. foetida* (Díaz-Ruiz et al. 2009 a/b/2010; Gutiérrez et al. 2013; Atienza et al. 2016) en estas dos poblaciones. Este estudio ha detectado la mayoría de estos QTLs aunque el restringido número de marcadores génicos comunes en las dos poblaciones ha limitado la comparación directa de todos los resultados.

Otras posibles diferencias pueden atribuirse al empleo de distintos programas en el análisis de ligamiento y de QTLs (Mapmaker y QTL Cartographer frente a JoinMap y MapQTL).

En la población 29H x Vf136, se han identificado cuatro zonas relacionadas con la resistencia a *A. fabae*, dos relacionadas con *O. crenata* y ninguna para *O. foetida*. Respecto a *A. fabae*, se ha detectado *Af2\_DSL* en el cromosoma II en diferentes ensayos (F\_DSL, Co99\_DSL y Lo98\_DSL) como explicaron los datos de Ávila et al. (2004) y Atienza et al. (2016). *Af3* situado en el cromosoma III solo fue detectado en los ensayos LO98\_DSL, LO98\_DSS y F\_DSP no detectándose los otros tres descritos por Atienza et al. (2016) para esta zona. En el cromosoma VI detectamos dos regiones, la primera (F\_DSP1, F\_DSS y Lo98\_DSL) concuerda con los datos de Atienza et al. (2016) y como se ha dicho se sitúa en la misma posición que *Oc7* ya identificado y validado para la resistencia a *O. crenata* en esta población. La segunda región formada por F\_DSP2 y Lo98\_DSL se identificó en este estudio al incluir el marcador RGA2\_97.

En el caso de **jopo**, se detectaron dos QTLs (*Oc7* y *Oc8*) de acuerdo con los datos de Gutiérrez et al. (2013). *Oc7* localizado en el cromosoma VI ha sido validado en diferentes ambientes y estudios explicando un alto porcentaje de la variación del rasgo lo que lo convierte en un interesante candidato para la MAS. Además el interés de saturación de esta zona aumenta al localizarse en la misma posición que F\_DSP1 detectado para *A. fabae* lo que indica la presencia de mecanismos de resistencia

comunes para ambos patógenos. *Oc8* (chr. V) solo se ha detectado en la campaña 07/08 por lo que es necesario seguir trabajando para definir su posición con mayor precisión y en los diferentes ambientes. No fueron detectados los QTLs *Oc9-Oc13* ni ninguno de los QTLs para *O. foetida* detectados por Gutiérrez et al. 2013.

En la población Vf6 x Vf136, se han identificado cinco regiones relacionadas con la resistencia a *A. fabae* confirmándose la estabilidad de *Af1* (chr. III) y *Af2* (chr. II) de acuerdo con los datos de Díaz-Ruiz et al. (2009a). La mayoría de QTLs descritos eran específicos de patógeno excepto *Af2* que se localiza muy cerca de *Oc3* identificando una región genómica con resistencia de amplio espectro y posibles mecanismos de defensa comunes ante *O. crenata* y *A. fabae*. Las resistencias de amplio espectro y duraderas son deseables en la mejora de cultivos ya que son efectiva en el tiempo, en distintos ambientes y ante distintos patógenos o enfermedades (Kou y Wang, 2010; Schweizer y Stein, 2011). Dichos autores sugieren la saturación de tales puntos calientes con marcadores génicos acompañados de estudios funcionales, para la identificación de genes de resistencia. Aunque las herramientas genómicas disponibles son aun limitadas, ejemplos de genes candidatos tales como GER4 en cebada, trigo y arroz (Himmelbach et al. 2010; Manosalva et al. 2008; Schweizer et al. 1999) sugieren que este enfoque puede tener un futuro éxito en nuestra especie.

Además, tres nuevas regiones (*Af3\_DSS*, *Af4\_DSS* y *Af5\_DSL*) no detectadas por Díaz-Ruiz et al. (2009a) fueron identificadas en este estudio. La incorporación de los nuevos marcadores PsMnSOD y P00098 han permitido detectar *Af3\_DSS* (II) y *Af5\_DSL* (LG12), respectivamente. P00098 sería un buen candidato para su futura validación en distintos fondos genéticos y empleo en MAS. Se trata de una secuencia de lenteja ortóloga al gen Medtr8g013640 de *M. truncatula* (PFAM: 4F5 protein family KOG: Small EDRK-rich protein H4F5), que codifica para proteínas cortas ricas en aspartato, glutamato, lisina y arginina. Aunque se expresan de modo ubicuo en los seres vivos, su función es aún desconocida. Sin embargo, uno de los marcadores flanqueantes de *Af5\_DSL*, LSSR9 descrito en *Arabidopsis thaliana* y genotipado por Ellwood et al. (2008) en habas, codifica para una lipasa (Gene Bank accession: CA411496; Putative lysosomal acid lipase). Chapman et al. (2012) y Benning et al. (2012) entre otros, sugieren que las lipasas son necesarias para la producción de una actividad antibiótica presente en los exudados de plantas y que limitan la infección por patógenos vegetales. En estudios previos, Louis et al. (2010)

demonstró que las lipasas ácidas están de algún modo generando señales implicadas en la respuesta inmune innata en plantas. Al estar *Af5\_DSL* en una zona con mayor número de marcadores génicos asociados a genes con posibles funciones de defensa, la identificación del gen responsable y su confirmación en distintos fondos genéticos para utilización en MAS, puede ser más inmediato.

Para el control de resistencia a *O. crenata* se identificaron dos QTLs (*Oc2* y *Oc4*). Tras incorporar un nuevo marcador (*SSAf\_5*), asociado con la resistencia a *Ascochyta* en otra población de habas (Madrid et al. 2013) *Oc2* en cromosoma VIb, fue detectado en Córdoba 2002/2003 (*Co03*) concordando con los datos de Díaz-Ruiz et al. (2010); *Oc4\_Co04* se detecta en el cromosoma IA.2 de acuerdo con Díaz-Ruiz et al. 2010. El resto de QTL (*Oc2\_Mg04*, *Oc3\_Cor03*, *Oc3\_Cor04*, *Oc5*) detectados por Díaz-Ruiz et al. (2010), no lo fueron en este estudio debido al empleo de criterios diferentes para determinar la presencia significativa de un QTL. Este autor identifica un QTL fijando como LOD crítico un  $LOD \geq 2$ , mientras que este estudio se determinó considerando valores que superaran el LOD crítico obtenido por permutaciones. Si empleamos los criterios previos todos los QTL detectados por Díaz-Ruiz et al. (2010), excepto *Oc5* también son detectados.

De igual modo, este estudio no detectó ningún QTL para *O. foetida* y sólo empleando los criterios utilizados por Díaz-Ruiz et al. 2009b se identifican dos QTL de resistencia. *Of1*, en Beja 2003 (cromosoma IA) y *Of2* en Beja 2004 (cromosoma III) cuya saturación fue mayor con la incorporación del nuevo marcador flaqueante (*mt\_Gen\_1017\_03\_3*).

Aunque nuestro estudio aporta nuevos marcadores génicos que co-localizan o flanquean algunos de los QTLs, e información sobre posibles mecanismos de resistencia para ambos patógenos, es necesario obtener una mayor saturación para estimar mejor la posición y efectos de dichos QTLs. Como ya se ha mencionado, el análisis de poblaciones más grandes o una evaluación más exacta de la resistencia, podría facilitar dicha tarea. Finalmente, el análisis para cambios de expresión de los marcadores más cercanos permitirá determinar si son responsables directos de la respuesta observada o si forman parte del complejo de rutas de reconocimiento y señalización necesario para dicha respuesta.

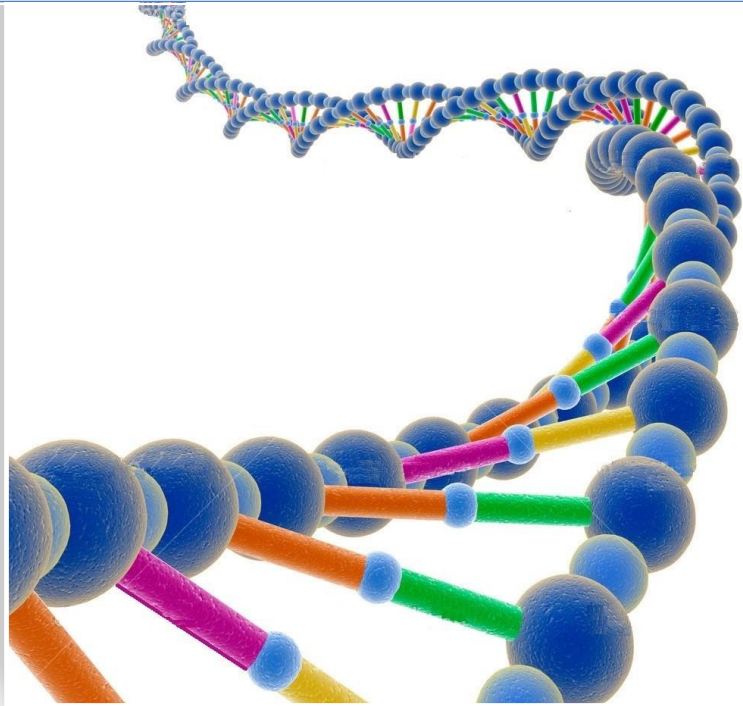
Al mismo tiempo, la saturación con marcadores comunes para ambas poblaciones podría permitir una comparación de sus resultados. Esta tarea se verá facilitada por el nuevo estudio de Ocaña et al. (2015), en el que se describe el

transcriptoma completo de la población 29H x Vf136 en su interacción con *Ascochyta fabae*. Dicho transcriptoma aporta una gran batería de SNP (single nucleotide polymorphism) que podría ayudar a saturar estas zonas de interés. Dicho trabajo se describe ampliamente en el capítulo III de esta tesis.

El análisis de estos nuevos marcadores junto a la información de los recientes mapas genéticos (Satovic et al. 2013, Webb et al. 2015), las ventajas de la colinealidad entre las habas y *M. truncatula* (Ellwood et al. 2008, Satovic et al. 2013, Webb et al. 2015) y el nuevo ensamblaje del genoma de *M. truncatula* (Mt 4.0) (Tang et al. 2014) ofrece una inmensa fuente de nuevas secuencias para el desarrollo de nuevos marcadores y la saturación de regiones de interés. La combinación de estudios genómicos y transcriptómicos debería aportar información suficiente para permitir la identificación de genes clave para los futuros programas de mejora en este cultivo.

# Capítulo III

## Large-Scale Transcriptome Analysis in Faba Bean (*Vicia faba* L.) under *Ascochyta fabae* Infection



**Este capítulo ha sido publicado como:**

Ocaña S, Seoane P, Bautista R, Palomino C, Claros GM, Torres AM, Madrid E (2015). Large-scale transcriptome analysis in faba bean (*Vicia faba* L.) under *Ascochyta fabae* infection. PLoS ONE 10(8): e0135143. doi:10.1371/journal.pone.0135143



### III. 1. Abstract

Faba bean is an important food crop worldwide. However, progress in faba bean genomics lags far behind that of model systems due to limited availability of genetic and genomic information. Using the Illumina platform the faba bean transcriptome from leaves of two lines (29H and Vf136) subjected to *Ascochyta fabae* infection have been characterized. De novo transcriptome assembly provided a total of 39,185 different transcripts that were functionally annotated, and among these, 13,266 were assigned to gene ontology against *Arabidopsis*. Quality of the assembly was validated by RT-qPCR amplification of selected transcripts differentially expressed. Comparison of faba bean transcripts with those of better-characterized plant genomes such as *Arabidopsis thaliana*, *Medicago truncatula* and *Cicer arietinum*, revealed a sequence similarity of 68.3%, 72.8% and 81.27%, respectively. Moreover, 39,060 single nucleotide polymorphism (SNP) and 3,669 InDels were identified for genotyping applications. Mapping of the sequence reads generated onto the assembled transcripts showed that, 393 and 457 transcripts were overexpressed in the resistant (29H) and susceptible genotype (Vf136), respectively. Transcripts involved in plant-pathogen interactions such as leucine rich proteins (LRR) or plant growth regulators involved in plant adaptation to abiotic and biotic stresses, were found to be differently expressed in the resistant line. The results reported here represent the most comprehensive transcript database developed so far in faba bean providing valuable information that could be used to gain insight into the pathways involved in the resistance mechanism against *A. fabae* and to identify potential resistance genes to be further used in marker assisted selection.

**Keywords:** Transcriptome, faba bean, *Ascochyta*, SNP, qPCR, candidate genes, genomics.





---

## III.2. Introduction

Faba bean (*Vicia faba* L.), one of the first domesticated plant species in Old World agriculture, is an important food crop worldwide and a source of dietary protein in developing countries. It ranks fourth in terms of cultivation after key food legumes such as chickpea, pea and lentil (<http://faostat.fao.org>). World production of dry faba beans remained stable in the last five years (4.2 million tons), with China and Ethiopia producing almost half. Traditionally grown in the Mediterranean basin, Middle East, China and Latin America the crop is currently achieving particular relevance in Australia (third world producer), Europe and North America. The main faba bean breeding objectives today include yield and seed quality improvement as well as resistance to biotic and abiotic stresses. Among them, broomrape (*Orobanche crenata*), a highly damaging parasitic weed, and Ascochyta blight, caused by *Ascochyta fabae* Speg., are the most serious threats to faba bean cultivation in most parts of the world.

The wild faba bean progenitor is unknown, and crosses with other *Vicia* species have proved unsuccessful (Cubero, 1984; Maxted et al. 1991), thus limiting genetic diversity to the primary gene pool. Faba bean has a diploid set of 12 chromosomes and a genome of 13,000 Mbp (Bennett et al. 1995) with a high proportion of repetitive DNA elements (Flavell et al. 1974). It represents, one of the largest legume genomes, exceeding 25 times that of the model legume *Medicago truncatula*. These biological aspects are the major challenges towards faba bean genome sequencing and associated marker development for efficient genomics-assisted breeding. In spite of these limitations, saturated genetic maps built with RAPDs (Random Amplified Polymorphic DNA), SSRs (Simple Sequence Repeats) and gene-based SNPs (Single Nucleotide Polymorphism) markers from *M. truncatula*, pea (*Pisum sativum* L.), lupin (*Lupinus albus* L.), lens (*Lens culinaris* Medik.) and soybean (*Glycine max* L. Merr), are now available (Torres et al. 2010; 2012; Satovic et al. 2013; Ma et al. 2013; Khazaei et al. 2014). These maps allowed the location of genes and QTLs (Quantitative Trait Loci) controlling faba bean resistance to different pathogens such as broomrape (Díaz-Ruiz et al. 2010; Gutierrez et al. 2013; Román et al. 2002), *A. fabae* (Avila et al. 2004; 2003; Díaz-Ruiz et al. 2009; Kaur et al. 2014; Román et al. 2003), rust (Avila et al. 2003) and other important agronomic traits as yield or drought adaptation (Khazaei et al. 2014; Cruz-Izquierdo et al. 2012).

Nevertheless, the genomic regions controlling those characters are still not well-characterized and no diagnostic markers have been developed either. So, the identification of markers with complete association with the QTL will boost the development of “perfect” markers in pulses (Kumar et al. 2011). Such markers are extremely useful for guiding the introgression of multiple resistant genes, because they increase selection efficiency and avoid recombination events between markers and QTLs (Hospital et al. 2009).

The success of marker assisted selection (MAS) for agronomically important traits will depend on the extent to which the underlying genes and pathways are identified and further used for the development of candidate gene markers (Madrid et al. 2012). However, faba bean genomics lags far behind that of model systems. This is reflected by the low number of faba bean ESTs (Expressed Sequence Tags) present in the dbEST database at NCBI (release 130101; 1 January 2013). Only 5,510 ESTs are available for this crop, compared with 1.5 million and 270,000 public entries for *Arabidopsis* and *M. truncatula* ESTs, respectively (<http://www.ncbi.nlm.nih.gov/dbEST>). Therefore, much effort is still needed to develop comprehensive faba bean genomic resources and to generate fast and cost-effective gene-based markers useful for molecular breeding or comparative genomic analysis.

The emergence of next generation sequencing (NGS) technology in the latest decade, is rapidly changing this scenario (Schuster et al. 2008). Compared to traditional sequencing methods, NGS yields a large number of sequences suitable to develop fast and cost-effective gene-based markers such as EST-based SSRs, SNPs or Indels (short insertions and deletions) and therefore, high-density genetic maps. Examples in faba bean include 2,397 (Kaur et al. 2012) and 28,503 SSRs (Yang et al. 2012) that are already available. Moreover, SNP assays using the KASPar™ (KBiosciences Competitive Allele Specific PCR) platform (LGC, UK) have been described (Cottage et al. 2012), and have been used to build a fairly dense genetic map for genomic-assisted breeding (Khazaei et al. 2014). Besides marker development, NGS offers an efficient way for wide transcriptome assemblies in crops with large and poorly characterized genomes, as faba bean is (Varshney et al. 2013). This approach provides an excellent resource of gene sequence information and expression data for identification of candidate genes associated with relevant traits. A preliminary faba bean transcriptome analysis (Kaur et al. 2012) identified a collection

of EST-derived SNPs that were used to develop a genetic map in a faba bean population segregating for ascochyta blight resistance (Kaur et al. 2014).

In order to increase the publicly available genomic data for this crop and to further assist in the development of allele-specific markers efficient in breeding selection, we performed a comprehensive characterization of the faba bean transcriptome derived from leaves of two lines (29H and Vf136) subjected to *A. fabae* infection, using the Illumina platform. Illumina reads together with the 105,094 faba bean sequences available in public databases were used for de novo transcriptome assembly. Results reported here will aid in identifying potential gene targets for use in marker assisted selection and represent the most comprehensive transcripts database reported so far in this crop species.

### **III.3. Material and methods**

#### **III.3.1. Plant material and *A. fabae* inoculation**

Two faba bean genotypes (29H and Vf136) were used for de novo transcriptome assembly. Line 29H has been described as resistant to *A. fabae* in several studies (Tivoli et al. 1998; Maurin et al. 1992; Bond et al. 1994; Sillero et al. 2001) and Vf136 belongs to the germplasm collection of the IFAPA Centro Alameda del Obispo in Córdoba (Spain), and is susceptible to ascochyta blight (Román et al. 2003).

Seeds of both genotypes were pre-germinated and sown with three replicates; three plants each, in 14 cm-diameter pots, using a 1:1 mixture of sand and peat. All plants were grown in a controlled condition chamber at 20-22°C. Inoculation was performed as described by Madrid et al. (2013) using a monoconidial isolate of *A. fabae*, CO99-01, originating from Córdoba, Spain. Non-inoculated replicated plants were included in the assay as controls. To confirm that *A. fabae* infection had been effective, inoculated plants were checked for expected disease symptoms at 15 days after inoculation and compared to controls.

#### **III. 3. 2. RNA extraction, library preparation and DNA sequencing**

As early defence responses occur shortly after contact with a pathogenic organism (Gururani et al. 2012; Pritchard et al. 1989), entire leaf tissue was taken from plants at 4, 8 and 12 hours after inoculation, and immediately frozen in liquid nitrogen. An

identical collection of leaf tissue was accomplished in the non-inoculated plants of the same lines. Total RNA isolation and cDNA synthesis were performed as described in a previous study (Madrid et al. 2013). Two cDNA libraries were generated from the pooled time point RNAs of each faba bean line (29H and Vf136). The library construction and sequencing was performed by GenXPro, Frankfurt am Main, Germany. Normalization of cDNA was performed as described by Zhulidov et al. 2004 and Bogdanova et al. 2008. The normalized cDNA was random fragmented and adapters for Illumina sequencing ligated. Products were amplified and size selected on agarose gel. Insert size of the fragments was 150-300 bp. Sequencing was performed on an Illumina GAII machine, using 1 x 100 bp reads. The sequence data generated in this study have been deposited at SRA (Short Read Archive) database under Experiment Accessions SRX690543 (29H) and SRX690544 (Vf136).

### **III. 3. 3. Sequence pre-processing, de novo assembly and annotation**

With the aim of constructing a faba bean transcriptome that integrates all known sequences in this species, 19,067 entries from European Nucleotide Archive (ENA), and 79,657 singletons and 6,370 contigs from Kaur et al. 2012 were incorporated. Illumina reads and these additional faba bean sequences were assembled following an updated version of the workflow described in Seoane et al. 2015. Differences with respect to the published pipeline are that (i) more ENA entries were incorporated, (ii) the pre-processing included more putative contaminant sources of microorganisms, (iii) Full-LengtherNext analyses can now detect and split chimeric transcripts, (iv) UniProtKB database for Full-LengtherNext analyses was updated on September 4th, 2014, and (v) the prediction of species-specific coding transcripts was carried out with TransDecoder (<http://transdecoder.github.io>). The resulting assembling was named as the v 1.1 of the *V. faba* transcriptome.

In addition to the annotations provided by Full-LengtherNext in the pipeline, the reconstructed transcripts were compared against the nucleotide sequence database TAIR 10 of *Arabidopsis* orthologue using BLASTN (Altschul et al. 1990) with a threshold *E-value* of  $10^{-10}$ . These orthologue identifiers were used for functional analyses using the AgriGo tool (<http://bioinfo.cau.edu.cn/agriGO>) (Du et al. 2010). A nucleotide comparison using BLASTN was also performed against genetically related legume species (*M. truncatula*, and *Cicer aritenum*). Finally, the protein sequences derived from the faba bean transcriptome were compared against the nucleotide

sequence database of chickpea using the tBLASTX (E-value $10^{-10}$ ) to derive other putative annotations. The resulting annotated transcriptome is provided in Supplemental File 1.

### **III. 3. 4. Variant analyses**

Identification of nucleotide variations (SNPs and InDels) was performed by separated mapping data from each genotype against the tentative transcripts using Bowtie v2.1.0 (Langmead et al. 2002). SNPs were called using bcftools and filtered using vcftools from the Li et al. 2009, using the default parameters, to obtain a subset of high quality SNPs and InDels.

### **III. 3. 5. Differential gene expression and functional enrichment**

The sequence reads obtained from resistant and susceptible genotypes were mapped separately against the reference transcriptome and then counted using Sam2counts.py. The count chart was loaded into the RobiNa software (Lohse et al. 2012) and analyses with the statistical method edgeR (Robinson et al. 2010). Reliability of differential expression required a P-value < 0.05. Finally, a functional enrichment analysis was conducted using KOBAS 2.0 (<http://kobas.cbi.pku.edu.cn/home.do>) with the Fisher's exact test based on the differentially expressed genes detected and known KEGG pathways, in order to identify genes potentially associated with phenotypic differences between lines.

### **III. 3. 6. Expression profiles via RT-qPCR**

Two  $\mu\text{g}$  of total RNA was reverse transcribed in duplicates from separate tissue samples for each time point from both genotypes and inoculated/non-inoculated plants using the M-MLV reverse transcription enzyme (Invitrogen, Carlsbad, CA, USA), in combination with oligodT (dT12–18) according to manufacture's instructions. cDNA synthesis and quality controls were performed as described previously by Madrid et al. 2013. Primers were designed based on their differential expression between both genotypes using the software package Primer 3 (<http://frodo.wi.mit.edu/primer3/>; Table 1) (Rozen et al. 1999).

**Table 1.** Primer pairs designed to validate RNAseq data by RT-qPCR and number of reads detected in each of the faba bean genotypes.

Gene homology	Primer ID	Forward primer (5'-3')	Reverse primer (5'-3')	Vf 136 Reads	29H Reads
Transcription factor NAI1	C-3026	GGAATCCGGAGAAAAT TGGCC	ATTCGAGCCAGGAA TGGTGG	2175	0
LEA-18 protein	C-15319	GGAACCATTGAAGGG CTTGC	GAGTAGGTGCCTCA GTTGCA	110	0
Jasmonate O-methyltransferase	L-45566	GTGCAACACCAGGCA GTTTT	TGAGCAAATTTTCT GGCGCC	1	121
F-box/LRR-repeat protein At3g59250	L-57706	CGGITCACCACCTTGG GTGT	ATGCATTGCCGAAA CCACAC	1	291

In order to normalize the data two reference genes, actin1 (ACT1) and cyclophilin (CYP2), were used (Gutierrez et al. 2011). The real-time quantitative PCR (qPCR) reactions used the iTaq™ Universal SYBR® (Biorad) according to manufacturer's instructions. qPCR amplifications were carried out in a 7500 HT sequence detection System (Applied Biosystems, Foster City, CA, USA) with the following temperature profile: initial denaturation at 95°C for 5 min, followed by 30 cycles of 95°C for 30 s and 60°C for 30 s (annealing and elongation). No-template controls were included. Amplicon quality was checked by an additional melting curve gradient with fluorescence measures after each temperature step. The amplification of the target genes at each cycle was monitored by SYBR green fluorescence. The Ct, defined as the PCR cycle at which a statistically significant increase of reporter fluorescence is first detected, was used as a measure for the starting copy numbers of the target gene. The efficiency of each primer pair was checked for all templates using LinReg software. The qPCR data were normalized with the relative efficiency of each primer pair.

### III. 4. Results

#### III. 4. 1. A new faba bean transcriptome

The construction of a complete faba bean transcriptome is the first step towards the identification of putative genes associated with resistance to *A. fabae* in this crop. Using RNA obtained from leaf tissue taken at 4, 8 and 12 hours after inoculation, a total of 33,023,160 raw sequences, with a read length of approximately 100 pb were generated. Approximately half of the sequences (16,567,244) belong to the 29H

parental line and the other half (16,455,916) to line Vf136. A total of 30,907,202 (93.59%) useful reads were submitted to different assembly procedures as described in Seoane et al. 2015. Additionally, 105,094 sequences from ENA and Kaur et al. 2012 were pre-processed to provide 87,269 useful reads to be assembled and merged with the previous assemblies. As a result, 98,195 tentative transcripts were obtained, with the longest contig having 5,439 nucleotides in length (Table 2). This figure is overestimating the number of genes in *V. faba*, although the number has been reduced from the previous version (v 1.0), indicating that the new ENA sequences have served to assemble previously fragmented contigs. The percentage of transcripts with annotation (40.12% in v 1.1; S1 File) did not differ significantly between the three transcriptomes, but the number of non-redundant orthologues increased, possibly as a result of the higher input of reads and the improvement of Full-LengtherNext. The 21,243 non-redundant orthologues in Table 2 are closer to the number of total genes in plants, as suggested for *Arabidopsis* (~ 27,200) (Lamesch et al. 2012).

From the 39,185 annotated transcripts, 44% (9,325) contained a complete open reading frame for a known protein, 6,593 being non-redundant full-length proteins. The numbers recorded are much higher than those provided by Kaur et al. 2012, but do not differ from v 1.0. A detailed analysis indicated that v 1.0 was contaminated with transcripts from microorganisms that are absent in v 1.1 (results not shown).

Although RNA-seq libraries were not designed to detect any ncRNA, due to the deep analysis provided by the Illumina Platform 2,254 putative ncRNAs (including microRNA precursors) were detected by comparing against miRBase (Kozomara et al. 2012), although these were not confirmed. Interestingly, a previous transcriptome report identified 134 putative ncRNAs by the same comparison criteria (Kaur et al. 2012).

The proportion of unknown transcripts was 60% in v 1.1, compared to 65% in v 1.0 and 52% in Kaur et al. 2012. This indicates that the update of sequences and software in the workflow described in Seoane et al. 2015 clearly improved the faba bean transcriptome. These transcripts can be faba/legume-specific transcripts (without a clear orthologue in the databases) or assembling artefacts. The TransDecoder analysis included in Full-LengtherNext revealed that 270 transcripts likely are faba/legume-specific and coding for a complete, unknown protein, while 579 transcripts could be coding for a complete protein or a protein fragment.



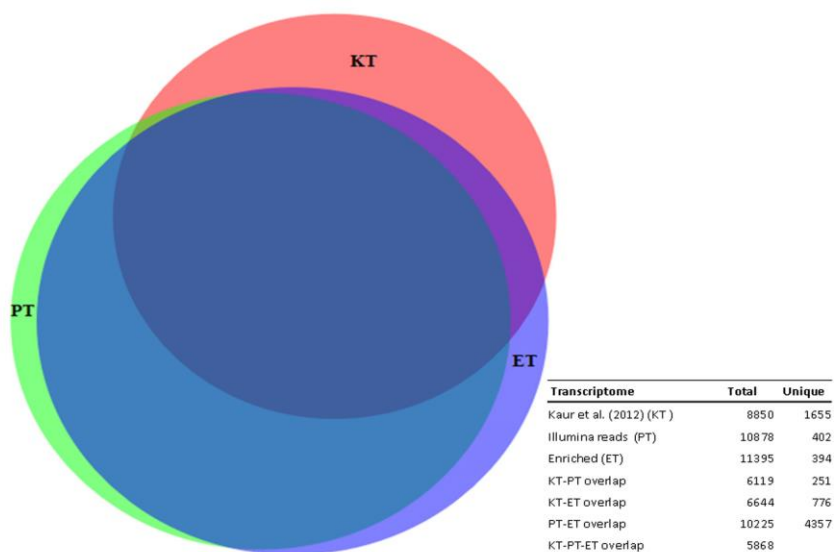
**Table 2.** Summary of the attributes assemblies obtained using three different datasets after the analysis using Full-LengtherNext. Transcriptomes of Kaur et al. 2012 and version 1.0 were analysed with older releases of Full-LengtherNext, while the current transcriptome (version 1.1) was analysed with a new release of this software.

Transcriptome attributes	Kaur et al. [23]	v 1.0 [36]	v 1.1 (this work)
Tentative transcripts	85,844	118,188	98,195
Transcripts with annotation	41,049	38,004	39,185
Unique IDs	18,871	20,413	21,243
Transcripts including a complete ORF	1,578	10,516	9,325
Unique complete ORFs	1,424	6,787	6,593
ncRNAs	134	2,789	2,254
Transcripts without orthologue	44,661	77,395	59,010
Coding	0	3,314	270
Putative coding	0	-	579
BA index	0,95	0,86	0,91

Based on the characteristics of the transcriptome (Table 1), the non-redundant *V. faba* protein-encoding transcriptome consists of 22,092 transcripts (21,243 unique orthologues, 6,593 of them coding for a complete, known protein and 270 + 579 for unknown transcripts). The subset of 21,243 non-redundant known transcripts was selected for further studies. This subset corresponded to the longest transcripts with different orthologous IDs, including those that code for a complete open reading frame (ORF).

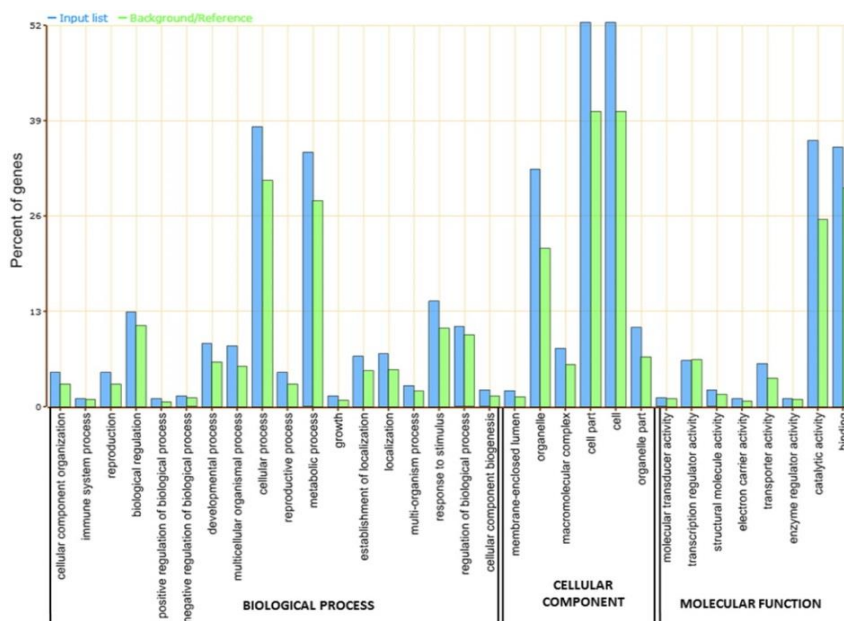
#### III. 4. 2. Functional classification of the *V. faba* transcriptome

A comparison among previous transcriptome data (KT) (Kaur et al. 2012) with the present transcripts assembled only from Illumina reads (PT) or enriched with database sequences (ET) was performed. These data were obtained from intermediary steps of the assembly workflow (results not shown). The *Arabidopsis* orthologues in the three subsets (KT, PT and ET) were used to assess the overlapping information with BioVenn (Hulsen et al. 2008). This comparison, revealed that the ET transcriptome represents 75% of KT (Kaur et al. 2012) and 95.5% of PT. The Venn diagram depicts the number of overlapping sequences and genes unique among the three transcriptomes (Figure. 1).



**Figure 1.** Venn diagram depicting the number of overlapping sequences and unique genes present among the three transcriptomes. KT: transcriptome developed by Kaur et al. (2012); PT: Illumina reads from the present study; ET: enriched transcriptome (Illumina reads and database sequences).

Faba bean transcripts were categorised by predicted function using AgriGO against *A. thaliana*. A total of 13,266 transcripts were classified based on the Gene Ontology (GO) hierarchy and assigned at least to one GO term. Transcripts were classified into 401 significant GO terms, of which 237, 83 and 81 belonged to the categories biological process, molecular function and cellular component, respectively. Among the biological process terms, cellular (38.2%) and metabolic processes (34.7%) were the most represented. Other biological processes such as biological regulation (12.9%), developmental process (8.7%), response to stimulus (14.4%) and regulation of biological process (10.9%) were also found (Figure 2). In the molecular function category a significant percentage of genes were assigned to binding and catalytic activity (35.4% and 36.3%, respectively). Other molecular functions found were transcription regulator activity (6.2%) and transporter activity (5.9%) (Figure 2). Finally, the higher number of cellular component annotations fell into cell part (52.4%), organelle (32.4%) and organelle part (10.8%) (Figure 2).



**Figure 2.** Distribution of the GO categories assigned to the faba bean transcriptome. Unique transcripts were annotated in three categories: biological process, cellular components and molecular functions.

### III. 4. 3. Sequence similarity of *V. faba* transcripts with other plants

The 21,243 faba bean transcripts were analysed for similarity/sequence conservation against the transcript data sets of *A. thaliana* and the legume species *M. truncatula* and *C. arietinum*, using BLASTX. The largest number of faba bean transcripts (17,265 accounting for 81.3%), showed significant similarity with the chickpea unigenes, followed by *Medicago* (15,457 transcripts; 72.8%) and *Arabidopsis* (14,506 transcripts; 68.3%). As reported previously (Doyle and Luckow, 2003) the analysis confirms the close phylogenetic relationship between faba bean, chickpea and *Medicago*. All of them are cool-season legumes, members of the Papilionoid subfamily and diverged from a common ancestor ~60 million years ago.

A tBlastx search was performed with chickpea, whose genome was recently sequenced and annotated, using the chickpea CDS information (GA\_v1.0) (Varshney et al. 2013) and the faba bean transcripts described in this study. We found that 17,129 (80.63%) of the faba bean transcripts are conserved in the chickpea transcript data set. Considering the high degree of conservation among legumes, it may be assumed that the assembly of the faba bean transcriptome may further be improved as more sequence data from other species become available.

#### III. 4. 4. Discovery of SNPs and InDels variants

A total of 44,145 putative variants were identified in the transcriptome dataset. After filtering, high confidence differences were obtained resulting in 39,060 SNP and 3,669 Indels (S2 File). The highest number of SNPs detected were C/T (7,320, 18.7%), followed by A/G (6,443, 16.5%), T/C (5,969, 15.3%) and G/A (5,566, 14.2%). The remaining SNP types account for less than 5%. Transitions were the most common SNPs (64.8%) compared to transversions (35.2%).

#### III. 4. 5. Transcripts involved in blight resistance

A comparison between the transcript data derived from the resistant and susceptible genotypes was performed by mapping the sequence reads generated onto the assembled transcripts. Out of the 21,243 transcripts, 850 (4%) showed significant differences between the two genotypes, 393 and 457 transcripts being overexpressed in the resistant and susceptible genotype, respectively. Moreover, 290 transcripts were only unique to the resistant line while 278 were only found in the susceptible line (S3 File).

In order to gain insights into the pathways involved in *Ascochyta* resistance in faba bean, transcripts that consistently showed significant expression differences between resistant and susceptible genotypes were grouped by KEGG IDs. Sixty-five KEGG pathways were identified in the resistant genotype (29H) and 94 in the susceptible one (Vf136). In addition, transcripts involved in 24 pathways related to pathogen resistance in both genotypes were identified, including those involved in biosynthesis of secondary metabolites, ethylene, phenylpropanoid and isoflavonoids.

In the resistant genotype (29H) several transcripts involved in plant-pathogen interaction were differently expressed (S3 File). These genes are leucine rich proteins (LRR) such as RGA2 or FEI1 and plant growth regulators with documented roles in adaptation to abiotic and biotic stresses such as abscisic acid, ethylene, jasmonic acid and salicylic acid. Examples include jasmonate O-methyltransferase, 1-aminocyclopropane-1-carboxylate oxidase, abscisic acid 8'-hydroxylase 3 and salicylic acid carboxyl methyltransferase, as well as several abscisic acid-induced proteins (HVA22 and stress-induced receptor-like kinase) or heat shock proteins (Hsp90, DNAJ/Hsp40 or Hsp70). The signaling pathway genes TGF-beta, p53 and EGF receptor (histone acetyltransferase and rac-like GTP binding protein) as regulators of reactive oxygen species production and cell death in plant species were

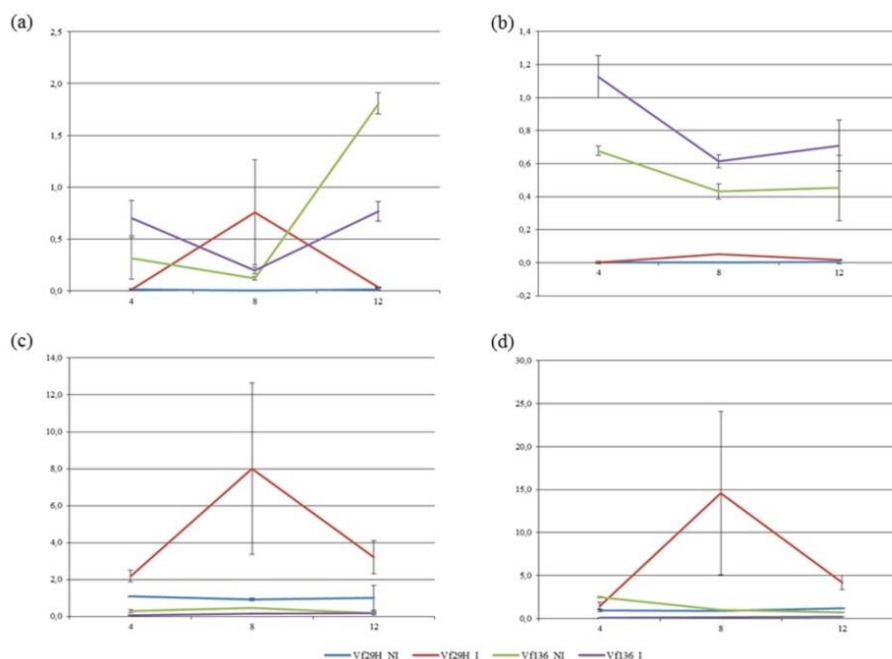
also represented. Enzymes involved in biosynthesis of secondary metabolites such as chlorogenic acid, scopoletin, suberin and phenylpropanoids; caffeoyl-CoA 3-O-methyltransferase, flavonoids (dihydroflavonol reductase) and chitin elicitor-binding protein were either identified.

A number of genes encoding proteins with essential roles in disease resistance and response were differentially expressed between the resistant and susceptible genotypes. Transcripts encoding NBS-LRR proteins (RGA2 and protein 10), enzymes involved in jasmonate and ethylene pathways (methyl jasmonate esterase and 1-aminocyclopropane-1-carboxylate oxidase) and heat shock proteins (Hsp90, Hsp70) were found in different read numbers. Remarkable is the detection of a transcript encoding MLO, a gene conferring broad spectrum powdery mildew resistance in monocots and dicots (Humphry et al. 2011), as well as a MYB-related transcription factor, several pathogenesis-related proteins (PR1a, PR5, a disease resistance response protein 206), a gibberellin induced protein and a CNGC5-like protein. Differential expression of calmodulin and aldehyde dehydrogenase 7a, two key regulators of the plant immune response, was also observed.

