



TESIS DOCTORAL

**DESARROLLO Y CARACTERIZACIÓN DE HERRAMIENTAS GENÓMICAS
EN *FRAGARIA* DIPLOIDE PARA LA MEJORA DEL CULTIVO DE FRESA**

AUTOR:

JULIO BONET GIGANTE

Memoria de Investigación presentada al
Departamento de Bioquímica y Biología Molecular
para la obtención del grado de

DOCTOR EN BIOTECNOLOGÍA

por la Universidad Autònoma de Barcelona

Memoria de la investigación realizada en el
Laboratorio de Marcadores Moleculares
del Programa de Genómica y Biotecnología del
Institut de Recerca i Tecnologia Agroalimentàries (IRTA)
perteneciente al
Centre de Recerca en Agrigenòmica (CRAG),
financiada por el
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
y **Talència (Catalunya Recerca)**

VºBº

El candidato:

La directora:

La tutora:

Julio Bonet Gigante

Dra. Amparo Monfort Vives
(IRTA-CRAG)

Dra. Carmen Martínez Gómez
(UAB)

Bellaterra, Barcelona,
3 de Noviembre de 2010

Introducción General	1
I El género <i>Fragaria</i> : Taxonomía, sistemática y distribución geográfica	3
II Morfología	5
III La fresa cultivada	6
IV Historia del cultivo de fresa	8
V Importancia económica de la fresa	9
VI Importancia social de la fresa	10
VII Mejora genética en plantas	11
VIII Métodos de mejora	12
IX La mejora asistida por marcadores	12
X La mejora genética de la fresa	14
Objetivos	16
PRIMERA PARTE	
Desarrollo y caracterización de herramientas moleculares para el estudio del genoma de <i>Fragaria</i> diploide	19
1. Introducción	21
1.1 Resumen	23
1.2 Molecular markers in <i>Fragaria</i>	25
1.2.1 Isoenzymes	25
1.2.2 RFLP	26
1.2.3 Arbitrarily-primed markers	26
1.2.4 SCAR	28
1.2.5 CAPS	28
1.2.6 Microsatellites	29
1.2.7 Gene-specific markers	30
1.3 Linkage mapping in <i>Fragaria</i>	33
1.3.1 The <i>Fragaria</i> reference map FV x FB	33
1.3.2 Bin mapping in <i>Fragaria</i>	35
1.3.3 Octoploid mapping	36
1.4 Comparative mapping	38
1.5 Future perspectives	40
2. Desarrollo y caracterización de un mapa de ligamiento de alta resolución en <i>Fragaria</i> diploide	41
2.1 Introducción	42
2.2 Material y métodos	45
2.2.1 Conservación de la población FV x FB	45
2.2.2 Recolección de las muestras y obtención de ADN genómico	46
2.2.3 Selección de marcadores	46
2.2.4 Localización de secuencias <i>in silico</i>	47
2.2.5 Diseño de cebadores específicos y reacción de PCR	47
2.2.6 Visualización de productos de PCR	48
2.2.7 Secuenciación de ADN genómico	49
2.2.7.1 Preparación de los amplicones a secuenciar	49
2.2.7.2 Reacción de secuenciación	49
2.2.7.3 Purificación de los productos de la reacción de secuenciación	51
2.2.7.4 Electroforesis capilar y tratamiento de los datos	51
2.2.7 Análisis genético de los nuevos marcadores y construcción de un mapa de ligamiento	51
2.3 Resultados	52
2.3.1 Diseño de cebadores e identificación de polimorfismos	52
2.3.2 Localización de marcadores mediante <i>Bin mapping</i>	53
2.3.4 Análisis de la segregación de los nuevos marcadores en FV x FB	54
2.3.5 Construcción del mapa de ligamiento FV x FB	57
2.3.6 Revisión del mapa de <i>bins</i> de FV x FB	60
2.4 Discusión	63
2.4.1 Optimización de marcadores moleculares mediante el uso de bases de datos	63
2.4.2 Mapeo genético de nuevos marcadores en el mapa de referencia FV x FB	64

3.	Desarrollo y caracterización de una genoteca de BACs en <i>Fragaria vesca</i>	69
3.1	Resumen	71
3.2	Introduction	72
3.3	Methods	74
3.3.1	High molecular weight DNA preparation	74
3.3.2	Treatment of the plugs	74
3.3.3	Partial digestion and size selection of the high molecular weight DNA	75
3.3.4	Ligation and transformation	75
3.3.5	Estimation of BAC clone insert sizes	75
3.3.6	Pool and superpool development	77
3.3.7	Library screening and identification of BAC clones containing markers from the diploid <i>Fragaria</i> reference map	77
3.4	Results and discussion	78
3.4.1	BAC clone insert sizes and library genome coverage	78
3.4.2	PCR analysis of BAC library pools and superpools	78
3.5	Conclusion	79
3.6	Authors' contribution	83
4.	Localización de genes candidatos para la calidad de fruto en el genoma de <i>Fragaria</i> diploide	85
4.1	Resumen	87
4.2	Introducción	89
4.3	Material and methods	91
4.3.1	Plant material	91
4.3.2	Fruit cDNA libraries and EST sequencing	92
4.3.3	EST processing and annotation	92
4.3.4	<i>P. persica</i> candidate gene analyses	96
4.3.5	Single nucleotide polymorphisms analysis	97
4.3.6	Candidate gene mapping	97
4.4	Results	100
4.4.1	Peach cDNA library and EST sequences	100
4.4.2	Candidate gene selection	100
4.4.3	Primer design to optimize polymorphism identification	101
4.4.4	SNP detection and candidate gene mapping	103
4.4.4.1	Mapping in peach	103
4.4.4.2	Bin mapping in <i>Fragaria</i>	103
4.5	Discussion	108
4.5.1	Peach cDNA library and EST processing	108
4.5.2	Selection of the candidate genes	109
4.5.3	Primer design	109
4.5.4	SNP detection and candidate gene mapping	110
4.5.4.1	Mapping in peach	110
4.5.4.2	Bin mapping in strawberry	110
4.5.5	Synteny analysis between <i>Prunus</i> and <i>Fragaria</i>	111
4.6	Conclusion	111

SEGUNDA PARTE

Desarrollo y caracterización preliminar de una colección de líneas de introgresión en <i>Fragaria</i> diploide	113
1. Introducción	115
1.1 Conocimiento actual de las bases genéticas y moleculares del desarrollo y maduración del fruto de fresa	115
1.1 Mapeo genético de la diversidad fenotípica en plantas	117
1.1.1 Líneas recombinantes consanguíneas	119
1.1.2 Líneas de sustitución	120
1.1.3 Líneas de introgresión por retrocruzamiento	121
1.1.4 Líneas casi isogénicas	122
1.2 Construcción de líneas casi isogénicas	124
1.3 Construcción de una colección de NILs en <i>Fragaria</i> diploide	126
2. Material y métodos	128
2.1 Obtención y mantenimiento de las plantas	128
2.2 Inducción floral y polinizaciones	129
2.3 Recolección de las muestras	130
2.4 Extracción y preparación de ADN genómico	131
2.5 Selección de marcadores	131
2.6 Reacción de PCR	134
2.7 Digestiones enzimáticas	134
2.8 Visualización de productos de PCR	135
2.9 Análisis de caracteres cuantitativos en las líneas de introgresión	136
2.10 Métodos probabilísticos y de análisis estadístico	136
3. Resultados	137
3.1 Diseño experimental del desarrollo de una colección de líneas de introgresión en <i>Fragaria</i> diploide	137
3.2 Obtención de la generación F ₁	138
3.3 Caracterización y selección asistida por marcadores de una población <i>Backcross</i> 1	139
3.4 Caracterización y MAS en 9 familias <i>Backcross</i> 2	143
3.5 Caracterización y MAS en 7 familias generadas mediante autofecundación de las selecciones BC2	147
3.6 Caracterización genómica de las líneas BC2 seleccionadas	151
3.7 Selección de una colección de líneas de introgresión	154
3.8 Mapeo de caracteres cualitativos en la colección de líneas de introgresión	158
3.9 Estudio preliminar de algunos caracteres cuantitativos en la colección	159
3.9.1 Tiempo de floración	159
3.9.2 Resistencia a oidio	160
3.9.3 Tamaño de los folíolos	161
3.9.4 Tamaño del fruto	163
3.9.4.1 Longitud del fruto	163
3.9.4.2 Anchura del fruto	164
3.9.5 Forma del fruto	165
3.9.6 Evaluación de otros caracteres	166
4. Discusión	169
4.1 Estudio genético de las poblaciones generadas	169
4.2 Estructura genómica de la colección de líneas de introgresión	170
4.3 Diseño experimental	171
4.4 Caracterización preliminar de la colección de líneas de introgresión	173

Discusión General	177
V.1 Genética de las especies diploides de <i>Fragaria</i>	179
V.2 Herramientas genómicas en las especies diploides de <i>Fragaria</i>	182
V.3 Potencial de la colección de líneas de introgresión para el estudio de caracteres fenotípicos	184
V.4 Perspectivas de futuro en la genómica del género <i>Fragaria</i>	186
Conclusiones	187
Bibliografía General	191
Anexos	217
Anexo I Segregación de marcadores moleculares en FV x FB	219
Anexo II Abreviaciones	226
Anexo III Lista de figuras y tablas	228
Agradecimientos	231

La fresa es uno de los frutos con mayor cuota en la exportación de frutas y hortalizas en nuestro país. Sin embargo, la naturaleza octoploide de su genoma ha lastrado históricamente las investigaciones genéticas en esta especie. La fresa diploide constituye un excelente modelo para el estudio de las bases genéticas que gobiernan la calidad del fruto de fresa y de otras especies de la familia *Rosaceae*, dado que es una planta pequeña, de ciclo de vida corto, posee un genoma muy pequeño, es susceptible de ser transformada y regenerada con facilidad, se reproduce tanto de manera sexual como asexual y produce gran cantidad de descendientes en una sólo generación. Además, los genomas diploide y octoploide del género *Fragaria* presentan un alto grado de sintenia y colinearidad. Es por ello que en este trabajo de Tesis Doctoral se ha escogido *Fragaria vesca* como modelo para el desarrollo de herramientas genómicas encaminadas a localizar en su genoma los factores genéticos que determinan las principales características agronómicas y nutricionales que más interesan al sector fresero, de cara a la mejora genética de la especie.

Este trabajo se compone de dos partes. En la primera parte se detalla el desarrollo de herramientas moleculares para el estudio del genoma de *Fragaria* diploide como la mejora del mapa genético de referencia del género, mediante la transformación de algunos marcadores moleculares en marcadores de tipo PCR y la incorporación de nuevos marcadores, la generación y caracterización de una genoteca de BACs para la confección de un mapa físico del género y el mapeo genético de genes candidatos, seleccionados para la calidad del fruto y obtenidos mediante la transferencia de marcadores del género *Prunus* al género *Fragaria*. En la segunda parte, se detalla el desarrollo de una serie de poblaciones segregantes generadas con objeto de crear una colección de líneas de introgresión a partir de las especies diploides *Fragaria vesca* var. 'Reine des Vallées' (conocida como Reina de los Valles en nuestro país) y la especie silvestre *Fragaria bucharica*. Esta colección de líneas ha sido obtenida tras dos rondas de retrocruzamientos seguidas de una autofecundación, mediante selección asistida por marcadores. La colección consta de 30 líneas cuyo genoma es esencialmente igual al del parental recurrente, *F. vesca*, a excepción de un segmento definido por marcadores moleculares, heredado del parental donante, *F. bucharica*. Cada una de las líneas posee un segmento o introgresión diferente, de manera que la suma de todas las introgresiones cubre más del 97% del genoma de *Fragaria* diploide. Las introgresiones tienen un tamaño medio de 43 cM, proporcionando una resolución media para el mapeo de 16,9 cM. Dado que la mayoría de las líneas ha heredado más del 90% del genoma del parental recurrente, y puesto que éste es un cultivar comercial, la genoteca presentada aquí es, además, un recurso precompetitivo que puede ser usado en mejora genética. La caracterización preliminar de algunas de las líneas obtenidas mediante el análisis de algunos caracteres agronómicos como el tiempo de floración, la fenología de la planta, la resistencia a

enfermedades, así como el tamaño y la forma de sus hojas y frutos ha permitido identificar ciertas regiones genómicas que parecen estar controlando estos caracteres en el género *Fragaria*, demostrando que ésta colección de líneas de introgresión es una herramienta muy eficaz para la identificación y disección de caracteres complejos. Además, se ha localizado un nuevo locus implicado en la autoincompatibilidad asociada a algunas especies del género y se han desarrollado líneas que pueden propagarse vegetativamente ya que, pese a ser comercial, la variedad 'Reine des Vallées' no produce estolones.

INTRODUCCIÓN GENERAL

I. EL GÉNERO *FRAGARIA*: TAXONOMÍA, SISTEMÁTICA Y DISTRIBUCIÓN GEOGRÁFICA

El género *Fragaria* (*Fragaria spp.*, del latín *fragans*, oloroso) pertenece a la familia *Rosaceae*, que agrupa unas 3000 especies de 107 géneros diferentes, distribuidas en su mayor parte en las zonas cálidas del hemisferio norte. Se trata de una de las familias con mayor importancia económica en el mundo, ya que, además de la fresa, incluye otras especies frutales pertenecientes a diversos géneros como *Prunus*, *Malus* (con más de 2000 variedades registradas), *Pyrus* o *Rubus*, así como especies de uso ornamental como las pertenecientes al género *Rosa* (con más de 5000 cultivares registrados), *Potentilla* o *Sorbus*.