#### **III. 4. 6. Validation of Transcriptome by RT-qPCR**

To confirm the transcript abundance differences identified by the read counts, expression levels of four transcripts (transcription factor NAI1, LEA-18, Jasmonate O-methyltransferase and F-box/LRR-repeat protein At3g59200, Table 1) were measured by real time RT-qPCR. Both faba bean genotypes were analyzed independently at each time point and condition (inoculated and non-inoculated leaf samples). Transcripts were selected to represent putative candidate genes related to *Ascochyta* resistance and a wide range of expression profiles. NAI1 belongs to the helix-loop-helix (HLH) family and regulates ER body formation. ER bodies are enriched at the epidermis cells to protect plants from pathogens/herbivores that may enter or feed them. This is supported by the fact that leaf wounding triggers local and systemic de novo formation of ER bodies in a jasmonic acid (JA)-dependent manner (Nakano et al. 2014). LEA proteins have functional properties related to their presumed role as cellular stabilizers under stress conditions (Hundertmark et al. 2008). Jasmonic acid is a phytohormone involved in plants response against necrotrophic pathogens (Pieterse et al. 2009). Finally, the F-box genes are involved in the control of many crucial processes such as pathogen resistance, embryogenesis,

hormonal responses, etc. Relative transcript levels obtained by RT-qPCR analysis for the selected transcripts were consistent with those previously observed by RNA-seq. The differential expression between the two genotypes was confirmed in all cases (Figure 3) being the expression higher in the inoculated leaves.



**Figure 3.** Analysis of the kinetics of 4 transcripts analyzed by RT-qPCR in the resistant and susceptible genotype. (a) LEA-18; (b) Transcription factor NAI1; (c) Jasmonate O-methyltransferase; (d) F-box/LRR-repeat protein At3g59200 analyzed at 4, 8 and 12 hours after inoculation with *A. fabae*. Relative mRNA quantification was performed using ACT1 and CYP2, as reference genes for normalization.

### III. 5. Discussion

Faba bean genetic and genomic studies have been limited by the lack of genomic resources. Transcriptome assemblies allow the detailed comparative analysis across different genera and the discovery of functionally relevant markers. In crops such as faba bean, with a large and poorly characterized genome, comprehensive transcriptome assemblies offer a way to directly access the genes and the causative functional polymorphisms, yielding valuable insights about genome organization (Bohra et al. 2014).

The goal of this study was to generate a comprehensive faba bean transcriptome assembly with the final aim of detecting potential resistance genes for *Ascochyta fabae*. For this purpose, sequences from three different data sets were combined: newly obtained transcripts from Illumina sequencing, those obtained from FLX/454 reads (Kaur et al. 2012) and the ESTs sequences available at public databases. The completeness and quality of the new assembly was superior to earlier faba transcriptome assemblies (Kaur et al. 2012; Seoane et al. 2015). For instance, when the datasets were analyzed individually, a wide range of counts were reported: 85,844 transcripts from 304,680 FLX/454 reads (Kaur et al. 2012) and 98,195 transcripts from 30,907,202 Illumina reads. Hybrid assembly using a combination of different datasets was previously shown to be superior to that generated from a single sequencing platform (Schatz et al. 2010). Drawbacks of a single sequencing platform can be compensated by different characteristics of sequences obtained from other platforms, and the combination of both may help to correct sequence errors/biases improving the quality of draft assembly (Schatz et al. 2010). The enriched transcriptome obtained in the present study increased the number of annotated genes by 27.8% as compared with the previous report Kaur et al. 2012. Although a single pooled sample was used for each genotype and condition, the differences in gene expression of four tested transcripts determined by RT-qPCR were consistent with those obtained by RNA-seq, confirming the validity of the expression data. Further research is required to determine the potential function of the identified candidate genes in faba bean disease resistance.

BLAST searches against public databases allowed annotation for the 98,195 faba bean transcripts, yielding a set of 21,243 (54.21%) non-redundant transcripts with an orthologue gene ID. As reported in other legumes crops (Kudapa et al. 2014), comparison across species favors understanding of the biology and identification of genes in under-studied crops such as faba bean. This information may facilitate gene expression analysis and deliver evidence about gene content and function both of particular interest for the identification of candidate genes and the development of molecular markers.

The number of unigenes described in this study (21,243) is close to those estimated for diploid plants such as *A. thaliana* (25,000) (Bevan and Walsh, 2005) or *C. arietinum* (28,269) (Varshney et al. 2013). Interestingly, 270 transcripts likely encoding complete proteins did not show significant homology with any other

sequence in the database, indicating that these transcripts may represent faba bean specific genes. Lineage and species-specific genes have been identified previously in other plant species, including legumes (Graham et al. 2004; Lin et al. 2010). These genes could be interesting for further functional studies and may reveal novel legume-specific pathways. Moreover, knowledge on these genes may help to dissect species-specific cellular processes and to understand evolutionary processes such as speciation and adaptation.

The functional annotation obtained in faba bean was similar to that of *C. arietinum* (Hiremath et al. 2011) in terms of GO descriptions. The 14.4% of the annotated genes fell into “response to stimulus” subcategory. This result could contribute to our understanding of the global transcriptional changes occurring during infection by *A. fabae*. However, changes at the gene expression level are not necessarily a direct indication of the involvement of a gene in a biological process. Therefore, further genetic and functional analyses of differentially expressed transcripts are required to understand the biological significance of these changes in gene expression and their role in plant immune response.

Analysis of sequence conservation might help in the transfer knowledge from model plants to faba bean for functional genomic studies. As expected, lower similarity of faba bean transcripts was found with *Arabidopsis* compared with model legume species *Medicago* and chickpea. As previously reported (Kaur et al. 20014; Cruz-Izquierdo et al. 2012; Varshney et al. 2013), our results confirm the close phylogenetic relationship and genome conservation among the three galegoid cool-season legumes that derived from a common ancestor (Doyle et al. 2003). A large number (81%) of predicted faba bean proteins showed significant similarity with chickpea, indicating that their function might be conserved. Given the high degree of conservation among legumes, assembly of the faba bean transcriptome could be further improved as additional legume sequence data become available.

The present study identified a large set of potential SNPs between the lines 29H and Vf136 (44,146). Almost double rates of transitions in comparison with transversions were identified, comparable with results obtained in other plant species (Picoult-Newberg et al 1999; Nelson et al. 2011). Conversion of these SNPs into GoldenGate (Rostoks et al 2006) or KASP genotyping assays (<http://www.lgcgroup.com/products/kasp-genotyping-chemistry/>) will provide a low-cost and high-throughput marker genotyping system for accelerating their use in genetics



and breeding programmes. Mapping and colocalization of these SNPs in previously reported *A. fabae* resistance QTLs (Avila et al. 2004; Díaz-Ruiz et al. 2009; Román et al. 2003) will validate the functional relationship of these candidate genes and their future application in molecular breeding approaches.

Since the two sequenced genotypes showed differential resistance to *A. fabae*, and as DNA libraries were obtained from inoculated and non-inoculated leaves, genes with differential expression were identified in both genotypes. The results provided by this study are of interest to understand the transcriptional regulation of defense-associated genes as a first step toward understanding faba bean resistance to *Ascochyta blight* and improvement of disease resistance in plants (Dangl et al. 2001). Nevertheless, the nature of the signaling systems involved in resistance to major diseases in faba bean is still not well understood. The necrotrophic nature of *A. fabae* further complicates the elucidation of the resistance mechanism acting against this pathogen. As a result, only few studies have been reported to reveal the genes and metabolic pathways involved in this resistance (Madrid et al. 2013; Almeida et al. 2015; Fondevilla et al. 2011). In other pathosystems, however, a wide range of defense responses are known to be induced. These include preformed structural and chemical components, activation of the phytoalexin biosynthetic pathway, production of PR proteins, cell wall reinforcement mediated by hydrogen peroxide and detoxification of fungal toxins (Fondevilla et al. 2011). Our results suggest that several of these mechanisms may contribute to resistance to *A. fabae* in faba bean line 29H.

After perception and recognition of a pathogen, constitutive basal defense mechanisms lead to an activation of complex signaling cascades of defense. Thus, ion channels and kinase cascades are activated, reactive oxygen species (ROS), phytohormones like abscisic acid (ABA), salicylic acid (SA), jasmonic acid (JA), and ethylene (ET) accumulate in concert to reduce infection, and a reprogramming of the genetic machinery lead to defense reactions to minimize the biological damage caused by the stress (Pieterse et al. 2009; Rejeb et al. 2014; Derksen et al. 2013). The comparison between the resistant (29H) and the susceptible (Vf136) genotypes revealed significant differences in the expression of several genes related to these defence mechanisms. Thus, several phytoalexins (Dihydroflavonol-4-reductase) and a chitin elicitor-binding protein (CEBiP) were only expressed in line 29H (S3 File). Chitin recognition results in the activation of defense signaling pathways and CEBiP

is known to play a critical role in plant cells to mediate chitin perception and plant disease resistance (Wan et al. 2008). There was also evidence of alterations in cell wall metabolism indicated by overexpression of cellulose synthase biosynthesis genes. Other genes involved in the synthesis of methyl-jasmonate (jasmonate-O-methyltransferase), ethylene (1-O-aminocyclopropano-1-carboxylato), abscisic acid (abscisic acid 8-hydroxylase3) and salicylic acid (adenosyl-L-methionine: benzoic/salicylic acid carboxyl methyltransferase) were overexpressed or expressed only in the resistant genotype. Additional examples are provided by the ACC oxidase (only expressed in the resistance line) that catalyzes the final step of ethylene synthesis after induction by biotic or abiotic stresses (S3 File). Some of these findings were verified by RT-qPCR. Thus the enzyme Jasmonate O-methyltransferase, and a F-box/LRR-repeat protein were clearly activated only in line 29H. Conversely, two differentially expressed genes, NAI1 and LEA-18, which were found to be expressed in the susceptible faba bean line were not activated in the resistant line. Differentially expressed genes involved in these pathways point towards potential *Ascochyta* disease-resistance genes for future use in precision breeding.

Several pathogenesis related (PR) proteins were as well mainly expressed in the resistant genotype, suggesting that they may be key players in the faba bean-*ascochyta* interaction. PR proteins are indicators of a gene-for-gene resistance, associated with an immune response known as systemic acquired resistance (SAR) (Van et al. 1999; 2006). Resistance (R) genes are important for plant breeding purposes being ultimately responsible for activation of plant defense mechanisms (Dangl et al. 2001). Among these, the nucleotide binding site-leucine rich repeat (NBS-LRR) class is the most abundant and found in all types of plants (Kang et al. 2012). Our analysis identified a number of CC-NBS and NBS-LRR resistance proteins (such as RGA2, FEI1) that were highly expressed in the resistant parent. In addition, genes involved in secondary metabolism were also found to be expressed upon fungal infection. These metabolites are responsible in higher plants for generating metabolic intermediates directed into several pathways, including the lignin and flavonoid pathways (Dixon et al. 1995; Fofana et al. 2005). Legumes utilize flavonoids, notably isoflavones and isoflavanones, to defend themselves against physical injury and pathogens. Many of these effects appear to be related to their ability to modulate cell-signaling pathways (Rojas-Molina et al. 2007).

### III. 6. Conclusions

The present study contributes a non-redundant set of 21,243 transcripts in *V. faba* and provides a global view of genes expressed during the interaction with *A. fabae*. The assembly described here is the most comprehensive transcript database developed so far in this crop. Data set analysis revealed important features of the faba bean transcriptome such as gene annotation, assignment of functional categories and identification of SNPs. The large number of SNPs and InDels identified offer a cost-effective way to further develop functional markers for assisting breeding purposes. Furthermore, transcripts identified under specific categories like response to stimulus and enzyme classification, represent a valuable resource for faba bean crop improvement by identifying stress-response genes and genes involved in key metabolic pathways.

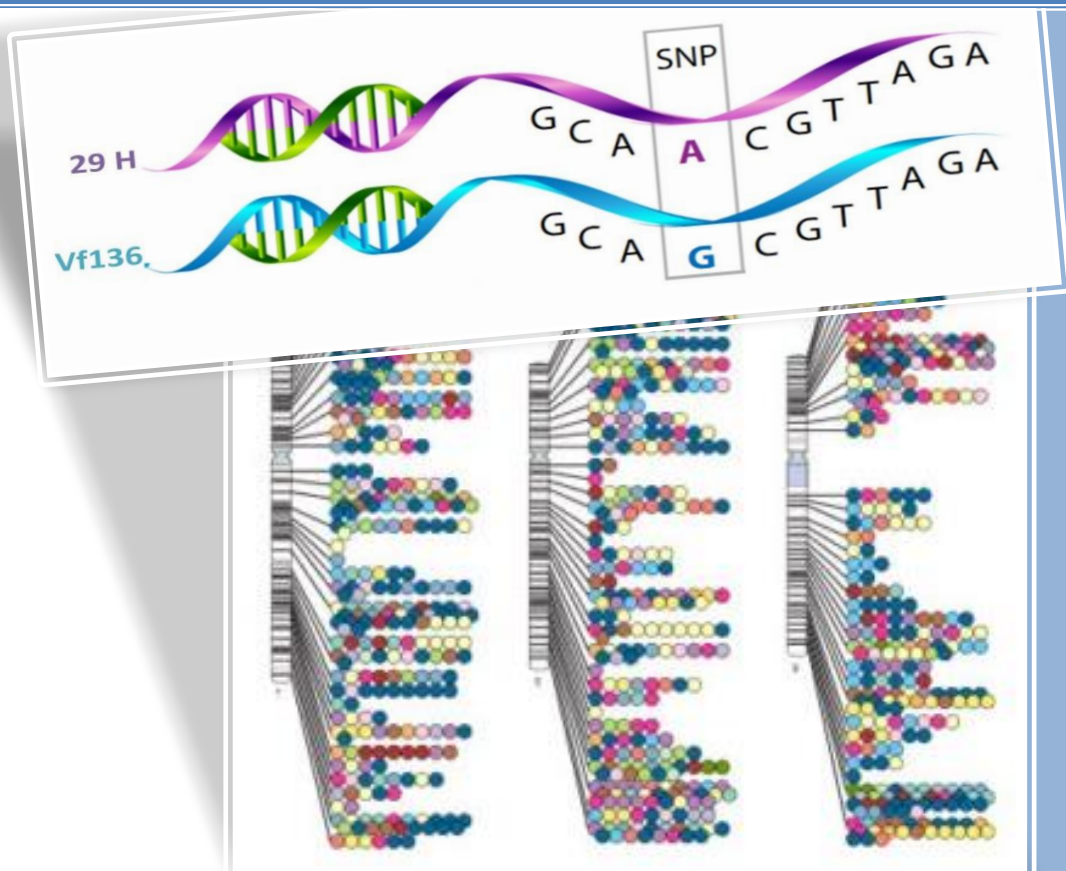
Different mechanisms and pathways involved in the resistance to pathogens such as chitin recognition, production of phytoalexins and PR proteins were present. We also identified putative candidate genes related to pathogen resistance that will be future targets for molecular breeding of *A. fabae* resistance in faba bean. The next steps will include the identification of polymorphisms in the candidate resistance genes to facilitate gene mapping and validation in forthcoming marker-assisted breeding programs.

### III. 7. Acknowledgements

The authors thankfully acknowledge GenXPro GmbH for RNA sequencing and the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Malaga. E. Madrid acknowledges the ‘Juan de la Cierva’ programme from the Spanish MINECO.

# Capítulo IV

## Saturation mapping of regions determining resistance to *Ascochyta* blight and broomrape in faba bean using transcriptome-based SNP genotyping



**Este capítulo ha sido enviado a Theoretical and Applied Genetics como:**

Ocaña S, Gutiérrez N, Torres AM, Madrid E (2017) Saturation mapping of regions determining resistance to *Ascochyta* blight and broomrape in faba bean using transcriptome-based SNP genotyping



#### IV. 1. Abstract

Faba bean is an important food crop worldwide. Marker assisted selection for disease resistance is a top priority in current faba bean research programs, with pathogens such as *Ascochyta fabae* and broomrape (*Orobanche crenata*) being among the major constraints in global faba bean production. However, progress in genetics and genomics in this species has lagged behind that of other grain legumes. Although genetic maps are available, most markers are not in or are too distant from target genes to enable an accurate prediction of the desired phenotypes. In this study, a set of SNP markers located in gene coding regions were selected using transcriptomic data. Ninety two new SNP markers were genotyped to obtain the most complete map reported so far in the 29HxVf136 faba bean population. Most of the QTL regions previously described in this cross were saturated with SNP markers. Two QTLs for *O. crenata* resistance (*Oc7* and *Oc8*) were confirmed. *Oc7* and *Oc10* colocalized with *A. fabae* resistance QTLs, suggesting that these genomic regions might encode common resistance mechanisms and could be targets for selection strategies against both pathogens. We also confirmed three regions in chromosomes II (*Af2*), III (*Af3*) and VI associated with ascochyta blight resistance. The QTLs ratified in the present study are now flanked by or include reliable SNP markers in their intervals. This new information provides a valuable starting point in the search for relevant positional and functional candidates underlying both types of resistance.

**Keywords:** *Vicia faba* L., disease resistance, transcriptome, differentially expressed genes, candidate genes, marker assisted selection.



## IV. 2. Introduction

Faba bean (*Vicia faba* L.) is an important legume crop cultivated on a global basis as a source of dietary protein, and ranks fourth in terms of cultivated area after the key food legumes chickpea, pea and lentil (<http://faostat.fao.org>). This crop plays an important role in management of soil fertility through crop rotation with cereals by fixing atmospheric nitrogen in the soil, thus providing disease breaks and contributing to agricultural sustainability. Major priorities in faba bean breeding programs include resistance/tolerance to biotic and abiotic stresses. Among the former, broomrape (*Orobanche crenata* Forsk.) and *Ascochyta fabae* Speg. represent the major constraints for this crop.

Broomrape is a highly aggressive parasitic plant that compromises faba bean production in the Mediterranean region, North Africa and the Middle East, with losses ranging from 50 to 80% (Gressel et al. 2004). Heavy field infestations force farmers to abandon the growing in traditional legume growing areas (Abu-Irmaileh 1994). *Orobanche* is a holoparasitic weed that obtains nutrients through haustoria, which connect to the host vascular system. Its control remains challenging because no fully satisfactory and economic crop management measures are available. The best long-term strategy for limiting damage is the development of *Orobanche*-resistant cultivars. Nevertheless, resistance against parasitic weeds is difficult to assess, scarce, of complex nature and of low heritability, making breeding for resistance a challenging task (Rubiales 2003). On the other hand, ascochyta blight caused by the fungal pathogen *A. fabae* is a disease prevalent in the Middle East, Europe, Canada, New Zealand, and Australia, that provokes between 35 to 90% yield losses in regions with cool and wet climate. Symptoms of infection include reduction of the photosynthetic area, pod and seed infection, seed staining and seed abortion, leading to yield reduction and reduced market value. Although the disease can be managed by application of chemicals and crop rotation, breeding of resistant cultivars is the most effective and sustainable means of control (Sillero et al. 2010).

The development of new varieties for complex traits controlled by QTLs (Quantitative Trait Loci) is a cumbersome task using traditional breeding methods. Marker-assisted selection (MAS) provides a potential solution to replace the difficult or expensive phenotype-based selection systems with markers linked to the genes of interest. The first faba bean genetic maps were developed with RAPD (Random



Amplified Polymorphic DNA) markers, which are difficult to transfer between laboratories and/or genetic studies. Subsequently, the use of ESTs (Expressed Sequence Tags) and the development of public databases have increased the possibility of transferring information from closely related legume species such as *Medicago truncatula*, *Pisum* sp., *Lens* sp. or *Cicer arietinum* to faba bean (Satovic et al. 2013). Recently, large sets of gene-based markers have been developed in this crop (Kaur et al. 2014; Webb et al. 2016), representing a breakthrough in the expansion of the faba bean genetic map.

Although several QTLs related to broomrape and *Ascochyta* resistance have been described and validated (Avila et al. 2004; Díaz-Ruiz et al. 2009; 2010; Gutiérrez et al. 2013; Kaur et al. 2014; Atienza et al. 2016), their utilization in breeding programs has not been implemented. It is imperative to saturate and narrow down these QTL intervals to identify the underlying candidate genes and most importantly to identify allele specific markers linked to the trait of interest. However, the large and repetitive nature of the faba bean genome has slowed the development of maps with sufficient density for these purposes. Moreover, most of the markers tagging QTLs are RAPDS, a type of marker not reliable enough to warrant MAS approaches.

The advances in sequencing technology with high-throughput SNPs genotyping platforms at low cost are promoting the development of high resolution maps in crop species with less genomic resources such as faba bean (Webb et al. 2016). SNP alleles have several advantages over other markers in terms of gene mapping. They are extremely stable and mostly identical-by-descent, thereby preventing scoring errors associated to homoplasy. Transcriptome profiling under biotic stresses (e.g *Ascochyta* infection) has been recently implemented in this crop (Kaur et al. 2012; Madrid et al. 2013; Ocaña et al. 2015), providing gene based SNPs that can be applied towards targeted genomic studies of disease resistance/tolerance.

The objectives of the present study were: (i) to construct a more detailed and informative genetic map of the faba bean recombinant inbred line (RIL) population 29H x Vf136, which varies in *ascochyta* blight and broomrape resistance and (ii) to refine the position of relevant trait-influencing QTLs with a more reliable and discriminatory marker set. The final aim was investigate the genetic basis of QTLs of interest by mapping differentially expressed candidate genes. To achieve this, a selection of SNPs obtained from recent transcriptome

analyses of two faba bean parental lines subjected to *A. fabae* infection (Madrid et al. 2013; Ocaña et al. 2015) were used to construct an improved map and to narrow down the genomic regions of the QTLs. Besides refining the position of target regions, this approach has potential to identify relevant candidate genes determining disease resistance.

### **IV. 3. Material and methods**

#### **IV. 3. 1. Plant material**

A mapping population comprising 119 F<sub>7:8</sub> individuals developed by single seed descendent from the cross between 29H × Vf136 was used in this study. The genotype 29H is a *minor* type faba bean developed at INRA (France) and it was originally provided by Drs Berthelem and Le Guen (Station d'Amélioration des Plantes, INRA-Rennes, France). This genotype has been described as resistant to *A. fabae* in different studies (Tivoli et al. 1987; Maurin et al. 1992; Bond et al. 1994; Sillero et al. 2001, 2010) and susceptible to broomrape (Gutiérrez et al. 2013). The male parent, Vf136, is an *equina* type obtained at IFAPA (Córdoba) from the cross Vf1071 × Alameda (Cubero et al. 1992). This genotype is susceptible to *A. fabae* and resistant to broomrape (Roman et al. 2003; Avila et al. 2004; Diaz-Ruiz et al. 2009, 2010; Gutierrez et al. 2013).

#### **IV. 3. 2. SNP selection and genotyping**

In order to identify SNPs likely linked to the disease resistance QTLs, we exploited a recent transcriptome analysis from the faba bean parental lines 29H and Vf136 in response to ascochyta blight (Ocaña et al. 2015). A set of putative SNPs associated with expressed sequence tags was chosen based on the following criteria: (i) 127 SNPs belonged to transcripts differentially expressed between the two parental lines upon infection ( $p < 0.05$ ); (ii) a second set of 40 SNPs were located in genes/proteins associated with defense/resistance response to pathogens, although the corresponding transcripts were not differentially expressed between the parental lines; (iii) 61 SNPs were chosen from a group of 162 tags previously described as differentially expressed in response to *A. fabae* infection (Madrid et al. 2013) and used for blast analysis against annotated genes in the transcriptome developed by Ocaña et al. (2015).

Detailed information of the SNPs selected is shown in additional file1. A total of 228 SNP were finally selected to be genotyped using Kompetitive Allele Specific PCR (KASPar) assays provided by LGC genomics ([www.lgcgenomics.com/genotyping/kasp-genotyping-chemistry](http://www.lgcgenomics.com/genotyping/kasp-genotyping-chemistry)) and the MassArray iPLEX (Sequenom) SNP typing platform at the Spanish National Genotyping Center facility of the University of Santiago de Compostela (<http://www.cegen.org>). KASP is a homogenous, fluorescence-based genotyping technology based on allele-specific oligo-extension and fluorescence resonance energy transfer for signal generation (Semagn et al. 2014). MassArray iPLEX consists of an initial locus-specific polymerase chain reaction (PCR), followed by single-base extension using mass-modified dideoxynucleotide terminators of an oligonucleotide primer that anneals immediately upstream of the polymorphic site (SNP) of interest. The distinct mass of the extended primer identifies the SNP allele (Gabriel et al. 2009).

#### **IV. 3. 3. Linkage map construction and QTL analysis**

The SNP markers genotyped in the study were incorporated to a previous dataset (Gutierrez et al. 2013; Atienza et al. 2016). Segregation of loci was analyzed for goodness of fit to the expected 1:1 ratio in the RIL population using the Chi square test. The new set of markers was analyzed for linkage using JoinMap 4.1 (Van Ooijen 2006). Markers were grouped using Maximum likelihood option at a minimum LOD score of 4.0 and maximum recombination fraction of 0.25 as general linkage criteria to establish linkage groups (LG). Recombination fractions were converted to centimorgans (cM) using the mapping function of Kosambi (1994). LGs were assigned, when possible, on the basis of commonality with already published linkage maps (Gutierrez et al. 2013; Satovic et al. 2013; Atienza et al. 2016). To overcome the lack of common markers with previous QTLs analyses, transcripts bearing mapped SNPs were blasted against the *M. truncatula* genome assembly version Mt4.0 (Tang et al. 2014).

QTL analysis was performed using the MAPQTL 5 software (Van Ooijen 2006). Interval mapping (IM) was used to identify putative disease resistance QTLs in each linkage group, considering the evaluations made in the 29H x Vf136 RIL F<sub>7:8</sub> population (Gutierrez et al. 2013; Atienza et al. 2016). Threshold for QTL significance at  $\alpha=0.05$  was estimated by permutation analysis using 1000 replicates (Churchill and Doerge 1994). The genome-wide LOD scores, corresponding to  $P =$

0.05, were considered as significance thresholds for the detected QTLs. The limits of the confidence interval of QTLs position were estimated at the positions where the LOD drop-off was equal to 2 (Li 2011). The presence of a QTL was established based on its P-value and co-location with QTLs for the corresponding phenotypic trait in another year or environment and when the difference with the LOD threshold was < 0.5. The coefficient of determination ( $R^2$ ) value of the best marker in a linkage group associated with resistant reaction was used to estimate the percentage of the total phenotypic variation explained by the QTL.

#### **IV. 3. 4. Field trials and Resistance scoring**

We used the *Orobanche crenata* and Ascochyta blight disease resistance scores previously reported by Gutierrez et al. (2013) and Atienza et al. (2016). Broomrape resistance was evaluated in naturally infested fields in Cordoba (Spain) in three consecutive seasons (2006-2009) and at Kafr El-Sheikh (Egypt) in 2007-2008. In addition, the population was evaluated for *Orobanche foetida* resistance at Beja (Tunex) during two growing seasons (2006-2008). Simple regression was carried out using the broomrape score in susceptible checks as an independent variable and the broomrape score in the RILs as a dependent variable to remove any statistically significant effects of field infestation variability. Regression corrected values (residuals) were then calculated to correct for differences in broomrape seed density in the soil between plots (Román et al. 2002; Diaz-Ruiz et al. 2010).

Ascochyta blight resistance was determined in three different experiments. The first two tests were carried out at seedling stage in growth chamber and against two monocolonial isolates, CO99 and LO98, collected in Córdoba and Logroño (Spain), respectively (Avila et al. 2004). Fifteen days after inoculation, disease scoring was performed separately on leaves (DSL, disease severity on leaves) and stems (DSS, disease severity on stems) based on the percentage of symptomatic leaf area. The third disease test was performed in field trials and adult plant at Córdoba during the season 2005-06 using a local isolate. The disease infection was determined on leaves (DSL), stems (DSS) and pods (DSP).

## IV. 4. Results

### IV. 4. 1. SNP selection and genotyping

A total of 228 SNP were initially selected for genotyping in the 29H x Vf136 RIL population. However, 136 had to be discarded for the following reasons: 62 SNPs (2 KASPar and 60 MassArray iPLEX data) were polymorphic in the parental lines but monomorphic in the RILs; 49 SNPs (2 KASPar and 47 MassArray iPLEX) showed skewed segregation or ambiguity, likely due to genotyping errors; finally, 25 SNPs did not meet the requirements for primer design preventing further analysis (Table 1). The remaining 92 SNPs (20 KASPar and 72 MassArray iPLEX) could be precisely genotyped in the mapping population (Additional file 1). The two most common SNP variants were A/G and C/T, representing 34,93% and 30,57% of the changes, respectively. The other SNP variants (T/G, C/G, A/C and A/T) accounted for less than 10% of the total (Additional file 1).

**Table 1.** Origin and characteristics of the SNPs assayed in the faba bean RIL population Vf29H x Vf136.

SNP transcripts	Genotyped	Homozygous	Genotyping errors	Discarded by design	Total
DE/Orthologous to <i>Cicer</i> or <i>Mt</i> CDS <sup>1</sup>	24	15	6	2	<b>47</b>
DE/ Not orthologous to <i>Cicer</i> or <i>Mt</i> CDS	13	15	34	18	<b>80</b>
DE/ Annotated <sup>2</sup>	36	17	6	2	<b>61</b>
Related with resistance processes <sup>1</sup>	19	15	3	3	<b>40</b>
<b>TOTAL</b>	<b>92</b>	<b>62</b>	<b>49</b>	<b>25</b>	<b>228</b>

DE: Differentially expressed. Mt: *Medicago truncatula*. CDS: coding sequences.

<sup>1</sup>Origin: SNP transcripts (Ocaña et al. 2015).

<sup>2</sup>Origin: SNPs identified in SuperSAGE tags (Madrid et al.2013).

From the 92 SNPs genotyped, 73 (79,35 %) belong to the set of SNPs differentially expressed upon ascochyta infection while 19 (20,65%) were not differentially expressed but associated with resistance processes (Table 1). Eighty-eight of these SNPs had homologues in the *M. truncatula* genome (Mtr4.0), while the remaining four may represent specific faba bean sequences and/or targets for resistance candidates. Moreover, five putative transmembrane proteins and four hypothetical proteins annotated in the *M. truncatula* genome but not assigned to known functional groups might correspond to genes involved in resistance.

#### IV. 4. 2. Linkage analysis

To build a more saturated map and refine the position of the QTLs, we considered the most recent linkage maps and the assessments for *A. fabae* (Atienza et al. 2016) and broomrape (Gutierrez et al. 2013) resistance reported in this population. The 92 SNPs markers were combined with the previous data set. Additional file 2 shows the list of these SNPs, their annotation in the *M. truncatula* genome (Mtr4.0) and their assignment to the different faba bean LGs or chromosomes (chr.). The new linkage analysis thus involved 307 markers of which 14,33% weren't fitting the expected 1:1 ratio. Of these, 215, many of which are RAPDs (136), have been genotyped in previous studies. Eighty SNPs were mapped and could be assigned to the six chromosomes of the species. The final map includes 257 loci distributed over 19 LGs and spanning 2796.91 cM (Fig. 1, Table 2). LGs with two markers were discarded. LG identity and orientation were determined by comparison with the *M. truncatula* genome, as well as from the use of previously map-assigned anchoring markers (Gutierrez et al. 2013; Satovic et al. 2013; Webb et al. 2016; Atienza et al. 2016). The largest group corresponding to chr. VI included 37 loci and spanned almost 445 cM, while the smallest group, with only three loci, covered 39.3 cM.

**Table 2.** Linkage groups, chromosome (Chr.) assignment and marker distribution in the 29H x Vf136 map.

LGs ID <sup>1</sup>	Chr.	Length (cM)	Number of mapped markers	Number of mapped SNP	Average marker density (cM)
1 (LG8)	I	98.98	11	4	9.00
2 (LG12+LG19)	I	121.887	13	5	9.38
3 (LG11)	I	59.147	8	2	7.39
4 (LG5+LG6+LG10)	I	384.548	32	6	12.02
5 (LG7+LG17)	I	171.516	13	2	13.19
6	I	29.353	5	4	5.87
7 (LG14)	IB	38.585	6	2	6.43
8 (LG4)	II	261.702	30	14	8.72
9 (LG29)	IIA	59.063	7	4	8.44
10 (LG2+LG9)	III	512.052	38	10	13.48
11 (LG13+LG16)	III	152.613	13	6	11.74
12 (LG18)	III	27.602	5	2	5.52
13 (LG3)	V	206.491	16	3	12.91
14 (LG1)	VI	444.134	37	8	12.00
15 (LG15)	-	40.609	6	3	6.77
16 (LG26)	-	39.302	3	0	13.10
17 (LG27)	-	57.408	5	2	11.48
18 (LG20)	-	48.262	5	2	9.65
19 (LG22)	-	43.659	4	1	10.91
<b>Total</b>		<b>2796.91</b>	<b>257</b>	<b>80</b>	<b>9.89</b>

LGs ID <sup>1</sup>: LGs identified in this study (19). Between brackets: correspondences with LGs described by Gutierrez et al. 2013.

The incorporation of new SNP markers allowed closing previous gaps between 11 minor LGs (Gutiérrez et al. 2013). These LG could now be combined into 5 chromosomal regions (Table 2). Accordingly, the large metacentric chr. I is now represented by six LGs, chr. III by three LGs and the remaining chrs. IB, II, IIA, V and VI by a single LG (Fig. 1). Five LGs (15, 20, 22, 26 and 27) could not be assigned to specific chromosomes. In agreement with Satovic et al. (2013), LG8 and LG6 were assigned to chr. I, while LG29 corresponds to chr. IIA. The resulting map is the most gene-enriched map published to date in this population.

#### **IV. 4. 3. QTL analysis for *Ascochyta fabae***

In agreement with Atienza et al. (2016), three QTL regions in chr. II, III and VI were related with *Ascochyta* resistance. The QTL in chr. II (*Af2*) was identified taking into account the disease severity scores in leaves (DSL), and was consistently detected in three years in both field and growth chamber evaluations (Field\_DSL, CO99\_DSL and LO98\_DSL), explaining between 10.6% (CO99\_DSL) and 19.1% (LO98\_DSL) of the total phenotypic variation (Fig. 1, Table 3). Chr. II was saturated with 14 new gene markers, and the QTL interval was enriched with two new SNPs, Contig 17163 (homologue to Medtr3g102120) and Contig 9100 (homologue to Medtr3g099010). Contig 17163 is annotated as a RYanodine receptor (SPRY) domain protein, which is likely involved in immune processes (D'Cruz et al. 2013), while Contig 9100 corresponds to a TB2/DP1, HVA22 family protein. Marker 14-3-3P (Medtr3g099380) was the nearest locus to the maximum LOD peak for this QTL. This gene is belonging to a family of conserved regulatory molecules reported in all eukaryotic cells with the ability to bind a multitude of functionally diverse signaling proteins (Fu et al. 2000).

According with previous studies (Avila et al. 2004, Atienza et al. 2016), a second region related with *Ascochyta* resistance (*Af3*) was detected in chr. III. Nevertheless, from the six QTLs reported considering field and growth chamber tests, only three (LO98\_DSL, LO98\_DSS, Field\_DSP) were significant in this study (Fig. 1, Table 3), explaining 15.7%, 11.2% and 12.4% of the phenotypic variation, respectively. Although chr. III was saturated with 10 markers, only one (Contig 19144/Medtr1g089810), was significant for *Af3* that remains flanked by OPD16\_1732/OPI01\_1985. Contig 19144 corresponds to a susceptible endoplasmic reticulum auxin-binding protein.

Finally, three QTL regions were detected in chr. VI explaining between 12% and 21.4% of the phenotypic variation (Fig. 1, Table 3). The former QTL region included Field\_DSP1 and LO98\_DSL evaluations and was enriched with two new SNPs, Locus 10591 (Medtr8g096020) a putative transmembrane protein and Contig 13425 (Medtr8g093250) corresponded to a cystathionine beta-lyase family protein.

**Table 3.** Putative QTLs for *O. crenata* (Oc) and *A. fabae* (Af) detected in the faba bean cross 29H x Vf136 by Interval mapping (IM) analysis.

QTLs	Environment	Chr.	QTL Interval <sup>1</sup>	LOD Threshold <sup>2</sup>	Max. LOD score	% Phenotypic variation (R <sup>2</sup> )
Oc7	Cor06/07	VI	OPU09_831/ OPI10_1178	3	7.86	34
Oc7	Cor07/08	VI	Contig9900/ OPA17_633	3.1	5.89	27.1
Oc7	Cor08/09	VI	Contig9900/ OPI10_1178	3.1	6.9	31
Oc8	Cor07/08	V	mtmt_GEN_00866_02 _1/ OPAC06_1034	3.1	3.25	17.7
Oc10	Kafr El-Shikh 07/08	III	OPB19_644/ Contig19144	3.2	2.02	12.4
Field_DSL (Af2)	Cor05/06	II	OPJ12_828/ Contig17163	3.3	3.23	13.1
Co99_DSL (Af2)	Cor99/00	II	OPJ12_828/ Contig9100	2.9	2.47	10.6
LO98_DSL (Af2)	Lo98/99	II	OPB13_1133/ Contig9100	3.2	3.5	19.1
LO98_DSL (Af3)	Lo98/99	III	OPP17_944/ SECA	3.2	3.54	15.7
LO98_DSS (Af3)	Lo98/99	III	OPD16_1732/ OPI01_1985	2.8	2.45	11.2
Field_DSP (Af3)	Cor05/06	III	OPP17_944/ OPI01_1985	3.1	2.66	12.4
Field_DSP1	Cor05/06	VI	OPI10_1178/ Contig17534	3.1	3.53	16.3
LO98_DSL	Lo98/99	VI	RNAR/ Locus_10591	3.2	2.99	12
Field_DSP2	Cor05/06	VI	RGA2_97/ OPL12_622	3.1	3.65	21.4

<sup>1</sup> Flanking markers considering 2-LOD support intervals except for Kafr El-Sheikh where 1-LOD support intervals was applied.

<sup>2</sup> LOD threshold was determined using permutations in all environments.  
Environment. Cor: Córdoba; Lo: Logroño.