La fresa es una planta dicotiledónea del género *Fragaria*. Su clasificación sistemática es:

- Reino: *Plantae*
- Subreino: *Embryobionta*
- División: *Magnoliophyta*
- Clase: *Magnoliopsida*
- Subclase: *Rosidae*
- Superorden: *Rosanae*
- Orden: *Rosales*
- Familia: *Rosaceae*
- Subfamilia: *Rosoideae*
- Tribu: *Potentilleae*
- Subtribu: *Fragariinae*
- Género: *Fragaria*

El género *Fragaria* comprende 25 especies e híbridos que poseen diferentes niveles de ploidía (tabla 1) y se distribuyen por toda la zona templada del hemisferio norte, ocupando diferentes regiones climáticas que van de los climas templados a los subtropicales (Staudt, 2008). Sólo *F. vesca* ($2n=2x=14$) se distribuye de manera homogénea a lo largo de todo este rango (Hancock, 1999) y al oeste de Sudamérica (Fig. 1). El resto de especies diploides están confinadas en Eurasia (Fig. 2), las tetraploides al este de Asia y, *F. moschata*, la única especie hexaploide, a Europa central y Siberia (Staudt, 1989). Las especies octoploides de *Fragaria* son nativas de América (*F. virginiana* y *F. chiloensis*) desde donde colonizaron la isla de Iterup, al nordeste de Japón (*F. iterupensis*). Como la mayor parte de las especies de *Fragaria* se localizan en Asia, esta región está considerada como el centro de diversidad del género, e incluso ha sido postulada como su lugar de origen (Staudt, 1999).

Tabla 1. Especies del género *Fragaria*, ploidía y distribución geográfica. De Hancock, 1999; Rousseau-Gueutin et al. 2009, Hummer et al. 2009

Especie	Ploidía	Distribución geográfica
<i>F. bucharica</i>	2x	Himalaya occidental
<i>F. daltoniana</i>	2x	Himalaya
<i>F. iinumae</i>	2x	Japón occidental
<i>F. mandshurica</i>	2x	Nordeste asiático
<i>F. nilgerrensis</i>	2x	Asia central y China
<i>F. hayatai</i>	2x	Taiwan
<i>F. nipponica</i>	2x	Islas de Honshu y Yakushima, Japón
<i>F. nubicola</i>	2x	Asia central, de Bokhara al Himalaya
<i>F. pentaphylla</i>	2x	China y Tibet
<i>F. chinensis</i>	2x	China
<i>F. vesca</i>	2x	Eurasia y América
<i>F. ×bifera</i>	2x	Europa
<i>F. viridis</i>	2x	Euro-Siberia
<i>F. corymbosa</i>	4x	Norte de China
<i>F. gracilis</i>	4x	Noroeste de China
<i>F. moupinensis</i>	4x	Suroeste de China
<i>F. orientalis</i>	4x	Nordeste de Asia
<i>F. tibetica</i>	4x	Himalaya oriental
<i>F. moschata</i>	6x	Europa y Rusia
<i>F. chiloensis</i>	8x	Oeste, norte y sur de América, Hawaii
<i>F. virginiana</i>	8x	Norteamérica
<i>F. iturupensis</i>	8x/10x	Iturup, Japón/Monte Atsunupuri
<i>F. cuneifolia</i>	8x	Costa oeste de norteamérica
<i>F. ×ananassa</i>	8x	Híbrido, cultivado en todo el mundo
<i>F. ×bringhurstii</i>	8x	Híbrido, costa oeste de los EEUU



Figura 1. Distribución de *Fragaria vesca*. Basado en los datos de Hancock y Luby (1993).

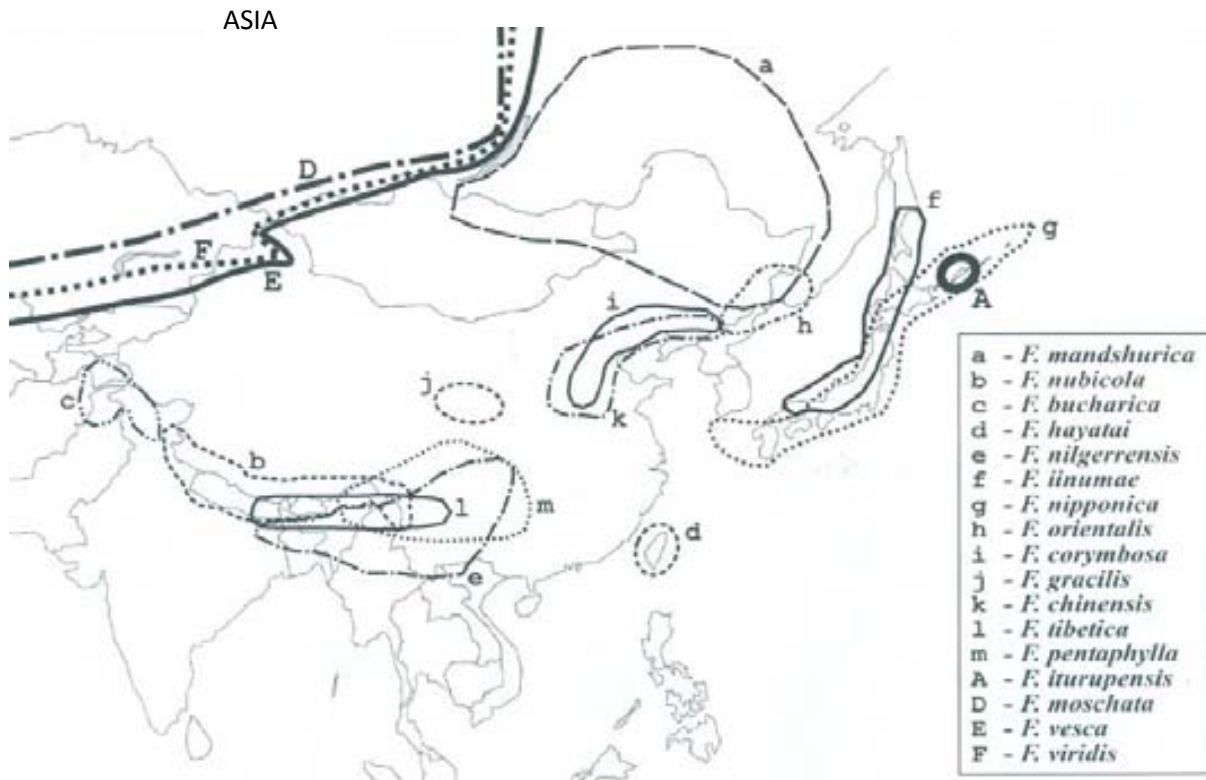


Figura 2. Distribución de las especies asiáticas descritas en el género *Fragaria* (Staudt, 2008).

II. MORFOLOGÍA

La fresa es una planta perenne de pequeño porte, que se reproduce de manera sexual y asexual (mediante el desarrollo de estolones). Aunque tradicionalmente se considera como planta herbácea, no es tal, sino que en realidad se trata de una especie leñosa y perenne con las mismas o similares pautas fisiológicas que los árboles y arbustos frutales de hoja caduca (López-Aranda, 2008). Su ciclo de vida es corto (de doce a veinte semanas por generación). El tallo está comprimido en una roseta basal o corona, de la que surgen las hojas en muy estrechos intervalos, trifoliadas dentadas de haz glabrescente y envés con pelos aplicados, cuyos peciolo pueden alcanzar los 20 cm de altura (Fig. 3). En las axilas de las hojas se desarrollan yemas o meristemos axilares. Estas yemas, dependiendo del estado nutricional y de las condiciones ambientales, evolucionan de diferente manera: Permanecen aletargadas o desarrollan estolones, ramas o escapos florales. Los estolones, o tallos rastreros, producen raíces adventicias, de las que pueden surgir eventualmente nuevas plantas. Se reproduce sexualmente mediante la formación de inflorescencias generalmente hermafroditas, pequeñas, en cima dicásial o monocásial, de pétalos blancos y receptáculo amarillo. Los receptáculos terminan desarrollando poliaquenos o 'eterios' que contienen los verdaderos frutos (aquenos) en su superficie. Los eterios, denominados fresas, son ovoides o subglobosos, jugosos, dulces y muy aromáticos, con aquenos de 0,6-1,5 mm, glabros, no hundidos en alvéolos del pseudocarpo. El fruto

de fresa pertenece a la categoría de los no climatéricos, por lo que no completará su madurez comercial una vez recolectado. La forma y tamaño de los frutos es una característica varietal, aunque los factores ambientales afectan en gran medida a este carácter (Navarro y Muñoz-Garmendia, 2005).

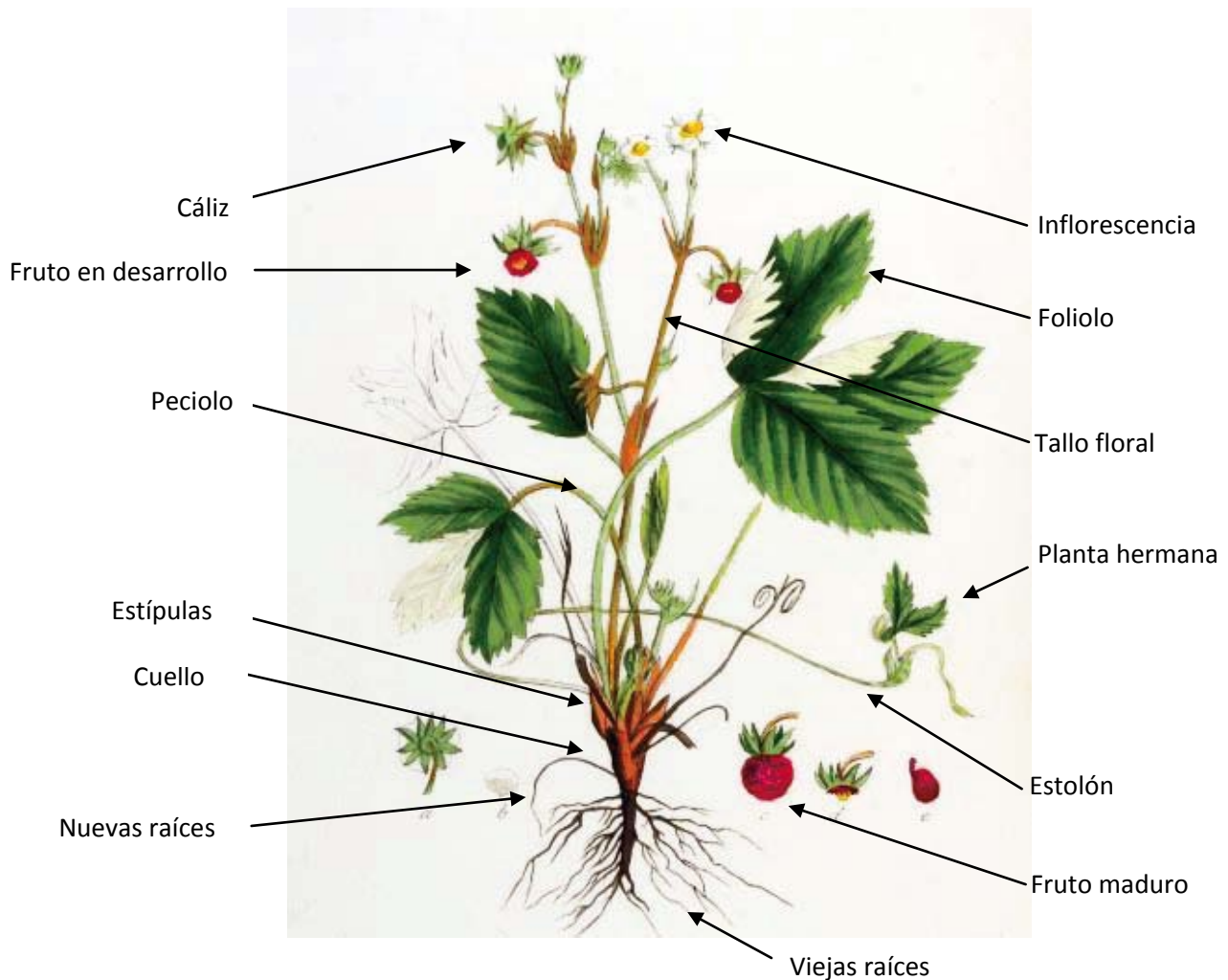


Figura 3. Morfología general de una planta de fresa. Imagen de Kops *et al.* (1844)

III. LA FRESA CULTIVADA

La fresa que normalmente se comercializa (fresón) es el híbrido *Fragaria x ananassa*, con caracteres intermedios entre *Fragaria chiloensis* y *F. virginiana*. Es robusta (10-40 cm), con rizomas muy ramificados y sin estolones o con éstos muy cortos, hojas verde oscuro, coriáceas, de haz glabro y envés peloso. El eterio, de 20-50 mm, es subcónico, de un rojo intenso brillante, a veces rosado o de un rojo pálido, que se separa con dificultad del tálamo. (Navarro y Muñoz-Garmendia, 2005).