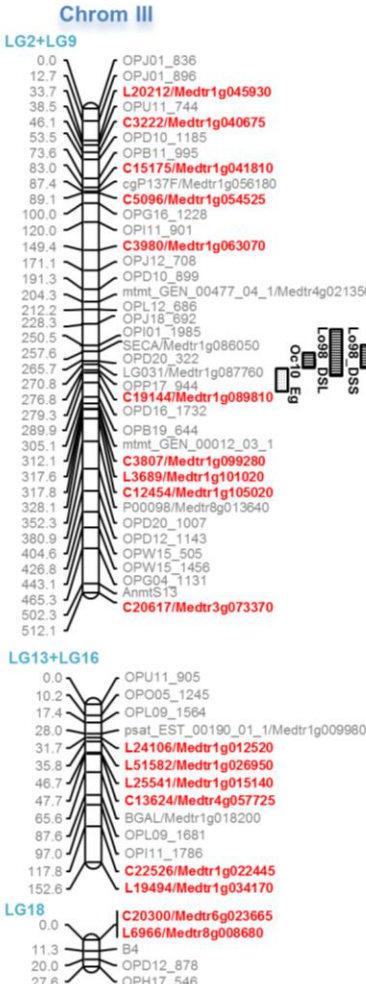
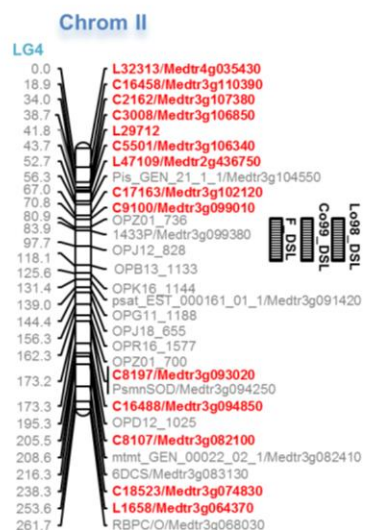
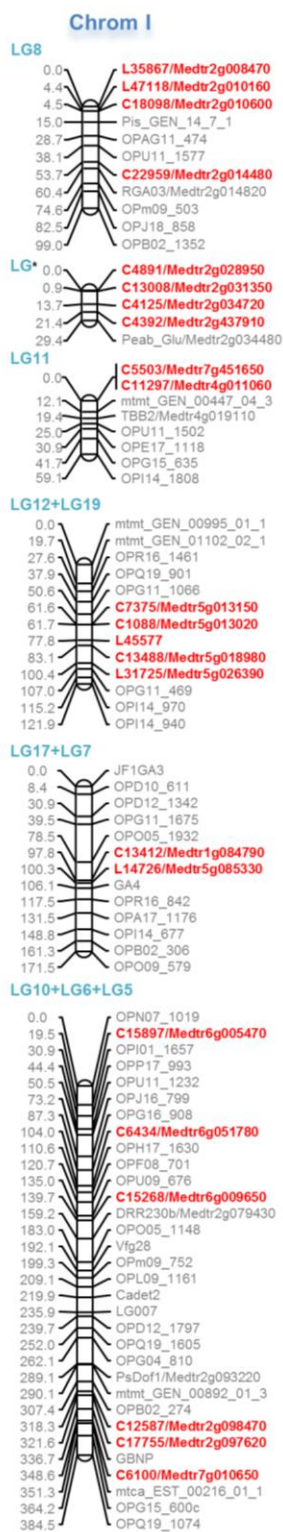
Finally, Field\_DSP2 was identified for the first time in this population. This QTL was located between RGA2\_97 (Medtr6g079800) and OPL12\_622. Within the QTL interval, we also located mmK1a (Medtr4g087620) annotated as a MAPK-like Ntf4 protein and mtmt\_GEN\_01109\_01\_1 (Medtr4g091610) coding for a hydrolyzing O-glycosyl compounds hydrolase.

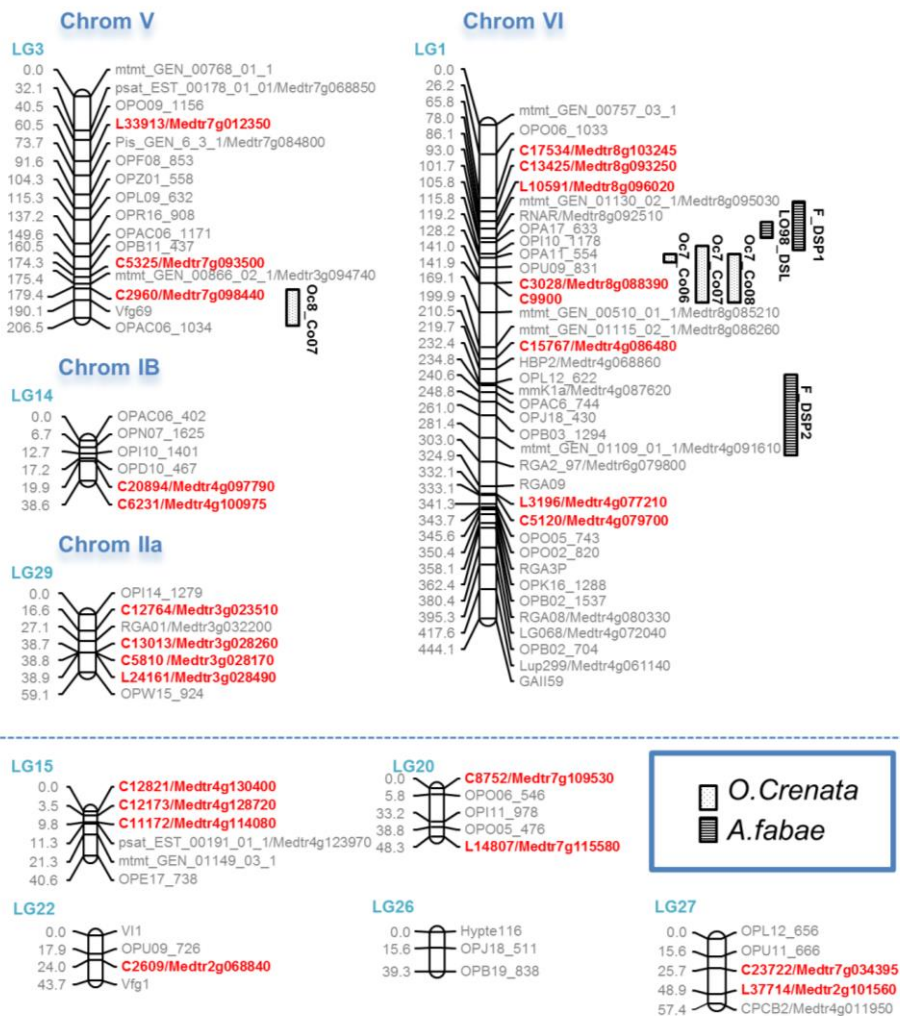
#### **IV. 4. 4. QTL analysis for *Orobanche* sp.**

As expected, previous stable QTLs for *O. crenata* (Gutierrez et al. 2013) were identified in this study (Table3). The former *Oc7* detected in chr. VI associated with *O. crenata* resistance was verified in three seasons (2006/07, 07/08 and 08/09), explaining between 27.1% and 34% of the phenotypic variation (Table 3, Fig. 1). The present analysis saturated chr. VI with eight SNP markers, two of which were placed within the QTL interval. Thus, the previous *Oc7* interval covered exclusively by RAPD markers has been enhanced by two gene coding sequences, Contig 9900 and Contig 3028. Contig 9900 was undetected in the *M. truncatula* genome suggesting that it might be part of a specific faba bean disease resistance gene. In contrast, Contig 3028 showed a significant match with the Medtr8g088390 gene, annotated as a protein that physically interacts with CBL-interacting serine/threonine-protein kinase 1 (CIPK1).

Another QTL for *O. crenata* resistance (*Oc8*) was detected in chr. V, but was limited to the season 2007/08. *Oc8* is flanked by two markers (mtmt\_GEN\_00866\_02\_1 and OPAC06\_1034) and accounted for 17.7% of the variation (Table 3). Contig 2960 (Medtr7g098440), one of the three new SNP markers included in chr. V, encodes a ras GTPase-activating protein involved in cellular organization and signaling processes related to cell growth, differentiation and apoptosis (Rojas et al. 2012).

Regarding the remaining *O. crenata* (*Oc9* to *Oc13*) and *O. foetida* (*Of3*, *Of4* and *Of5*) QTLs previously reported (Gutiérrez et al. 2013), none of them reached the genome-wide significance level in our analysis. Only *Oc10* in chr. III, associated to OPD16\_1732 was considered in this study, although the maximum LOD score did not reach the significant threshold obtained by permutation.





**Figure 1:** Genetic map and summary of the *Orobancha crenata* and *Ascochyta fabae* resistance QTL in the faba bean RIL population derived from the cross 29H x Vf136. *O. crenata* QTLs were named using the initials Oc followed by a number assigned in previous works and the identifier for each location (Eg, Egypt and Cor, Córdoba). For Cordoba analyses, a season identifier was also included (06, season 2006/2007; 07, season 2007/2008; 08, season 2008/2009). In the case of *A. fabae*, the QTLs were named considering the experiment (field, isolate Co99 in growth chamber, isolate Lo98 in growth chamber) and trait (DSL, disease severity on leaves; DSS, disease severity on stems; DSP, disease severity on pods). QTL positions are represented as 2-LOD bars.

## IV. 5. Discussion

### IV. 5. 1. Genetic linkage mapping and synteny analysis

The saturation of the 29H x Vf136 faba bean genetic map with new SNP markers located in coding sequences related with the response to *A. fabae* is described. From the 228 SNPs initially selected, only 92 could be successfully genotyped, in most cases due to the loss of polymorphism in the RIL population. Inaccurate SNP calling together with accidental cross-contamination by pollinators in any of the parental lines might be among the causes. Honeybees, bumblebees and large solitary bees are the main faba bean pollen vectors. Except from the honeybees that use rock cavities and hollow trees as nesting sites, the rest of pollinators usually construct nests below ground in holes or within the remains of previous fibrous root system (Mader and Hopwood 2013). Since faba bean multiplications are normally performed in insect-proof cages, the presence of some pollinator in the ground might have affected in this outcome.

The new map includes 257 loci distributed over 19 LGs. The merging process allowed us to join 11 previous minor LGs (Gutiérrez et al. 2013), that are now combined into five chromosomal regions. In agreement with the faba bean consensus map (Satovic et al. 2013), LG9 and LG16 which were previously associated to chr. I (Gutierrez et al. 2013), are now placed in chr. III. In addition, we could assign LG29 to chr. IIa, and LG8 together with a new LG to chr. I. The new linkage map has significantly longer cumulative genetic map coverage (2796.91 cM versus 1402.1 cM reported by Gutierrez et al. 2013). Map expansion is a common phenomenon in genetic research (Liu et al. 1996; Knox et al. 2002), that can be ascribed to several factors such as mapping strategies, number and type of mapped loci together with potential genotyping errors (Leonforte et al. 2013; Liu et al. 2015).

Resistance to ascochyta blight and broomrape are difficult traits in terms of genetics and breeding. The complexity of the disease evaluation and the polygenic nature of resistance to both pathogens make breeding for resistance a challenging process and an ideal candidate for MAS. Previous studies have described different stable QTLs for *A. fabae*, *O. crenata* and *O. foetida* resistance in different populations (Díaz-Ruiz et al. 2009; Kaur et al. 2014; Gutierrez et al. 2013; Atienza et al. 2016). However, most QTLs were flanked by RAPDs and is imperative narrow down these

QTL intervals with more reliable and transferrable molecular markers. Here we exploited the high degree of colinearity between faba bean and *M. truncatula* (Ellwood et al. 2008; Satovic et al. 2013; Webb et al. 2016) and submitted the selected SNP markers to homology searches against the *M. truncatula* genome (Mt4.0) using BLASTn (Tang et al. 2014). This approach detected significant matches in homologous regions that will facilitate future QTL comparisons between faba bean and different legume crops. QTLs controlling resistance to *Ascochyta* and *O. crenata* have been reported in chickpea and pea (Prioul-Gervais et al. 2007; Fondevilla et al. 2010; Millan et al. 2015). However, the lack of common markers in the corresponding maps has so far prevented the establishment of clear homologies. The macrosyntenic relationships reported between faba bean, *Medicago*, *Pisum*, *Lens* and *Cicer* suggest that some of the QTLs identified in the present study could be conserved between the Galegoid legumes.

#### **IV. 5. 2. QTL detection and candidate genes identification**

Three regions related with *Ascochyta* blight resistance were identified in chr. II, III and VI. *Af2* was located on chr. II using the phenotypic data for Field\_DSL, CO99\_DSL and LO98\_DSL. In a previous report (Atienza et al. 2016), *Af2* was flanked by RAPD markers, while the present work implemented the map with two new SNP markers (Contig 9100 and Contig 17163), that are now flanking the QTL. This led to the displacement of the QTL towards a position closer to QTL3 described in the population Icarus x Ascot (Kaur et al. 2014), and to *Af2* described in the RIL population Vf6 xVf136 (Diaz-Ruiz et al. 2009). The relative position of the orthologous genes in *M. truncatula* suggests that all these studies identified the same QTL region. The new gene markers in the *Af2* interval, Contig 17163 (Medtr3g102120), Contig 9100 (Medtr3g099010) and the already mapped1433P (Medtr3g099380), open the possibility for fine mapping and future development of MAS for this QTL. The area between Medtr3g102120 and Medtr3g099380 appears to be a good target for the identification of candidate genes for *ascochyta* blight resistance, narrowing down the region pinpointed by Atienza et al. (2016). Contig 17163 located in the QTL3 interval (Kaur et al. 2014) encodes a RYanodine receptor (SPRY) domain protein that regulates intracellular signaling (Ponting et al. 1997) and is likely involved in immune processes (D'Cruz et al. 2013). The human genome also encodes SPRY/B30.2 domains, several of which are involved in the immune

response. The role and function of the majority of SPRY/B30.2 domains remain unknown in plants but might include specific protein-protein interactions. Given the function of these proteins in humans, it is tempting to speculate that they might be involved in innate immunity against different infectious agents (Perfetto et al. 2013; Yap et al. 2005).

On the other hand, Contig 9100 corresponds to a TB2/DP1, HVA22 family protein. HVA22 is a unique abscisic acid (ABA)/stress-induced protein first isolated from barley aleurone cells and located in the endoplasmic reticle and Golgi apparatus, with a proposed role in seed germination and seedling growth regulation (Guo and Ho 2008). HVA22 is induced by environmental stresses such as dehydration, salinity, extreme temperatures and by the plant stress hormone ABA. ABA induces the accumulation of HVA22 protein that inhibits gibberellin (GA)-induced formation of large digestive vacuoles, an important process in GA-induced programmed cell death. Also, inhibition of vesicular trafficking involved in nutrient mobilization to delay coalescence of protein storage vacuoles has been proposed as part of its role in regulating seed germination and seedling growth (Guo and Ho 2008). Finally, 14-3-3P belongs to a family of conserved regulatory proteins that have the ability to bind a multitude of functionally diverse signaling proteins, including kinases, phosphatases, and transmembrane receptors, playing important roles in a wide range of vital regulatory processes such as mitogenic signal transduction, apoptotic cell death and cell cycle control (Fu et al. 2000). It is noteworthy that the three markers most closely associated to *Af2* correspond to expressed genes associated with cell death or stress response. Although additional studies are required for validation, the fact that both Contig 9100 and Contig 17163 were differentially expressed in response to *A. fabae* (Ocaña et al. 2015) is an indication that the genes identified within this QTL-containing region are interesting candidates for further studies and development of allele-specific markers.

The second QTL reported in chr. III (*Af3*) was detected in both growth chamber (LO98\_DSL and LO98\_DSS) and field evaluation (Field\_DSP). The remaining three regions reported by Atienza et al. (2016) were not identified, likely due to the different parameters used for QTL detection. Consistent with previous results (Avila et al. 2004; Atienza et al. 2016), *Af3* was flanked by OPD16\_1732/OPI01\_1985 and the QTL region included a new SNP marker (Contig 19144/Medtr1g089810). Contig 19144 was annotated as a susceptible endoplasmic

reticulum auxin-binding protein that regulates cell division, cell expansion, meristem activities, and root development (Gao et al. 2015). The lack of additional gene-derived markers in the QTL region prevents the discovery of potential candidate genes. It remains a challenge to reduce the  $Af3$  confidence intervals to enable the identification of strong candidates for *A. fabae* resistance.

Finally, three independent QTLs were detected in chr. VI. The first, involving Field\_DSP1 and LO98\_DSL evaluations is the same as described by Atienza et al. (2016). As mentioned above, this region is adjacent to *Oc7*, the main QTL identified for *O. crenata* resistance. Correlated responses to multiple diseases have been reported in different plant species (Jamann et al. 2014). Co-localization of Ascochyta blight and *Orobanche* resistance QTLs reinforces the idea of gene clusters with common mechanisms for resistance against both pathogens. This QTL region contains markers that could be directly involved in the resistance process, including *mtmt\_GEN\_01130\_02\_1* (Medtr8g095030) corresponding to a LRR receptor-like kinase (LRR-RKs) involved in the regulation of a wide variety of developmental and defense-related processes (Torii 2004) or RNAR (Medtr8g092510), an oxidoreductase involved in DNA synthesis. The region was also enriched with two new SNPs, Contig 13425 (Medtr8g093250) an aluminium resistance protein and Locus\_10591 (Medtr8g096020) encoding a putative transmembrane protein which might be involved in pathogen recognition or signaling in local defense responses.

Interestingly, the new QTL identified (Field\_DSP2), flanked by RGA2\_97 (Medtr6g079800) and OPL12\_622, likely corresponds to QTL4, a homologous region controlling *A. fabae* resistance in the Icarus x Ascot population (Kaur et al. 2014). RGA2\_97 is a putative NBS-LRR type disease resistance protein (Timmerman-Vaughan et al. 2000) reported as a candidate gene for quantitative resistance to *Mycosphaerella pinodes* in pea (Prioul-Gervais et al. 2007) while *mmK1a* (Medtr4g087620) was annotated as a MAPK-like Ntf4 protein. MAPK cascades play important roles in regulating plant growth, development, and responses to stress or defense stimuli (Pitzschke et al. 2009; Rodriguez et al. 2010; Tena et al. 2011; Meng and Zhang 2013). They function downstream of sensors/receptors to convert the signals into cellular responses. Biochemical studies in various plant species have demonstrated that MAPKs are activated rapidly in cultured cells or plants treated with elicitors or after pathogen infection, representing one of the earliest responses after the perception of invading pathogens. The rapid activation of MAPKs allows them to

regulate a variety of early, intermediate and late defense responses (Lee et al. 2004; Del Pozo et al. 2004). Within the QTL interval, we also located *mtmt\_GEN\_01109\_01\_1* (Medtr4g091610), a hydrolyzing O-glycosyl compounds hydrolase. Conversely *O. crenata* QTLs described previously (Gutierrez et al. 2013) were saturated in this study. *Oc7* explaining a substantial part of the variation for this trait in chr. VI was nearby of a QTL underlying ascochyta resistance and related with field disease resistance (Field\_DSP). This QTL co-localization is consistent with the idea that quantitative disease resistance in plants is conditioned by a range of mechanisms and could have considerable overlap with basal resistance (Poland et al 2011). The *Oc7* QTL interval was saturated with new SNP markers (Contig 9900 and Contig 3028). Contig 9900, undetected in the *M. truncatula* genome, could be associated specifically with *A. fabae* resistance, although this function remains to be elucidated. Contig 3028 (Medtr8g088390) is homologue to a *M. truncatula* gene coding for an evolutionarily carboxy-terminal region protein that physically interacts with CIPK1 (serine-threonine protein kinase) that interact with CBL proteins. CBL proteins represent important relays in plant calcium signaling (Kolukisaoglu et al. 2004). They form complex networks with their targets, the CBL-interacting protein kinases (CIPKs). Even though the function of most CIPKs in crop plants remains unknown, they are thought to play an important role in stress signal transduction and stress tolerance (Romeis et al. 2001; Tai et al. 2016).

According to Gutierrez et al. (2013), *Oc8* was detected in the distal part of chr. V but in our study the genome wide threshold was only reached in the season 2007/08. *Oc8* was flanked by *mtmt\_GEN\_00866\_02\_1/OPAC06\_1034*. Contig 2960 (Medtr7g098440), one of new SNP markers included in this QTL region, corresponds to a putative ras GTPase-activating binding-like protein that could be involved in signal transduction related to cell growth, differentiation and apoptosis. *Vfg69*, a microsatellite located in the QTL interval, derives from an inbred line of the *Orobanche* resistant cultivar Giza 402 (Zeid et al. 2009). Some evidence for the third QTL (*Oc10*) reported by Gutierrez et al. (2013), was detected in chr. III associated to the RAPD marker *OPD16\_1732* and flanked by Contig 19144. Nevertheless, this QTL did not reach the threshold criteria to consider it as a QTL, so more markers are needed on this region in order to increase the density in the genetic map and determine the region carrying *Oc10* accurately, due to this QTL co-localized with *Af3* underlying *A. fabae* resistance (Avila et al. 2004). Although *Oc10* has not been



validated yet, these findings suggest that this genomic region may be an interesting target for selection against both diseases. Among the three QTLs previously reported for *O. foetida* resistance (*Of3*, *Of4* and *Of5*) none of them turned out to be significant in our analysis. The inconsistencies in QTL detection may be due to differences in the mapping strategies, together with the low phenotypic variation explained by these QTLs and the relatively small size of the mapping population.

Implementation of molecular markers in MAS has rarely been achieved for complex traits such a disease resistance. In this work we exploited the transcriptome data recently published in faba bean (Ocaña et al. 2015) to saturate genomic regions controlling broomrape and *Ascochyta* resistance. In spite of the relatively low proportion of polymorphic markers that could be mapped, the new SNP markers located within expressed genes fall within or nearby all the QTL intervals. As a result, previous RAPD markers flanking QTLs can now be replaced by robust and transferrable SNPs, and the chromosomal intervals are now smaller than those previously reported. Our study shows that the stable QTLs in chr. II, III and VI can become amenable for MAS after further fine mapping and physiological trait dissection. Since most of the SNP-flanking QTLs were described to be differentially expressed during *Ascochyta* infection, they represent a starting point to search for positional and functional candidates in these QTL intervals. The combination of genomic and transcriptomic studies will thus expedite the identification of key genes for faba bean molecular breeding programs. The approach will aid to elucidate the molecular and biochemical mechanisms involved in resistance to the different faba bean pathogens. Further exploitation of the macrosyntenic relationships among legume crops, along with the advances in genotyping capacity offer the potential to predict and map new candidate genes for this trait. This information will be used for the development of linked and diagnostic polymorphisms for MAS and targeted introgression programs. The results reported here will also assist other trait-dissection studies and facilitate the transfer of information from related legume crops for future faba bean enhanced breeding.

#### **IV. 6. Acknowledgements**

This research was financed by the Spanish MINECO (project RTA2010-00059) co-financed by the EU through the European Regional Development Fund (ERDF) 2014-

2020 "Programa Operativo de Crecimiento Inteligente" and the IFAPA project PR.AVA.AVA201601.17 co-financed by ERDF. Partial funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under the grant agreement n°FP7-613551, LEGATO project is also acknowledged. S Ocaña-Moral acknowledges the FPI-INIA fellowship. The authors would like to thank Dr. CM Avila and Dr. SG Atienza for their helpful and critical comments.



# Capítulo V

## Discusión General





El principal objetivo de los mejoradores en cualquier cultivo, es incrementar y estabilizar su rendimiento mediante el desarrollo de cultivos adaptados a estreses ambientales. La selección asistida por marcadores (“Markers assisted selection”, MAS) es una alternativa a los problemas que se plantean en la mejora clásica como los derivados del efecto ambiental, dominancia o interacciones de tipo epistático durante los procesos de selección, así como procedimientos de fenotipado complicados y costosos. La MAS se basa en la identificación de marcadores estrechamente ligados a los genes o regiones genómicas (QTLs) que controlan caracteres de interés y nos permite la selección de las plantas portadoras del alelo más favorable. De esta forma es posible sustituir o asistir a la selección fenotípica de forma que sea más eficiente y fiable, aunque también hay que tener en cuenta que su coste compense al compararla con la Mejora convencional.

En habas, uno de los factores más limitantes de su producción es la susceptibilidad a distintos patógenos. Entre ellos cabe destacar la planta parásita *Orobanche crenata*, en toda la cuenca mediterránea y el hongo patógeno *Ascochyta fabae*, en todas las zonas de cultivo húmedas y templadas. El desarrollo de variedades resistentes a estos dos patógenos, es una tarea muy complicada debido a la naturaleza cuantitativa de la resistencia, controlada por numerosos genes sujetos a la influencia del ambiente. Además, el genotipado de marcadores moleculares y la saturación de sus mapas está dificultado por el gran tamaño de su genoma (~ 13.000 Mb) con más del 85% de ADN repetitivo. A esto hay que añadir la compleja evaluación fenotípica de estas dos enfermedades y que *V. faba* es una especie parcialmente alógama, donde el polimorfismo entre líneas parentales puede perderse por algún cruzamiento incontrolado. Todos estos factores han originado que el desarrollo de herramientas genómicas y la información asociada a ellas, esté por detrás del de otras leguminosas que cuentan, además, con la información completa de sus genomas como *Medicago truncatula* (Young et al. 2011), *Lotus japonicus* (Sato et al. 2008), *Phaseolus vulgaris* (Schmutz et al. 2014), *Cajanus cajan* (Varshney et al. 2012), soja (Schmutz et al. 2010) y garbanzo (Jain et al. 2013; Varshney et al. 2013).

A pesar de todas estas dificultades, los nuevos avances genómicos, tecnológicos (herramientas moleculares) y bioinformáticos en otras leguminosas, han permitido un mejor conocimiento del cultivo de habas a través del mapeo comparativo entre especies con estrecha relación filogenética (Echenique et al. 2004). Así, el desarrollo de marcadores moleculares basados en estudios de genómica

estructural, comparada y funcional está abriendo una nueva etapa en los programas de mejora tradicional, facilitando la construcción de mapas genéticos y la identificación genes o QTLs relacionados con la resistencia a patógenos y a otros caracteres de interés agronómico (Román et al. 2002; 2003; Ávila et al. 2004; Díaz-Ruiz et al. 2009 a y b; 2010; Gutiérrez et al. 2013; Satovic et al. 2013; Kaur et al. 2014; Webb et al. 2015; Atienza et al. 2016).

Debido al desconocimiento de la base genética de los mecanismos de resistencia frente a estos dos patógenos (*A. fabae* y *O. crenata*), los objetivos de esta tesis se han centrado en ampliar los datos moleculares disponibles para el control de estos mecanismos y en la identificación y desarrollo de marcadores moleculares en genes situados en regiones genómicas portadoras de QTLs asociados a la resistencia a ambos patógenos (*A. fabae* y *O. crenata*). El objetivo final es facilitar el desarrollo de marcadores diagnóstico que permitan extrapolar los resultados derivados de QTLs estables de unas poblaciones a otras y faciliten la selección eficiente de materiales resistentes en los programas de mejora. Para ello en el capítulo II, se ha recurrido al principio de macrosintenia (Choi et al. 2004; Kaló et al. 2004; Zhu et al. 2005; Phan et al. 2006; Ellwood et al. 2008; Nayak et al. 2010; Cruz-Izquierdo et al. 2012) y al uso de herramientas de genómica estructural y comparada para analizar 116 marcadores moleculares procedentes de regiones homólogas en diferentes leguminosas bien caracterizadas (*P. sativum*, *M. truncatula*, *L. culinaris*, *C. arietinum* y *V. faba*) en dos poblaciones RILs (29H x Vf136 y Vf6 x Vf136) de habas que segregan para la resistencia a *Ascochyta* y jopo. La transferencia de información entre leguminosas nos ha ayudado a saturar, actualizar y refinar la posición de QTLs previamente identificados relacionados con la resistencia a ambos patógenos. Los nuevos marcadores introducidos en los mapa genéticos se han desarrollado en secuencias codificantes, por lo que son marcadores mucho más estables y transferibles entre laboratorios que los marcadores RAPD mayoritarios en los mapas previos de esta dos poblaciones y que han limitado su comparación con otras leguminosas.

De los 116 marcadores analizados solo se pudieron genotipar 12 en el cruce 29H x Vf136 y 22 en Vf6xVf136, probando que el genotipado de marcadores en *V. faba* es una tarea muy complicada y el nivel de transferencia se ve limitado por manejar secuencias que no están suficientemente conservadas entre especies. A pesar de esto, la incorporación de nuevos marcadores ha permitido la unión de distintos

grupos de ligamiento en otros de mayor tamaño y ha contribuido al desarrollo de un mapa consenso en habas (Satovic et al. 2013; Anexo I) que integra toda la información genómica generada hasta ahora por el grupo. La actualización de los dos mapas genéticos ha permitido ampliar su cobertura genética (4,970.92 cM en Vf6 x Vf136 y 2,125.39 cM en la población 29H x Vf136) y saturar con marcadores robustos la mayoría de QTLs descritos previamente. Concretamente en la población 29H x Vf136, la inclusión de nuevos marcadores ha permitido saturar tres QTL previos relacionadas con la resistencia a *A. fabae* (*Af2*, *Af3*, *F\_DSP1*) e identificar a *F\_DSS* y *Lo98\_DSL* en la misma región que *F\_DSP1*. Además, se ha identificado un nuevo QTL no descrito anteriormente en el cromosoma VI (*F\_DSP2* y *Lo98\_DSL*). Para *O. crenata* se ha validado *Oc7* (chr. VI) y *Oc8* (chr. V), no identificándose *Oc9-Oc13*, así como ningún QTLs descrito para *O. foetida*, detectados por Gutiérrez et al. (2013). La variación de parámetros en el análisis, el uso de distintos programas y la aplicación de diferentes criterios a la hora de detectar QTLs podrían explicar las diferencias en esta identificación. En la población Vf6 x Vf136, la inclusión de nuevos marcadores ha permitido validar dos regiones relacionadas con la resistencia a *A. fabae* (*Af1* y *Af2*) e identificar tres nuevas (*Af3\_DSS*, *Af4\_DSS* y *Af5\_DSL*) no detectadas por Díaz-Ruiz et al. (2010). Para *O. crenata* solo se han detectado dos QTLs (*Oc2* y *Oc4*) y para *O. foetida* ninguno, debido a la aplicación de diferentes criterios en la identificación de QTLs en ambos trabajos.

Aunque nuestro estudio aporta nuevos marcadores genéticos que colocan o flanquean algunos QTLs, el restringido número de marcadores comunes introducidos en las dos poblaciones limita la comparación directa de los resultados, por lo que se requiere una mayor saturación con marcadores relacionados con procesos de resistencia que ayuden a unificar y asignar todos los grupos de ligamiento e identificar posibles genes candidatos de resistencia. Además, la saturación con nuevos marcadores ayudará a detectar dichos QTLs con valores más altos de LOD. Concretamente, en la población 29H x Vf136 es de gran interés saturar el chr. VI para unificar los dos grupos de ligamiento que forman este cromosoma, por contener a *Oc7*, validado en diferentes ambientes con un alto porcentaje de la variación fenotípica y por colocalizar con *F\_DSP1* detectado para *A. fabae* lo que sugiere mecanismos de resistencia comunes para ambos patógenos.

Por otro lado, el uso de herramientas de genómica funcional en habas (Capítulo III) ha permitido caracterizar el transcriptoma completo en respuesta a la



infección de *A. fabae* en dos líneas parentales de habas (29H y Vf136), dicho transcriptoma ha permitido la caracterización de un gran número de transcritos (39.185) que podrían ayudarnos a identificar genes asociados con la resistencia a este patógeno. La comparación de estos transcritos con genomas vegetales mejor caracterizados (*A. thaliana*, *M. truncatula* y *C. arietinum*) nos ha permitido profundizar en la interacción planta-patógeno, así como identificar marcadores moleculares funcionalmente relevantes que podrán ayudar al futuro desarrollo de marcadores útiles en MAS. De los 39.185 transcritos, 21.243 se identificaron como transcritos no-redundantes con un gen ortólogo en las bases de datos. Este número se aproxima bastante al número de genes estimado en plantas diploides como *A. thaliana* (25.000) (Bevan y Walsh, 2005; Lamesch et al. 2012) o *C. arietinum* (28.269) (Varshney et al. 2013). Además, se identificaron 6.593 transcritos que codifican proteínas completas conocidas y 270 que codifican proteínas completas desconocidas que podrían corresponder a proteínas específicas de habas. Estos genes pueden ser interesantes para estudios funcionales ya que podrían revelar procesos celulares específicos de la especie y de su adaptación al medio.

Como se ha comentado anteriormente, en habas se desconoce la base genética de los mecanismos de resistencia frente a estos dos patógenos. En otras especies, se conoce que tras el reconocimiento de un patógeno se activan complejas cascadas de señalización de defensa que activan canales iónicos, cascadas de proteínas MAPK, especies reactivas de oxígeno, óxido nítrico, distintas fitohormonas (ácido abscísico, ácido salicílico, ácido jasmónico y etileno) y se expresan proteínas relacionadas con patogénesis (Pieterse et al. 2009; Rejeb et al. 2014; Derksen et al. 2013; Zipfel et al. 2008). Con la finalidad de caracterizar genes y mecanismos relacionados con la resistencia a *A. fabae*, se realizó un análisis de expresión diferencial entre el genotipo resistente y susceptible que identificó diferencias significativas en 850 transcritos. Entre los resultados, encontramos transcritos que codifican proteínas relacionadas con la activación de mecanismos de defensa (Dangl et al. 2001) como son las proteínas PR (PR1a, PR5, FEI1, RGA2), un factor de transcripción relacionado con MYB, varias enzimas y proteínas implicadas en la síntesis de fitohormonas (jasmonato-O-metiltransferasa, 1-O-aminociclopropano-1-carboxilato oxidasa, ACC oxidasa, la ácido abcísico 8-hidroxilasa3, ácido salicílico carboxil metiltransferasa, HVA22 y quinasa similar a un receptor inducido por estrés), fitoalexinas (Dihidroflavonol-4-reductasa) y una proteína de unión al elicitor de

quitina (CEBiP) necesaria para que el reconocimiento de quitina active vías de señalización de defensa y resistencia a enfermedades (Wan et al. 2008). También se identificaron transcritos que intervienen en la formación de la pared celular (sintasa de celulosa), se encontraron flavonoides (dihidroflavonol reductasa) y enzimas implicadas en la biosíntesis de metabolitos secundarios (ácido clorogénico, escopoletina, suberina y fenilpropanoides; Caffeoyl-CoA 3-O-metiltransferasa); además de un transcripto que codifica MLO, un gen que confiere resistencia al oídio en monocotiledóneas y dicotiledóneas (Humphry et al. 2011), una proteína inducida por giberelina, una proteína similar a CNGC5, dos reguladores de la respuesta inmune en plantas (calmodulina y aldehído deshidrogenasa 7) y genes de la vía de señalización TGF-beta, p53 y EGF receptor (histona acetiltransferasa y proteína de unión GTP) reguladores de la producción de especies reactivas de oxígeno y de la muerte celular en plantas.

Todos estos datos han proporcionado una información muy valiosa sobre posibles genes candidatos y las vías involucradas en los mecanismos de resistencia contra *A. fabae*, que podrán ser utilizados en futuros estudios de selección asistida por marcadores. Además, el nuevo transcriptoma ha identificado un alto número de SNPs (39.060) que ha servido en el capítulo IV para seleccionar un subconjunto de ellos (228) expresados diferencialmente entre ambos genotipos (resistente y susceptible) y construir un nuevo mapa de ligamiento mucho más saturado que ha refinado la posición de QTLs detectados en estudios previos, relacionados con la resistencia a *A. fabae* y *O. crenata* (Atienza et al. 2016; Gutierrez et al. 2013).

Debido a la pérdida de polimorfismo detectada en la población RIL, debida a posibles errores informáticos en la detección de SNPs y a una posible contaminación accidental por polinizadores en alguna de las líneas parentales, solo se han podido genotipar 92 SNPs. Así, el nuevo mapa de ligamiento en la población 29H x Vf136 contiene 257 marcadores de los cuales 92 (80SNPs + 12 ESTs) fueron incluidos en este estudio, obteniéndose así el mapa más completo descrito hasta ahora en esta población. Dichos marcadores están distribuidos en 19 GL y cubren una distancia bastante mayor (2796,91 cM) que el mapa previo (1402,1 cM, Gutierrez et al. 2013) seguramente debido a las distintas estrategias de cartografía, número y tipo de loci mapeados o la acumulación de sobrecruzamientos durante la meiosis (Leonforte et al. 2013 Liu et al. 2015). Los nuevos SNPs genotipados (92) se compararon con el genoma de *M. truncatula* (Mtr4.0) identificándose 88 genes ortólogos. Los cuatro

marcadores no detectados en *M. truncatula* podrían representar secuencias específicas de habas y/o ser futuros genes candidatos de resistencia. Las coincidencias entre regiones homólogas facilitarán futuras comparaciones de QTL entre habas y leguminosas, ya que hasta el momento, la mayoría de QTLs descritos en habas para la resistencia a *A. fabae* y *O. crenata* (Díaz-Ruiz et al. 2009, Kaur et al. 20143, Gutiérrez et al. 2013, Atienza et al. 2016) estaban flanqueados por marcadores RAPD que dificultaban las comparaciones directas.

Este estudio confirmó dos QTLs descritos para *O. crenata* (*Oc7* y *Oc8*; Gutiérrez et al. 2013) y cuatro para *A. fabae* (Atienza et al. 2016) y consigue una mayor saturación de estas zonas al introducirse en ellas marcadores que podrían estar relacionados con mecanismos de resistencia. Además, se sugiere la región donde se localiza *Oc7* (chr. VI) como una buena diana para comenzar el desarrollo de marcadores moleculares para MAS debido al alto porcentaje de variación fenotípica explicado por esta región durante tres temporadas consecutivas y por colocalizar con otro QTL identificado asociado a la resistencia a *Ascochyta* (Field\_DSP), lo que podría confirmar la existencia de mecanismos de resistencia comunes frente a distintos patógenos en plantas (Jamann et al. 2014). Además esta zona podría corresponder a *Oc2* detectado en otra población de habas (Román et al. 2012; Díaz-Ruiz et al 2010). Por último, este QTL ha pasado de estar formado por marcadores RAPD a tener dos marcadores desarrollados en secuencias codificantes (Contig 9900 y Contig3028) que podrían estar relacionadas con la resistencia. Resulta interesante destacar que uno de los marcadores, Contig 9900, no tiene homología con ningún gen caracterizado en el genoma de *M. truncatula*, por lo que podría ser un gen específico de resistencia a la enfermedad en habas. El otro marcador identificado en la región (Contig3028) codifica una proteína evolutivamente conservada que interactúa con CIPK1 y que podría desempeñar un papel importante en la transducción de señales de estrés y tolerancia al mismo (Rome et al. 2001).

El interés de *Oc8* radica en su posición (parte distal del chr. V) y por tener en su intervalo de confianza un microsatélite (Vfg69) derivado de una línea consanguínea del cultivo resistente a *Orobanche* (Giza 402; Zeid et al. 2009). El interés de esta zona aumenta con la introducción de un nuevo marcador (Contig2960) que codifica una proteína activadora de ras GTPasa que podría estar implicada en la transducción de señales relacionadas con el crecimiento, la diferenciación y la apoptosis celular (Rojas et al. 2012).

En el caso de resistencia a *A. fabae*, sería interesante seguir saturando los cuatro QTLs detectados para refinar su posición. La inclusión de dos nuevos marcadores SNPs (Contig9100 y Contig17163) en *Af2* (chr. II; Field\_DSL, CO99\_DSL y LO98\_DSL) ha originado el desplazamiento de este QTL hacia una posición más cercana al QTL3 descrito en la población Icarus x Ascot (Kaur et al. 2014) y hacia *Af2* descrito en la población Vf6 x Vf136 (Diaz-Ruiz et al. 2009). La posición de los genes ortólogos en *M. truncatula* sugiere que los tres estudios identifican la misma región por lo que habría que saturar la región entre Medtr3g102120 y Medtr3g099380 para identificar posibles genes candidatos relacionados con la resistencia a *A. fabae*. Los marcadores génicos de este intervalo corresponden al Contig9100 que codifica para una proteína inducida por estrés ambiental y acumulo de la hormona ABA y que regula la germinación de semillas y el crecimiento de plántulas (Guo y Ho et al. 2008); el Contig17163 que codifica una proteína que podría estar implicada en procesos inmunitarios (D'Cruz et al. 2013) en humanos (Perfetto et al. 2013, Yap et al. 2005) y el marcador 14-3-3P que pertenece a una familia de proteínas reguladoras con papeles importantes en la transducción de señales mitogénicas, apoptosis y control del ciclo celular (Fu et al. 2000). *Af3* (LO98\_DSL, LO98\_DSS y Field\_DSP) localizado en el chr. III, se sitúa en la misma zona que *Oc10* para la resistencia *A. fabae* lo que aumenta el interés por la saturación de esta zona. Nuestro estudio solo permitió incluir un marcador (Contig19144) en esta región, anotado como una proteína de unión a la auxina que regula la división celular, las actividades meristemáticas y el desarrollo radicular (Gao et al. 2015). Finalmente, en el chr. VI se detectaron dos zonas. La primera (relacionada con Field\_DSP1 y LO98\_DSL) descrita por Atienza et al. (2016), co-localiza con *Oc7* identificado para la resistencia a *O. crenata*. Esta región está compuesta por los marcadores mtmt\_GEN\_01130\_02\_1 (quinasa similar a receptores LRR (LRR-RKs) implicados en la regulación de procesos relacionados con el desarrollo y la defensa (Torii 2004)), RNAR (oxidorreductasa implicada en la síntesis de ADN) y dos nuevos SNPs (Locus\_10591 y Contig13425) que codifican para una proteína de resistencia al aluminio y para una proteína transmembrana que podría estar involucrada en el reconocimiento de patógenos o en la señalización en respuestas de defensa local. La segunda zona detectada corresponde a un nuevo QTL (Field\_DSP2) identificado en este estudio que podría corresponder al QTL4 descrito para la resistencia *A. fabae* en la población Icarus x Ascot (Kaur et al. 2014).