Las variedades de fresa muestran diferentes tipos de respuestas a la longitud del día (fotoperiodo) y a las temperaturas (termoperiodo). Cada tipo de respuesta ha dado lugar a un tipo de cultivares; variedades de día corto, remontantes (o reflorecientes) y de día neutro.

Las variedades de día corto diferencian yemas florales cuando los días llegan a ser cortos y las temperaturas son bajas a finales de verano o principio de otoño. Las diferentes fases de su desarrollo tienden a ser separadas: floración, fructificación y emisión de estolones, en esta secuencia. Hay un acuerdo general en aceptar que las denominadas variedades de día corto o variedades de cosecha única tienen su inducción floral en condiciones de día corto y temperaturas bajas (finales de verano/otoño). Producen estolones vigorosamente hasta que los días se acortan hasta 12/14 horas. En condiciones de día corto, incluso con el estolonado ya parado, el desarrollo de coronas y hojas continúa mientras que las condiciones sean favorables.

Las variedades remontantes diferencian yemas de flor más libremente en días largos (durante todo el verano) que en días cortos. Los estolones se producen desde principios de la primavera hasta que se paran por las bajas temperaturas otoñales; en contraste con las plantas de día corto, las variedades remontantes producen menos estolones y tienden a formar múltiples coronas. Las variedades de día neutro se ven relativamente poco afectadas por el fotoperiodo y por el termoperiodo; fructificarán siempre que las temperaturas sean suficientemente altas para mantener el crecimiento. Los estolones se reducen durante el verano y continúan hasta los primeros días cortos; no entran en reposo bajo condiciones de día corto si prevalecen las temperaturas favorables. Los tipos de día neutro florecen, fructifican y estolonan simultáneamente. Producen una cosecha continuada desde primavera hasta otoño, en función del clima, con varios picos de cosecha a lo largo del periodo de cultivo. Los términos día neutro y remontante se suelen usar como sinónimos. Mientras que las variedades remontantes son producto de una selección natural en zonas continentales y nórdicas, las variedades de día neutro con comportamiento fisiológico y productivo similar son producto de la mejora genética tradicional que busca la ruptura de la dependencia del fotoperiodo para incrementar la producción durante todo el año.

El fotoperiodo impone su influencia sobre la formación de yemas florales, elongación de estolones, tamaño de la hoja y longitud del peciolo. Por otro lado, el termoperiodo puede modificar e incluso anular los efectos de la longitud del día, y debe ser considerado de igual importancia en la determinación de la adaptación de la fresa a ambientes concretos. Las variedades de día corto finalizan el estolonado en otoño junto con la diferenciación floral y la iniciación de la latencia, con

fotoperiodo y temperatura decrecientes. En invierno (fotoperiodo y temperatura mínimos) ocurre la parada vegetativa; en primavera (fotoperiodo y termoperiodo crecientes) se reanuda la actividad vegetativa y se desencadena el proceso de floración-fructificación, además del estolonado. En verano (fotoperiodo y temperatura máximos) disminuye el proceso de floración-fructificación y aumenta el proceso de estolonado.

La cantidad de frío que la planta recibe durante la latencia hibernal es muy importante en el comportamiento fenológico de la planta. La fresa necesita satisfacer unas necesidades concretas de horas de frío, entendidas como horas por debajo de los 7°C, que varían de unas variedades a otras, y que conlleva el paso de la planta por viveros en altura, donde tiene lugar la producción de estolones.

IV. HISTORIA DEL CULTIVO DE FRESA

Las noticias más creíbles acerca del inicio del uso y cultivo de la fresa se remontan a la época romana, según constatan autores clásicos de los siglos I a III a.C. como Cato o Virgilio, u otros como Ovidio o Plinio, ya en el siglo I d.C. Todos ellos citan en sus escritos las fresas como plantas de frutos muy apreciados por su sabor y fragancia. Con el paso del tiempo, su utilización como planta medicinal intensificó su cultivo, pues a su zumo se le atribuyeron propiedades curativas para combatir las afecciones de garganta, la fiebre y la transpiración.

El cultivo en Europa no comenzó hasta el siglo XIV. Las primeras referencias hablan de plantas silvestres de *F. vesca* traídas desde sus hábitats naturales a los jardines de la corte francesa. La primera descripción botánica data de 1484, realizada en el Herbario Latino de Mainz, publicado en Alemania. En España, existen referencias a su cultivo desde 1539. Gabriel Alonso de Herrera dijo que las fresas eran plantas favoritas en pequeños huertos y jardines, donde crecían exuberantemente. Es muy probable que las especies cultivadas en España fuesen *F. vesca* y *F. moschata*.

A finales del S. XVI e inicios del XVII, los primeros colonizadores introdujeron simultáneamente en Francia e Inglaterra una especie americana, *F. virginiana* o fresa de Virginia. Los detalles de la introducción de *F. virginiana* son desconocidos, pero es probable que Jacques Cartier, descubridor del río San Lorenzo en Canadá (1534), fuese el primero en traerla del viejo continente. La fresa de Virginia resultó ser más productiva y mostraba un fruto de mayor tamaño, por lo que su cultivo se extendió por toda Europa.

Poco después, tras la conquista de Chile por Pedro de Valdivia, los españoles conocieron una nueva especie, *F. chiloensis* que, como su 'hermana' *F. virginiana*, presentaba grandes frutos y era conocida y utilizada desde los albores del segundo milenio por los nativos. Su llegada a Europa se produjo *via* Marsella en 1714, de la mano de François Frézier, oficial de Luis XIV, que trajo consigo 5 plantas vivas traídas desde Concepción. Desde París, la fresa chilena se distribuyó a jardines botánicos de Holanda, Inglaterra, Bélgica y Alemania, desde donde pronto se recibieron informes negativos, en tanto que muestra importantes problemas de esterilidad.

En Brest, y en especial en la cercana comunidad de Plougastel, los jardineros descubrieron que la esterilidad de la fresa chilena podía ser superada mediante polinización cruzada, de manera que la famosa fresa de Bretaña pasó a ser un híbrido entre *F. chiloensis* y *F. virginiana*. Phillip Miller fue el primero en describir la nueva fresa en el *Gardener's Dictionary* de 1759. Pocos años después, en 1766, Duchesne afirmó en su libro *Historia Natural de las Fresas*, que la nueva fresa es un híbrido entre *F. virginiana* y *F. chiloensis*, y nombra a este híbrido como fresa-ananás o fresa-piña, ya que encuentra en ella que su olor es como el de la piña tropical (*Ananas spp.*). Así, poco después la clasifica como *Fragaria x ananassa*, híbrido que hoy se cultiva en todo el mundo (Medina, 2008).