A pesar de la baja proporción de marcadores polimórficos mapeados inicialmente (capítulo II), la identificación de SNPs a partir del nuevo transcriptoma, ha permitido un mayor conocimiento de los mecanismos moleculares y bioquímicos involucrados en la resistencia a patógenos en habas. Además, la explotación de las relaciones macrosintéticas entre leguminosas, junto con los avances en técnicas de genotipado masivo, nos han permitido saturar regiones portadoras de QTLs con nuevos marcadores ortólogos relacionados con mecanismos de resistencia. Estos resultados facilitará la transferencia de información a otros cultivos de leguminosas, permitiendo la comparación directa de resultados y ayudando a estudios futuros de mejora en habas.

# Capítulo VI

## Conclusiones





A partir del trabajo realizado se han obtenido las siguientes conclusiones:

1. La explotación de la sintenia entre leguminosas y su aplicación en el desarrollo de los mapas genéticos de las habas ha permitido comparar distintos mapas y saturar regiones portadoras de QTLs relacionados con la resistencia a *A. fabae* y *O. crenata* en dos poblaciones RIL.
2. El grado de polimorfismo detectado (92 de los 228 SNPs seleccionados) es comparable al de trabajos previos. La pérdida de polimorfismo puede deberse a una identificación errónea de SNPs o a la polinización cruzada accidental por insectos, en alguno de los parentales.
3. Los marcadores ESTs y SNP derivados de secuencias codificantes han proporcionado mapas más robustos y completos e identificado nuevas zonas asociadas con la resistencia.
4. Algunos de los ESTs genotipados han resultado útiles como marcadores puente para el desarrollo de un mapa consenso en la especie.
5. El transcriptoma de habas en respuesta a la infección por *Ascochyta fabae*, caracterizado en este estudio (98,195 transcritos) es uno de los más completos hasta la fecha y proporciona información valiosa sobre genes y vías involucradas en la resistencia contra patógenos que podrán ser utilizados en el futuro desarrollo de marcadores para selección.
6. El número de transcritos no-redundantes con gen ortólogo, identificados (21,243), se aproxima al número de genes estimados en plantas diploides como *Arabidopsis thaliana* o *Cicer arietinum*.
7. 270 transcritos que codifican proteínas completas desconocidas podrían ser específicos de habas y ser, además, futuros genes candidatos de resistencia.
8. El uso de una batería de SNPs en genes relacionados con la resistencia a patógenos ha permitido ratificar QTLs de resistencia a *A. fabae* y *O. crenata*



validados en trabajos previos y reducir sus intervalos de confianza pudiéndose identificar, en la mayoría de ellos, candidatos posicionales y funcionales.

9. *Af2* (chr. II) se ha desplazado hacia una posición más cercana al QTL3, descrito en la población Icarus x Ascot, y hacia *Af2* descrito en la población Vf6 x Vf136 de habas, lo que podría sugerir que los tres estudios identifican la misma región genómica.
10. *Af3* co-localiza con *Oc10*, sugiriendo la posible existencia en habas de mecanismos de resistencia comunes frente a distintos patógenos.
11. EL nuevo QTL (Field\_DSP2) identificado en el cromosoma VI podría corresponder al QTL4 descrito para la resistencia *A. fabae* en la población Icarus x Ascot de habas.
12. Es necesario continuar con el genotipado de nuevos candidatos para conseguir el mapeo fino de los QTLs estudiados, especialmente de *Oc7* que explica el mayor porcentaje de la variación fenotípica para la resistencia a *Orobanche* y podría corresponder a *Oc2* detectado en otra población de habas.

# Capítulo VII

## Bibliografía





- 
- Abbo S, Berger J, Turner C (2003). Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct Plant Biol* 30: 1081-1087.
- Abu-Irmaileh BE (1994). Nitrogen reduces branched broomrape (*Orobanche ramosa*) seed germination. *Weed Sci* 42:57-60.
- Adams MD, Kelley JM, Gocayne JD, Dubnick M Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie RWR, Venter JC (1991). *Science* 252: 1651–1656.
- Alfano JR, Collmer A (2004). Type III secretion system effector proteins: Double Agents in Bacterial Disease and Plant Defense. *Annual Review of Phytopathology*, 42, 385-414.
- Almeida NF, Krezdorn N, Rotter B, Winter P, Rubiales D, Vaz Patto MC (2015). *Lathyrus sativus* transcriptome resistance response to *Ascochyta lathyri* investigated by deepSuperSAGE analysis. *Front Plant Sci.* 6:178.
- Al-Menoufi OA. (1994). The *Orobanche* Problem and Management in Egypt. In: *Biology and Management of Orobanche*, Proceedings of the Third International Workshop on *Orobanche* and related Striga research. Pieterse AH, Verkleij JAC and Ter Borg SJ (eds). Royal Tropical Institute, Amsterdam, pp. 663-671.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol.* 215(3):403-10.
- Atienza SG, Palomino C, Gutiérrez N, Alfaro CM, Rubiales D, Torres AM, Ávila CM. (2016). QTLs for ascochyta blight resistance in faba bean (*Vicia faba L.*): validation in field and controlled conditions. *Crop and Pasture Science* 67(2) 216-224.
- Aubert G, Morin J, Jacquin F, Loidon 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(6):1024-1041.
- Avila CM, Atienza S, Moreno M, Torres AM (2007). Development of a new diagnostic marker for growth habit selection in faba bean (*Vicia faba L.*) breeding. *Theor Appl Genet* 115(8):1075-1082.
- Avila CM, Nadal S, Moreno MT, Torres AM (2006). Development of a simple PCR-based marker for the determination of growth habit in *Vicia faba L.* using a candidate gene approach. *Mol Breed* 17:185-190.

- Avila CM, Satovic Z, Sillero JC, Rubiales D, Moreno MT, Torres AM (2004). Isolate and organ-specific QTLs for ascochyta blight resistance in faba bean. *Theor Appl Genet* 108:1071-078.
- Avila CM., Sillero JC., Rubiales D., Moreno MT., Torres AM (2003). Identification of RAPD markers linked to the Uvf-1 gene conferring hypersensitive resistance against rust (*Uromyces viciae-fabae*) in *Vicia faba L.* *Theor Appl Genet.* 107(2):353-8.
- Beaumont A (1950). On the *Ascochyta* spot disease of broad beans. *Trans. Br. Mycol. Soc.* 33: 345-349.
- Bechinger C, Giebel K-F, Schnell M, Leiderer P, Deising HB, Bastmeyer M(1999). Optical Measurements of Invasive Forces Exerted by Appressoria of a Plant Pathogenic Fungus. *Science*, 285, 1896-9.
- Beck D. Duc G (1991). Improving N<sub>2</sub>-fixation in faba bean: Rhizoiium inoculation and N nutrition. CIHEAM. Options Méditerranéennes. Série Seminaires n°10, pp. 97-103.
- Bennett MD, Leitch IJ (1995). Nuclear DNA Amounts in Angiosperms. *Ann Bot.* 76(2):113-76.
- Benning UF, Tamot B, Guelette BS, Hoffmann-Benning S (2012). New aspects of phloem-mediated long-distance lipid signaling in plants. *Front Plant Sci* 3:53.
- Bent AF, Mackey D (2007). Elicitors, Effectors, and R Genes: The New Paradigm and a Lifetime Supply of Questions. *Annual Review of Phytopathology*, 45, 399-436.
- Bevan M, Walsh S (2005). The *Arabidopsis* genome: A foundation for plant research. *Genome Res.* 15(12):1632-42.
- Bogdanova EA, Shagin DA, Lukyanov SA (2008). Normalization of full-length enriched cDNA. *Mol BioSyst.* 4(3):205-12.
- Bohra A, Pandey M, Jha U, Singh B, Singh I, Datta D et al. (2014). Genomics-assisted breeding in four major pulse crops of developing countries: present status and prospects. *Theor Appl Genet.* 1-29.
- Boller T, Felix G (2009). A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annual Review of Plant Biology*, 60, 379-406.

- 
- Boller T, He SY (2009). Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors in Microbial Pathogens. *Science*, 324:742-4.
- Bond DA and Poulsen MH (1983). Pollination. In: *The faba bean (Vicia faba L.)*, P.D. Hebblethwaite (ed.). Butterworths, London, pp. 77-101.
- Bond DA, Jellis GJ, Rowland GG, Le Guen J, Robertson LD, Khalil SA, Li-Juan L (1994). Present status and future strategy in breeding faba beans (*Vicia faba L.*) for resistance to biotic and abiotic stresses. In: Muehlbauer FJ, Kaiser WJ, editors. *Expanding the Production and Use of Cool Season Food Legumes*. *Curr Plant Sci Biotechnol Agric* 19: Springer Netherlands. p. 592-616.
- Bond DA, Pope M (1980). *Ascochyta fabae* on winter beans (*Vicia faba*): pathogen spread and variation in host resistance. *Plant Pathol* 29: 59-65.
- Bonierbale MW, Plaisted RL, Tanksley SD (1998) *Genetics* 120: 1095-1103.
- Boyd LA, Ridout C, ÓSullivan DM, et al. (2013) Plant-pathogen interactions: disease resistance in modern agriculture. *Trends Genet.* 29:233-240.
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F et al. (2006). Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proc. Natl. Acad. Sci. USA* 103: 14959-14964.
- Chapman GP (1983). Cytogenetics. En: *The Faba Bean (Vicia faba L.)*. PD Hebblethwaite (ed). Butterworths, London. Pp.197-216.
- Chapman KD, Dyer JM, Mullen RT (2012). Biogenesis and functions of lipid droplets in plants: Thematic Review Series: Lipid Droplet Synthesis and Metabolism: from Yeast to Man. *J. Lipid Res.* 53: 215-226.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006). Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell*, 124, 803-14.
- Choi H K, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kaló P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004a). A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *Medicago sativa*. *Genetics*.166:1463-1502.
- Choi HK, Luckow MA, Doyle J, Cook DR (2006.) Development of nuclear gene-derived molecular markers linked to legume genetic maps. *Mol Genet Genomics* 276(1):56-70.

- Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004b). Estimating genome conservation between crop and model legume species. *Proc Natl Acad Sci USA*. 101:15289-94.
- Chooi WR (1971). Comparison of the DNA of six *Vicia* species by the method of DNA-DNA Hybridisation. *Genetics* 68, 213-230.
- Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Collard BC, Mackill DJ,(2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R SocLond B BiolSci* 363:557-72.
- Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker assisted selection for crop improvement: The basic concepts. *Euphytica* 142:169-196.
- Cook DR (1999). *Medicago truncatula* a model in the making! *Curr Op Plant Biol* 2: 301-304.
- Cook DR, VanderBosch K, de Bruijn FJ, Huguet T (1997). Model legumes get the nod. *Plant Cell* 9:275-281.
- Corné M J Pieterse, Antonio Leon-Reyes, Sjoerd Van der Ent, Saskia C M Van Wees (2009). Networking by small-molecule hormones in plant immunity. *Nat Chem Biol* 5:308 - 316
- Cottage A, Gostkiewicz K, Thomas JE, Borrows R, Torres AM, O’Sullivan DM., (2012). Heterozygosity and diversity analysis using mapped single nucleotide polymorphisms in a faba bean inbreeding programme. *Mol Breed*. 30(4):1799-809.
- Cronquist A (1981). An integrated system of classification of flowering plants. Columbia University Press, NY, EEUU, p. 1262.
- Cruz-Izquierdo S, Avila CM, Satovic Z, Palomino C, Gutierrez N, Ellwood SR, Phan HTT, Cubero JI, Torres AM (2012). Comparative genomics to bridge *Vicia faba* with model and closely-related legume species: stability of QTLs for flowering and yield-related traits. *TheorAppl Genet* 125:1767-1782.
- Cubero JI (1991). Breeding for resistance to *Orobanche* species: a review. In: Webmann K, Musselman LJ (eds) *Progres in Orobanche* research. Ederhard-Karls-Universität, Tübingen, pp. 257-277.

- 
- Cubero JI, Moreno MT, Hernández L (1992). A faba bean cultivar resistant to *Orobanche crenata* Forsk. In 'Proceedings 1st European Conference on Grain Legumes'. Angers, France. pp. 41–42 (AEP, European Association for Grain Legume Research).
- Cubero JI (1974). On the evolution of *Vicia faba* L. Theor Appl Genet 45, 47-51.
- Cubero JI (1983). Parasitic diseases in *Vicia faba* L. with special referente to broomrape (*Orobanche crenata* Forsk). In: The Faba Bean (*Vicia faba* L.). Hebblethwaite PH (ed). Cambridge, Great Britain, pp. 493-521.
- Cubero JI (1984). Utilization of wild relatives and primitive forms of food legumes. Witcombe JRaE, W., editor. The Hague, The Netherlands and ICARDA: Martinus Nijhoff/Dr W Junk Publishers.
- Cubero JI (1973). Resistance to *Orobanche crenato* Forsk. in *Vicia faba* L. Proc. Eur. Weed Res. Coun. Parasitic Weeds, 205-217.
- Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature. 411(6839):826-33.
- D'Cruz AA, Babon JJ, Norton RS, Nicola NA, Nicholson SE (2013). Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity Protein. Science 22:1-10.
- De Lorenzo G, Brutus A, Savatin DV, Sicilia F, Cervone F (2011). Engineering plant resistance by constructing chimeric receptors that recognize damage-associated molecular patterns (DAMPs). FEBS Lett 585:1521-1528.
- Del Pozo O, Pedley KF, Martin GB (2004). MAPKKK $\alpha$  is a positive regulator of cell death associated with both plant immunity and disease. EMBO J. 23:3072–3082. 10.1038/sj.emboj.7600283.
- Denarie J (2002). Genomics and genetics of *Medicago truncatula*: resourses for legume crop improvement. Grain Legumes 38:20-21.
- Derksen H, Rampitsch C, Daayf F (2013). Signaling cross-talk in plant disease resistance. Plant Sci. 207(0):79-87.
- Devos KM, Gale MD (1997). Comparative genetics in the grasses. Plant Mol Biol 35(1-2):3-15.
- Díaz R, Torres AM, Satovic Z, Gutierrez MV, Cubero JI, Román B (2010). Validation of QTLs for *Orobanche crenata* resistance in faba bean (*Vicia faba* L.) across environments and generations. Theor Appl Genet, 120:909-919.



- Díaz-Ruiz R, Satovic Z, Avila CM, Alfaro CM, Gutierrez MV, Torres AM, Román B, (2009). Confirmation of QTLs controlling *Ascochyta fabae* resistance in different generations of faba bean (*Vicia faba L.*). *Crop Pasture Sci*, 60:353–361.
- Díaz-Ruiz R, Torres A, Gutierrez MV, Rubiales D, Cubero JI, Kharrat M, Satovic Z, Román B (2009). Mapping of quantitative trait loci controlling *Orobanche foetida Poir.* Resistance in faba bean (*Vicia faba L.*). *Afr J Biotechnol* 8:2718-2724.
- Dixon RA, Paiva NL(1995). Stress-Induced Phenylpropanoid Metabolism. *The Plant Cell*. 7(7):1085-97.
- Dong X (2001). Genetic dissection of systemic acquired resistance. *Curr Opin Plant Biol*, 4, 309-14.
- Doyle JJ (2000). Leguminosae. In *Encyclopedia of Genetics*. Edited by Brenner S, Miller JH. San Diego: Academic Press 1081-1085.
- Doyle JJ, Luckow MA (2003). The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol* 131:900-910.
- Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010). AgriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res*. 38(suppl 2):W64-W70.
- Duc G (1997). Faba bean. *Field Crops Res* 53, 99-109.
- Durrant WE, Dong X (2004). Systemic acquired resistance. *Annual Review of Phytopathology*, 42, 185-209.
- Echenique V, Schrauf G, Selva JP (2004). Genómica. In: *Biotecnología y Mejoramiento Vegetal*. INTA. Buenos Aires, Argentina. 212-228.
- Edwards D,Forster JW, Chagne D, Batley J (2007). What is SNPs? In: N.C. Oraguzie, E.H.A. Rikkerink, S.E. Gardiner and H.N. de Silva (eds.), *Association Mapping in Plants*. Springer, Berlin, pp. 41-52.
- Egan A, Schleiter J, Spooner DM (2012). Applications of next-generation sequencing techniques in plant biology. *American Journal of Botany*. 99(2):175-185.
- Eizenberg H, Tanaami Z, Jacobsohn R and Rubin B (2001). Effect of temperature on the relationship between *Orobanche spp.* and carrot (*Daucus carota L.*). *Crop Protection*, 20: 415-420.
- 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.
- Esposito A, Colantuono C, Ruggieri V, Chiusano ML (2016). Bioinformatics for agriculture in the Next-Generation sequencing era. *Chemical and Biological Technologies in Agriculture*, 3(1), 9.
- FAOSTAT (2014). <http://www.fao.org/faostat/es/>.
- Flavell RB, Bennett MD, Smith JB, Smith DB (1974). Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem Genet.* 12(4):257-69.
- Fofana B, Benhamou N, McNally DJ, Labbé C, Séguin A, Bélanger RR (2005). Suppression of Induced Resistance in Cucumber Through Disruption of the Flavonoid Pathway. *Phytopathology.* 95(1):114-23.
- Fondevilla S, Fernández-Aparicio M, Satovic Z, Emeran AA, Torres AM, Moreno MT, Rubiales D (2010). Identification of quantitative trait loci for specific mechanisms of resistance to *Orobanche crenata Forsk.* in pea (*Pisum sativum L.*). *Mol Breeding* 25: 259-272.
- Fondevilla S, Kuster H, Krajinski F, Cubero J, Rubiales D (2011). Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology. *BMC Genomics.* 12(1):28.
- Francia E, Tacconi G, Grosatti C, Barabaschi D, Bulgarelli D, Dall'Aglio, Valè G (2005). Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture* 82, 317-342.
- Frugoli J, Harris J. *Medicago truncatula* on the move (2001) *The Plant Cell.* ;13:458-463.
- Gabriel S, Ziaugra L, Tabbaa D (2009). SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet.* Chapter 2: Unit 2.12.
- Gao Y, Zhang Y, Zhang D, Dai X, Estelle M, Zhao Y (2015). Auxin binding protein 1 (ABP1) is not required for either auxin signaling or *Arabidopsis* development. *Proc. Natl. Acad. Sci. USA*, 112, pp. 2275-2280.
- Gao Z, Johansen E, Eyers S, Thomas CL, Ellis THN, Maule AJ (2004). The potyvirus recessive resistance gene, sbm1, identifies a novel role for translation initiation factor eIF4E in cell-to-cell trafficking. *Plant J* 40:376-385.

- García P, Neyra CD, Rey-Baños R, Miliani AS, García-López M, González-Cordero AI, Vaquero F, Vences FJ, Pérez de la Vega M, Sáenz de Miera LE, Polanco, C (2016). Actas de Horticultura 74: 225-226.
- García P, Sáenz de Miera LE, Vaquero F, Vences FJ, Polanco C, González AI, Jüngling R, Frank A, Horres R, Krezdorn N, Rotter B, Winter P, Kahl G; Pérez de la Vega M (2012). Análisis de los cambios en el transcriptoma de lenteja en respuesta a la infección por *Ascochyta lentis* mediante SuperSAGE. Actas de Horticultura 62: 133-134.
- García-Torres L and Lopez-Granados F. (1991). Control of broomrape (*Orobanche crenata Forsk*) in broad bean (*Vicia faba L.*) with imidazolinones and other herbicides. Weed Research, 31: 227-235.
- Garg N, Geetanjali (2009). Symbiotic nitrogen fixation in legume nodules: Process and signaling: A review. In: Sust Agri. Springer Netherlands, pp. 519-531.
- Gaunt RE (1983). Shoot disease caused by fungal pathogens. En: The Faba Bean (ed.: PD Hebblethwaite), pp. 463-492. Butterworths, Londres, 1983.
- Gaunt RE, Liew SS (1981). Control strategies for *Ascochyta fabae* in New Zeland field and broad bean crops. Seed Science and Technology 9: 707-715.
- Gepts P, William D, Beavis E, Brummer C, Randy C. Shoemaker H, Stalker T, Weeden NF, Young ND (2005). Legumes as a Model Plant Family. Genomics for food and feed report of cross-Legume Advances through Genomics Conference. Plant Physiol 137:1228-1235.
- Gérard Duc, Jelena M. Aleksić, Pascal Marget, Aleksandar Mikic, Jeffrey Paull, Robert J. Redden, Olaf Sass, F.L. Stoddard, Albert Vandenberg, Margarita Vishniakova, Ana María Torres.(2015). Faba bean. En: Antonio M. De Ron (Ed.) Grain Legumes, Series Handbook of Plant Breeding. Springer Science+Business Media, New York.
- Gómez-Gómez L, Felix G, Boller T (1999). A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. Plant J., 18, 277-84.
- Graham MA, Silverstein KAT, Cannon SB, VandenBosch KA (2004). Computational Identification and Characterization of Novel Genes from Legumes. Plant Physiol. 135(3):1179-97.
- Graham PH, Vance CP (2003). Legumes: Importance and constraints to greater use. Plant Physiol 131:872-877.

- Gressel J, Hanafi A, Head G, Marasas W, Obilana AB, Ochanda J, Souissi T, Tzotzos G (2004). Major heretofore intractable biotic constraints to African food security that may be amenable to novel biotechnological solutions. *Crop Prot* 23:661-689.
- Gujaria-Verma N, Vail SL, Carrasquilla-Garcia N, Penmetsa RV, 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.* 2014;5:676.
- Guo WJ and Ho TH (2008). An abscisic acid-induced protein, HVA22, inhibits gibberellin-mediated programmed cell death in cereal aleurone cells. *Plant Physiol* 147:1710-1722.
- Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW (2012). Plant disease resistance genes: Current status and future directions. *Physiol Mol Plant Pathol.* 78(0):51-65.
- Gutierrez N, Avila CM, Rodriguez-Suarez C, Moreno MT, Torres AM (2007). Development of SCAR markers linked to a gene controlling absence of tannins in faba bean. *Mol Breed* 19:305-314.
- Gutierrez N, Giménez MJ, Palomino C, Avila CM (2011). Assessment of candidate reference genes for expression studies in *Vicia faba L.* by real-time quantitative PCR. *Mol Breed.* 28(1):13-24.
- Gutiérrez N, Palomino C, Satovic Z, Ruiz-Rodríguez MD, Vitale S, Gutiérrez MV, Rubiales D, Kharrat M, Amri M, Emeran A, Cubero JI, Atienza SG, Torres AM, Avila CM (2013). QTLs for *Orobanche spp.* resistance in faba bean: Identification and validation across different environments. *Mol Breed* 32:909-922.
- Fu H, Subramanian RR, Masters SC (2000). 14-3-3 Proteins: Structure, Function, and Regulation. *Annu Rev Pharmacol* 40:617-647.
- Hanounik SB, Robertson LD (1989). Resistance in *Vicia faba* germplasm to blight caused by *Ascochyta fabae*. *Plant Dis* 73: 202–205.
- He P, Shan L, Sheen J (2007). Elicitation and suppression of microbe-associated molecular pattern-triggered immunity in plant-microbe interactions. *Cellular Microorganismology* 9:1385-96.
- Heath MC (2000). Nonhost resistance and nonspecific plant defenses. *Curr Opin Plant Biol*, 3, 315-9.

- Hewett PD (1973). The field behaviour of seed-borne *Ascochyta fabae* and disease control in field beans. *Annals of Applied biology* 74: 287-295.
- Hieter P, Boguski M (1997). Functional genomics: it's all how you read it. *Science* 278: 601-602.
- Himmelbach A, Liu L, Zierold U, Altschmied L, Maucher H, Beier F, Muller D, Hensel G, Heise A, Schutzendubel A, Kumlehn J, Schweizer P (2010). Promoters of the barley germin-like GER4 gene cluster enable strong transgene expression in response to pathogen attack. *Plant Cell* 22:937-952.
- Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, Tuteja R, et al. (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum L.*), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol J.* 9(8):922-31.
- Hospital F (2009). Challenges for effective marker-assisted selection in plants. *Genetica.* 136(2):303-10.
- Schmutz J , SB Cannon , J Schlueter , J Ma , T Mitros , W Nelson , DL Hyten , Q Song , JJ Thelen , J Cheng , D Xu , U Hellsten , GD May , Y Yu , T Sakurai , T Umezawa , MK Bhattacharyya , D Sandhu , B Valliyodan , E Lindquist , M Peto , D Grant , S Shu , D Goodstein , K Barry , M Futrell-Griggs , B Abernathy , J Du , Z Tian , L Zhu , N Gill , T Joshi , M Libault , A Sethuraman , XC Zhang , K Shinozaki , HT Nguyen , RA Wing , P Cregan , J Specht , J Grimwood , D Rokhsar , G Stacey , RC Shoemaker and SA Jackson (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.
- Hulsen T, de Vlieg J, Alkema W (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* 9(1):488.
- Humphry M, ReinstÄDler A, Ivanov S, Bisseling TON, Panstruga R (2011). Durable broad-spectrum powdery mildew resistance in pea er1 plants is conferred by natural loss-of-function mutations in PsMLO1. *Mol Plant Pathol* 12(9):866-78.
- Hundertmark M, Hinch D (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics.* 9(1):118.

- 
- Jain M, Misra G, Patel RK, Priya P, Jhanwar S, Khan AW, Shah N, Singh VK, Garg R, Jeena G, Yadav M, Kant C, Sharma P, Yadav G, Bhatia S, Tyagi AK, Chattopadhyay D (2013). Plant J. 74(5):715-29. doi: 10.1111/tpj.12173.
- Jamann T, Poland J, Kolkman J, Smith L, Nelson R (2014). Unraveling genomic complexity at a quantitative resistance locus in maize. Genetics 198:333-344.
- Jiménez-Díaz RM, Porta-Puglia A, Tivoli B (1998). New approaches in the integrated disease management of legume disease: toward sustainable crop health. 3rd
- Jones JD, Dangl JL (2006). The plant immune system. Nature 444, 323-9.
- Kaiser WJ (1997). Inter- and intranational spread of *ascochyta* pathogens of chickpea, faba bean and lentil. Canadian Journal of Plant Pathology 19: 215-224
- Kaló P, Choi HK, Ellis N, Kiss GB. (2011). Synteny and Comparative Genomics between Model and Cool Season Grain Legumes. In Genetics, Genomics and Breeding in Crop Plants: Cool Season Food Legumes. Edited by Perez de la Vega M, Torres AM, Cubero JI, Kole C. New Hampshire, Jersey, Plymouth: Science Pubs Inc; 2011:285-302.
- Kaló P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB (2004). Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genom 2004, 272:235-246.
- Kamoun S (2007). Groovy times: filamentous pathogen effectors revealed. Current Opinion in Plant Biology, 10, 358-65.
- Kang Y, Kim K, Shim S, Yoon M, Sun S, Kim M, et al. (2012). Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. BMC Plant Biol. 12(1):139.
- Kaur S, Kimber RBE, Cogan NOI, Materne M, Forster JW, Paull JG (2014). SNP discovery and high-density genetic mapping in faba bean (*Vicia faba* L.) permits identification of QTLs for ascochyta blight resistance. Plant Sci. 217–218(0):47-55.
- Kaur S, Pembleton L, Cogan N, Savin K, Leonforte T, Paull J, Materne M, Forster JW (2012). Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics. 13(1):104.
- Kharbanda PD, Bernier CC (1979). Effectiveness of seed and foliar application of fungicides to control *Ascochyta blight* in faba beans. Canadian Journal of Plant science 59: 661-666.

- Kharrat M (1999). *Orobanche* research activities on faba bean in Tunisia. In: Cubero JI, Moreno MT, Rubiales D, Sillero JC (eds) Resistance to broomrape, the state of the art. Junta de Andalucía, Sevilla, pp. 77-81.
- Kharrat M, Halila MH, Linke KH, Haddar T (1992). First report of *Orobanche foetida* Poir. on faba bean in Tunisia. FABIS Newslett 30:46-47.
- Khazaei H, O'Sullivan D, Sillanpää M, Stoddard F (2014). Use of synteny to identify candidate genes underlying QTL controlling stomatal traits in faba bean (*Vicia faba L.*). Theor Appl Genet. 127(11):2371-85.
- Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997). Towards map-based cloning of the barley stem rust resistance genes Rpg1 and rpg4 using rice as an intergenomic cloning vehicle. Plant Mol Biol 35:187-195.
- Kliebenstein DJ (2004). Secondary metabolites and plant environmental interactions: A view through *Arabidopsis thaliana* tinted glasses. Plant Cell Environmental 27: 675-684.
- Knox M R, Ellis THN (2002) Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. Genetics 162: 861-873.
- Kolukisaoglu U, Wein, S, Blazevic D, Batistic O, Kudla J (2004). Calcium sensors and their interacting protein kinases: genomics of the *Arabidopsis* and rice CBL-CIPK signaling networks. Plant Physiol 134: 43-58.
- Kosambi DD (1944). The estimation of map distance from recombination values. Ann Eugen 12:172-175.
- Kou Y, Wang S (2010). Broad-spectrum and durability: understanding of quantitative disease resistance. Current Opinion in Plant Biology 13(2): 181-185.
- Kozomara A, Griffiths-Jones S (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42(D1):D68-D73.
- Kudapa H, Azam S, Sharpe AG, Taran B, Li R, Deonovic B, et al. (2014). Comprehensive Transcriptome Assembly of Chickpea (*Cicer arietinum L.*) Using Sanger and Next Generation Sequencing Platforms: Development and Applications. PLoS ONE. 9(1):e86039.
- Kumar J, Choudhary AK, Solanki RK, Pratap A (2011). Towards marker-assisted selection in pulses: a review. Plant Breed. 130(3):297-313.

- 
- Kwon C, Panstruga R, Schulze-Lefert P (2008) Les liaisons dangereuses: immunological synapse formation in animals and plants. *Trends Immunology*, 29(4):159-166.
- Ladizinski G. (1975). On the origin of the broadbean, *Vicia faba L.* *Isr J Bot* 24 80-88.
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40(D1):D1202-D10.
- Lander E and Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199.
- Lander ES, Green P, Abramson J, Barlow A, Dali MJ, Lincoln DE, Newburg L, (1987). MAPMAKER: an interactive computer program for constructing genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181.
- Langmead B (2002). Aligning Short Sequencing Reads with Bowtie. *Curr Protoc Bioinformatics*: John Wiley & Sons, Inc.
- Lawes DA, Bond DA, Poulsen MH (1983). Classification, origin, breeding methods and objectives. En: *The Faba Bean (Vicia faba L.)*. PD Hebblethwaite (ed). Butterworths, London, pp. 23-76.
- Lee J, Rudd JJ, Macioszek VK, Scheel D (2004). Dynamic changes in the localization of MAPK cascade components controlling pathogenesis-related (PR) gene expression during innate immunity in parsley. *J Biol Chem* 279:22440–22448. 10.1074/jbc.M401099200.
- Lee M (1995). DNA markers and plant breeding programs. *Adv Agron* 55: 265-343.
- Leonforte A, Sudheesh S, Cogan NO, Salisbury PA, Nicolas ME, Materne M, Forster JW and Kaur S (2013). SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum L.*). *BMC Plant Biol*, 13: 161.
- Lewis G, Schrirer B, Mackinder B, Lock M (2005). *Legumes of the world*; Royal Botanical Gardens: Kew, UK.
- Li H (2011). A quick method to calculate QTL confidence interval. *J Genet* 90:355–360.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078-9.



- Libault M and Dickstein R (2013). “Functional genomic research in legumes” In: “Legumes in the Omic Era” (Nagasamy Nadarajan, Sanjeev Gupta and Debjyoti Sen Gupta Eds.), Springer, USA.
- Lin H, Moghe G, Ouyang S, Iezzoni A, Shiu S-H, Gu X, et al. (2010). Comparative analyses reveal distinct sets of lineage-specific genes within *Arabidopsis thaliana*. *BMC Evol Biol.* 10:41.
- Linke KH, Scheibel C, Saxena MC, Sauerborn J (1992). Fungi occurring on *Orobanche spp.* and their preliminary evaluation for *Orobanche* control. *Tropical Pest Management*, 38:127-130.
- Liu H, Niu Y, Gonzalez-Portilla PJ, Zhou H, Wang L, Zuo T, Qin C, Tai S, Jansen C, Shen Y, Lin H, Lee M, Ware D, Zhang Z, Lübberstedt T, Pan G (2015). An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. *BMC Genomics* 16:1078
- Liu J, Maldonado I, Lopez M, Cheung F, Town C, Harrison M (2007). Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. *Plant J.* 50 (3):529-544.
- Liu S-C, Kowalski SP, Lan T-H, Feldmann KA, Paterson AH (1996) Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* 142:247–58.
- Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, et al. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40(W1):W622-W7.
- López Bellido L (1988). El papel de las leguminosas en la agricultura de secano. *El Campo* 108:7-12.
- López-Granados F, García-Torres L (1998). Short and long-term implications of controlling crenata broomrape (*Orobanche crenata*) in broad bean (*Vicia faba*). *Crop Protection*, 17:131-143.
- López-Granados F, García-Torres L. (2002). Incidencia agronómica de especies parásitas (*Orobanche* y *Cuscuta*) en España. In: Plantas parásitas de la Península Ibérica e Islas Baleares. López-Sáez JA, Catalán P y Sáez Ll (Eds). Ediciones Mundi-Prensa. España, pp. 453-476.

- López-Granados F, García-Torres L. (1993). Population dynamics of crenata broomrape (*Orobancha crenata*) in faba bean (*Vicia faba*). Weed Science, 41:563-567.
- López-Granados F, García-Torres L. (1999). Longevity of crenate broomrape (*Orobancha crenata*) seed under soil and laboratory conditions. Weed Science, 47:161-166.
- Louis J , K Lorenc-Kukula, V Singh, J Reese, G Jander, Shah J (2010). Antibiosis against the green peach aphid requires the *Arabidopsis thaliana* myzus persicae-induced lipase1 gene. Plant J. 64:800-811.
- Ma Y, Bao S-Y, Yang T, Hu J-G, Guan J-P, He Y-H, et al. (2013). Genetic linkage map of Chinese native variety faba bean (*Vicia faba L.*) based on simple sequence repeat markers. Plant Breeding. 132(4):397-400.
- Mader E and Hopwood J (2013). Pollinator management for organic seed producers. 28 pp. Portland, Oregon: The Xerces Society.
- Madrid E, Gil J, Rubiales D, Krajinski F, Schlereth A, Millán T (2010). Transcription factor profiling leading to the identification of putative transcription factors involved in the *Medicago truncatula-Uromyces striatus* interaction. Theor Appl Genet 121:1311-1321.
- Madrid E, Palomino C, Plötner A, Horres R, Jüngling R, Rotter B, Winter P, Kahl G, Torres AM (2013). DeepSuperSage analysis of the *Vicia faba* transcriptome in response to *Ascochyta fabae* infection. Phytopathologia Mediterranea 52:166-182.
- Madrid E, Rajesh PN, Rubio J, Gil J, Millán T, Chen W (2012). Characterization and genetic analysis of an EIN4-like sequence (CaETR-1) located in QTL AR1 implicated in ascochyta blight resistance in chickpea. Plant Cell Reports 31:1033-1042.
- Manosalva, PM, Davidson RM, Liu B, Zhu XY, Hulbert SH, Leung H, Leach JE (2008). A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol. 149:286-296.
- Manschadi AM, Sauerborn J, Stützel H. (2001). Quantitative aspects of *Orobancha crenata* infestation in faba beans as affected by abiotic factors and parasite soil seedbank. European Weed Research, 41: 311-324.

- Maurin N, Tivoli B (1992). Variation in the resistance of *Vicia faba* to *Ascochyta fabae* in relation to disease development in field trials. *Plant Pathol.* 41:737-44.
- Mauromicale G, Restuccia G, Marchese M (2001). Soil solarization, a non-chemical technique for controlling *Orobanche crenata* and improving yield of faba bean. *Agronomie*, 21: 757-765.
- Maxted N, Callimassia MA, Bennett MD (1991). Cytotaxonomic studies of Eastern Mediterranean *Vicia* species (Leguminosae). *Pl Syst Evol.*177(3-4):221-34.
- McCarty MF, Barroso-Aranda J, Contreras F (2009). The low-methionine content of vegan diets may make methionine restriction feasible as a life extension strategy. *Medical hypotheses* 72:125-128.
- Menancio-Hautea D, Fatokum CA, Kumar L, Danesh D, Young ND (1993). Comparative genome analysis of mungbean (*Vigna radiata* (L.) Wilczek) and cowpea (*V. unguiculata* (L.) Walpers) using RFLP mapping data. *Theor Appl Genet* 86:797-810.
- Meng X, Zhang S (2013). MAPK cascades in plant disease resistance signaling. *Annu. Rev. Phytopathol.* 51, 245–266. doi: 10.1146/annurev-phyto-082712-102314.
- Mesa-García J and García-Torres L (1984). A competition index for *Orobanche crenata* Forsk. Effects on broad bean (*Vicia faba* L.). *Weed Research*, 24: 379-382.
- Millán T, Madrid E, Cubero JI, Amri M, Castro P, Rubio J (2015). Chickpea. In: De Ron A (ed) *Grain Legumes*, pp. 85-109.
- Mittler R, Shulaev V (2013) Functional genomics, challenges and perspectives for the future. *Physiol Plant* 148:317-321.
- Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, Young ND (2005). Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biol.* 5:15-25.
- Mun JH , Kim DJ , Choi HK, Gish J, Debellé F, Mudge J, Denny R , Endré G, Saurat O , Dudez AM, Kiss GB , Roe B , Young ND and Cook DG (2006). Distribution of Microsatellites in the Genome of *Medicago truncatula*: A Resource of Genetic Markers That Integrate Genetic and Physical Maps. *Genetics* 172:2541-2555.

- 
- Muzquiz M, Varela A, Burbano C, Cuadrado C, Guillamón E, Pedrosa MM (2012). Bioactive compounds in legumes: pronutritive and antinutritive actions. Implications for nutrition and health. *Phytochemistry Review*, 11, 227–244
- Nadal S, Moreno MT, Cubero JI (2004). *Las leguminosas grano en la agricultura moderna*. Madrid: MundiPrensa.
- Nakano RT, Yamada K, Bednarek P, Nishimura M, Hara-Nishimura I (2014). ER bodies in plants of the Brassicales order: Biogenesis and association with innate immunity. *Front Plant Sci*. 5:73. doi: 10.3389/fpls.2014.00073.
- Nassib AM, Ibrahim AA, Khalil SA (1982). Breeding for resistanceto *Orobanche*. In: Hawting G, Webb (eds) *Faba bean improvement*. Martinus Nijhoff, The Hague, pp. 199-206.
- Nayak SN, Zhu HY, Varghese N (2010). Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor Appl Genet*, 120:1415-1441.
- Nelson J, Wang S, Wu Y, Li X, Antony G, White F, et al. (2011). Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. *BMC Genomics*. 12(1):352.
- Nelson M, Phan H, Ellwood S, Moolhuijzen P, Hane J, Williams A, O'Lone C, Fosun-Nyarko J, Scobie M, Cakir M, Jones M, Bellgard M, Ksiazkiewicz M, Wolko B, Barker S, Oliver R, Cowling W (2006). The first gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved synteny with *Medicago truncatula*. *Theor Appl Genet*, 113(2):225-238.
- Nickrent D (2002). Plantas parásitas en el mundo. In: *Plantas parásitas de la Península Ibérica e Islas Baleares*. López-Sáez JA, Catalán P y Sáez Ll. Ediciones Mundi-Prensa. España, pp. 7-27.
- Nutman PS (1976). IBP field experiments on nitrogen fixation by nodulated legumes. En: *Symbiotic Nitrogen Fixation in Plants*. PS Nutman (ed). Cambridge University Press, pp. 211-237.
- Ocaña S, Seoane P, Bautista R, Palomino C, Claros GM, Torres AM, Madrid E (2015). Large-scale transcriptome analysis in faba bean (*Vicia faba* L.) under *Ascochyta fabae* infection. *PLoS One* 10, e0135143. doi:10.1371/journal.pone.0135143.

- Ondrej M (1993). Response of resistant lines of horse bean to pathogenous fungus *Ascochyta fabae* Speg. Plant Genet Resour 2: 45-48.
- Pandey, M.K., Roorkiwal, M., Singh, V.K., Ramalingam, A., Kudapa, H., Thudi, M., Chitikineni, A., Rathore, A., Varshney, R.K. (2016), 'Emerging genomic tools for legume breeding: Current status and future prospects', Frontiers in Plant Science, 07 (455). 01-18. ISSN 1664-462X.
- Parker C and Riches CR (1993). Parasitic Weeds of the World: Biology and control. CAB International, UK, 332 p.
- Parker C (1994). The present state of the *Orobanche* problem In: Biology and Management of *Orobanche*, Proceedings of the Third International Workshop on Orobanche and related Striga research, Pieterse AH, Verkleij JAC and Ter Borg SJ. Amsterdam, The Netherlands, Royal Tropical Institute
- Parker JE (2009). The Quest for Long-Distance Signals in Plant Systemic Immunity. Sci. Signal., 2, pe31.
- Peck SC, Nuhsu TS, Hess D, Iglesias A, Meins F and Boller T (2001). Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. Plant cell, 13, 1-11.
- Pérez-de-Luque A, Eizenberg H, Grenz JH, Sillero JC, Ávila C, Sauerborn J, Rubiales D (2010). Broomrape management in faba bean. Field Crops Res 115:319-328.
- Pérez-de-Luque A, Lozano MD, Moreno MT, Testillano PS, Rubiales D (2007). Resistance to broomrape (*Orobanche crenata*) in faba bean (*Vicia faba*): cell wall changes associated with prehaustorial defensive mechanisms. Ann Appl Biol 151:89-98.
- Perfetto L, Gherardini PF, Davey NE, Diella F, Helmer-Citterich M, Cesareni G (2013). Exploring the diversity of SPRY/B30.2-mediated interactions. Trends biochem sci 38:38-46.
- Pflieger S, Lefebvre V, Causse M (2001). The candidate gene approach in plant genetics: a review. Mol Breed 7:275-291.
- Phan HTT, Ellwood SR, Adhikari K, Nelson MN, Oliver RP (2007). The first genetic and comparative map of white lupin (*Lupinus albus* L.): identification of QTLs for anthracnose resistance and flowering time, and a locus for alkaloid content. DNA Res. 14, 59-70. 10.1093/dnares/dsm00.

- Phan HTT, Ellwood SR, Ford R, Thomas S, Oliver RP (2006). Differences in syntenic complexity between *Medicago truncatula* with *Lens culinaris* and *Lupinus albus*. *Funct. Plant Biol.* 33(8):775-782.
- Phan HTT, 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(3):549-558.
- Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, et al. (1999). Mining SNPs From EST Databases. *Genome Res.* 9(2):167-74.
- Pieterse AH (1991). The effect of nitrogen fertilizers on the germination of seeds of striga hermonthica and *Orobanche crenata*. In: Wegmann K and Musselman LJ (eds). *Progress in Orobanche Research. Proceedings, International Workshop on Orobanche Research, Obermarchtal. Eberhard-Karls-Universität, Tübingen*, pp. 115-124.
- Pieterse CMJ, Leon-Reyes A, Van der Ent S, Van Wees SCM (2009). Networking by small-molecule hormones in plant immunity. *Nat Chem Biol.* 5(5):308-16.
- Pitzschke A, Schikora A, Hirt H (2009). MAPK cascade signaling networks in plant defence. *Curr Opin Plant Biol* 12: 421–426.
- Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc Natl Acad Sci USA* 108:6893–6898.
- Ponting CP, Schultz J and Bork P (1997) SPRY domains in ryanodine receptors (Ca<sup>2+</sup> release channels). *Trends Biochem Sci* 22:193–194.
- Prioul-Gervais S, Deniot G, Receveur EM, Frankewitz A, Fourmann M, Rameau C, Pilet-Nayel ML, Baranger A (2007). Candidate genes for quantitative resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum L.*). *Theor Appl Genet*, 114(6): 971-984.
- Pritchard PR, Rowe PS, Rossall S. (1989) A comparison of infection of resistant and susceptible lines of field bean (*Vicia faba*) by *Ascochyta fabae*. *Plant Pathol.* 38(2): 266-70.
- Pujadas-Salvá AJ (1999). Species of the family *Orobanchaceae* parasitic of cultivated plants and its relatives growing on wild plants, in the South of the Iberian Peninsula. In: Cubero JI, Moreno MT, Rubiales D, Sillero JC (eds) *Resistance to broomrape, the state of the art. Junta de Andalucía, Sevilla*, pp 187–193.

- Rabey JM, Vered Y, Shabtai H, Graff E, Harsat A, Korezyn AD (1993). Broad bean (*Vicia faba*) consumption and Parkinson's disease. En: Advances in Neurology, Volume 6. H Narabayashi, T Nagatsu, N Yanagisawa, Y Mizuno (eds). Raven Press, New York. pp. 681-684.
- Raina SN, Ogihara Y (1995). Ribosomal DNA repeat unit polymorphism in 49 *Vicia* species. Theor Appl Genet 90: 477-486.
- Randhir R, Shetty P, Shetty K (2002). l-DOPA and total phenolic stimulation in dark germinated fava bean in response to peptide and phytochemical elicitors. Process Biochemistry. Volume 37, Issue 11, pp. 1247-1256.
- Rashid KY, Bernier CC, Conner RL (1991a). Evaluation of faba bean for resistance to *Ascochyta fabae* and development of host differential for race identification. Plant Dis 75: 852-855.
- Rashid KY, Bernier CC, Conner RL (1991b). Genetic of resistance in faba bean inbred lines to five isolates of *Ascochyta fabae*. Can J Plant Pathol 13: 218-225.
- Ray H, Bock C, Georges F. (2015). Faba Bean: Transcriptome Analysis from etiolated seedling and developing seed coat of key cultivars for synthesis of proanthocyanidins, phytate, raffinose family oligosaccharides, vicine, and convicine. Plant Genome 8:1 10.3835/plantgenome2014.07.0028
- Rejeb I, Pastor V, Mauch-Mani B (2014) Plant responses to simultaneous biotic and abiotic stress: molecular mechanisms. Plants. 3(4):458.
- Robinson MD, McCarthy DJ, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26(1):139-40.
- Rodriguez MC, Petersen M, Mundy J (2010). Mitogenactivated protein kinase signaling in plants. Annu Rev Plant Biol 61: 621-649
- Rojas AM, Fuentes G, Rausell A, Valencia A (2012). The Ras protein superfamily: evolutionary tree and role of conserved amino acids. J Cell Biol 196:189-201.
- Rojas-Molina MDM, Rubiales D, Prats E, Sillero JC. (2007). Effects of Phenylpropanoid and Energetic Metabolism Inhibition on Faba Bean Resistance Mechanisms to Rust. Phytopathology 97(1):60-5.

- 
- Román B, Satovic Z, Avila CM, Rubiales D, Moreno MT, Torres AM (2003). Locating genes associated with *Ascochyta fabae* resistance in *Vicia faba*. Aust J Agric Res 54:85-90.
- Román B, Torres AM, Rubiales D, Cubero JI, Satovic Z (2002). Mapping of quantitative trait loci controlling broomrape (*Orobanche crenata* Forsk.) resistance in faba bean (*Vicia faba* L.) Genome, 45:1057-63.
- Romeis T, Ludwig A, Martin R, Jones JDG (2001). Calcium dependent protein kinases play an essential role in a plant defence response. EMBO J 20:5556–5567.
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, et al. (2006). Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA. 103(49):18656-61.
- Roy F, Boye JI, Simpson BK (2010). Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. Food Research International 43:432-442.
- Rozen S, Skaletsky H. (1999). Primer3 on the WWW for General Users and for Biologist Programmers. In: Misener S, Krawetz S, editors. Bioinformatics Methods and Protocols. Methods in Molecular Biology. 132: Humana Press; p. 365-86.
- Rubiales D (2003). Parasitic plants, wild relatives and the nature of resistance. New Phytol 160:459–461.
- Rubiales D, Alcántara C, Pérez-De-Luque A, Gil J and Sillero JC (2003). Infection of chickpea (*Cicer arietinum*) by crenate broomrape (*Orobanche crenata*) as influenced by sowing date and weather conditions. Agronomie, 23: 359-362.
- Rubiales D, Perez-de-Luque A, Fernández-Aparicio M, Sillero J, Román B, Kharrat M, Khalil S, Joel D, Riches C (2006). Screening techniques and sources of resistance against parasitic weeds in grain legumes. Euphytica 147:187–199
- Rubiales D, Sadiki M, Roman B (2005). Report of *Orobanche foetida* on common wetch (*Vicia sativa*) in Morocco. Plant Dis 89:528.
- Rubiales D., Fernández-Aparicio M. (2012). Innovations in parasitic weeds management in legume crops. A review. Agronomy for Sustainable Development, 32: 433-449.
- Salt GA (1983) Root diseases of *Vicia faba* L. In: PD Hebblethwaite (ed) The Faba Bean (*Vicia faba* L.). Butterworths, London, UK, pp. 393-420.



- Sato S, Nakamura Y, Kaneko T et al. (2008). "Genome structure of the legume, *Lotus japonicus*". DNA Research 15 (4): 227-239.
- Satovic S, Avila CM, Cruz-Izquierdo S, Díaz-Ruíz R, García-Ruíz GM, Palomino C, Gutiérrez-Leiva N, Vitale S, Ocaña-Moral S, Gutiérrez MV, Cubero JI and Torres AM (2013). A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba L.*). BMC Genomics, 14:932.
- Sauerborn J (1991). The economic importance of the phytoparasites *Orabanche* and *Striga*. In: Proceedings, 5th International Symposium on Parasitic Weeds, Ransom JK, Musselman LJ, Worsham AD and Parker C, Nairobi, Kenya, pp. 137-143.
- Schatz MC, Delcher AL, Salzberg SL (2010). Assembly of large genomes using second-generation sequencing. Genome Res. 20(9):1165-73.
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46:707-13.
- Schuster SC (2008). Next-generation sequencing transforms today's biology. Nat Meth. 5(1):16-8.
- Schweizer P, Stein N (2011). Large-Scale Data Integration Reveals Colocalization of Gene Functional Groups with Meta-QTL for Multiple Disease Resistance in Barley Published by APS PRESS in cooperation with the International Society for Molecular Plant-Microbe Interactions, Volume 24, Number 12, Pages 1492-1501.
- Schweizer P, Christoffel A, Dudler, R (1999). Transient expression of members of the germin-like gene family in epidermal of wheat confers disease resistance. Plant J. 20:541-552.
- Semagn K, Babu R, Hearne S, Olsen M (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breeding 33:1-14.
- Seoane P, Ocaña S, Carmona R, Bautista R, Madrid E, Torres A, Claros G (2016). AutoFlow, a versatile workflow engine illustrated by assembling an optimised de novo transcriptome for a non-model species, such as Faba Bean (*Vicia faba*). Curr. Bioinform. 11, 440-450.

- Sillero JC, Rubiales D, Moreno MT (1999). New sources of resistance to broomrape (*O. crenata*) in a collection of *Vicia* species. 45–54 in J. I. Cubero, M. T. Moreno, D. Rubiales, and J. C. Sillero, eds. Resistance to Broomrape, the State of the Art. Sevilla, Spain: Junta de Andalucía.
- Sillero JC, Avila CM, Moreno MT, Rubiales D (2001). Identification of resistance to *Ascochyta fabae* in *Vicia faba* germplasm. *Plant Breed* 120: 529–531.
- Sillero JC, Villegas-Fernández AM, Thomas J, Rojas-Molina MM, Emeran AA, Fernández-Aparicio M, Rubiales D (2010). Faba bean breeding for disease resistance. *Field Crops Res* 115: 297–307.
- Spanu P, Grosskopf D.G, Felix G. and Boller T (1994). The Apparent Turnover of 1-Aminocyclopropane-1-Carboxylate Synthase in Tomato Cells Is Regulated by Protein Phosphorylation and Dephosphorylation. *Plant Physiol.* 106, 529-35.
- Tai F, Yuan Z, Li S, Wang Q, Liu F, Wang W (2016). ZmCIPK8, a CBL-interacting protein kinase, regulates maize response to drought stress. *Plant Cell Tiss Organ Cult* 124:459-469.
- Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, Gentzbittel L, Childs K, Yandell M, Gundlach H, Mayer K, Schwartz D, Town C (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics* 15:312.
- Tena G, Boudsocq M, Sheen J (2011). Protein kinase signaling networks in plant innate immunity. *Curr. Opin. Plant Biol* 14: 519–529
- Thaler JS, Bostock RM (2004). Interactions between abscisic-acidmediated responses and plant resistance to pathogen and insects. *Ecology* 85:48-58
- Thoquet P, Ghérardi M, Journet EP, Kereszt A, Ané JM, Prospero JM and Huguet T (2002). The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biology* 2:1.
- Timmerman-Vaughan GM, Frew TJ and Weeden NF (2000). Characterization and linkage mapping of R-gene analogous DNA sequences in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 101: 241-247.
- Tivoli B, Berthelem P, LeGuen J, Onfroy C (1988). A study of the performance of certain faba bean genotypes in relation to *Botrytis fabae* and *Ascochyta fabae* in France. *FABIS Newsl* 21: 36-39.

- Tivoli B, Reynauld B, Maurin N, BerthelemP, Le Guen J (1987). Comparison of some methods for evaluation of reaction of different faba bean genotypes to *Ascochyta fabae*. FABIS Newsletter 17, 35–38.
- Torii Keiko U (2004). Leucine-Rich Repeat Receptor Kinases in Plants: Structure, Function, and Signal Transduction Pathways. International Review of Cytology. Volume 234, 1-46.
- Torres AM, Avila CM, Gutierrez N, Palomino C, Moreno MT, Cubero JI (2010). Marker-assisted selection in faba bean (*Vicia faba L.*). Field Crops Res. 115(3):243-52.
- Torres AM, Avila CM, Stoddard FL, Cubero JI. (2012) Faba bean. In: Pérez de la Vega M, Torres AM, Cubero JI, Kole C, editors. Genetics, Genomics and Breeding in Crop Plants: Cool Season Food Legumes. New Hampshire, Jersey, Plymouth: Science Pubs Inc; 2012.
- Torres AM, Millaín T, Cubero JI (1993). Identifying Rose cultivars using random amplified polymorphic DNA markers. HortSci 28:333–334.
- Udvardi M, Poole PS (2013.) Transport and metabolism in legume-rhizobia symbioses. Annu Rev Plant Biol 64:781-805.
- Van Loon L, Bakker P, Pieterse C (1998). Systemic resistance induced by rhizosphere bacteria. Ann Rev Phyt. 36:453-483.
- Van Loon LC, Rep M, Pieterse CMJ (2006). Significance of Inducible Defense-related Proteins in Infected Plants. Annu Rev Phytopathol. 44(1):135-62.
- Van Loon LC, Van Strien EA (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol;55(2):85-97.
- Van Ooijen JW (2006). JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V. Wageningen, Netherlands.
- Varshney RK, Chen W, Li Y et al. (2012). "Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers". Nature Biotechnology 30 (1): 83–89.
- Varshney RK, Dubey A (2009). Novel Genomic Tools and Modern Genetic and Breeding Approaches for Crop Improvement. J Plant Biochemistry & Biotechnology 18:127-138.
- Varshney RK, Mohan SM, Gaur PM, Gangarao NVPR, Pandey MK, Bohra A, Sawargaonkar SL, et al. (2013). Achievements and prospects of genomics-

- assisted breeding in three legume crops of the semi-arid tropics', *Biotechnology Advances*, 31(8):1120-1134.
- Varshney RK, Chi Song, Rachit K Saxena, Sarwar Azam, Sheng Yu, Andrew G Sharpe, Steven Cannon, Jongmin Baek, Benjamin D Rosen, Bunyamin Tar'an, Teresa Millan, Xudong Zhang, Larissa D Ramsay, Aiko Iwata, Ying Wang, William Nelson, Andrew D Farmer, Pooran M Gaur, Carol Soderlund, R Varma Penmetsa, Chunyan Xu, Arvind K Bharti, Weiming He, Peter Winter, Shancen Zhao, James K Hane, Noelia Carrasquilla-Garcia, Janet A Condie, Hari D Upadhyaya, Ming-Cheng Luo, Mahendar Thudi, CLL Gowda, Narendra P Singh, Judith Lichtenzveig, Krishna K Gali, Josefa Rubio, N Nadarajan, Jaroslav Dolezel, Kailash C Bansal, Xun Xu, David Edwards, Gengyun Zhang, Guenter Kahl, Juan Gil, Karam B Singh, Swapan K Datta, Scott A Jackson, Jun Wang, Douglas R Cook (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol.* 31:240–246.
- Vlot AC, Klessig DF, Park S-W (2008). Systemic acquired resistance: the elusive signal(s). *Current Opinion in Plant Biology* 11:436-42.
- Vogel JP, Raab TK, Schiff C, Somerville SC (2002). PMR6, a Pectate Lyase-Like Gene Required for *Powdery Mildew* Susceptibility in *Arabidopsis*. *PLANT CELL*, 14, 2095-106.
- Vogel JP, Raab TK, Somerville CR, Somerville SC (2004). Mutations in PMR5 result in *powdery mildew* resistance and altered cell wall composition. *The Plant Journal*, 40, 968-78.
- Walbot V (1985). On the life strategies of plants and animals. *Trends Genet* 1:165-169.
- Wan J, Zhang X-C, Stacey G (2008). Chitin signaling and plant disease resistance. *Plant Signal Behav.* 2008;3(10):831-3.
- Wang S, Basten CJ, Zeng ZB (2007). Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh. <http://stagen.ncsu.edu/qtlcart/WQTLCart.htm>.
- Wang TL, Domoney C, Hedley CL, Casey R, Grusak MA (2003). Can we improve the nutritional quality of legume seeds? *Plant Physiol* 131: 886–891.

- Wang Z, Luo Y, Li X, Wang L, Xu S, Yang J, Weng L, Sato S, Tabata S, Ambrose M et al (2008). Genetic control of floral zygomorphy in pea (*Pisum sativum L.*). Proc Natl Acad Sci USA 105:10414–10419.
- Webb A, Cottage A, Wood T, Khamassi K, Hobbs D, Gostkiewicz K, White M, Khazaei H, Ali M, Street D, Duc G, Stoddard FL, Maalouf F, Ogbonnaya FC, Link W, Thomas J, O’Sullivan DM (2015). A SNPbased consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba L.*). Plant Biotechnology Journal doi:10.1111/pbi.12371.
- Weeden NF, Muehlbauer FJ, Ladizinsky G (1992). Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83: 123-129.
- Xu Y, Crouch JH (2008). Marker-assisted selection in plant breeding: from publications to practice. Crop Science 48:391 – 407.
- Xu, Y (2010). Molecular plant breeding. CAB International.
- Yang S, Gao M, Xu C, Gao J, Deshpande S, Lin S, Roe BA, Zhu H (2008). Alfalfa benefits from *Medicago truncatula*: the RCT1 gene from *M. truncatula* confers broad-spectrum resistance to anthracnose in alfalfa. Proc Natl Acad Sci USA 105:12164-12169.
- Yang T, Bao S-y, Ford R, Jia T-j, Guan J-p, He Y-h, et al. (2012). High-throughput novel microsatellite marker of faba bean via next generation sequencing. BMC Genomics. 13(1):602.
- Yap MW, Nisole S, Stoye JP (2005). A single amino acid change in the SPRY domain of human Trim5 alpha leads to HIV-1 restriction. Curr Biol 15:73-78.
- Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005). Sequencing the Genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiology. 137: 1174-1181.
- Young ND, Steiner KE, Depamphilis CW (1999). The evolution of parasitism in *Scrophulariaceae/Orobanchaceae*: plastid gene sequences refute an evolutionary transition series. Ann Missouri Bot Garden 86:876–893.
- Young ND, Debellé F, Oldroyd G, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, et al 2011 (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480, 520–524.

- Zahran MA (1982). Control of parasitic plants (broomrape and dodder) in different crops in Egypt. Final Technical Report, Agricultural Research Program, PL, 53 p.
- Zeid M, Mitchell S, Link W, Carter M, Nawar A, Fulton T, Kresovich S (2009). Simple sequence repeats (SSRs) in faba bean: New loci from *Orobanche*-resistant cultivar 'Giza 402'. *Plant Breeding* 128:149-155.
- Zhu H , Kim DJ ,Baek JM, Choi HK, Ellis LC, Kuester H, McCombie WR, Peng HM, Cook DR (2003). Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization, *Plant Physiol*, vol. 131 (pg. 1018-1026).
- Zhu H, Choi H, Cook DR, Shoemaker RC (2005). Bridging model and crop legumes through comparative genomics. *Plant Physiology*. 137:1189-1196.
- Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, et al. (2004). Simple cDNA normalization using kamchatka crab duplex specific nuclease. *Nucleic Acids Res.* 32(3):e37.
- Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JDG, Felix G and Boller T, (2004). Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature*, 428, 764-7.
- Zipfel C (2008) Pattern-recognition receptors in plant innate immunity. *Current Opinion in Immunology* 20: 10-16.



# Anexos

## Otras publicaciones y comunicaciones del autor







## RESEARCH ARTICLE

## Open Access

# A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.)

Zlatko Satovic<sup>1,3</sup>, Carmen M Avila<sup>1</sup>, Serafin Cruz-Izquierdo<sup>1,4</sup>, Ramón Díaz-Ruiz<sup>1,5</sup>, Gloria M García-Ruiz<sup>1</sup>, Carmen Palomino<sup>1</sup>, Natalia Gutiérrez<sup>1</sup>, Stefania Vitale<sup>1</sup>, Sara Ocaña-Moral<sup>1</sup>, María Victoria Gutiérrez<sup>1</sup>, José I Cubero<sup>2</sup> and Ana M Torres<sup>1\*</sup>

**Abstract**

**Background:** Faba bean (*Vicia faba* L.) is among the earliest domesticated crops from the Near East. Today this legume is a key protein feed and food worldwide and continues to serve an important role in culinary traditions throughout Middle East, Mediterranean region, China and Ethiopia. Adapted to a wide range of soil types, the main faba bean breeding objectives are to improve yield, resistance to biotic and abiotic stresses, seed quality and other agronomic traits. Genomic approaches aimed at enhancing faba bean breeding programs require high-quality genetic linkage maps to facilitate quantitative trait locus analysis and gene tagging for use in a marker-assisted selection. The objective of this study was to construct a reference consensus map in faba bean by joining the information from the most relevant maps reported so far in this crop.

**Results:** A combination of two approaches, increasing the number of anchor loci in diverse mapping populations and joining the corresponding genetic maps, was used to develop a reference consensus map in faba bean. The map was constructed from three main recombinant inbred populations derived from four parental lines, incorporates 729 markers and is based on 69 common loci. It spans 4,602 cM with a range from 323 to 1041 loci in six main linkage groups or chromosomes, and an average marker density of one locus every 6 cM. Locus order is generally well maintained between the consensus map and the individual maps.

**Conclusion:** We have constructed a reliable and fairly dense consensus genetic linkage map that will serve as a basis for genomic approaches in faba bean research and breeding. The core map contains a larger number of markers than any previous individual map, covers existing gaps and achieves a wider coverage of the large faba bean genome as a whole. This tool can be used as a reference resource for studies in different genetic backgrounds, and provides a framework for transferring genetic information when using different marker technologies. Combined with syntenic approaches, the consensus map will increase marker density in selected genomic regions and will be useful for future faba bean molecular breeding applications.

**Keywords:** Faba bean, Consensus map, Gene based markers, Quantitative trait loci, Comparative mapping, Molecular breeding, Marker-assisted selection, Genomics

\* Correspondence: [anam.torres.romero@juntadeandalucia.es](mailto:anam.torres.romero@juntadeandalucia.es)  
<sup>1</sup>IFAPA, Centro Alameda del Obispo, Área de Mejora y Biotecnología, Avda. Menéndez Pidal s/n, Apdo. 3092, Córdoba 14080, Spain  
Full list of author information is available at the end of the article



© 2013 Satovic et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

## Background

Faba bean (*Vicia faba* L.) is an important food and fodder crop worldwide and a staple in Middle East, Central and East Asia and North Africa. In terms of cultivation area, faba bean ranks fourth among the cool-season food legumes (close to 2.5 million hectares per year) after chickpea, pea and lentil (<http://faostat.fao.org>). Its agricultural role is currently increasing as the crop is receiving a renewed interest in European countries, North America and Australia.

Faba bean is a diploid legume crop ( $2n = 2 \times = 12$ ) with facultative cross-pollination and has one of the largest described genomes among legumes. The genome size is ~13,000 Mb, more than 25 times larger than that of the model *Medicago truncatula*, and contains more than 85% of repetitive DNA [1]. The large chromosome size has allowed faba bean to become one of the cytogenetically best characterised plant species. However, it has also complicated the development of saturated linkage maps and the identification of relevant genes/QTLs (Quantitative Trait Loci). Moreover, it precludes whole-genome shotgun assembly with next generation sequencing technologies.

Genetic linkage maps are essential tools for a wide range of genetic and breeding applications, including the

study of inheritance of qualitative and quantitative traits and the identification of markers linked to relevant agronomic traits. The availability of high-density maps enhances the breeding process through the application of association analyses, map-based cloning or marker-assisted approaches. Table 1 presents a summary of the faba bean linkage maps reported so far.

Prior to 1990, only a few morphological and isozyme loci were mapped in the *V. faba* genome and no extended linkage groups (LGs) had been reported. Primary trisomics and translocation stocks allowed the first assignment of genes and LGs to specific chromosomes [3,16-19]. This approach was further explored to develop physically localized markers and microsatellites (or SSR-Simple Sequence Repeats) from specific chromosomal regions [20]. This led to the integration of the first genetic and physical maps and allowed the unambiguous assignment of LGs to their respective chromosomes.

Preliminary maps constructed with  $F_2$  populations were mostly based on dominant markers such as RAPDs together with morphological, isozyme, seed storage protein genes and microsatellites, which saturated different areas of the genome [5-8,21,22]. First attempts to map genes/QTLs for seed weight [6] and resistance to a parasitic plant (*Orobanche crenata*) and fungal diseases (*Ascochyta*

**Table 1 Information of relevant faba bean mapping populations highlighting the ones used to construct this composite map**

Reference	Cross	Mapping population	No. individuals	No. markers	No. LGs <sup>a</sup>	Length (cM)	Uses <sup>b</sup>
Van de Ven et al. [2]		BC		17	7 (-)	231	
Torres et al. [3]		2 $F_2$	20	51	11 (1)	~300	
Ramsay et al. [4]		BC		23	7 (-)	~300	
Satovic et al. [5]	c	7 $F_2$	813	157	48 (6)	~850	T/C
Vaz Patto et al. [6]	Vf6 × Vf27	3 $F_2$	175	116	13 (7)	~1200	T/C
Román et al. [7,8]	Vf6 × Vf136	$F_2$	196	121	16 (9)	1446	Q
Román et al. [9]	d	11 $F_2$	654	192	14 (5)	1559	T/C
Ávila et al. [10]	29H × Vf136	$F_2$	159	103	18 (6)	1308	Q
Ellwood et al. [11]	Vf6 × Vf27	RIL	96	135	12 (-)	1686	
Arbaoui et al. [12]	Côte d'Or × BPL14628	RIL	101	131	21 (-)	~980	Q
Diaz et al. [17,19]	Vf6 × Vf136	RIL	165	277	21 (9)	2857	Q
Cruz-Izquierdo et al. [13]	Vf6 × Vf27	RIL	124	258	16 (8)	1874	Q
Ma et al. [14]	91825 × K1563	$F_2$	129	128	15 (-)	1587	
Gutiérrez et al. (in press) [15]	29H × Vf136	RIL	119	171	29 (15)	1402	Q
<b>This study</b>	e	3 RIL	408	587 <sup>f</sup>	6 (6)	3515	C
				151 <sup>g</sup>	37 (7)	1171	C

<sup>a</sup>Between brackets no. of linkage groups (LGs) assigned to chromosomes.

<sup>b</sup>T: Assignment of linkage groups to chromosomes by trisomic segregation; C: Development of a composite map; Q: QTL analysis.

<sup>c</sup>Vf6 × Vf2; Vf6 × Vf33; Vf6 × Vf159.

<sup>d</sup>Vf6 × Vf2; Vf6 × Vf27; Vf6 × Vf33; Vf6 × Vf136; Vf6 × Vf136; Vf6 × Vf159.

<sup>e</sup>Vf6 × Vf27; Vf6 × Vf136; 29H × Vf136.

<sup>f</sup>Data of the six main LGs ascribed to chromosomes.

<sup>g</sup>Data of the minor LGs.

Studies carried out by the IFAPA group and considered in this study in bold.

*fabae* and *Uromyces viciae-fabae*) were reported [7,8,10,23]. Using a F<sub>2</sub> population from the cross Vf6 × Vf136, a linkage map was developed to locate QTLs controlling crenate broomrape (*O. crenata*) [7] and *A. fabae* resistance [8]. Nine of the 16 LGs reported could be assigned to specific chromosomes thanks to markers that were common with those of previous studies. Subsequently, a linkage map of an F<sub>2</sub> population from the cross 29H × Vf136, segregating for resistance to the two pathogens, was constructed in which 6 of the 18 LGs were assigned to chromosomes [10].

These faba bean maps did not allow wider mapping comparisons, since they mostly shared dominant and anonymous markers such as RAPDs, with scarce transferrability between genotypes and legume species. Despite this limitation, marker data of 11 F<sub>2</sub> populations (Table 1), all sharing the common female parent Vf6, were used to construct a composite linkage map [9]. After joint segregation analysis of 501 markers in 654 individuals, 192 markers were included in 14 major LGs, of which 5 were unambiguously assigned to specific chromosomes (Table 1). This composite map covered 1,559 cM and was one of the most comprehensive faba bean genetic map published to date [9].

These maps with dominant markers in F<sub>2</sub> were followed by more precise maps constructed in the corresponding RIL populations, using co-dominant markers. In addition to microsatellites, expressed sequence tags (EST) from other legume species emerged as efficient tools in faba bean. A large number of intron-targeted primer pairs (ITAPs), developed within the Grain Legumes Integrated Project-GLIP (<http://www.pcg.in.org/GLIP/pubrep.pdf>), was tested and mapped in two faba bean inbred populations (Vf6 × Vf136, 29H × Vf136). These were used to validate QTLs underlying broomrape and *Ascochyta* resistance in different environments and genetic backgrounds [15,24-26]. A third RIL population derived from cross Vf6 × V27 was used to construct the first exclusively gene-based genetic map in faba bean. It contained 135 ITAPs joined in 12 unassigned LGs, that spanned 1,685 cM, and allowed for the first time the study of macrosyntentic relationships between *V. faba*, *M. truncatula*, *Lens culinaris* and other legume species [11]. After further saturation, the map was used to identify and validate QTLs controlling flowering time and other yield-related traits [13]. Recently, a new map was reported based on the F<sub>2</sub> population from the cross 91825 × K1563, which includes 128 SSRs markers arranged in 15 unassigned LGs [14]. Unfortunately the lack of common markers prevented comparisons with previous mapping studies.

To date 14 major genetic maps have been constructed in faba bean (Table 1). Integrating the information of multiple populations from diverse genetic backgrounds offers several advantages over individual genetic maps:

(i) a larger number of loci is mapped than in single crosses, (ii) the relative position of common markers can be determined across the mapping populations, (iii) better genome coverage and opportunities to validate marker order, (iv) better assignment of LGs to chromosomes, (v) it allows comparison of genes/QTLs of interest across maps and, (vi) it provides the basis for comparing genomes between related species [27-29]. Consensus genetic maps have been developed in many crops such as wheat [30], maize [31], barley [32] and rice [33], and in the legume crops soybean [34], pea [35], chickpea [36], phaseolus [37], pigeonpea [38], cowpea [39], groundnut [40] and red clover [41].

With the development of genome sequencing projects and expression studies in different model and crop legumes, the construction of a faba bean consensus genetic map has become possible. The objectives of this study were to: (1) saturate the faba bean maps developed in RIL progenies with common gene based markers to facilitate anchoring of linkage groups from different populations, (2) update the position of the most relevant faba bean QTLs controlling resistance and yield related traits using Bulk Segregant Analysis (BSA), and (3) construct a reference map integrating all the genomic information reported so far in this crop. To this aim, we fused information of 11 F<sub>2</sub> populations and marker data of three RIL genetic maps to derive a consensus map including 729 markers and covering 4,602 cM. The six main LGs could be unambiguously assigned to their corresponding faba bean chromosomes. The map represents a significant improvement over single-population genetic maps and provides a new tool of reference for faba bean breeding and genomic approaches.

## Results

### Individual maps and QTL analysis

#### Cross Vf6 × Vf27 (RIL1)

The first RIL1 map [13] included 258 markers joined in 16 LGs and covering 1,875 cM. The linkage groups were composed of 2-45 loci with an average marker interval of 7.3 cM. The map allowed to identify and validate QTLs controlling 5 flowering and reproductive traits [13]: days to flowering (DF), flowering length (FL), pod length (PL), number of seeds per pod (NSP) and number of ovules per pod (NOP), located mainly in chromosomes (chr.) V and VI (Additional file 1: Table S1).

For the extended RIL1 map constructed herein, 313 polymorphic markers were used in the global analysis (Table 2). Of these, 273 were assembled in 19 LGs, 11 of which could be assigned to specific chromosomes. The distance covered by the map was 2,183 cM with an average marker interval of 10 cM. Sixty five of the markers are common with the other two RIL populations, 25

**Table 2 Number and type of markers genotyped in each inbred population**

Type of marker	RIL population		
	Vf6 × Vf27	Vf6 × Vf136	29H × Vf136
ITAP	176	59	46
RAPD	107	327	145
SSR	16	6	6
Isozyme	5	4	0
RGA	4	5	5
Morphological trait	2	0	0
Seed storage protein	2	2	2
DR	1	1	1
Total	313	404	205

with RIL2 (Vf6 × Vf136) only, 15 with RIL3 (29H × Vf136) only and 25 with both (Table 3).

#### Cross Vf6 × Vf136 (RIL2)

The previously published RIL2 map [24,25] was based on 277 marker loci assembled in 21 LGs (16 consisting of 3 or more markers) that span 2,857 cM with an average marker interval of 12.7 cM. In this population, 2 QTLs controlling ascochyta blight resistance (*Af1* and *Af2*) were identified on chr. III and chr. II [24]. In parallel studies, 2 QTLs (*Oj1* and *Oj2*) controlling *O. foetida* resistance and 4r QTLs controlling *O. crenata* resistance (*Oc2-Oc5*) were detected [25]. *Oc2* and *Oc3* were stable in at least two of the three environments, while *Oc4* and *Oc5* were only detected in one environment and thus appeared to be environment-dependent.

In an attempt to saturate the regions bearing the *O. crenata* and *A. fabae* QTLs, a BSA approach based on RAPD markers was applied. Two-hundred and eight of the 748 RAPD primers assayed in the cross revealed promising polymorphisms between at least one pair of bulks and were subsequently used to screen 14 plants individually. Only 41 of the RAPD primers maintained the expected pattern of polymorphism and were used to screen the entire RIL2 population, resolving 39 scorable polymorphic markers. Of these, 31 markers exhibited the expected segregation pattern and were thus included in the RIL data set for mapping and QTL analysis. Thirty of the 31 RAPD markers were mapped, 24 to target regions [11 to chr. VI (*Oc2*), 8 to chr. II (*Oc3* and *Af2*) and 5 to chr. III (*Af1*)] and 6 to other LGs (Table 4; Figure 1). To increase the number of common markers between different faba bean crosses, additional ITAP markers were scored and the map was reconstructed using 404 segregating loci.

The map obtained in this study consists of 364 mapped loci assembled into 21 LGs, of which 13 were assigned to specific chromosomes. The genome distance

covered by the map is 3,537 cM, with an average density of one marker locus every 12.6 cM (Table 4). Fifty four markers included in RIL2 are common with the other two populations, 25 with RIL1 only, 4 with RIL3 only and 25 with both (Table 3).

As mentioned above, QTLs *Oc2* and *Oc3*, that confer broomrape resistance, were previously validated in 2 of the 3 field trials assayed [26]. By saturating the target regions we were able to identify significant QTLs in each trial. Thus, the conservation of QTLs both across generations and environments was confirmed. In case of *Oc2* (Additional file 1: Table S1B) a newly added marker, OPAG11<sub>956</sub>, was the closest to LOD peak value in both Córdoba 2003 and Córdoba 2004 datasets. Three additional QTLs were identified in chr. II (*Oc14\_C3* and *Oc15\_M4*) and V (*Oc15\_C4*). None of these QTLs was stable in the different field assays or years, pointing towards an environment-dependent expression. In the case of ascochyta blight resistance, the analysis which was based on a more saturated map yielded higher LOD scores and narrower confidence intervals for both QTLs (*Af1* and *Af2*). The new marker OPZ08<sub>530</sub> was the closest to LOD peak value of both *Af1* QTLs identified in leaves and stems (Additional file 1: Table S1B). Addition of new markers in target areas of the map was an efficient method to increase the genome coverage and to obtain more defined QTLs.

#### Cross 29H × Vf136 (RIL3)

The third map used in the present study was reported recently and includes 172 markers spanning 1402 cM [15]. The linkage groups were composed of 2 to 25 loci with a marker interval of 9.87 cM. Seven QTLs for *O. crenata* (*Oc7* to *Oc13*) and 3 QTLs for *O. foetida* (*Oj3* to *Oj5*) were identified in this map. *Oc7* was detected along three years, explaining between 22% and 33% of the phenotypic variation. It has been suggested that *Oc2* (previously reported in RIL2) and *Oc7*, which are both located in chr. VI and validated in different environments and genetic backgrounds, might correspond to the same QTL region (Figure 1; Additional file 1: Table S1). The new analysis considered 205 marker loci segregating in this population, of which 25 were common to both other RIL populations, 25 to RIL1 only and 4 to RIL3 only (Table 3).