V. IMPORTANCIA ECONÓMICA DE LA FRESA

La fresa cultivada ha adquirido una gran importancia económica y social en Europa, ya que es uno de los productos con mayor cuota en la exportación de frutas y hortalizas. En menor medida, la fresa diploide *F. vesca* f. *semperflorens*, también se cultiva para usos industriales. En 2007 la producción mundial de fresa superó los 3.5 millones de toneladas, de las cuales 263.900 (el 7.2%), se cultivaron en España (fig 4). Si bien no existen datos oficiales referentes a la producción china, las estimaciones del departamento estadounidense de agricultura cifraron la producción en 1.950.000 millones de toneladas en la temporada 2009/2010 (<http://www.usdachina.org>). La producción fresera española se concentra principalmente en Huelva (donde en 2007 se cultivaron 6.326 hectáreas) y genera un volumen de facturación de unos 300 millones de euros al año. Después de la aceituna, la fresa es el segundo producto agrícola más exportado por España, que es además, segundo productor mundial. Parece evidente que cualquier esfuerzo encaminado a mejorar el cultivo de la fresa es beneficioso para la modernización del sector agrícola en nuestro país (FAOStat <http://faostat.fao.org/>).

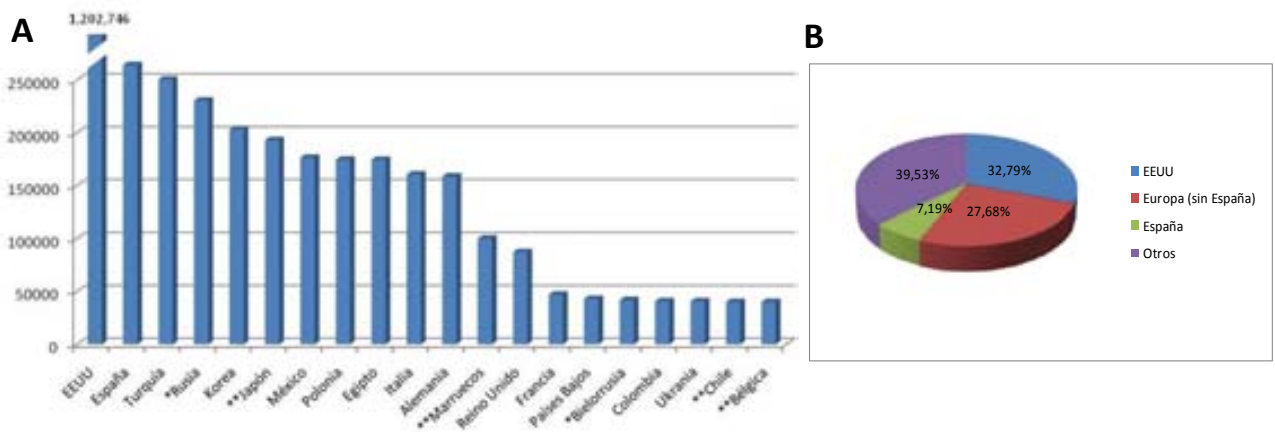


Figura 4. Producción mundial de fresa en 2007. A: Producción de los 20 mayores productores de fresa del mundo en millones de toneladas. *Datos no oficiales. **Estimación de la FAO. B: Producción fresera española frente a la producción europea, estadounidense y mundial. Fuente: <http://faostat.fao.org/>.

VI. IMPORTANCIA SOCIAL DE LA FRESA

Existen numerosas publicaciones recientes que señalan a la fresa como una excelente fuente de polifenoles antioxidantes, vitamina C y fibra alimentaria soluble e insoluble de buena calidad dietética (Hannum, 2004; Zafra-Stone *et al.* 2007). Los pigmentos antocianicos y taninos hidrolizables (elagitaninos) y no hidrolizables (proantocianidinas) son sus principales constituyentes interesantes para la salud (Atkinson *et al.* 2006). Estos compuestos parecen aportar beneficios disminuyendo el riesgo de padecer determinados tipos de cáncer (sobre todo aquellos del aparato digestivo (Chung *et al.* 2002; Damianaki *et al.* 2000)) y enfermedades cardiovasculares (Díaz *et al.* 1997). Las fresas son una fuente de polifenoles que pueden presentar propiedades positivas para la prevención y erradicación de enfermedades neurodegenerativas (Andres-Lacueva *et al.* 2005; Bickford *et al.* 2000), la diabetes tipo II y el envejecimiento en general (Miller *et al.* 1998). La mejora de las variedades comerciales de que disponemos hoy permitirá seleccionar nuevas variedades de fresa con mayores contenidos en estos y otros principios activos, y por tanto, ayudar a mejorar la calidad de vida de nuestra sociedad.

VII. LA MEJORA GENÉTICA EN PLANTAS

Históricamente, la mejora genética de plantas ha tenido como objetivos principales el aumento de la productividad, la obtención de resistencias a plagas y enfermedades o la mejora de características agronómicas. Se estima que durante la segunda mitad del S. XX, el 50% de los aumentos de cosecha en los cultivos más importantes para la humanidad son atribuibles a la introducción de nuevas variedades obtenidas mediante mejora genética (Hayward *et al.* 1993). En los países donde tuvo más éxito la denominada 'segunda revolución verde' (Pakistán, India o México), la producción de alimentos básicos se duplicó en pocos años gracias a la introducción de variedades mejoradas.

La mejora genética de la calidad es, salvo excepciones, un objetivo relativamente reciente; surge a partir del momento en que para muchos cultivos hay superproducción en el mundo desarrollado y que, el éxito en el mercado de una nueva variedad, sobre todo en el campo de las frutas y hortalizas, está cada vez más ligado a la calidad sensorial, en primer lugar, y a la calidad nutricional y sanitaria en segundo lugar. Si bien es cierto que en el mundo hay todavía muchos millones de personas desnutridas (OMS, 2010) la solución no sólo debe pasar por la obtención de mayor cantidad de alimentos, sino también por el incremento de su calidad nutricional.

Sin embargo, la mejora genética de la calidad en plantas avanza muy lentamente, debido a una serie de problemas que dificultan este avance: Es frecuente que las exigencias de los mercados no se correspondan con caracteres intrínsecos de calidad. Los criterios de calidad varían y a menudo dependen de las grandes cadenas de distribución, cuyo criterio a menudo difiere con el del consumidor final (Llácer *et al.* 2005). En frutales, por ejemplo, la mayoría de esfuerzos se han dedicado a la apariencia externa más que en la calidad interna o nutricional. Por otro lado, los objetivos de mejora de la calidad son a veces contrarios a otros objetivos que no pueden soslayarse. Así, la obtención de mutantes de maíz con buena calidad proteica aportó, secundariamente, una serie de características agronómicas poco deseables, como menor rendimiento, mala germinación y mayor índice de plagas (Oikeh *et al.* 2004). Igualmente, la selección de productos de larga conservación y tolerancia al transporte o la introducción de resistencias a plagas y enfermedades a menudo modifica negativamente sabores y texturas en perjuicio de la calidad sensorial.

VIII. MÉTODOS DE MEJORA

Los métodos más elementales de mejora se basan en la selección de mutaciones espontáneas a partir de variedades originales. Prácticamente todas las variedades de cítricos cultivadas en la actualidad tienen este origen. El método más utilizado para generar nueva variabilidad es, sin duda, la mejora clásica o convencional mediante cruzamientos. En la actualidad, la mejora clásica se combina con la mejora asistida por marcadores (MAS, por sus siglas en inglés, *Marker Assisted Selection*) cuando éstos están disponibles, ya que permiten realizar la selección en fases juveniles de individuos con caracteres específicos. En los últimos años se está imponiendo cada vez más la producción de híbridos F₁ desarrollados por empresas privadas mediante el cruzamiento de dos o más líneas puras, que no tienen valor agrícola si se siembran por separado (Cuartero y Rodríguez, 1998). La mutagénesis inducida por métodos físicos o químicos es otra vía bastante utilizada para aumentar la variabilidad de caracteres de calidad. Así, en mandarinos se ha logrado obtener, mediante tratamiento con colchicina, plantas autotetraploides estables que luego se emplean como parentales femeninos para la obtención de híbridos triploides, productores de frutos sin semillas (Aleza *et al.* 2009). Los métodos basados en cultivo *in vitro* tienen aplicaciones en la generación de variabilidad; la inducción de variación somaclonal, seguida de la regeneración de plantas *via* embriogénesis somática es una de las fuentes de variabilidad aprovechada en programas de mejora genética. Por otra parte, la hibridación somática permite cruzamientos interespecíficos o incompatibles que no serían posibles por métodos convencionales. Por último, y a pesar de los problemas políticos y de desinformación pública que impiden su aplicación en la industria agroalimentaria europea, se han realizado grandes progresos en los métodos de transformación genética que permiten introducir en una especie genes procedentes de otra con la que no se puede cruzar. Estos métodos podrían tener mucha importancia en un futuro próximo para productos con usos industriales.

IX. LA MEJORA ASISTIDA POR MARCADORES

Un marcador genético es un polimorfismo heredable de forma mendeliana simple, cuya variabilidad puede ser detectada y estudiada en análisis genéticos. Los primeros marcadores descritos fueron marcadores morfológicos, basados en diferencias fenotípicas debidas a genes con efectos mayores que presentan herencia simple y son de expresión estable en diferentes ambientes (Immer y Henderson, 1943; Allard, 1953). Posteriormente se desarrollaron los marcadores isoenzimáticos, basados en que dos variantes de una misma enzima (isoenzimas) pueden presentar diferencias de

carga y por lo tanto pueden ser separadas por electroforesis en geles de almidón (Hunter y Market, 1975). La principal limitación de estos marcadores es que se encuentran en cantidades muy reducidas. En cambio, los marcadores moleculares, basados en polimorfismos del ADN, se distribuyen homogéneamente por todo el genoma y no están influenciados por el ambiente (Tanskley *et al.* 1989).

Los marcadores moleculares pueden utilizarse tanto para el estudio de la variabilidad genética o identificación varietal como para elaborar mapas genéticos, también conocidos como mapas de ligamiento. El concepto de ligamiento entre marcadores genéticos en la herencia fue propuesto hace casi un siglo por Morgan (1911) y el primer mapa genético lineal, el del cromosoma X de *Drosophila melanogaster* fue publicado por Sturtevant en 1913. Los mapas genéticos son representaciones gráficas del orden de los marcadores genéticos de un cromosoma o grupo de ligamiento. El mapeo de marcadores moleculares se basa en el principio de que los genes y marcadores segregan *via* recombinación cromosómica durante la meiosis, lo que permite su análisis en la progenie. Los genes o marcadores que están próximos en el cromosoma, son transmitidos juntos a la descendencia más frecuentemente que los genes o marcadores que están más alejados. En una población segregante, hay una mezcla de genotipos de tipo parental y recombinante. La frecuencia de genotipos recombinantes para dos loci puede ser empleada para calcular la fracción de recombinación, que se usa para inferir la distancia genética entre esos loci. Cuantos más genotipos recombinantes aparecen entre dos loci, mayor es la distancia genética, medida en centiMorgans (cM), entre ellos. Un centiMorgan equivale a una frecuencia de recombinación del 1%. La fracción de recombinación no es aditiva a lo largo del cromosoma y la desviación respecto de la aditividad aumenta con la distancia entre los loci debido a la presencia de eventos de doble recombinación. De este modo, se han desarrollado funciones de mapa tales como Haldane (1919) o Kosambi (1944) para caracterizar la separación entre dos loci.

El uso de marcadores moleculares proporciona varias ventajas respecto a la mejora clásica o selección masal, dada la independencia de estos marcadores respecto a las condiciones ambientales y del estadio en el que se encuentra la planta. Este fenómeno permite una selección temprana, disminuyendo los gastos del cultivo. La eficacia de esta estrategia depende de varios factores como el modo de herencia del carácter de interés, el grado de asociación entre el marcador empleado y el gen de interés, el número de individuos a analizar y el entorno genético al que se transfiere el gen de interés (Francia *et al.* 2005).

X. LA MEJORA GENÉTICA DE LA FRESA

Actualmente existen más de mil variedades de *F. x ananassa* en el mundo. Las más cultivadas son las californianas. En España, las variedades californianas han dominado el mercado tradicionalmente, si bien en los últimos años algunas variedades desarrolladas en Huelva están alcanzando cuotas de mercado crecientes. Los principales problemas a los que se enfrenta el cultivo de la fresa, concretamente en la región de Huelva, son los siguientes (Bombarely, 2007):

- El mercado europeo se abastece de fresas procedentes de mercados del norte de África, en especial de Marruecos, donde la producción es más precoz y presenta un menor coste.
- La eliminación de los sistemas tradicionales de desinfección del suelo en Europa (bromuro de metilo) por agotar la capa de ozono, sin que existan alternativas ecológicas de eficacia similar.
- La baja vida post-cosecha de los frutos y su vulnerabilidad al daño mecánico durante el transporte.
- La mejora de la calidad organoléptica del fruto
- Las patologías fúngicas del fruto, como la podredumbre gris, causada por *Botrytis cinerea*, la antracnosis, causada por *Colletotrichum spp.* y, en menor medida, el moho pulverulento, *Spaerotheca sp.*
- Las patologías fúngicas foliares, como el oidio (*Oidium fragariae* y *Spaerotheca masularis*), la mancha púrpura (*Mycosphaerella fragariae*) y la quemadura sobre la hoja (*Diplocarpon earliana*). Existen otros hongos foliares menos extendidos como *Fusarium sp.*, *Phytophthora sp.*, *Rhizoctonia sp.*, *Rhizopus sp.*, *Pythium sp.*, *Cladosporium sp.*, *Alternaria sp.* y *Penicillium sp.*
- Las patologías causadas por insectos. El insecto más dañino es sin duda la araña roja (*Tetranychus urticae* Koch), seguido de los *thrips* (*Frankliella occidentalis*)
- Las patologías bacterianas. La más común, causada por *Xanthomonas fragariae*, afecta a las hojas.
- Las patologías víricas: Clorosis del borde de la hoja moteado (SMV), el retorcimiento de la hoja (SCV) causado por un *Rhabdovirus* citoplasmático, el virus del curvamiento de la hoja y la clorosis intervenal (SVBV), causado por un *Caulimovirus*.