#### Consensus linkage map

Three sets of faba bean mapping data were used in the construction of a consensus map connecting information of 11 F<sub>2</sub> populations, marker data of 3 RILs, as well as new markers genotyped in the present study (Tables 1 and 2). The number of individual marker loci ranged from 313 in RIL1 to 404 in RIL2 and 205 in RIL3

**Table 3 Common markers used as anchors for map integration**

No.	Marker	Marker type	RIL population		
			Vf6 × Vf27	Vf6 × Vf136	29H × Vf136
1	PeaβGlu	DR	x	x	
2	Ppx-1	Isozyme	x	x	
3	Sod-1	Isozyme	x	x	
4	1433P	ITAP	x		x
5	6DCS	ITAP	x	x	x
6	AIGPb	ITAP	x	x	
7	AnMt513	ITAP	x	x	x
8	AnMt537	ITAP	x	x	
9	BGAL	ITAP	x		x
10	cgP137F	ITAP	x		x
11	GBNP	ITAP	x		x
12	GLIP171b	ITAP	x	x	
13	GLIP651	ITAP	x	x	
14	HBP2	ITAP	x		x
15	LG007	ITAP	x		x
16	LG031	ITAP	x	x	x
17	LG041	ITAP	x	x	
18	LG054	ITAP	x	x	
19	LG068	ITAP	x	x	x
20	Lup066	ITAP	x	x	x
21	Lup299	ITAP	x	x	x
22	MMK1a	ITAP	x		x
23	mtmt_GEN_00012_03_1	ITAP	x	x	x
24	mtmt_GEN_00022_02_1	ITAP		x	x
25	mtmt_GEN_00024_04_1	ITAP	x	x	
26	mtmt_GEN_00032_01_1/a	ITAP	x	x	
27	mtmt_GEN_00036_02_1/a	ITAP	x	x	x
28	mtmt_GEN_00103_01_1	ITAP	x	x	
29	mtmt_GEN_00447_04_3	ITAP	x		x
30	mtmt_GEN_00477_04_1	ITAP	x		x
31	mtmt_GEN_00510_01_1	ITAP	x		x
32	mtmt_GEN_00757_03_1	ITAP	x	x	x
33	mtmt_GEN_00861_03_1	ITAP	x		x
34	mtmt_GEN_00866_02_1	ITAP		x	x
35	mtmt_GEN_00892_01_3	ITAP	x	x	x
36	mtmt_GEN_00995_01_1	ITAP	x	x	x
37	mtmt_GEN_01017_03_3	ITAP	x	x	
38	mtmt_GEN_01102_02_1	ITAP	x	x	x
39	mtmt_GEN_01109_01_1	ITAP	x	x	x
40	mtmt_GEN_01115_02_1	ITAP	x		x
41	mtmt_GEN_01130_02_1	ITAP	x	x	x
42	mtmt_GEN_01951_11_1a	ITAP	x	x	x
43	Pis_GEN_14_7_1	ITAP		x	x

**Table 3 Common markers used as anchors for map integration (Continued)**

44	Pis_GEN_20_1_1	ITAP	x	x	
45	Pis_GEN_23_5_6_1	ITAP	x	x	
46	Pis_GEN_25_2_3_1	ITAP	x	x	
47	Pis_GEN_5_4_5_1	ITAP	x	x	
48	Pis_GEN_57_1_2_1	ITAP	x	x	
49	Pis_GEN_6_3_1	ITAP	x	x	x
50	Pis_GEN_7_1_2_1	ITAP	x	x	
51	psat_EST_00180_01_2	ITAP	x	x	
52	psat_EST_00190_01_1	ITAP	x		x
53	PsMnSOD	ITAP		x	x
54	psmt_EST_00196_01_1	ITAP	x	x	
55	RBPC/O	ITAP	x		x
56	RNAR	ITAP	x		x
57	SAT	ITAP	x	x	
58	TBB2	ITAP	x	x	x
59	UNK28	ITAP	x	x	
60	RGA01	RGA	x	x	x
61	RGA03	RGA	x	x	x
62	RGA08	RGA	x	x	x
63	RGA09	RGA	x	x	x
64	B3	Seed storage protein	x	x	
65	B4	Seed storage protein	x	x	x
66	GA4	SSR	x	x	x
67	GAlI30	SSR	x	x	
68	GAlI59	SSR	x	x	x
69	JF1GA3	SSR	x	x	x
Total number of common markers			65	54	44

(Table 2). Chi-square test was performed on new marker genotyping data for individual mapping populations to test the null hypothesis of segregation ratios of 1:1. A variable percentage of distorted markers ( $P < 0.01$ ) was observed, ranging from 2.93% in RIL3 to 7.35% in RIL1. A list of the marker loci is provided in Additional file 2: Table S2. QTL regions characterized in previous studies were also covered by the present consensus map.

The number of anchor markers for pairwise comparisons was initially small, with less than 20 markers in common. However, after the new marker analysis the number triplicated to 69 (2 isozymes, 4 SSRs, 2 leguminines, 4 RGAs, 1 DR gene and 56 ITAPs). Sixty five of these, present in RIL1, were primarily used as bridges to integrate the individual maps into a single consensus map (Table 3). Twenty five markers were common in the 3 maps, while the remaining 44 were shared by at least 2 mapping populations. As a result, marker segregation data were assembled for a total of 828 marker

loci, 759 of which were unique i.e. mapped only in one population (Table 5; Additional file 2: Table S2).

A total of 729 marker loci were assembled into the 43 LGs constituting the consensus map (Figure 1; Table 5), while 99 markers remained unlinked. The 6 major LGs contained between 53 (chr. V) and 165 marker loci (chr. I), and were assigned to the corresponding chromosomes. Seven additional LGs (Ia to IVa) could also be assigned thanks to the presence of loci previously located in individual chromosomes. Thirty one LGs consisted of 2–5 markers, and the remaining 6 LGs contained between 6 and 21 loci (Table 5; Additional file 2: Table S2). The total length of the consensus genetic linkage map was 4,613 cM, of which 3,442 cM were covered by the 6 main LGs/chromosomes. The length of these major LGs ranged from 323 cM (chr. V) to 1041 cM (the large metacentric chr. I). The entire consensus map had an average marker density of one marker per 10.7 cM, which was reduced to 6 cM when considering only the 6

**Table 4 Linkage map of Vf6 × Vf136 (RIL2)**

No.	Chromosome/ LG*	No. markers	New markers**	Length (cM)	Intermarker distance (cM)
1	I-1	69	1 / 7	730.18	10.74
2	I-2	15	2 / 2	164.43	11.75
3	I-3	2	0 / 2	21.18	21.18
4	Ic	2	0 / 0	8.67	8.67
5	II	53	8 / 11	522.80	10.05
6	IIa	11	0 / 1	80.19	8.02
7	III-1	53	5 / 12	462.63	8.90
8	III-2	4	0 / 1	33.88	11.29
9	IV	38	2 / 2	346.53	9.37
10	V-1	27	0 / 5	290.67	11.18
11	V-2	7	0 / 1	55.24	9.21
12	VI-1	28	0 / 8	284.94	10.55
13	VI-2	24	11 / 3	228.73	9.94
14	LG01	8	0 / 0	61.63	8.80
15	LG04	6	0 / 0	86.05	17.21
16	LG05	5	1 / 0	48.55	12.14
17	LG06	4	0 / 0	37.40	12.47
18	LG11	2	0 / 0	20.27	20.27
19	LG21	2	0 / 0	22.14	22.14
20	LG22	2	0 / 2	11.93	11.93
21	LG23	2	0 / 0	18.80	18.80
	Mapped	364	30 / 57	3536.86	12.60
	Unmapped	40			
	Total	404			

\*Linkage groups correspond to those of the composite map.

\*\*Markers mapped in addition to already published linkage map [24,25]: first number represents the number of markers added by saturation mapping targeted at the regions conferring resistance to *O. crenata* and *Ascochyta fabae* (II, III-1, VI-2), while the second is the number of markers added to increase the number of common markers among different faba bean crosses.

main LGs. The marker order of the integrated map was largely collinear with the three individual maps, although a few local inversions and marker rearrangements over short intervals were observed.

#### Integration of QTL information

The number of QTL studies in faba bean is relatively low compared to other major legume species. Most traits have been genetically characterized in only one or two different mapping experiments, which limits the meta-analysis of QTLs in this species. Moreover, QTL intervals did not always include the minimum of two anchor markers, which is required for their projection onto the consensus map. Nevertheless, by comparing the maps published to date we provide a synthetic view of the most relevant loci controlling polygenic traits in faba bean. Further mapping of common markers between

maps will be crucial to enhance the comparison of QTL positions from different mapping studies and to refine the localization of hot-spot genomic regions.

The 5 faba bean mapping experiments in the 3 RIL populations reported so far, identified 37 QTLs for 9 traits (Additional file 1: Table S1). Most of the QTL analyses focused on biotic stresses (e.g. broomrape and ascochyta blight resistance). The number of QTLs for broomrape resistance is 15 for *O. crenata* and 5 for *O. foetida*. Meanwhile the *A. fabae* resistance QTLs were reduced to 2, *Af1* and *Af2*, that were conserved among populations ( $F_2$  and RIL) and environments. Regarding flowering and yield related traits, the number of stable QTLs reported were 1 (NOP), 2 (FL), 4 (PL), 3 (NSP) and 5 (DF) (Additional file 1: Table S1). The distribution of these 37 QTLs varied from 9 in chr. I, 8 in chr. II, 4 in chr. III and VI, 3 in chr. IV, 6 in chr. V. The last three QTLs remained unassigned.

#### Discussion

During the last decade, significant progress was made in the development of genotyping tools. This allowed the addition of a large number of robust and transferrable marker loci in the genetic maps of relevant crop species. In faba bean, mapping studies were initiated in the 90's with the development of the first maps in  $F_2$  populations using mostly RAPDs together with SSRs, isozymes and morphological markers. Previous studies comparing these linkage maps have been reported [9]. The use of a recurrent parent (Vf6) in all the  $F_2$  populations then allowed to join data from different progenies by means of common markers in the female parent. Moreover, the use of trisomic families for chromosomes III, IV, V and VI allowed allocation of LGs to chromosomes [9]. Ever since, attempts have been made to increase marker density using new SSRs and gene-based markers in RIL populations. The main objective of the present study was the development of a high density consensus genetic map that integrates all the relevant maps reported so far and serves as a reference map for the international faba bean community.

Building a consensus map is not possible without common or bridge loci on each LG or chromosome. For this reason, a number of additional markers was genotyped in each mapping population to increase the number of common markers among them. A bridge marker was considered as such when its name and position were the same in the different mapping populations. The genetic map was created combining two approaches: (a) increasing the number of anchor marker loci in the different populations, (b) merging the resulting genetic maps through markers common to three RIL populations with MergeMap [42] as reported in many other crop species [39,43-47].



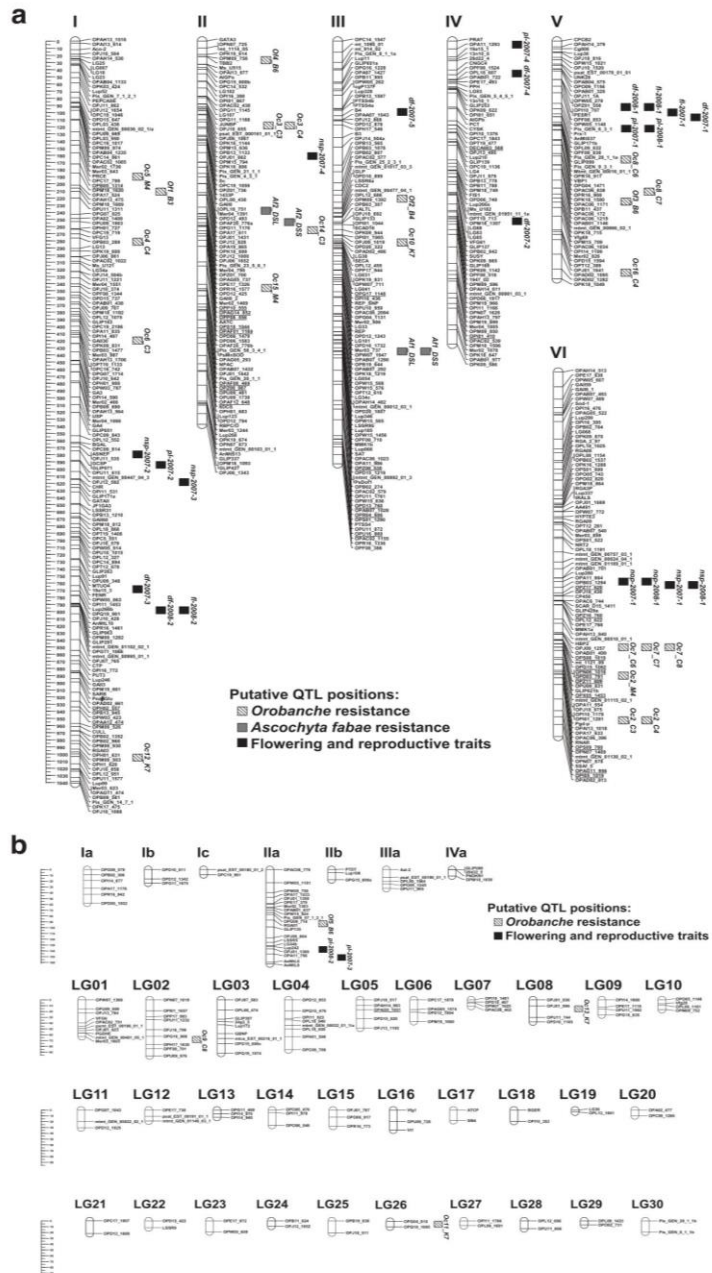


Figure 1 (See legend on next page.)

(See figure on previous page.)

**Figure 1 Faba bean consensus map showing: (a) the six main linkage groups or chromosomes (578 loci) and (b) minor linkage groups (30) or small chromosome fragments (7).** Additional markers derived from the BSA analysis are underlined. Boxes represent putative locations of QTLs. Black boxes were used for flowering and yield related traits: days to flowering (*df*), flowering length (*fl*), pod length (*pl*), number of ovules per pod (*nop*) and number of seeds per pod (*nsp*). Grey boxes: *Ascochyta fabae* (*Af*) QTLs. Stripped boxes: *Orobanche crenata* (*Oc*) and *Orobanche foetida* QTLs. Marker distance is given in cM.

Using this approach, segregation data for 729 marker were assembled on 43 LGs. In terms of marker order, the consensus map contains few changes compared to the individual maps. Small discrepancies in the marker order or position in some LGs might be due to (i) different population sizes used, (ii) weak linkages existing in the different maps, or (iii) missing or poor quality data, rather than to real chromosome rearrangements. As reported in previous studies in *Vitis vinifera* L. [47], phaseolus [37] or *Brassica napus* [44], the faba bean consensus marker order is significantly more reliable than that of the individual maps because of the higher number of individuals and recombination events occurring across three or more populations.

Based on previous LG/marker allocations, 13 of the larger LGs could be assigned to specific chromosomes while 30 LGs remained unassigned. Considering the enormous size of the faba bean genome, unassigned LGs may be due to recombination gaps at the distal ends of the main LGs because of a lack of marker loci. None of the main LGs differed considerably in marker density. The length of our core map was 3,431 cM, which is higher than the single RIL maps. In many other species the increased size of the composite map was attributed to an improved coverage of the chromosome ends [37,48-50].

All the individual maps reported to date allocate LGs to 5 of the 6 faba bean chromosomes, excluding chr. IV. After acknowledging the erroneous assignment of LG I. B, which actually corresponds to chr. IV [51], the new integrated map anchors for the first time the main LGs to the whole chromosome complement of the species. This information was used here to update the reported large-scale synteny between LGs and/or chromosomes of *M. truncatula* and cool season grain legumes such as pea, chickpea, lens and faba bean [52]. Figure 2 shows the main syntenic blocks and rearrangements among these species and their correspondence to the six faba bean chromosomes.

The integration of a high number of coding regions in the consensus map provides an excellent framework for downstream analyses, including comparisons between the locations of major genes for important traits or QTL positions between populations from different crosses. Moreover, combining syntenic studies with a consensus map will contribute to increase marker density in genomic regions of interest for indirect selection or for map

based cloning [37]. Successful application of consensus maps for synteny based candidate gene identification and/or definition of QTL location has been extensively used both in cereals [59-62] and in legumes [35,37,63,64].

The RIL mapping populations considered in the consensus map were used previously for detecting QTLs of agronomically important traits. These are displayed in Figure 1, together with the QTLs identified in the improved Vf6 × Vf136 map and detected in the present study. In order to increase the density of loci around the OTLs for broomrape and ascochyta blight resistance, we used BSA in contrasted DNA pools. The BSA approach has been applied in numerous studies and provides a platform for high-resolution genetic analysis [65]. In the present study, 24 of 31 RAPD markers were mapped to the major linkage groups and allowed more accurate determination of QTL locations and effects. These results highlight the usefulness of BSA based on markers flanking QTLs, as an efficient tool for saturation of targeted regions, opening the possibility of future marker-assisted selection for these traits.

Faba bean has been considered a “genomic orphan” crop with a huge and complex genome and limited availability of genetic and genomic resources. At present, the situation has greatly improved thanks to the technological advances in high-throughput sequencing and genotyping, together with the access to genomic and transcriptomic tools. Genome-wide transcription profiling by deepSuperSAGE was recently used for quantifying the transcriptional changes elicited by *A. fabae* and to identify candidate resistance genes governing faba bean responses to this fungal pathogen [66]. Several genome libraries have been constructed and characterized for putative SSR sequences using the Roche 454 GS FLX Titanium Sequencing Platform [67,68]. These transcriptomic studies provide a foundation for the identification of novel regulators associated with faba bean-pathogen interactions and also a valuable source of markers for molecular breeding applications in this crop.

Translation of genomic resources from the model species *M. truncatula* or other sequenced related legume species such as chickpea, should be further exploited to raise the prospects in molecular faba bean breeding programs. The availability of large sets of conserved ESTs from model or related species constitutes a valuable source of markers that are physically associated with coding regions. These are good candidates for gene cloning or faba bean

**Table 5 Composite map of faba bean (*Vicia faba* L.) genome**

No.	Chromosome/LG	No. markers	Length (cM)	Intermarker distance (cM)
1	I	165	1041.20	6.35
2	II	93	537.60	5.84
3	III	109	593.14	5.49
4	IV	70	425.20	6.16
5	V	53	333.60	6.42
6	VI	88	511.22	5.88
7	Ia	6	57.80	11.56
8	Ib	3	23.25	11.63
9	Ic	2	8.67	8.67
10	IIa	21	162.46	8.12
11	IIb	3	17.48	8.74
12	IIIa	5	32.63	8.16
13	IVa	4	11.02	3.67
14	LG01	10	63.25	7.03
15	LG02	9	96.45	12.06
16	LG03	9	91.99	11.50
17	LG04	8	86.04	12.29
18	LG05	5	48.55	12.14
19	LG06	4	37.40	12.47
20	LG07	4	11.07	3.69
21	LG08	4	37.86	12.62
22	LG09	4	25.98	8.66
23	LG10	4	16.75	5.58
24	LG11	3	30.36	15.18
25	LG12	3	18.75	9.38
26	LG13	3	12.63	6.31
27	LG14	3	26.43	13.22
28	LG15	3	28.03	14.01
29	LG16	3	34.35	17.18
30	LG17	2	18.24	18.24
31	LG18	2	19.48	19.48
32	LG19	2	3.48	3.48
33	LG20	2	9.80	9.80
34	LG21	2	22.14	22.14
35	LG22	2	11.93	11.93
36	LG23	2	18.80	18.80
37	LG24	2	9.34	9.34
38	LG25	2	20.67	20.67
39	LG26	2	9.37	9.37
40	LG27	2	8.41	8.41
41	LG28	2	13.43	13.43
42	LG29	2	6.89	6.89
43	LG30	2	19.37	19.37

**Table 5 Composite map of faba bean (*Vicia faba* L.) genome (Continued)**

Mapped	729	4612.52	10.73
Unmapped	99		
Total	828		
Main LGs (No. 1–6)	578	3441.96	6.02

marker assisted selection. This is the approach used in this study to integrate all the faba bean genomic information so far reported, and to generate a new tool of reference for faba bean breeding and genomics approaches.

### Conclusions

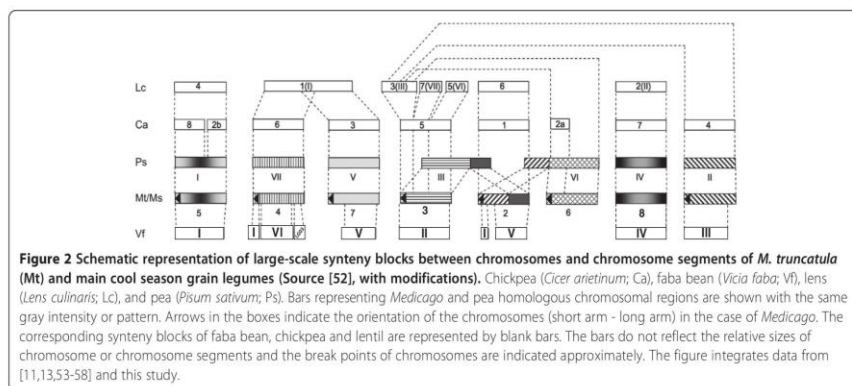
We have constructed the first marker consensus genetic linkage map for faba bean by integrating segregation data from three recombinant inbred line populations, together with new common markers genotyped in this study. The final integrated map has allowed to (i) join a larger number of markers than in any previous individual map, (ii) obtain a more complete coverage of the faba bean genome, (iii) fill a number of gaps in previous independent maps, and (iv) improve the resolution of key QTLs. The colinearity of the consensus map was well maintained and will serve as reference for future faba bean multiple-line cross QTL mapping studies. Since 60% of the markers in the most developed map (RIL1) corresponded to coding regions, this consensus enhanced-density faba bean map provides a functional framework for candidate gene studies, expression analysis, comparative genomics, evolution studies and anchoring of the future faba bean genome sequences.

### Methods

#### Mapping populations

The most recent maps of three RIL mapping populations, Vf6 × Vf27 (RIL1), Vf6 × Vf136 (RIL 2) and 29H × Vf136 (RIL3), were used to develop an integrated map of faba bean (Table 1). Vf6 was a common female parent in two progenies, Vf6 × Vf136, which segregates for broomrape and *Ascochyta* resistance [7,8,24–26] and Vf6 × Vf27, first reported by [6] and further used to construct the first exclusively gene-based genetic map in the species [11] and to identify and validate QTLs controlling flowering time and other yield-related traits [13]. Vf136 was the common male parent with the third population, 29H × Vf136, segregating for resistance to broomrape and *A. fabae* [10,15]. The populations consisted of 124 RILs for RIL1, 165 individual lines for RIL2, and 119 for RIL3.

The female parents, Vf6 and 29H are equina medium-seeded field beans with beige seed coat and resistance to *A. fabae*, Vf136 is also an equina type with reported resistance to broomrape and Vf27 is a black and small-



seeded paucijuga form, supposedly close to a putative wild faba bean progenitor [69].

#### Marker analyses

Genomic DNA was extracted from young leaves using liquid nitrogen and the procedure was as described by [3]. To increase marker density and to provide common markers to anchor the LGs from different populations, new markers from different sources were assayed. A set of SSRs, resistant gene analogs (RGAs), defence related (DR) genes and ITAPs designed from different legume ESTs (*M. truncatula*, pea, lentil, lupin and soybean), were tested in the parental lines and the polymorphic ones genotyped in the corresponding RIL population.

#### SSR markers

Fifty four faba bean SSRs [70] and 41 pea SSRs [71], were assayed for polymorphism among the parental lines using their respective protocols. In case of pea SSRs the reaction mixture was modified slightly by using 2,5 mM of MgCl<sub>2</sub> and 1U Taq polymerase instead, to facilitate the orthologous amplification. SSRs revealing consistent and easily scorable bands were genotyped in the whole populations after electrophoresis in 2.5% - 3% agarose gels.

#### RGAs and DR genes

Ten RGA classes were tested using PCR conditions described by [72]. To reveal polymorphism, amplification products for each RGA class were digested with a set of restriction enzymes according to the manufacturer's instructions to obtain CAPS (Cleaved Amplified Polymorphic Sequences). Twelve additional RGAs [73,74] along with 12 DR genes, cloned and mapped in different legume species and mapped in pea [74], were also

assayed. Amplifications and PCR conditions were as described by [72].

#### Intron-targeted amplified polymorphic markers (ITAPs)

A total of 635 EST derived markers developed within the Grain Legumes Integrated Project (GLIP-Food-CT-2004-506223), were tested for polymorphism among the parental lines using the amplification protocols reported by [13,26]. Special efforts were focused on genotyping the ITAPs previously mapped in the most advanced Vf6 × Vf27 map [11,13]. As mentioned above, when no polymorphism was detected on agarose gels, PCR products amplified from both parents were digested with a range of restriction endonucleases in order to detect a SNP as a CAPS (Cleaved Amplified Polymorphic Sequence) marker that was further genotyped in the corresponding population.

#### Saturation mapping

In order to saturate targeted regions conferring broomrape or ascochyta blight resistance we applied the BSA [75] based on previous QTL mapping information in cross Vf6 × Vf136 [24,26]. BSA has been widely adopted as a method to rapidly identify molecular markers in specific genome regions. The BSA principle consists in pooling DNAs from individuals from a segregating population according to two phenotypic classes. The resulting DNA bulks are equivalent to those from two Near Isogenic Lines (NILs) for which is assumed to generate a random genetic background at all other unlinked loci. In this study segregating individuals were grouped according to the genotype of markers flanking already localized QTLs. The contrasting pools were then screened with new markers in order to identify recombinants within each QTL interval.

QTLs underlying resistance to *O. crenata* and *A. fabae* were named *Oc* and *Af*, respectively [24,26]. Accordingly,

bulks of plants fixed for alleles of the two markers flanking four QTL regions were selected from the RIL population: OPN07<sub>1409</sub> and OPA113<sub>1018</sub> (flanking *Oc2* on linkage group VI.B), OPC19<sub>1059</sub> and OPD12<sub>425</sub> (*Oc3* on LG II.A), OPF08<sub>710</sub> and OPW15<sub>576</sub> (*Afl* on chr. III), OPAG5<sub>737</sub> and MERO2<sub>1469</sub> (*Af2* on chr. II). A total of 748 RAPD primers was used in search for polymorphisms between the two bulks. For a given target region, markers showing expected differences between the pair of bulks were subsequently used to screen 14 plants individually. Markers that maintained the expected pattern of polymorphisms, were then used to screen the entire RIL population.

#### Quantitative traits

Traits and QTL information were selected from seven published works [7,8,10,13,15,24,26] and supplemented by the bulked segregant analysis (in cross VF6 × VF136) and the saturation process described above. Trait descriptions, evaluation methods and abbreviations were assigned according to the previous references. Thus, the nine traits considered (Figure 1; Additional file 2), were the following: (1) *Orobanche crenata* resistance (trait abbreviation OC); (2) *Orobanche foetida* resistance (OF); (3) *Ascochyta fabae* resistance: disease severity on leaves (DSL); (4) *Ascochyta faba* resistance: disease severity on stems (DSS); (5) Days to flowering (DF); (6) Flowering length (FL); (7) Pod length (PL); (8) Number of ovules per pod (NOP); (9) Number of seeds per pod (NSP). Most of these QTLs showed to be stable as were identified and validated in different environments or genetic background. Each QTL was treated independently, making it possible to notice the number of times that a QTL is reported in a similar genomic location across independent experiments (Figure 1; Additional file 2).

#### Data analysis

##### Map construction and QTL analysis in VF6 × VF136 (RIL 2)

MAPMAKER 3.0 [76] was used to identify linkage groups using an LOD score of four as the threshold for considering significant linkage. MSTMap software [77] was used to determine marker orders by finding the minimum spanning tree of a graph for each linkage group. MAPMAKER was used to confirm marker orders determined by MSTMap and to convert the recombination fractions to centiMorgans (cM) using the mapping function of Kosambi [78].

QTL analysis was conducted using composite interval mapping (CIM) and multiple interval mapping (MIM) in Windows QTL Cartographer V2.5 [79]. Markers to be used as cofactors for CIM were selected by forward-backward stepwise regression. The number of markers controlling the genetic background in CIM was set to five.

The thresholds for the detection of QTLs were estimated by permutations analysis [80] using 1,000 permutations.

#### Consensus map construction

Three individual genetic maps were used to generate a consensus map using MergeMap [42] by converting the individual maps into directed acyclic graphs (DAGs) that are then merged in consensus graph on the basis of their shared vertices [81]. As MergeMap tends to inflate genetic distances in the consensus genetic map [42,82], marker data from different mapping populations were pooled together and the order of each consensus linkage group as established by MergeMap was set, in order to calculate genetic distances using MAPMAKER. The consensus map for each linkage group was visualized by MapChart [83].

#### Additional files

**Additional file 1: Table S1.** (A) Putative QTLs for flowering time and yield related traits detected in the faba bean RIL population VF6 × VF27 (from Cruz-Izquierdo et al., 2012 with modifications). (B) Putative QTLs for *Ascochyta fabae*, *Orobanche crenata* and *Orobanche foetida* resistance detected in the faba bean RIL population VF6 × VF136 (Díaz-Ruiz et al., 2009a; 2009b; 2010 and this study). (C) Putative QTLs for *Orobanche crenata* and *Orobanche foetida* resistance detected in the faba bean RIL population 29H × VF136 (from Gutierrez et al., 2013 with modifications).  
**Additional file 2: Table S2.** Information on the markers used in this study and mapped in the three faba bean RIL populations.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

ZS performed the statistical analyses for map merging, integrated QTL information, prepared tables and graphic representations and helped to draft the manuscript. CMA generated marker and QTL data for the mapping populations, participated in the design of the study and assisted with manuscript preparation and editing. SCI, RDR performed additional marker genotyping and provided new QTL data. GGR, CP, NG, SV, SOM and MVG provided new marker data for mapping saturation and anchoring points for map integration. JJC assisted with manuscript editing. AMT coordinated the map integration study, contribute to the analysis and interpretation of data, provided the marker and QTL data and drafted the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

This research has been funded by the Spanish projects AGL2008-02305/AGR, RTA2010-00059 and IPT-2011-1259-010000, co-financed by FEDER, and by the EU project GLIP (FP6- FOOD-CT-2004-506223). The authors thank Juan Prieto, Gloria de Lara and Ana Pozo for excellent technical assistance in the laboratory, greenhouse and field. We are grateful to A. Di Pietro for carefully reading the manuscript.

#### Author details

<sup>1</sup>IFAPA, Centro Alameda del Obispo, Área de Mejora y Biotecnología, Avda. Menéndez Pidal s/n, Apdo. 3092, Córdoba 14080, Spain. <sup>2</sup>Departamento de Mejora Genética, IAS-CSIC, Apdo. 4084, Córdoba 14080, Spain. <sup>3</sup>Present address: Department of Seed Science and Technology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia. <sup>4</sup>Colegio de Postgraduados, Recursos Genéticos y Productividad – Genética, Campus Montecillo, Km 36.5 Carretera México-Tezcoco, CP, Texcoco, Edo. de México 56230, México. <sup>5</sup>Colegio de Postgraduados, Campus Puebla, Km 125.5 Carretera México-Puebla, CP, Puebla, Pue 72760, México.

Received: 16 September 2013 Accepted: 12 December 2013  
 Published: 30 December 2013

## References

- Flavell RB, Bennett MD, Smith JB, Smith DB: **Genome size and proportion of repeated nucleotide sequence DNA in plants.** *Biochem Genet* 1974, **12**:257–269.
- van de Ven WTG, Waugh R, Duncan N, Ramsay G, Dow N, Powell W: **Development of a genetic linkage map in *Vicia faba* using molecular and biochemical techniques.** *Aspects Appl Biol* 1991, **27**:49–54.
- Torres AM, Weeden NF, Martin A: **Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*.** *Theor Appl Genet* 1993, **85**:937–945.
- Ramsay G, van de Ven W, Waugh R, Griffiths DW, Powell W: **Mapping quantitative trait loci in faba beans.** In *Improving production and utilisation of grain legumes*. Edited by AEP. Copenhagen, Denmark: 2nd European Conference on Grain Legumes; 1995:444–445.
- Satovic Z, Torres AM, Cubero JI: **Genetic mapping of new morphological, isozyme and RAPD markers in *Vicia faba* L. using trisomics.** *Theor Appl Genet* 1996, **93**:1130–1138.
- Vaz Patto MC, Torres AM, Koblikzova A, Macas J, Cubero JI: **Development of a genetic composite map of *Vicia faba* using F<sub>2</sub> populations derived from trisomic plants.** *Theor Appl Genet* 1999, **98**:736–743.
- Román B, Torres AM, Rubiales D, Cubero JI, Satovic Z: **Mapping of quantitative trait loci controlling broomrape (*Orobanche crenata* Forsk.) resistance in faba bean (*Vicia faba* L.).** *Genome* 2002, **45**:1057–1063.
- Román B, Satovic Z, Avila CM, Rubiales D, Moreno MT, Torres AM: **Locating genes associated with *Ascochyta fabae* resistance in *Vicia faba* L.** *Aust J Agric Res* 2003, **54**:85–90.
- Román B, Satovic Z, Pozarkova D, Macas J, Dolezel J, Cubero JI, Torres AM: **Development of a composite map in *Vicia faba* L. breeding applications and future prospects.** *Theor Appl Genet* 2004, **108**:1079–1088.
- Avila CM, Satovic Z, Sillero JC, Rubiales D, Moreno MT, Torres AM: **Isolate and organ-specific QTLs for ascochyta blight resistance in faba bean.** *Theor Appl Genet* 2004, **108**:1071–1078.
- Ellwood SR, Phan HTT, Jordan M, Torres AM, Avila CM, Cruz-Izquierdo S, Oliver RP: **Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*.** *BMC Genomics* 2008, **9**:380.
- Arbaoui M, Link W, Satovic Z, Torres AM: **Quantitative trait loci of frost tolerance and physiologically related traits in faba bean (*Vicia faba* L.).** *Euphytica* 2008, **164**:93–104.
- Cruz-Izquierdo S, Avila CM, Satovic Z, Palomino C, Gutierrez N, Ellwood SR, Phan HTT, Cubero JI, Torres AM: **Comparative genomics to bridge *Vicia faba* with model and closely-related legume species: stability of QTLs for flowering and yield-related traits.** *Theor Appl Genet* 2012, **125**:1767–1782.
- Ma Y, Bao S, Yang T, Hu J, Guan J, He Y, Wang X, Wan Y, Sun X, Jiang J, Gong C, Zong X: **Genetic linkage map of Chinese native variety faba bean (*Vicia faba* L.) based on simple sequence repeat markers.** *Plant Breeding* 2013. doi:10.1111/pbr.12074.
- Gutiérrez N, Palomino C, Satovic Z, Ruiz-Rodríguez MD, Vitale S, Gutiérrez MV, Rubiales D, Kharat M, Arni M, Emeran A, Cubero JI, Atienza SG, Torres AM, Avila CM: **QTLs for *Orobanche* spp. resistance in faba bean: Identification and validation across different environments.** *Molecular Breeding* 2013. in press.
- Cabrera A, Cubero JI, Martín A: **Genetic mapping using trisomics in *Vicia faba* L.** *FABIS Newslett* 1989, **23**:5–7.
- Macas J, Dolezel J, Lucretti S, Pich U, Meister A, Fuchs J, Schubert I: **Localization of seed genes on flow-sorted field bean chromosomes.** *Chromosome Res* 1993, **1**:107–115.
- Macas J, Weschke W, Baumlein H, Pich U, Houben A, Wobus U, Schubert I: **Localization of vicilin genes via polymerase chain reaction on microisolated field bean chromosomes.** *Plant J* 1993, **3**:883–886.
- Fuchs J, Pich U, Meister A, Schubert I: **Differentiation of field bean heterochromatin by *in situ* hybridization with a repeated *FokI* sequence.** *Chromosome Res* 1994, **2**:25–28.
- Pozarkova D, Koblikzova A, Román B, Torres AM, Lucretti S, Lysak M, Dolezel J, Macas J: **Development and characterization of microsatellite markers from chromosome 1-specific DNA libraries of *Vicia faba*.** *Biol Plantarum* 2002, **45**:337–345.
- Torres AM, Satovic Z, Cánovas J, Cobos S, Cubero JI: **Genetics and mapping of new isozyme loci in *Vicia faba* L. using trisomics.** *Theor Appl Genet* 1995, **91**:783–789.
- Torres AM, Vaz Patto MC, Satovic Z, Cubero JI: **New isozyme loci in Faba bean (*Vicia faba* L.): Genetic analysis and mapping using trisomics.** *J Hered* 1998, **89**:271–274.
- Avila CM, Sillero JC, Rubiales D, Moreno MT, Torres AM: **Identification of RAPD markers linked to *Uvf-1* gene conferring hypersensitive resistance against rust (*Uromyces viciae-fabae*) in *Vicia faba* L.** *Theor Appl Genet* 2003, **107**:353–358.
- Díaz-Ruiz R, Satovic Z, Avila CM, Alfaro CM, Gutierrez MV, Torres AM, Román B: **Confirmation of QTLs controlling *Ascochyta fabae* resistance in different generations of faba bean (*Vicia faba* L.).** *Crop Pasture Sci* 2009, **60**:353–361.
- Díaz-Ruiz R, Torres A, Gutierrez MV, Rubiales D, Cubero JI, Kharat M, Satovic Z, Román B: **Mapping of quantitative trait loci controlling *Orobanche foetida* Poir. resistance in faba bean (*Vicia faba* L.).** *African J Biotechnol* 2009, **8**:2718–2724.
- Díaz R, Torres AM, Satovic Z, Gutierrez MV, Cubero JI, Román B: **Validation of QTLs for *Orobanche crenata* resistance in faba bean (*Vicia faba* L.) across environments and generations.** *Theor Appl Genet* 2010, **120**:909–919.
- Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort S, Jeuken MJW, Kesseli RV, Lindhout P, Michelmore RW: **A high-density, integrated genetic linkage map of lettuce (*Lactuca spp.*).** *Theor Appl Genet* 2007, **115**:735–746.
- Varshney RK, Marcel TC, Ramsay L, Russell J, Röder MS, Stein N, Waugh R, Langridge P, Niks RE, Graner A: **A high density barley microsatellite consensus map with 775 SSR loci.** *Theor Appl Genet* 2007, **114**(6):1091–103.
- Shirasawa K, Bertoli DJ, Varshney RK, Moretzsohn MC, Leal-Bertoli SC, Thudi M, Pandey MK, Rami JF, Foncêka D, Gowda MV, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S: **Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes.** *DNA Res* 2013, **20**(2):173–184. doi:10.1093/dnares/dss042.
- Somers JD, Isaac P, Edwards K: **A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.).** *Theor Appl Genet* 2004, **109**:1105–1114.
- Davis GL, McMullen MD, Baysdorfer C, Musket T, Grant D, Staebell M, Xu G, Polacco M, Koster L, Mella-Hancock S, Houchins K, Chao S, Coe EH Jr: **A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map.** *Genetics* 1999, **152**:1137–1172.
- Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svenson JT, Wanamaker S, Bozdogan S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhoft A, Muehlbauer GJ, DeYoung J, Marshall DF, Madisetty K, Fenton RD, Candamine P, Graner A, Waugh R: **Development and implementation of high-throughput SNP genotyping in barley.** *BMC Genomics* 2009, **10**:582.
- Antonio BA, Inoue T, Kajiya H, Nagamura Y, Kurata N, Minobe Y, Yano M, Nakagahra M, Sasaki T: **Comparison of genetic distance and order of DNA markers in five populations of rice.** *Genome* 1996, **39**:946–956.
- Hwang T-Y, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia Z, Harada K, Kitamura K, Ishimoto M: **High-density integrated linkage map based on SSR markers in soybean.** *DNA Res* 2009, **16**:213–225. doi:10.1093/dnares/dsp010.
- Bordat A, Savoie V, Nicolas M, Salse J, Chauveau A, Bourgeois M, Potier J, Houtin H, Rond C, Murat F: **Translational genomics in legumes allowed placing in silico 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. G3-Genes genomes.** *Genetics* 2011, **13**(2):93–103.
- Millan T, Winter P, Jüngling R, Gil J, Rubio J, Cho S, Cobos MJ, Iruela M, Rajesh PN, Tekeoglu M, Kahl G, Muehlbauer FJ: **A consensus genetic map of chickpea (*Cicer arietinum* L.) based on 10 mapping populations.** *Euphytica* 2010, **175**:175–189.
- Galeano CH, Fernandez AC, Franco N, Cichy K, McClean P, Vanderleyden J, Blair MW: **Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean.** *PLoS One* 2011, **6**(12):e28135. doi:10.1371/journal.pone.0028135.

38. Bohra A, Saxena RK, Gnanesh BN, Kulbhusan Saxena K, Byregowda M, Rathore A, KaviKishor PB, Cook DR, Varshney RK: **An intra-specific consensus genetic map of pigeonpea (*Cajanus cajan* (L.) Mills) derived from six mapping populations.** *Theor Appl Genet* 2012, **125**(6):1325–1338.
39. Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ: **A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs.** *Proc Natl Acad Sci U S A* 2009, **106**:18159–18164.
40. Gautami B, Foncéka D, Pandey MK, Moreszohn MC, Sujay V, Qin H, Hong Y, Faye I, Chen X, Prakash AB, Shah TM, Gowda MVC, Nigam SN, Liang X, Hoesington DA, Guo B, Bertoli DJ, Ram JF, Varshney RK: **An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.).** *PLoS One* 2012, **7**(7):e41213. doi:10.1371/journal.pone.0041213. 2012.
41. Isobe S, Kölliker R, Hisano H, Sasamoto S, Wada T, Klimenko I, Okumura K, Tabata S: **Construction of a consensus linkage map for red clover (*Trifolium pratense* L.).** *BMC Plant Biol* 2009, **9**:57. doi:10.1186/1471-2229-9-57.
42. Wu Y, Close TJ, Lonardi S: **Accurate construction of consensus genetic maps via integer linear programming.** *IEEE/ACM Trans. Comput Biol Bioinform* 2011, **8**:381–394. doi:10.1109/TCBB.2010.35.
43. Rajaram E, Nepolean T, Senthilvel S, Varshney RK, Vadez V, Srivastava RK, Shah TM, Supriya A, Kumar S, Kumari BR, Bhanuprakash A, Narasu ML, Riera-Lizarazu O, Hash CH: **Pearl millet (*Pennisetum glaucum* (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs.** *BMC Genomics* 2013, **14**:159. doi:10.1186/1471-2164-14-15.
44. Wang J, Lydiate D, Parkin I, Falentin C, Delourme R, Carion P, King GJ: **Integration of linkage maps for the Amphidiploid *Brassica napus* and comparative mapping with *Arabidopsis* and *Brassica rapa*.** *BMC Genomics* 2011, **12**:101.
45. Khan MA, Han Y, Zhao YF, Troggio M, Korban SS: **A multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the apple genome.** *PLoS One* 2012, **7**(11):e47864. doi:10.1371/journal.pone.0047864.
46. Han Y, Zheng D, Virmolmangkang S, Khan MA, Beever JE, Korban SS: **Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome.** *J Exp Bot* 2011, **62**:5117–5130. doi:10.1093/jxb/err215.
47. Vezzulli S, Troggio M, Coppola G, Jermakow A, Cartwright D, Zharkikh A, Stefanini M, Grando MS, Viola R, Adam-Blondon AF, Thomas M, This P, Velasco R: **A reference integrated map for cultivated grapevine (*Vitis vinifera* L.) from three crosses, based on 283 SSR and 501 SNP-based markers.** *Theor Appl Genet* 2008, **117**:499–511.
48. Mace ES, Rami J-F, Bouchet S, Klein P, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR: **A consensus genetic map of sorghum that integrates multiple component maps and high-throughput diversity array technology (DArT) markers.** *BMC Plant Biol* 2009, **9**:13. doi:10.1186/1471-2229-9-13.
49. Spiller M, Linde M, Hibrand-Saint-Oyant L, Tsai C-J, Byrne DH, Smulders MJ, Foucher F, Debener T: **Towards a unified genetic map for diploid roses.** *Theor Appl Genet* 2011, **122**:489–500.
50. Marone D, Laido G, Gadaleta A, Colasuonno P, Ficco DBM, Giancaspro A, Giove S, Panio G, Russo MA, De Vita P, Cattivelli L, Papa R, Bianco A, Mastrangelo AM: **A high-density consensus map of A and B wheat genomes.** *Theor Appl Genet* 2012, **125**:1619–1638. doi:10.1007/s00122-012-1939-y.
51. Ruiz-Rodríguez MD, Avila CM, Torres AM, Fuchs J, Schubert I: **Anchoring of genetic linkage maps to the chromosome complement of *Vicia faba* L. Mole breed.** in press doi:10.1007/s11032-013-9969-y.
52. Kaló P, Choi HK, Ellis N, Kiss GB: **Syntenic and Comparative Genomics between Model and Cool Season Grain Legumes.** In *Genetics, Genomics and Breeding in Crop Plants: Cool Season Food Legumes*. Edited by de la Vega P, Torres AM, Cubero JI, Kole C. New Hampshire, Jersey, Plymouth: Science Pubs Inc; 2011:285–302.
53. Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR: **Estimating genome conservation between crop and model legume species.** *Proc Natl Acad Sci U S A* 2004, **101**:15289–15294.
54. Kaló P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB: **Comparative mapping between *Medicago sativa* and *Pisum sativum*.** *Mol Genet Genom* 2004, **272**:235–246.
55. Zhu H, Choi HK, Cook DR, Shoemaker RC: **Bridging model and crop legumes through comparative genomics.** *Plant Physiol* 2005, **137**:1189–1196.
56. Nelson MN, Phan HTT, Ellwood SR, Moolhuijzen PM, Hane J, Williams A, O'Lone CE, Fosu-Nyarko J, Scobie M, Cakir M, Jones MGK, Belgard M, Ksiarkiewicz M, Wolko B, Barker SJ, Oliver RP, Cowling WA: **The first gene-based map of *Lupinus angustifolius* L. -location of domestication genes and conserved synteny with *Medicago truncatula*.** *Theor Appl Genet* 2006, **113**:225–238.
57. Phan HTT, Ellwood SR, Ford R, Thomas S, Oliver R: **Differences in syntenic complexity between *Medicago truncatula* with *Lens culinaris* and *Lupinus albus*.** *Funct Plant Biol* 2006, **33**:775–782.
58. Nayak SN, Zhu HY, Varghese N: **Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome.** *Theor Appl Genet* 2010, **120**:1415–1441.
59. Marcel TC, Varshney RK, Barbieri M, Jafari H, Kock MJD, Graner A, Niks RE: **A high-density consensus map of barley to compare the distribution of QTLs for partial resistance to *Puccinia hordei* and of defence gene homologues.** *Theor Appl Genet* 2007, **114**:487–500. doi:10.1007/s00122-007-0611-1.
60. Cockram J, White J, Zuluaga DL, Smith D, Comadrán J, Macaulay M, Luo Z, Kearsey MJ, Werner P, Harrop D, Tapsell C, Liu H, Hedley PE, Stein R, Schulte D, Steuernagel B, Marshall DF, Thomas WT, Ramsay L, Mackay I, Balding DJ, Waugh R, O'Sullivan DM, The AGOUEB Consortium: **Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome.** *Proc Natl Acad Sci U S A* 2010, **107**(50):21611–21616. doi:10.1073/pnas.1010179107.
61. Sehgal D, Rajaram V, Vadez V, Hash CT, Yadav RS: **Integration of gene-based markers in pearl millet genetic map for identification of candidate genes underlying drought tolerance quantitative trait loci.** *BMC Plant Biol* 2012, **12**:9.
62. Swamy BP, Vikram P, Dixit S, Ahmed HU, Kumar A: **Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus.** *BMC Genomics* 2011, **12**:319. doi:10.1186/1471-2164-12-319.
63. Muchero W, Ehlers JD, Close TJ, Roberts PA: **Genic SNP markers and legume synteny reveal candidate genes underlying QTL for *Macrophomina phaseolina* resistance and maturity in cowpea [*Vigna unguiculata* (L) Walp.].** *BMC Genomics* 2011, **12**:8.
64. Pottorff M, Ehlers JD, Christian Fatokun, JD, Philip AR, Timothy JC: **Leaf morphology in Cowpea [*Vigna unguiculata* (L.) Walp.]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species.** *BMC Genomics* 2012, **13**:234.
65. Chen X, Hedley PE, Morris J, Liu H, Niks RE, Waugh R: **Combining genetical genomics and bulked segregant analysis-based differential expression: an approach to gene localization.** *Theor Appl Genet* 2011, **122**:1375–1383.
66. Madrid E, Palomino C, Plötner A, Horres R, Jungling R, Rotter B, Winter P, Kahl G, Torres AM: **DeepSuperSage analysis of the *Vicia faba* transcriptome in response to *Ascochyta fabae* infection.** *Phytopathologia Mediterranea* 2013, **52**(1):166–182.
67. Yang T, Bao S, Ford R, Jia T, Guan J, He Y, Sun X, Jiang J, Hao J, Zhang X, Zong X: **High-throughput novel microsatellite marker of faba bean via next generation sequencing.** *BMC Genomics* 2012, **13**:602.
68. Kaur S, Pemberton L, Cogan N, Savin K, Leonforte T, Paull J, Mateme M, Forster J: **Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers.** *BMC Genomics* 2012, **13**:104.
69. Cubero JI: **On the evolution of *Vicia faba* L.** *Theor Appl Genet* 1974, **45**:47–51.
70. Zeid M, Mitchell S, Link W, Carter M, Nawar A, Fulton T, Kresovich S: **Simple sequence repeats (SSRs) in faba bean: new loci from *Orobanchaceae* resistant cultivar 'Giza 402'.** *Plant Breeding* 2009, **128**:149–155.
71. Loidion K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne C, Lejeune-Henaut I, Burstin J: **Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.).** *Theor Appl Genet* 2005, **111**:1022–103.
72. Palomino C, Fernández-Romero MD, Rubio J, Torres A, Moreno MT, Millan T: **Integration of new CAPS and dCAPS-RGA markers into a composite chickpea genetic map and their association with disease resistance.** *Theor Appl Genet* 2009, **118**:671–682.
73. Timmerman-Vaughan GM, Frew TJ, Weerden NF: **Characterization and linkage mapping of R-Gene analogous DNA sequences in pea (*Pisum sativum* L.).** *Theor Appl Genet* 2000, **101**:241–247.

74. Prioul-Gervais S, Deniot G, Receveur EM, Frankewitz A, Fourmann M, Rameau C, Pilet-Nayel ML, Baranger A: **Candidate genes for quantitative resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.).** *Theor App Gen* 2007, **114**:971–984.
75. Michelmore RW, Paran I, Kesseli RV: **Identification of markers linked to disease-resistance genes by bulked segregate analysis: A rapid method to detect markers in specific genomic regions by using segregating populations.** *Proc Natl Acad Sci U S A* 1991, **88**:9828–9832.
76. Lander ES, Green P, Abramson J, Barlow A, Daly MJ, Lincoln DE, Newburg L: **MAPMAKER: an interactive computer program for constructing genetic linkage maps of experimental and natural populations.** *Genomics* 1987, **1**:174–181.
77. Wu Y, Bhat PR, Close TJ, Lonardi S: **Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph.** *PLoS Genet* 2008, **4**:e1000212.
78. Kosambi DD: **The estimation of map distance from recombination values.** *Ann Eugen* 1944, **12**:172–175.
79. Wang S, Basten CJ, Zeng ZB: **Windows QTL Cartographer 2.5.** Department of Statistics, North Carolina State University, Raleigh, NC. 2010. <http://statgen.uncs.edu/qtlcart/WQTLCart.htm>.
80. Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** *Genetics* 1994, **138**:963–971.
81. Yap IV, Schneider D, Kleinberg J, Matthews D, Cartinhour S, McCouch SR: **A graph-theoretic approach to comparing and integrating genetic, physical and sequence-based maps.** *Genetics* 2003, **165**:2235–2247.
82. Muñoz-Amatriain M, Moscou MJ, Bhat PR, Svensson JT, Bartoš J, Suchánková P, Šimková P, Endo TR, Fenton RD, Lonardi S, Castillo AM, Chao S, Cistué L, Cuesta-Marcos A, Forrest KL, Hayden MJ, Hayes PM, Horsley RD, Makoto K, Moody D, Sato D, Vallés MP, Wulff BBH, Muehlbauer GJ, Doležel J, Close TJ: **An improved consensus linkage map of barley based on flow sorted chromosomes and SNP markers.** *The Plant Genome* 2011, **4**:238–249.
83. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs.** *J Hered* 2002, **93**:7–8.

doi:10.1186/1471-2164-14-932

**Cite this article as:** Satovic et al: A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). *BMC Genomics* 2013 **14**:932.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)







## RESEARCH ARTICLE

## AutoFlow, a Versatile Workflow Engine Illustrated by Assembling an Optimised *de novo* Transcriptome for a Non-Model Species, such as Faba Bean (*Vicia faba*)

Pedro Seoane<sup>1</sup>, Sara Ocaña<sup>2</sup>, Rosario Carmona<sup>3</sup>, Rocío Bautista<sup>3</sup>, Eva Madrid<sup>4</sup>, Ana M. Torres<sup>2</sup> and Manuel Gonzalo Claros<sup>1,3,\*</sup>

<sup>1</sup>Departamento de Biología Molecular y Bioquímica, Universidad de Málaga, E-29071, Málaga, Spain; <sup>2</sup>Área de Mejora y Biotecnología, IFAPA Centro "Alameda del Obispo", Apdo 3092, E-14080 Córdoba, Spain; <sup>3</sup>Plataforma Andaluza de Bioinformática, Universidad de Málaga, E-29071 Málaga, Spain and <sup>4</sup>Institute for Sustainable Agriculture, CSIC, Apdo 4084, E-14080 Córdoba, Spain

## ARTICLE HISTORY

Received: July 20, 2015  
 Revised: September 30, 2015  
 Accepted: October 2, 2015

DOI:  
 10.2174/1574893616061602122351  
 17

**Abstract:** The use of workflows to automate routine tasks is an absolute requirement in many bioinformatics fields. Current workflow manager systems usually compromise between providing a user-friendly interface and constructing complex, scalable pipelines. We present AutoFlow, a Ruby-based workflow engine devoid of graphic interface and tool repository, that is useful in most computer systems and most workflow requirements in any scientific field. It accepts any local or remote command-line software and converts one workflow into a series of independent tasks. It has been supplied with control patterns that allow for iterative task capability, supporting static and dynamic variables for decision-making or chaining workflows, as well as debugging utilities that include graphs, file searching, functional consistency and timing. Two proof-of-concept cases are presented to illustrate AutoFlow capabilities, and a case-of-use illustrates the automated construction of the best transcriptome for a non-model species (*Vicia faba*) after analysis of several combinations of Illumina reads and Sanger sequences with different assemblers and different parameters in a complex and repetitive workflow where branching and convergent tasks were used and internal, automated decisions were taken. The workflow finally produced an optimal transcriptome of 118,188 transcripts, of which 38,004 were annotated, 10,516 coded for a complete protein, 3,314 were putatively new faba-specific transcripts, and 23,727 were considered the representative transcriptome of *V. faba*.



**Keywords:** Workflow, pipeline, automation, Ruby, bash, transcriptome, assembly, non-model species.

## INTRODUCTION

The high-throughput technologies produce a large amount of data that require the combination of multiple bioinformatic tools to process data, validate results and provide scientific insights. For example, *de novo* transcriptome sequencing generates large files of reads that must be pre-processed, assembled, verified, and then annotated using multiple software tools and data sources where varying parameters provide different results that require further reconciliation or selection [1, 2]. Since different input data will require different software or parameters to provide the best result (i.e., assembly), this can become a long and cumbersome task. The development of sophisticated workflows, where the output of a tool is used as input for other tool(s) can save researchers time and effort

[3]. Workflows provide advantages related to effectiveness (by automation of repetitive tasks), reproducibility (analyses can be repeated over time), traceability (by storing intermediate results) and reusability (of complete workflows or of intermediate results). Many workflow manager systems can build and execute workflows (reviewed by [3-5]), with Taverna [6] and Galaxy [7] being the most popular and accepted.

Installation of workflow manager systems usually includes a large library or repository of tools and ready-made scripts for data processing. Ontology-guided exploitation of local tools, remote web services and computer grids by means of graphical, user-oriented interfaces aims to simplify the design of customised workflows. In return, interface usability limits flexibility and sometimes precludes the design of workflows with new software not included in repositories, or the use of data from non-model organisms, or even the use of large datasets, as occurs in high-throughput sequencing [8]. In spite of graphical front-ends, workflow managers still require knowledge of tools, data formats and programming skills, especially in complex computational

\*Address correspondence to this author at the Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, E-29071, Málaga, Spain; Tel: +34 95 213 72 84; Fax: +34 95 213 20 41; E-mail: [claros@uma.es](mailto:claros@uma.es)

workflows [3]. Finally, when the workflow includes big data analyses, a significant bottleneck occurs in data transfer between nodes. One way to overcome this limitation is the use of cloud computing [8, 9] or the exploitation of local, high-performance computers.

Building *de novo* transcriptomes for non-model organisms with poorly characterised genomes serves as a proxy for the genome sequence [10, 11]. From the biological point of view, this is a resource for studying genome diversity and identifying candidate genes associated with agronomic traits of interest (resistance to biotic or abiotic stresses, production, yield, etc.) [11, 12]. From the bioinformatic point of view, it may involve the use of Terabytes of sequences and local programs instead of web services [2]. Moreover, the use of this huge amount of data may require a procedure optimisation that can handle different, repetitive executions of the same program with different parameters. A decision-making task is also required when different programs or parameters have to be evaluated in a workflow for selecting the best result from a plethora of intermediate data. Therefore, building *de novo* transcriptomes may be a good example of the benefits of automated workflows.

Faba bean (*Vicia faba*), rich in proteins, carbohydrates and other nutrients, is one of the most important legume crops. It occupies an important nutritional and cultural role as a staple pulse crop all around the Mediterranean basin and in the Near and Middle East. In spite of being a break crop in sustainable agriculture and a key source of biologically-fixed nitrogen in land-based systems, the overall popularity of faba bean cultivation has declined in the last years due to different factors such as its high susceptibility to diseases and pests [13]. Faba bean is a non-model crop that has one of the largest genomes among legumes (~13,000 Mb). This has complicated the development of saturated linkage maps and the identification or location of important genes, making it an understudied legume in which genomic resources are still limited [a few thousand ESTs in dbEST (release 130101; 1 January 2013) and only a very preliminary transcriptome is available [14]], lagging far behind that of model systems. Consequently, the construction of its *de novo* transcriptome could provide opportunities for developing fast and cost-effective gene-based markers such as EST-derived SSRs (single sequence repeats) and SNPs (single nucleotide polymorphisms) for breeding selection [34] that could help in recovering this important crop for human and animal foodstuffs.

We present AutoFlow, a versatile, bash-based, workflow engine developed in Ruby that can cope with command-line software in dynamic workflows allowing decision-making. The lack of graphical front-end and supporting repository or ontology mean that users need computing skills and knowledge of the software in order to overcome limitations inherent to other workflow manager systems. Two simple cases of use are presented, and the true utility and convenience of AutoFlow is illustrated by the assembling and annotation of a new transcriptome of *Vicia faba* combining Illumina reads and Sanger sequences.

## METHODS

### Architecture and Launching

AutoFlow has been implemented as a Ruby 1.9 gem in SUSE Linux Enterprise Server and accepts any kind of

command line software or scripts, whether local or external (provided that an internet connection is available), even for very specific resources (such as graphics processing unit [GPU] or field programmable gate array [FPGA]). It is self-contained and the only dependencies are GNUplot (<http://www.gnuplot.info>) and GraphViz [15]. AutoFlow is available for download at <https://rubygems.org/gems/autoflow> and can also be installed as any other Ruby gem with the command `gem install autoflow`. It supports installation in servers where tasks are sent to a queue system and/or in clients (Windows, OSX or Linux), provided that Ruby and dependencies are installed.

The workflow description has not been implemented in a specific language (such as SCUFL) or format (such as JSON), but in a simple text file called «flow template». This text file is intended to contain bash (Bourne-again shell) commands which are the *de facto* standard in unix shells since it supports filename wildcarding, piping, command substitution, variables, control structures and iteration. AutoFlow reads the flow template, parses the description of workflow to tasks, and launches every task to the queue system (examples in this study use the SLURM [Simple Linux Utility for Resource Management] queue system). As a result, every task in the flow template can be independently launched step-by-step. Flow templates are launched as `AutoFlow -w my_flow_template`, where `-w` indicates that `my_flow_template` is the flow template to be executed. Note that `-w` is the only mandatory parameter. Help for AutoFlow can be invoked as `AutoFlow -h`.

### Implementation of Flow Templates

The workflow described in a flow template consisting of a concatenation of bash-based tasks. It can contain comments as lines starting by “#” that are thus ignored. Each task must start with a unique word ending with a ‘)’ that is used as task identifier (TID) for reference purposes and graphical representations. The general structure of a task is as follows:

```
listing){
# dummy task to list files within a folder
# no initialisation required
?
ls folderName
}
```

where *listing* is considered the TID. The task content, based on bash commands, is flanked by braces ‘{}’. It starts with the initialisation of the environment variables and/or the use of small scripts (in this example it is empty since the `ls` command does not require any initialisation). A line starting with the reserved character ‘?’ indicates that next line is launching the main command of the task (i.e. `ls`). Therefore, the first word of the first line after the ‘?’ is used as main task command for the output storage and graphical purposes. Since the syntax is bash-based, any command line software of script in any language (MatLab, C, C++, Ruby, Perl, Python, R, Java, etc.) can be used. This declaration allows the inclusion of any software, irrespective of the number of command-line parameters or the type of data required.

**Task Iteration**

This control pattern serves to iterate the same task with different parameters, defined within brackets '[']' and separated by semicolons alongside the TID. It saves coding lines and is very useful for software benchmarking or optimisation, as, for example, when testing different *k*-mers in assembling, or mapping with different mismatching parameters. A minor modification of the code presented above allows listing of files within three folders at different relative directories called *myData*, *myTemplates* and *myGraphs*:

```
listing_[myData;./myTemplates;./myGraphs]){
# example listing files within three different folders
# no initialisation required
?
ls (*)
}
```

Iterative tasks in the flow template are transformed to independent, queued tasks by replacing the reserved character '\*' by each parameter (i.e., the name and path of each folder).

**Task Dependences**

Since inter-dependence is a key feature of workflows, AutoFlow follows the data flow model of execution [4], where every task is launched only when all its input data are available, that is, when all leading tasks creating the dependences have been finished. In some workflow managers, such as Taverna, this requires a hard user intervention, while AutoFlow simply defines task dependence by reference to the leading task causing the dependence and not by reference to files or any other flagging system. In the following code, the *show* task will be executed only after *listing* has finished:

```
listing){
# this is the leading task causing the dependence
?
ls folder > previousList
# the results of the listing is stored in file 'previousList'
}
show){
# this task is depending on the existence of file 'previousList'
# 'cat' does not require initialisation
?
cat listing)/previousList
}
```

**Combining Iteration and Dependences**

When there is a dependence to an iterative task, instead of writing a new task for each dependence of the iteration, AutoFlow supports one single task merging the dependence

to results of all iterated tasks surrounding the TID with the '!' character. Following with the last examples, after *listing* several folders, one wants to *show* the contents of each list.

```
# iterative task that generates the list of files contained in
three folders
listing_[myData;./myTemplates;./myGraphs]){
# no initialisation required
?
ls (*) > previousList
# a file for each directory listing has been created
}
show){
# this task will list the contents of every 'previousList'
file
# 'cat' does not require initialisation
?
cat !listing_!)/previousList
}
```

In cases where the *show* task could overflow the time limit of queue systems, there is another syntax for creating one new task for each dependence, as follows:

```
show_[myData;./myTemplates;./myGraphs]){
?
cat !listing_*/previousList
# every task 'show_X' will be launched when the task
'listing_X' is finished
}
```

**Handling Variables**

Variables are an important control construct for, for example, taking decisions on-the-fly, changing parameters without flow template edition, or using the same flow template in different computers or environments. Variables are alphanumeric strings that begin with '\$' or '@' characters. Those starting with '\$' are static variables (having a constant value throughout the flow template) defined at workflow launching or in the first lines of code. Static variables enable launching the same flow template with different data, making it reusable and shareable.

AutoFlow also handles dynamic variables beginning with '@' (i.e. *@sentence* in the following example) that can be modified during the workflow execution. Dynamic variables are set as environment variables in bash and can be transferred from one task to another, changing its value and helping in decision-making. Modification of dynamic variables requires calling an environment manager (script *env\_manager*, included on AutoFlow gem).

```
# definition of the dynamic variable at the beginning with
an empty value
@sentence = NULL
message){
```



**AutoFlow, a Versatile Workflow Engine**

and the end of each task; (2) a file containing the relations between TIDs and where have they been saved; and (3) a file where all dynamic variables are defined. This file is loaded by all tasks and is created only if a dynamic variable is set in the template. Therefore, AutoFlow can run on a desktop computer, but also on a computer cluster using a queue system (such as SLURM, as used in this project, although other queue systems can be implemented). The execution of parallel tasks and other technical details is handled by queue system and enactment engine, so the user does not have to worry about invoking software, distribution, or scheduling.

**Debugging and Graphics**

When a new workflow is created, it must be checked for errors and inconsistencies before the first launch. AutoFlow contains several command line options for this. For example, `--verbose` or `-v` do not launch the flow template but list all TIDs in white, main commands in red, the final command in orange, output path in blue, and dependences in green. For example, the listing presented in Fig. 1A prompted the error «//» in the input paths. The command line option `-graph` or `-g` generates a directed acyclic graph, where TIDs are the nodes, and the dependences and data flows correspond to edges. The representation can be semantic (TID-based, Fig. 1B) or structural (using the main task command, Fig. 1C). These plots depict the task flow and can serve to detect relation inconsistencies (Fig. 2A and 2B).

**Flow template 1: A Simple Proof-of-Concept**

This proof-of-concept describes a very simple pipeline for mapping reads on a transcriptome and the subsequent determination of differentially expressed genes. It also serves to illustrate the combination of a C-compiled binary, a Python script and an R script.

RNA-seq reads were obtained with the Illumina HiSeq2000 platform from two different conditions (reference and treatment) with three experimental replicates each one. Experimental replicates have 2,457,983, 2,866,872 and 988,173 reads, while references (control) have  $\approx$  3,000,000 reads each, which serve as inputs for the flow template listed below (downloadable at [http://www.scbi.uma.es/web/pedro/workflows/flow\\_template\\_1.txt](http://www.scbi.uma.es/web/pedro/workflows/flow_template_1.txt)).

```
# Relative path to input files
$input_path=../

# if path is finished by '/', the error of double // in Fig. 1A
is obtained

# First task: creating a mapping index of the reference
using Bowtie2
Index_Ref(){
# environment initialisation of Bowtie2 based on
SLURM
module load bowtie
?
# loading the reference transcriptome
```

**Current Bioinformatics, 2016, Vol. 11, No. 3 5**

```
bowtie-build -f $input_path/Ref/reference.fasta ref
}
# Second task A: Mapping reads from treatment
replicates to the reference
Mapping_Treat_[1;2;3](){
# environment initialisation of Bowtie2 based on
SLURM
module load bowtie
?
# loading the three sequence sets
bowtie -q Index_Ref/ref
$input_path/reads/illumina(*)fastq -p [cpu] -S
out(*)sam
}
# Second task B: Mapping reads from reference (control)
replicates
Mapping_Ctrl_[1;2;3](){
module load bowtie
?
bowtie -q Index_Ref/ref
$input_path/reads/control(*)fastq -p [cpu] -S ctrl(*)sam
}
# Third task: Counting the mapped reads
Count(){
# environment initialisation of sam2counts.py based on
SLURM
module load sam2counts
?
sam2counts.py -o Ctrl_vs_treat.txt
Mapping_Ctrl_1)/ctrl1.sam Mapping_Ctrl_2)/ctrl2.sam
Mapping_Ctrl_3)/ctrl3.sam
Mapping_Treat_1)/out_1.sam
Mapping_Treat_2)/out_2.sam
Mapping_Treat_3)/out_3.sam
}
# Last task: Detection of differentially expressed
transcripts with DESeq
differential_expression_analysis(){
..-soft/initializes/init_R
?
RNASeq_Pipeline.R Count)/Ctrl_vs_treat.txt
$input_path/target.txt $input_path/annot_file.txt $value
}
```

After creating a static variable (`$input_path`) with the relative path to input files, the workflow starts creating a Bowtie2 mapping index (`Index_Ref` TID) which is then used in the mapping steps of the reads from the treatment replicates (`Mapping_Treat_[1;2;3]` TIDs) and the reads from the control condition (`Mapping_Ctrl_[1;2;3]` TIDs). Note

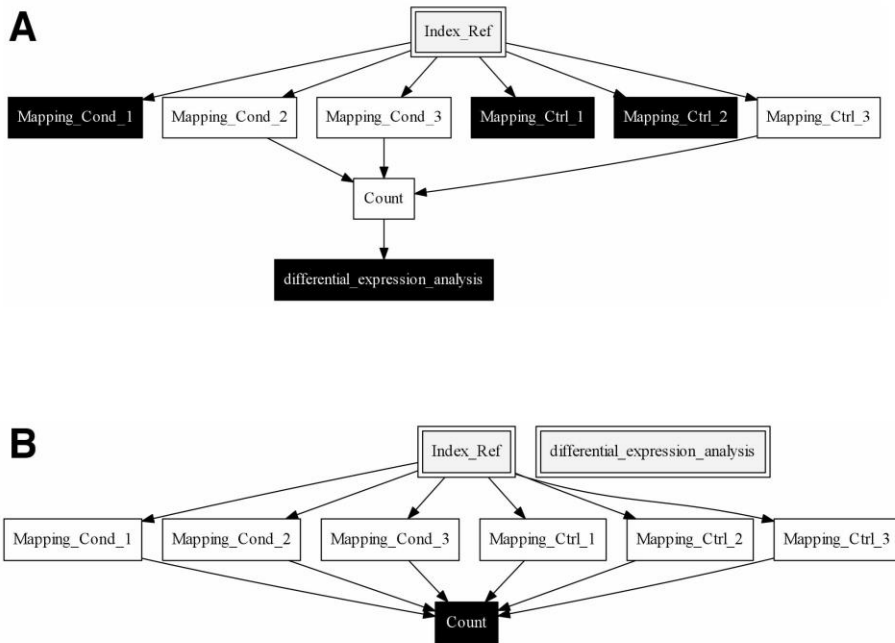
that these tasks are recursive for each library of reads. Since both tasks depend on *Index\_Ref*, they can be launched simultaneously (Fig. 1B and C) upon finishing *Index\_Ref*. The third task (*Count* TID) will only be launched once all mappings have been finished, since it is dependent on all of them. It uses a python script *sam2counts.py* (<https://github.com/vsbuffalo/sam2counts>) that constructs a table with the mapping results that will be used by an R script in the last task (*differential\_expression\_analysis* TID). The R script is a wrapper for the library DESeq [16] to find differentially expressed genes with the different *P*-values defined by the static variable *\$evalue*.

**Flow Template 2: Combining Data and Decision-Making**

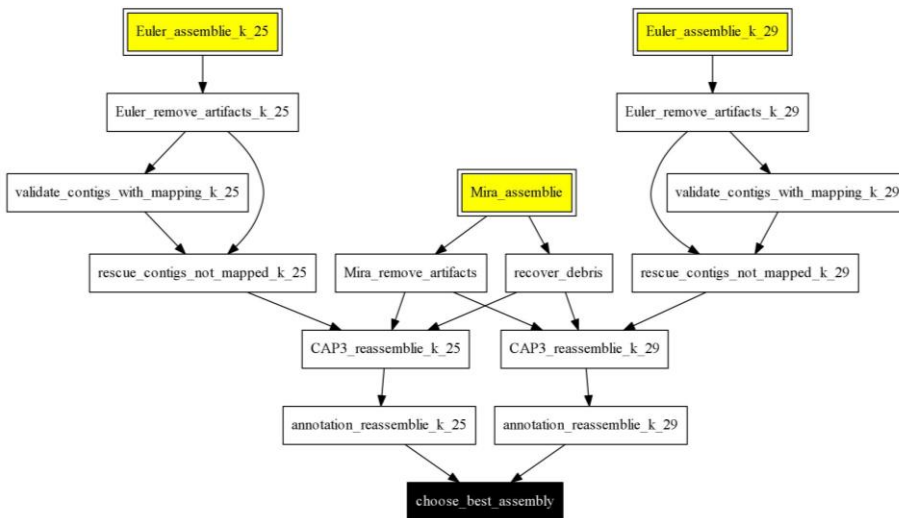
This proof-of-concept goes a step beyond the usability of AutoFlow. Its aim is to obtain the best assembly when using several sequence files as simultaneous input for two different assemblers using various parameters. The flow template ([http://www.scbi.uma.es/web/pedro/workflows/flow\\_template2.txt](http://www.scbi.uma.es/web/pedro/workflows/flow_template2.txt), whose semantic representation can be found in Fig. 3) was launched with two Roche/454 Titanium+ libraries containing cDNA sequencing from different tissues (A and

B) containing 372,750 and 429,909 useful reads, respectively. These reads were assembled separately and together using a de Bruijn algorithm for long reads (EULER-SR [17] in *Euler\_assembling* TIDs) with two different *k*-mers (25 and 29) and an overlap-layout-consensus algorithm (MIRA3; *MIRA3\_assembling* TID), which produced three different assemblies. EULER-SR contigs were validated (*validate\_contigs\_with\_mapping* TIDs) by mapping original reads with Bowtie2 [18]. Our software Full-LengtherNext (<http://www.scbi.uma.es/fulllengthernext>, Seoane et al. in preparation) was used to recover unmapped contigs from EULER-SR and debris reads from MIRA3 (*Euler\_remove\_artifacts* and *recover\_MIRA3\_debris* TIDs). Contigs and recovered reads were reconciled using CAP3 [19] with the MIRA3 contigs and each EULER-SR assembly (*CAP3\_reconciliation* TIDs).

The task *choose\_best\_assembly* (Fig. 3) analysed both reconciliations and determined the best one, assigning the corresponding value to the dynamic variable *@best\_assembly*, that allows the last task (*cp\_best\_assembly* TID) to copy only the best assembly in the final result folder. The best assembly (BA) index was empirically determined as the geometric mean of two ratios, one up-weighting the



**Fig. (2).** The semantic representation can help in debugging a flow template. A: The Flow template 1 contains a mistyping of the function dependences to *Mapping\_Cond1*, *Mapping\_Ctrl\_1* and *Mapping\_Ctrl\_2* in the *Count* TID as *Mapping\_Cnd1*, *Mapping\_Crl\_1* and *Mapping\_trl\_2*, making these TIDs appear as final tasks. B: The Flow template 1 with a syntax error, where the TID *differential\_expression\_analysis* contains a mistyped dependence to the *Count* TID, preventing its connection to the other pipeline tasks (last line starts with *RNASeq\_Pipeline.R count'* instead of *RNASeq\_Pipeline.R Count'*). The correct semantic representation of Flow template 1 is in Fig. 1B.



**Fig. (3).** Semantic representation, based on TIDs, of Flow template 2. Double line boxes correspond to three different, simultaneously starting tasks. The black box is the finishing TID. Right and left branches correspond to assembling with Euler using 25 and 29  $k$ -mers, respectively. Central branch corresponds to MIRA3 assembling. After the boxes corresponding to CAP3 reconciliation, the final decision on the best assembly is executed.

annotated transcripts (calculated as the product of the number of annotated transcripts and the number of full-length ORFs normalised by the number of transcripts longer than 500 bp and per 10,000 transcripts) and the other down-weighting the redundancy of annotations (calculated as the sum of unique full-length ORFs and unique orthologues normalised by the total number of orthologues and full-length ORFs).

### Flow Template 3: Searching for the Best *Vicia faba* Transcriptome

The *V. faba* transcriptome was mainly based on two faba bean genotypes, one resistant to *Ascochyta fabae* (Vf29H) and one susceptible (Vf136), in infected and non-infected stages [34]. Total RNA from entire leaves was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and was sent to GenXPro company, where they were normalised using double strand specific nuclease and PCR [20], and sequenced in a Illumina GAI platform. The 16,567,244 raw reads from Vf29H and the 16,455,916 from Vf136 have been deposited in the SRA (Short Read Archive) database under Experiment Accessions SRX690543 and SRX690544, respectively. These reads were complemented with 106,397 sequences of *V. faba* found in databases (3,435 ESTs and 13,935 nucleotide sequences from ENA, and 79,657 singletons and 6,370 contigs from [14]).

Figure in Supplemental file 1 shows the semantic representation of the workflow (available at [http://www.scbi.uma.es/web/pedro/workflows/flow\\_template3.txt](http://www.scbi.uma.es/web/pedro/workflows/flow_template3.txt)). In brief, Illumina reads were pre-processed using SeqTrimNext (<http://www.scbi.uma.es/seqtrimnext>; [21], *clean\_illumina* TID) to remove low quality, ambiguous and low complexity stretches, linkers, adaptors, vector fragments, organelle DNA, polyA/polyT tails, contaminated sequences and useful fragments shorter than 20 bp. The 28,627,470 resulting useful reads were assembled with two different *de novo* transcript assemblers (Oases 0.2.08 [22] and SOAPdenovoTrans [23]) using different  $k$ -mers (25, 31, 37, 45, 55 and 70) following the current approaches in which several algorithm and parameters are tested before obtaining the final assembling [24]. The resulting contigs were clustered with CD-HIT [25] in two ways, one in which all assemblies were clustered together (*clustering\_Oases\_kmers\_all* and *clustering\_SOAP\_kmers\_all* TIDs), and another where only those obtained for lower and higher  $k$ -mers (25 and 70, respectively) were clustered (*clustering\_Oases\_kmers\_25\_70* and *clustering\_SOAP\_kmers\_25\_70* TIDs). As a result, there were 8 different assemblies to evaluate per assembler, 6 from different  $k$ -mers and 2 from the clusterings. As in Flow template 2, the evaluation of the 16 assemblies was based on the BA ratio after a Full-LengtherNext analysis (*Assessing\_Oases\_assembly* and *Assessing\_SOAP\_assembly* TIDs in Supplemental file 1). The best assembly (*choose\_best\_illumina\_assembly* TID) was reconciled with the 89,408 useful, cleansed sequences from databases using



CAP3 [19] in *reduction\_with\_DBseqs* TID. An alternative reconciliation (simplification) of the best assembly without database sequences was also performed (Supplemental file 1, *sequence\_reduction* TID). Both reconciliations were compared also with the BA index and the best assembly (*best\_assembly\_and\_annotation* TID) was submitted for annotation with Sma3s [26] and then to an external KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway retrieval (*check\_kegg* TID, Supplemental file 1) to demonstrate that AutoFlow not only is able to pipeline local programs, but also external ones, enabling its use with web services. The resulting annotated transcriptome can be downloaded from [http://www.scbi.uma.es/web/pedro/vicia/Vicia\\_faba\\_transcriptome\\_1.0.tar.gz](http://www.scbi.uma.es/web/pedro/vicia/Vicia_faba_transcriptome_1.0.tar.gz).

**RESULTS AND DISCUSSION**

**Flow Template 1 Confirms that Increasing P-Value Stringency Decreases the Number of Differentially Expressed Transcripts**

This simple flow template (Fig. 1) was used to illustrate the main capabilities of AutoFlow, iteration, dependences and variables, by means of a basic RNA-seq analysis using different *P*-values. It was launched as *Autoflow -w flow\_template1.txt -c 4 -V '\$value=0.05'*, where *-c 4* indicates that four cores are to be assigned to each task for execution, (see the *[cpu]* variable within the Flow template 1) and the *-V* option sets internal static variable values (i.e. set *\$value* to 0.05). The *\$value* variable is the cut-off *P*-value used to determine the statistical significance of differential expression. This flow template was launched as many times as *P*-values to be tested, setting, for example, the static variable *\$value* to 0.01, 0.05, and 0.1. The results were 2,780, 2,937 and 2,995 differentially expressed genes, respectively. As expected, the lower the *P*-value, the greater the number of genes. Of course, the most reasonable way to do this is to take advantage of AutoFlow iteration and launch the flow template only once by changing the last task to:

```
differential_expression_analysis{0.01;0.05;0.1}{
  .-soft/initializes/init_R
  ?
  RNASeq_Pipeline.R          Count/Ctrl_vs_cond.txt
  $input_path/target.txt $input_path/annot_file.txt (*)
}
```

More sophisticated and iterative analyses with this flow template could help in, for example, the determination of the best *P*-value threshold without the need for human intervention.

**Flow Template 2 Facilitates the Selection of the Best Assembly of 454 Reads**

Flow template 2 was launched three times with different input 454 reads, two corresponding to separated tissues A and B and the third to the whole dataset, as *Autoflow -w flow\_template2.txt -c 100*, where *-c* now selects 100 cores for each task. Since computer nodes have less cores, instead of substituting the *[cpu]* tag in the template, the *[lcpu]* is used to

select cores belonging to different computer nodes when the software is not able *per se* to do that. The path to the input file was changed for execution, and is declared in the static variables *\$input\_mira\_454* and *\$input\_euler*. Also declared is the dynamic variable *@best\_assembly* initialised to *null*, where the best assembly path will be stored. Assembly evaluation is based on the BA parameter described above.

The three executions of this flow template provided 9 transient assemblies that were finally resolved into 3 optimal, reconciled assemblies, one for each dataset (last row of Table 1). Since AutoFlow has not implemented the recursive execution of a template, the user has to decide which is the best assembly from the results of Table 1. The BA parameter indicated that the best assembly uses the *k*-mer 29 assembling with all reads (Table 1, last column), saving computer time and load in a subsequent annotation step. In fact, its annotation with Sma3s [26] revealed putative functions for 34,971 transcripts (results not shown). In conclusion, Flow template 2 has illustrated its capability for exchanging results between tasks and its capability of taking decisions. However, in cases where there are many more assemblies to inspect, manual decision may be a cumbersome task. A new template based on *choose\_best\_assembly* task using as input all previously selected assemblies by Flow template 2 can be created and chained.

**Table 1. Key parameters of the CAP3 reconciled assemblies using Flow template 2. All cases include the MIRA3 assembly, and the *k*-mer refers to different EULER-SR assembling.**

	Tissue A		Tissue B		Tissues A + B	
	<i>k</i> -mer 25	<i>k</i> -mer 29	<i>k</i> -mer 25	<i>k</i> -mer 29	<i>k</i> -mer 25	<i>k</i> -mer 29
All transcripts	37,283	37,235	14,178	14,191	44,837	44,858
Transcripts > 500 bp	28,415	28,498	10,541	10,589	34,124	34,213
Transcripts with orthologue	24,908	24,925	9,375	9,391	29,020	29,077
Unique orthologues	13,775	13,786	5,962	5,955	15,328	15,307
Transcripts coding for a full-length ORF	5,654	5,732	2,114	2,127	7,292	7,347
Unique full-length ORFs	3,882	3,942	1,512	1,504	4,736	4,765
Annotation ratio	0.4956	0.5013	0.1880	0.1886	0.6201	0.6244
Redundancy ratio	0.5777	0.5782	0.6505	0.6476	0.5525	0.5777
BA index	0.5351	0.5384	0.3497	0.3495	0.5854	0.5866
<i>@best_assembly</i>		+	+			+

**Follow-up of Time Spent by Each Task**

The time spent by each task can be monitored with the command *flow\_time path\_to\_log\_file\_within\_exec\_folder* after the flow template execution is finished. The time analysis is provided as a histogram (Fig. 4) where the first bar is the time used by the complete workflow. The time saved by means of simultaneous execution of tasks is the difference between

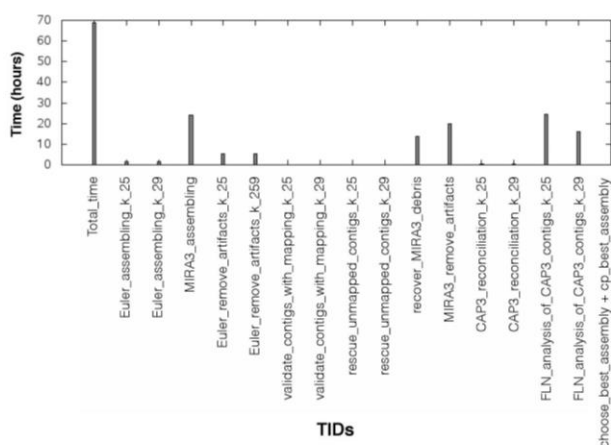


Fig. (4). Histogram illustrating the time spent by each task in the complete process of Flow template 2.

Table 2. Summary of attributes of significant assemblies obtained during the execution of Flow template 3 based on a Full-LengtherNext analysis.

Feature	Best Illumina read assembly	CAP3 reconciliation without DB sequences	CAP3 reconciliation including DB sequences
Number of transcripts	122,661	97,309	118,188
Mean length (nt)	-	-	-
Longest transcript (nt)	6,405	8,718	8,718
ncRNAs	3,242	2,733	2,789
Transcripts with annotation	39,473	33,451	38,004
Unique IDs	20,032	18,983	20,413
Transcripts including a complete ORF	11,637	9,912	10,516
Unique complete ORFs	6,669	6,410	6,787
Transcripts without orthologue	79,946	61,125	77,395
> 500 bp, likely coding for a protein	3,577	3,046	3,314
BA index	0.851	0.826	0.864
Representative transcripts	-	22,029	23,727

the total time and the sum of the times of every TID. This can help to refine tasks within the workflow; for example, when the workflow has been waiting for idle nodes due to overloading of computers, total time is greater than the sum of times (not shown). Fig. 4 reveals that MIRA3 assembling (that can use only 1 core, MIRA3\_assembling TID) and analyses involving Full-LengtherNext (recover\_MIRA3\_debris, MIRA3\_remove\_artifacts, FLN\_analysis\_of\_CAP3\_contigs\_k\_25 and FLN\_analysis\_of\_CAP3\_contigs\_k\_29) even if they are using 100 cores) are the most time-consuming tasks.

#### Flow Template 3 was Used to Construct the Optimal Vicia Faba Transcriptome with Current Data

The previous flow templates were presented with the aim of illustrating AutoFlow capabilities in simple proofs-of-concept, but do not provide any relevant biological information. The present case-of-use exploits AutoFlow capabilities in a complex flow template that provides the most complete and annotated transcriptome of *V. faba* to date. From the original 33,023,160 Illumina raw reads and the 106,397 sequences from databases, 28,627,470 (86.7%) and 89,408 (84.0%) were considered useful sequences, respectively. Of the 16 different

assemblies tested with different  $k$ -mers and assemblers (Supplemental file 1), the one with best BA index corresponded to the clustering of all  $k$ -mers from Oases (first column in Table 2). CAP3-based simplification [27] of this assembly alone (second column of Table 2) indicates that this step was necessary to reduce the overestimation of the original assembly, since the BA index decreased from 0.851 to 0.826. However, when reconciliation included the sequences from databases in the *reduction\_with\_DBseqs* task, the BA index increased from 0.826 to 0.864 (Table 2), indicating that these sequences served to improve the final faba bean assembly, and that the best transcriptome was obtained by combining all available sequences.

In the task *best\_assembly\_and\_annotation*, the best assembly is sent to the time-consuming job of Sma3s annotation. As a result, 11,845 (71.2%) transcripts received at least one Gene Ontology (GO) term, metabolic process (19.9%) and translation (12.4%) being the most represented functions. A total of 2,789 putative ncRNAs (Table 2) were also identified. Moreover, the 20,413 different types of transcripts found is close but below the ~27,200 protein-coding genes described for *A. thaliana* [28], suggesting that more tissue samples and development conditions need to be incorporated in future versions of this transcriptome. The high proportion of complete ORFs obtained with this approach (6,787 out of 20,413) supports the accuracy of the transcriptome. Regarding future molecular research in *V. faba*, a set of 3,314 faba-specific transcripts were identified based on the following properties: (1) they do not have any orthologous sequence in databases, (2) they are long enough (> 500 bp) to code for a complete or partial protein, and (3) they have a testcode index > 0.94 indicating that they are very likely coding for a protein [29]. From the complete collection of transcripts, the *FLNanalysis\_with\_DBseqs* task includes a Full-LengtherNext analysis that extracted a subset of 23,727 representative transcripts for further expression studies (last column, last row, in Table 2). These correspond to the longest transcript containing a full-length ORF with unique, different orthologous ID, and the longest, non-redundant faba-specific transcripts. The level of similarity between this transcriptome and the expressed genes of *Arabidopsis thaliana*, *Medicago truncatula* and *Cicer arietinum* is presented elsewhere [34], but preliminary results indicate that it ranges from 68.7% for *Arabidopsis* to 83.4% for *Cicer*.

## CONCLUSION

This paper introduces AutoFlow, a versatile and flexible standalone workflow engine for researchers with bash programming skills that combines the advantages of workflow manager systems [3] and allows the re-use of workflows between computers. Instead of facing cloud computing like other workflow engines [8, 9], it opts for high-performance computing architectures to cope with the increasing data size and complexity of current biological analyses. Our previous experience in developing [30] and using [6, 7, 31] ontology-based workflow manager systems demonstrated that most bioinformatic tools produce text files that abolished most ontology benefits. For this reason, AutoFlow is not based on ontologies or repositories but on bash-based scripts.

AutoFlow can (i) manage iterative tasks; (ii) handle dynamic workflows that decide which data can continue the workflow and which cannot (Supplemental file 1); (iii) pipeline independent flow templates based on static variables and dependences; (iv) monitor task timing (Fig. 4); (v) draw graphs for semantic and structural monitoring (Fig. 1); and (vi) debug the flow template based on previous graphs (Fig. 2) and parameter consistency, functional consistency and reachability of the workflow (Fig. 1A). Additionally, users do not need to worry about the technical details of invoking software components or using distributed computing resources since the queue system takes charge of this. Since AutoFlow is not limited by any repository or ontology of predefined tools [6, 7, 31] and does not rely on visual front-ends, it allows the incorporation of any desired local or remote software or scripts (as, e.g., Triana [32] or Kepler [33]) without regard to localisation (in contrast to Galaxy [7]). This integration capability was illustrated with results combining C compiled software (e.g., MIRA3, Oases, SOAPdenovoTrans, Bowtie2) with R scripts (DESeq), Ruby scripts (e.g., Full-LengtherNext, SeqTrimNext), and external database retrieval (KEGG request on Flow template 3). Flow templates can be shared between users provided that they have the same software. Although examples in this study were related to sequences and gene expression, AutoFlow can manage any command-line software from any scientific field (e.g., drug discovery, genomics, metagenomics, proteomics, etc.), and allows the use of data from non-model organisms, such as *V. faba*. A complex flow template (Supplemental file 1) provided the most complete faba bean transcriptome available to date, containing 38,004 annotated transcripts from which 10,516 coded for a complete protein and 3,314 were putatively new faba-specific transcripts. This transcriptome is currently being used for differential expression analyses and cDNA cloning [34].

## LIST OF ABBREVIATIONS

BA	= Best assembly index
EST	= Expression sequence tag
FPGA	= Field programmable gate array
GO	= Gene Ontology
GPU	= Graphics processing unit
KEGG	= Kyoto Encyclopedia of Genes and Genomes
ORF	= Open reading frame
SLURM	= Simple Linux Utility for Resource Management
SNP	= Single nucleotide polymorphism
SRA	= Short Read Archive
SSR	= Single sequence repeat
TID	= Task identifier

## CONFLICT OF INTEREST

None declared. This study has been funded by Plan Andaluz de Investigación, Desarrollo e Innovación (P10-CVI-6075), by European Union FP7 (LEGATO #613551)

**AutoFlow, a Versatile Workflow Engine**

and co-funded by the European Regional Development Fund and Ministerio de Economía y Competitividad (INNACTO IPT-2011-1259-010000).

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge Darío Guerrero for helpful discussions and programming support and GeneXPro for their efficient service of cDNA ultra-sequencing. The authors also acknowledge and thank the computer resources and the technical support provided by the Plataforma Andaluza de Bioinformática de the University of Málaga.

**SUPPLEMENTAL FILES**

Semantic (A) and structural (B) representation of Flow template 3. Double line boxes correspond to the two different starting points (Illumina reads [*clean\_illumina* TID] and Sanger sequences from databases [*trim\_DBseqs* TID]). The black box is the finishing TID.

**REFERENCES**

- [1] Canales J, Bautista R, Label P, *et al.* *De novo* assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. *Plant Biotechnol J* 2014; 12: 286-99.
- [2] Benzekri H, Armesto P, Cousin X, *et al.* *De novo* assembly, characterization and functional annotation of Senegalese sole (*Solea senegalensis*) and common sole (*Solea solea*) transcriptomes: integration in a database and design of a microarray. *BMC Genomics* 2014; 15: 952.
- [3] Romano P. Automation of in-silico data analysis processes through workflow management systems. *Brief Bioinform* 2008; 9: 57-68.
- [4] Curcin V, Ghanem M, editors. Scientific workflow systems - can one size fit all? Proceedings of the 2008 IEEE, CIBEC'08; 2008. London: Imperial College London; 2008.
- [5] Warr WA. Scientific workflow systems: Pipeline Pilot and KNIME. *J Comput Aided Mol Des* 2012; 26: 801-4.
- [6] Hull D, Wolstencroft K, Stevens R, *et al.* Taverna: a tool for building and running workflows of services. *Nucleic Acids Res* 2006; 34: W729-32.
- [7] Goecks J, Nekrutenko A, Taylor J, Galaxy T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010; 11: R86.
- [8] Liu B, Madduri RK, Sotomayor B, *et al.* Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses. *J Biomed Inform* 2014; 49: 119-33.
- [9] Abouelhoda M, Issa SA, Ghanem M. Tavaxy: integrating Taverna and Galaxy workflows with cloud computing support. *BMC Bioinformatics* 2012; 13: 77.
- [10] Ward JA, Ponnala L, Weber CA. Strategies for transcriptome analysis in nonmodel plants. *Am J Bot* 2012; 99: 267-76.
- [11] Hirsch CD, Evans J, Buell CR, Hirsch CN. Reduced representation approaches to interrogate genome diversity in large repetitive plant genomes. *Brief Funct Genomics* 2014; 13: 257-67.
- [12] Li CY, Deng GM, Yang J, *et al.* Transcriptome profiling of resistant and susceptible Cavendish banana roots following inoculation with *Fusarium oxysporum* f. sp. *cubense* tropical race 4. *BMC Genomics* 2012; 13: 374.
- [13] Akibode CS, Maredia M. Global and Regional Trends in Production, Trade and Consumption of Food Legume Crops. Rome: FAO, 2011.
- [14] Kaur S, Pembleton LW, Cogan NO, *et al.* Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 2012; 13: 104.
- [15] Zhao JH. Pedigree-drawing with R and graphviz. *Bioinformatics* 2006; 22: 1013-4.
- [16] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; 11: R106.
- [17] Pevzner PA, Tang H, Waterman MS. An Eulerian path approach to DNA fragment assembly. *Proc Natl Acad Sci U S A* 2001; 98: 9748-53.
- [18] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; 9: 357-9.
- [19] Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res* 1999; 9: 868-77.
- [20] Zhulidov PA, Bogdanova EA, Shcheglov AS, *et al.* Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res* 2004; 32: e37.
- [21] Falgueras J, Lara AJ, Fernandez-Pozo N, Canton FR, Perez-Trabado G, Claros MG. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* 2010; 11: 38.
- [22] Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 2012; 28: 1086-92.
- [23] Xie Y, Wu G, Tang J, *et al.* SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 2014; 30: 1660-6.
- [24] Melicher D, Torson AS, Dworkin I, Bowsher JH. A pipeline for the *de novo* assembly of the *Themira biloba* (*Sepsidae: Diptera*) transcriptome using a multiple k-mer length approach. *BMC Genomics* 2014; 15: 188.
- [25] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012; 28: 3150-2.
- [26] Muñoz-Mérida A, Viguera E, Claros MG, Trelles O, Perez-Pulido AJ, Sma3s: A Three-Step Modular Annotator for Large Sequence Datasets. *DNA Res* 2014; 21: 341-53.
- [27] Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. An optimized protocol for analysis of EST sequences. *Nucleic Acids Res* 2000; 28: 3657-65.
- [28] Lamesch P, Berardini TZ, Li D, *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res* 2012; 40: D1202-10.
- [29] Fickett JW. Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res* 1982; 10: 5303-18.
- [30] Ramirez S, Muñoz-Mérida A, Karlsson J, *et al.* MOWServ: a web client for integration of bioinformatics resources. *Nucleic Acids Res* 2010; 38: W671-6.
- [31] Kallio MA, Tuimala JT, Hupponen T, *et al.* Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics* 2011; 12: 507.
- [32] Taylor I, Shields M, Wang I, Harrison A. The Triana workflow environment: Architecture and Applications. *Workflows for e-Science*. Berlin: Springer; 2007; pp. 320-39.
- [33] Ludäscher B, Altintas I, Berkley C, *et al.* Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience* 2006; 13: 1039-65.
- [34] Seoane P, Bautista R, Palomino C, Claros MG, Torres AM, Madrid E. Large-Scale Transcriptome Analysis in Faba Bean (*Vicia faba* L.) under *Ascochyta fabae* Infection. *PLoS One*. 2015;10: e0135143.



**Genomics research for faba bean (*Vicia faba* L.) breeding applications****I-GRTM02**Avila C<sup>1</sup>, Satovic Z<sup>2</sup>, Madrid E<sup>3</sup>, Ocaña S<sup>1</sup>, Ruiz MD<sup>1</sup>, Cubero JI<sup>4</sup>, Torres AM<sup>1,\*</sup><sup>1</sup>IFAPA-Alameda del Obispo Córdoba, Spain<sup>2</sup>Faculty of Agriculture Svetosimunska, Zagreb, Croatia<sup>3</sup>University of Córdoba, Córdoba, Spain<sup>4</sup>CSIC-Instituto de Agricultura Sostenible Córdoba, Spain

\*E-mail: anam.torres.romero@juntadeandalucia.es

**Abstract**

Domesticated in ancient times, faba bean (*Vicia faba* L.) is now a major food and feed legume, particularly in Eurasia and North Africa. Despite being one of the cytogenetically best characterized plant species, subsequent progress in faba bean genomics has lagged behind other legume crops. Comprehensive genomic and post genomic resources are being developed but the very large and repetitive faba bean genome has so far restricted its sequencing and hampered genetic research and breeding applications. There are over 10 mapping populations established to study the inheritance of agronomic traits in different locations. Different molecular marker sets have been used to construct saturated linkage maps and identify genes/QTLs controlling resistance to crenate broomrape and several fungal major diseases. Recently, markers linked to growth habit and nutritional value of faba bean seeds were also identified. Despite these significant efforts, complete translation of marker discovery to faba bean breeding is still to be achieved. Recent efforts offer great promise for the development of highly accurate selective breeding tools. Thus, transcriptomic analysis (SuperSAGE and cDNA libraries) are enriching the scarce faba bean ESTs libraries providing additional resources for refining the maps with functional markers. Moreover, a combination of candidate gene and colinearity studies is being pursued to identify genes underlying agronomically important traits including seed size, flowering time, disease resistances and plant architecture. Finally, a faba bean functional consensus map is being constructed integrating all the genes and QTLs previously published. These new tools may help researchers to find candidate genes to accelerated improvement of future faba bean cultivars.

# Strategies to identify candidate genes controlling *Ascochyta* blight in different legume crops

**Ocaña-Moral S<sup>1</sup>, Madrid E<sup>2</sup>, Millán T<sup>3</sup>, García P<sup>4</sup>, Pérez de la Vega M<sup>4</sup>, Torres AM<sup>1</sup>**

<sup>1</sup> Área Mejora y Biotecnología, IFAPA-Centro Alameda del Obispo, Apdo. 3092, 14080, Córdoba, Spain  
<sup>2</sup> Instituto de Agricultura Sostenible, CSIC, Mejora Genética Vegetal, Apdo. 4084, 14080, Córdoba, Spain  
<sup>3</sup> Universidad de Córdoba, Dpto. Genética, Campus de Rabanales Edificio C5 2a planta, 14071 Córdoba, Spain  
<sup>4</sup> Universidad de León, Area de Genética, Facultad de Ciencias Biológicas y Ambientales, 24071 León, Spain


---

### INTRODUCTION

Until now complete resistance to *Ascochyta* blight has not been reported neither in chickpea nor in faba bean although several relevant QTLs have been described.

**CHICKPEA.** The most significant QTLs are in LGs 2 (QTL<sub>AR3</sub>) and 4 (QTL<sub>AR1</sub> and QTL<sub>AR2</sub>). Attempts to saturate QTL<sub>AR1</sub> have been performed by developing BAC clones or by identifying candidate genes from the model *Medicago truncatula*.

**FABA BEAN.** Two QTLs (*Af1* and *Af2*), assigned to chromosomes III and II were validated across environments and generations. In a second cross, six QTLs (*Af3* to *Af8*), were detected. *Af1* and *Af3* were both assigned to chromosome III and might correspond to the same genomic region.



---

### RESULTS and DISCUSSION

- The EST based markers assayed so far mapped in faba bean chromosomes, I, II, III and VI.
- Chromosomes II and III, bearing the most stable *Ascochyta* resistance QTLs reported to date, included 24 ESTs markers (12 each), that saturated the target QTL regions.
- Chromosomes I and VI included 7 and 4 new markers, respectively that might identify new genomic areas involved in the resistance.
- The inclusion of PS-MnSOD in chr. II revealed a new QTL (*Af3*), controlling DDS.

Our aim is to progress from basic mapping of QTLs to fine mapping or gene identification, thus allowing development of efficient markers for MAS in breeding.

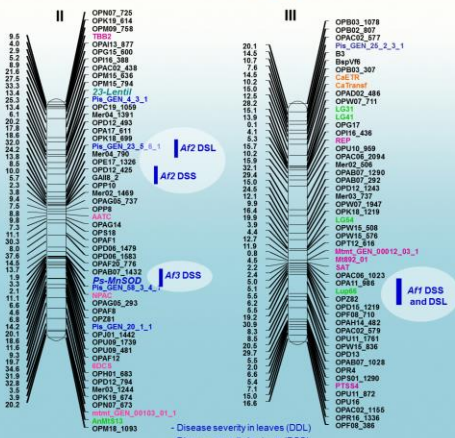
### MATERIALS and METHODS

**Comparative and functional genomic approaches** are being used to saturate regions bearing *Ascochyta* resistance QTLs in chickpea and faba bean, to establish possible homologies between them and to identify orthologous genes/QTLs underlying blight resistance.

- **New gene base markers** are being assayed from *M. truncatula* (51), pea (25) together with a number of **ITAP markers** (Eilwood et al 2008).
- **Candidate genes revealed by SuperSAGE** analysis and associated with the interactions **ascochyta-faba bean (36 tags)** and **ascochyta-lentil (18 Tags)** are being tested for polymorphism and mapping to identify potential targets for molecular breeding.
- Both, the *M. truncatula* and the recent *Cicer arietinum* whole genome sequences, are being exploited as well, to perform wide genome comparative mapping studies with the less-studied faba bean. New **chickpea markers** mapped in faba bean chromosomes are noted in **orange**.

---

**ACKNOWLEDGEMENTS.** This research is financed by the Spanish MICINN (project RTA2010-00059), co-financed with FEDER. S Ocaña-Moral acknowledges the INIA fellowship at IFAPA, belonging to the Campus de Excelencia Internacional Agroalimentario (ceiA3)



# Using **AutoFlow**, a framework to resolve workflows, to build a *de novo* plant transcriptome



Pedro Seoane<sup>1,2</sup>, Sara Ocaña<sup>1</sup>, Rosario Carmona<sup>2</sup>, Rocío Bautista<sup>2</sup>, Dario Guerrero-Fernández<sup>2</sup>, Eva Madrid<sup>1</sup>, Ana M. Torres<sup>3</sup>, and M. Gonzalo Claros<sup>1,2</sup>

<sup>1</sup>Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, Málaga, Spain  
<sup>2</sup>Plataforma Andaluza de Bioinformática-Centro de Supercomputación y Bioinformática, Universidad de Málaga, Málaga, Spain  
<sup>3</sup>Área de Mejora y Biotecnología, IFAPA Centro "Alameda del Obispo", and Institute for Sustainable Agriculture, CSIC, Córdoba, Spain



## Introduction

Building a transcriptome of an organism involves the use of several programs and doing many attempts to obtain the best result. This is usually affronted by designed workflows. There are well known workflow builders such as Galaxy<sup>1</sup> and Taverna<sup>2</sup> aiming to simplify the interface disregarding a complex use. This may lead to a lack of scalability or the impossibility of using some software. Here is presented AutoFlow, a workflow builder that supports most computer systems through a queue software and can manage very specific resources (such as GPU or FPGA). It has been developed in Ruby and accepts any kind of command line software. The user gives to AutoFlow a template with the desired workflow and it automatically launches all tasks to the queue system. The utility and convenience of AutoFlow is illustrated with a complex workflow for the assembling and annotation of a transcriptome of *Vicia faba* combining Illumina libraries and EST sequences.

## Executing a workflow

AutoFlow workflows must be written in a template. The template is then launched on the computer using the command line terminal.

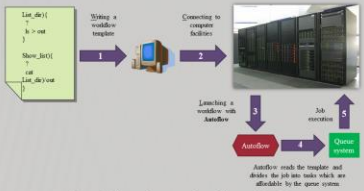


Figure 1. Schematic representation of the use of AutoFlow

## Inside AutoFlow

AutoFlow automatically traces the task dependencies, manages the local storage of the output of each job (and its firming tasks), and launches all tasks to the queue system.

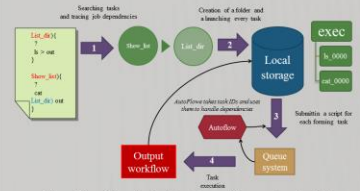


Figure 2. Description of each of stages in a workflow execution with AutoFlow

## AutoFlow can be used in ...

... both a server or a client side. AutoFlow is able to launch workflows (i) to servers based on different queue systems, such as SLURM or PBS, and (ii) within clients a bash terminal. Supported clients are those based on the three most popular OS: Linux, Windows and OSX.

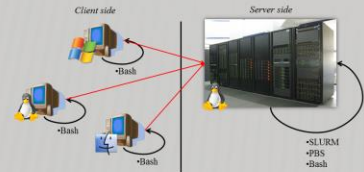


Figure 3. Representation of platforms, queue systems and modes in which AutoFlow can work.

## AutoFlow features

AutoFlow has many features to face from single to large and complex workflows. Some of these features are the iterative task managing, automatic assignment of dependencies, and data transfer between jobs.

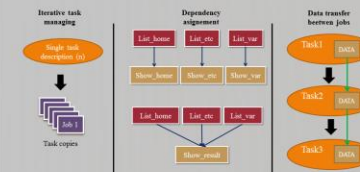


Figure 4. Some features of AutoFlow for easily building and launching of workflows

## Workflow for *V. faba* transcriptome assembling

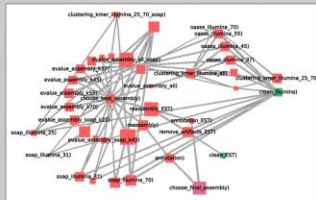


Figure 5. Workflow graph of *V. faba* transcriptome assembly. Green nodes are initial tasks, salmon nodes are intermediate tasks, and the pink node is the final task integrating the results.

## The best transcriptome

Several assemblies were tested to get the best result. The workflow was designed to perform all necessary assemblies (Fig. 5) and automatically choose the best one after checking annotations provided by Full-LengthNext for every assembly.

Input data of workflow	
Illumina reads	33,023,160
EST	106,397
Assemblies info	
Different assemblies evaluated	18
Used assemblers	Oases <sup>3</sup> , Soap <sup>4</sup> and CAP3 <sup>5</sup>
Final assembly attributes	
Assemblers	Primary assembler: Oases, Reassembler: CAP3
Kmer	25,37,45,55,70 merged with CD-hit
Reassembling with EST	Yes
Total transcripts	118,188
Unique transcripts	20413
Unique full-length transcripts	6787

Table 1. Main attributes of the automatically chosen assembly.

## Conclusions

It is shown that AutoFlow is a useful tool for workflow managing. This program can launch workflows automatically taking advantage of its capabilities, such as iterative task managing, dependency tracing and decision taking along the workflow. In addition, as a proof-of-concept, the first representative transcriptome of *V. faba* was obtained.

## Acknowledgements

Work supported by Junta de Andalucía's funding P2010-CV66075. The authors gratefully acknowledge the supercomputing capabilities of the SCBI of the Universidad de Málaga.

1. Holt D., Robinson J., Stoeckert P., Gruber A., Basore R., Lin P., Chen T., Turner J. A tool for building and running workflows of services. In: Next Acids Rev. Volume 34 (sept 2), (2006) 9729-9732  
 2. Garcia J., Nakorodan A., Taylor J. and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. In: Genome Biology. Volume 11 (June 8), (2010) E20  
 3. Hwang S. and Yehia A. CAP3: A DNA sequence assembly program. In: Genome Res. Volume 9, (1999) 166-173.  
 4. M.H. Schulz, D.H. Zentgraf, M. Yegorov and S. Brann Strany. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics, 2012.  
 5. Liu et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Genome Biology 2012, 13:118



# Large-scale transcriptome analysis in faba bean (*Vicia faba* L.)

45

Ocaña S<sup>1</sup>, Seoane P<sup>2</sup>, Claros MG<sup>2</sup>, Bautista R<sup>2</sup>, Torres AM<sup>1</sup>, Madrid E<sup>3</sup>

- (1) Área de Mejora y Biotecnología, IFAPA Centro "Alameda del Obispo", Apdo 3092, 14080 Córdoba, Spain
- (2) Departamento de Biología Molecular y Bioquímica, y Plataforma Andaluza de Bioinformática, Universidad de Málaga, 29071 Málaga, Spain
- (3) Institute for Sustainable Agriculture, CSIC, Apdo 4084, E-14080 Córdoba, Spain

## INTRODUCTION

Although faba bean is one of the biggest traded pulse crops, genomic resources are still limited. Recent advances in sequencing and genotyping technologies offer the opportunity to improve this status at relatively low cost. A comprehensive characterization of the faba bean transcriptome was performed using the Illumina platform. Two faba bean genotypes with differential reaction to *Ascochyta fabae* and 106.397 faba bean sequences found in different public databases were used for *de novo* transcriptome assembly. The study is facilitating new candidate genes for resistance to *Ascochyta* blight and for studying pathways involved in responses of faba bean to fungal pathogens.



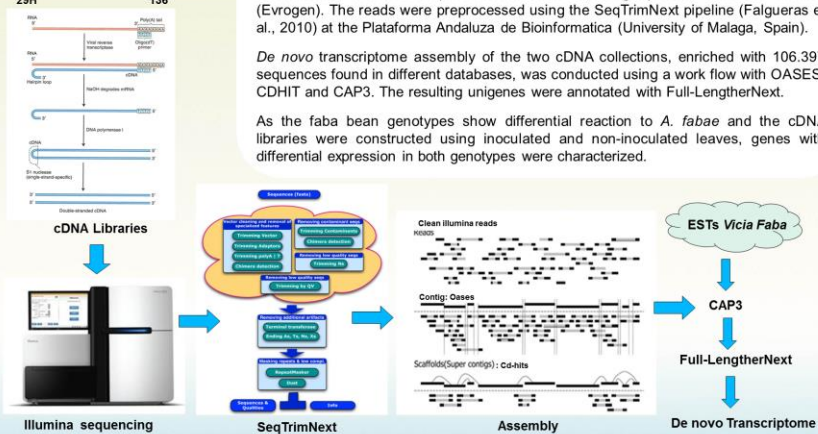
## MATERIALS and METHODS

Faba bean genotypes 29H and Vf136 (resistant and susceptible, respectively to *A. fabae*), were used for the analysis. Leaves were harvested at 4, 8 and 12 hours after inoculation. An identical collection was performed in the non-inoculated lines.

Normalization of cDNA was performed at GenXPro, using a protocol similar to "Trimmer" (Evrogen). The reads were preprocessed using the SeqTrimNext pipeline (Falgueras et al., 2010) at the Plataforma Andaluza de Bioinformática (University of Málaga, Spain).

*De novo* transcriptome assembly of the two cDNA collections, enriched with 106.397 sequences found in different databases, was conducted using a work flow with OASES, CDHIT and CAP3. The resulting unigenes were annotated with Full-LengtherNext.

As the faba bean genotypes show differential reaction to *A. fabae* and the cDNA libraries were constructed using inoculated and non-inoculated leaves, genes with differential expression in both genotypes were characterized.



## RESULTS and DISCUSSION

Based on orthologue ID 20,804 (52.98%) out of the 39,269 (32.47%) were non redundant transcripts. 2,752 putative ncRNAs and 7,165 unigenes lacking homology could be considered as putative legume-specific genes.

Despite of using a single source of tissue the alignment of the faba bean and the chickpea coding DNA sequences revealed a high level of conservation as 81.27% of them were present in the chickpea transcriptome.

The analysis of genes with differential expression revealed 348 and 327 transcripts unique in the *A. fabae* resistant and susceptible genotypes, respectively. Moreover, 212 and 282 transcripts were found to be overexpressed in both lines. These transcripts might represent candidate genes underlying the differences in resistance. The identification of SNPs through data mining is providing a database for fine mapping of target regions and rapid association of markers with *A. fabae* resistance QTLs.

These results represent an extensive databank of faba bean transcripts under stress which will constitute a rich source for gene discovery and molecular markers development to enhance the efficiency and cost effectiveness of faba bean breeding programs.

Reference: Falgueras J, Lara AJ, Fernandez-Pozo H, Canton FR, Perez-Trabado G and Claros MG. 2010. SeqTrim: a high-throughput pipeline for preprocessing any type of sequence read. *BMC bioinformatics*, 11, 38.

**ACKNOWLEDGEMENTS:** This research has been partly financed by the EU Seventh Framework Programme (FP7/ 2007-2013) grant agreement n°FP7-613551, LEGATO project. S Ocaña-Moral acknowledges the INIA fellowship at IFAPA, belonging to the ceiA3



# Análisis de la expresión diferencial de cDNAs y mapeo de genes candidatos para saturar QTLs de resistencia a *Ascochyta* en habas (*Vicia faba* L.)

Nº 87

Ocaña S<sup>1</sup>, Seoane P<sup>2</sup>, Claros MG<sup>2</sup>, Bautista R<sup>3</sup>, Madrid E<sup>3</sup>, Torres AM<sup>1</sup>

<sup>1</sup>Área Mejora y Biotecnología. IFAPA-Centro Alameda del Obispo. Apdo. 3092, 14080 Córdoba

<sup>2</sup>Departamento de Biología Molecular y Bioquímica y Plataforma Andaluza de Bioinformática. Univ. de Málaga. Apdo. 29071 Málaga

<sup>3</sup>Instituto de Agricultura Sostenible. CSIC, Mejora Genética Vegetal. Apdo. 4084, 14080, Córdoba

## INTRODUCCIÓN

Las habas son un importante cultivo y una valiosa fuente de proteínas en la dieta de países en desarrollo. Una de las enfermedades fúngicas que más le afecta es la ascoquitosis, provocada por el hongo *A. fabae* Speg que afecta a hojas, tallos y vainas. Es responsable de pérdidas del rendimiento que oscila entre el 30 y el 40% o incluso el 90% en cultivares muy susceptibles, cuando las condiciones ambientales favorecen el desarrollo de la infección. Se han desarrollado varios mapas genéticos donde se han identificado QTLs (Quantitative trait loci) de resistencia estables (1, 2, 3), pero aún se desconocen los genes responsables de su control. En este estudio se ha realizado una caracterización exhaustiva del transcriptoma de dos genotipos de haba (29H y VF136) con reacción diferencial frente a la infección por el hongo *A. fabae* utilizando la plataforma de secuenciación Illumina.

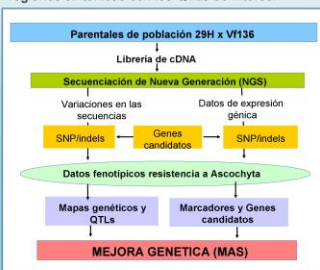


## MATERIALES Y METODOS

Tras el ensamblaje del transcriptoma se identificaron 21.243 unigenes de los que elegimos 850 expresados diferencialmente entre el parental resistente (29H) y el susceptible (VF136) a  $p < 0.05$ . Además se identificaron 39.060 SNPs y 3.669 INDELS.

De los 850 unigenes seleccionamos 128 por presentar algún SNP y se genotiparon en la progenie del este cruce mediante los sistemas de genotipado KASPar e iPLEX-Sequenom.

Asimismo incluimos genes ortólogos localizados en regiones sinténicas con los QTLs de interés.

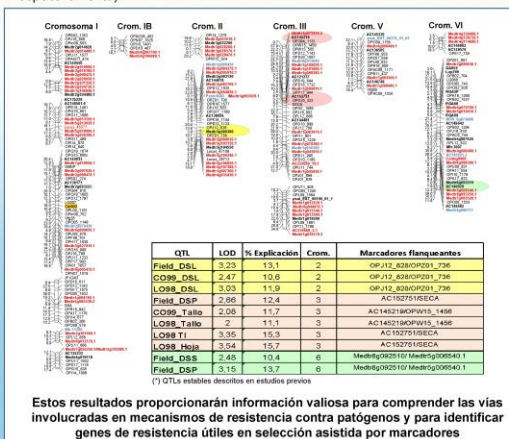


### Referencias

- (1) Avila CM, et al. (2004). Theor Appl Genet 108: 1071–1078
- (2) Diaz-Ruiz R, et al. (2009). Crop Pasture Sci 60: 353–361
- (3) Roman B, et al. (2003). Aust J Agric Res 54: 85–90

## RESULTADOS Y DISCUSIÓN

92 de los genes expresados diferencialmente tras la infección se han incluido en el mapa, saturando las regiones con QTLs de resistencia identificados en estudios previos. Asimismo se incluyeron 10 genes ortólogos (mostrados en rojo y azul, respectivamente).



Estos resultados proporcionarán información valiosa para comprender las vías involucradas en mecanismos de resistencia contra patógenos y para identificar genes de resistencia útiles en selección asistida por marcadores

**Agradecimientos.** Trabajo parcialmente financiado por los proyectos: EU (FP7/ 2007-2013) proyecto n°FP7-613551-LEGATO y 2014-2020 "Programa Operativo de Crecimiento Inteligente" (RTA2013-00025).

### Transcriptome analysis and mapping to identify candidate genes controlling *Ascochyta fabae* resistance in faba bean (*Vicia faba* L.)

Ocaña S<sup>1</sup>, Seoane P<sup>2</sup>, Claros MG<sup>2</sup>, Bautista R<sup>2</sup>, Torres AM<sup>1</sup>, Madrid E<sup>3</sup>

<sup>1</sup>Área Mejora y Biotecnología, IFAPA centro Alameda del Obispo, Apdo. 3092, 14080 Córdoba, Spain

<sup>2</sup>Departamento de Biología Molecular y Bioquímica, Plataforma Andaluza de Bioinformática, Universidad de Málaga, 29071 Málaga, Spain

<sup>3</sup>Institute for Sustainable Agriculture, CSIC, Apdo. 4084, E-14080 Córdoba, Spain

Current address: Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

Faba bean (*Vicia faba* L.) is the third most important food legume in the world being used as a protein source in human diets and animal feeding. Among biotic stresses, diseases are the key limiting factors for faba bean cultivation. One of the most important is ascochyta blight (*Ascochyta fabae* Speg.) which causes yield losses ranging from 35 to 90%. In this work, we aimed to saturate the main quantitative trait loci (QTLs) for *A. fabae* resistance identified in the 29H×Vf136 RIP (Recombinant Inbred Population).

We have exploited a recent transcriptome database obtained from lines 29H (one of the most resistant accessions) and Vf136 after their infection with *A. fabae* (Ocaña et al. 2015), to identify candidate genes related to pathogen resistance. From the 39,060 SNP found after the transcriptome assembling, 229 representing differentially expressed transcripts were selected to be genotyped using the Kaspars and the IPLEX-Sequenom platforms. Finally, 92 SNPs were genotyped in the 29H×Vf136 RIP and used to improve the last genetic linkage map (Atienza et al. 2016), saturate the regions bearing QTLs and refine their position.

Linkage analysis included 307 markers (215 genotyped in previous studies). Using a LOD score of 4, 287 markers mapped into 18 linkage groups. Eighty six of the 92 SNPs, were added in the new map being distributed in the 6 faba bean chromosomes. QTLs located on chromosome II, III, and VI reported previously were confirmed and flanked with new SNPs markers. These gene markers are a good starting point for expression studies and development of diagnostic markers. Apart of information to understand the pathways involved in the mechanism of resistance to *A. fabae*, this study provides the most comprehensive genetic map described in this RIP that will be used for marker-assisted selection in this crop.

#### References

1. Ocaña S, Seoane P, Bautista R, Palomino C, Claros GM, Torres AM, Madrid E (2015) Large-scale transcriptome analysis in Faba bean (*Vicia faba* L.) under *Ascochyta fabae* infection. PLoS ONE (2015) 10(8): 1-17; e0135143. doi:10.1371/journal.pone.0135143.
2. Atienza SG, Palomino C, Gutiérrez N, Alfaro CM, Rubiales D, Torres AM, Ávila CM (2016) QTLs for ascochyta blight resistance in faba bean (*Vicia faba* L.): validation in field and controlled conditions. Crop and Pasture Science 67:216-224.

**Acknowledgements.** This research has received funding the Ministerio de Innovación y Ciencia (MICINN) grant RTA2013-00025, co-financed by the EU through the ERDF 2014-2020 "Programa Operativo de Crecimiento Inteligente".