En los últimos años se han venido desarrollado programas de mejora genética de fresa en todo el mundo, encaminados a la obtención de variedades que se ajusten a la demanda de los diferentes sectores freseros en cuanto a calendario de producción, adaptación y características cualitativas y organolépticas del fruto. De entre los caracteres de interés agronómico, parece que el tamaño del fruto ha alcanzado su máximo comercial, sin que el consumidor se encuentre atraído por frutos mayores a los ya existentes. En cuanto a las necesidades de los productores, un elemento trascendental para la distribución es poder prolongar la vida postcosecha de los frutos. También son importantes la homogeneidad, la linealidad de la producción o el incremento de la productividad, así como la adecuación de las variedades existentes a las condiciones ambientales en las que se cultivan. Las características organolépticas son de importancia creciente para el consumidor (firmeza, color, aroma y contenido en azúcares). Igualmente, existe una demanda creciente de alimentos ricos en compuestos bioactivos de efecto beneficioso para la salud humana (compuestos fenólicos, capacidad antioxidante total y contenido en vitamina C) (Hureau *et al.* 2008; Soria *et al.* 2009).

Durante la última década, los mayores esfuerzos en la mejora genética de la especie se centraron en la resistencia a enfermedades como los hongos del suelo (principalmente *Phytophthora cactorum*, *Colletotrichum spp.* y *Verticillium spp.*) y hongos aéreos (*Botritis cynerea* y oidio especialmente). La reducción en el uso de pesticidas y la mejora de las características nutricionales del fruto de fresa son sin duda el gran reto para el futuro, puesto que los consumidores se han posicionado claramente a favor de alimentos más sanos y más seguros (López-Aranda *et al.* 2004).

En todos estos casos se constata la necesidad de conocer mejor los mecanismos, tanto bioquímicos como moleculares, relacionados con los procesos que determinan la maduración del fruto, no sólo por el conocimiento básico de dichos mecanismos, sino por la posible aplicación de técnicas biotecnológicas que redunden en la mejora de las propiedades comerciales de estos frutos. El desarrollo de técnicas de selección asistida por marcadores en los programas de mejora de fresa es fundamental para la rápida consecución de nuevas variedades que se ajusten a las necesidades del mercado, principalmente a las relacionadas con la reducción del uso de pesticidas y la mejora de las características nutricionales (Hidalgo *et al.* 2009).

En este trabajo de Tesis Doctoral se han desarrollado herramientas moleculares que pueden ser de gran utilidad para poner a punto las estrategias biotecnológicas que permitan avanzar en la mejora genética de la especie. Para ello, es imprescindible conocer la localización y naturaleza de los factores genéticos implicados. El desarrollo de una colección de líneas de introgresión es un importante paso

cuantitativo en el estudio de la variabilidad natural del germoplasma de *Fragaria*. Para el desarrollo de este tipo de poblaciones, es de vital importancia disponer de un mapa de ligamiento de alta resolución. Esta investigación incluye además el desarrollo de herramientas que permiten el aislamiento, clonaje y 'disección' de dichos factores, que sin duda permitirá conocer los mecanismos moleculares implicados en los procesos biológicos de mayor relevancia en la mejora de las variedades comerciales actuales.

OBJETIVOS

El objetivo principal de esta Tesis Doctoral es la creación de una colección de líneas casi isogénicas (NILs) que cubran todo el genoma de *Fragaria* diploide y que permitan localizar la posición de los loci responsables de las principales características agronómicas para el cultivo de fresa. Para desarrollar esta tarea, nos hemos planteado los siguientes objetivos:

Recopilar, evaluar y ampliar el conocimiento actual de la Genómica estructural, funcional y comparada del género *Fragaria*.

Mejorar el mapa genético de referencia de *Fragaria* diploide, mediante el desarrollo y localización de nuevos marcadores moleculares y localizar posibles genes candidatos para la calidad del fruto de fresa.

Iniciar el mapa físico de *Fragaria vesca* mediante el desarrollo y caracterización de una genoteca de BACs de *Fragaria vesca*.

Establecer las estrategias más adecuadas para el desarrollo de una colección de NILs en *Fragaria*.

Obtener y evaluar fenotípicamente la colección de NILs, para determinar la posición de los loci responsables de algunos caracteres interesantes a nivel agronómico.

PRIMERA PARTE

**Desarrollo y caracterización de herramientas moleculares
para el estudio del genoma de *Fragaria* diploide**

1. INTRODUCCIÓN

El contenido de esta introducción ha sido enviado a la revista
Genes, Genomes and Genomics
para su próxima publicación como capítulo del libro:
'Genomics, Transgenics, Molecular Breeding and Biotechnology of Strawberry'
Editado por Global Science Books, Ltd., UK, 2010, ISSN 1749-0383
Bajo el título:
'Structural genomic resources in *Fragaria* genus'
Julio Bonet y Amparo Monfort

La bibliografía citada en este capítulo está englobada en la bibliografía general.
La numeración de tablas y figuras no se corresponde con la versión pública, sino que responde al orden
establecido en esta Tesis.

1.1 RESUMEN

En este capítulo se recogen los principales avances que han tenido lugar en el campo de la Biotecnología y la Genómica estructural y comparada de la fresa durante los últimos años, así como las perspectivas para el futuro más próximo. A pesar de que el conocimiento de la estructura genómica del género ha avanzado bastante, poco se sabe sobre la función de la mayoría de genes y del control que estos ejercen sobre los caracteres agronómicos más interesantes, de cara a la implantación de sistemas de selección asistida por marcadores (MAS). Así como otras especies de la familia *Rosaceae* ya cuentan con programas de mejora que incluyen la caracterización genómica, la poliploidía de la fresa comercial ha supuesto un escollo que debería superarse en los próximos años con la publicación de la secuencia completa del genoma diploide.

El conocimiento de la Genómica estructural de las especies de *Fragaria* ha crecido de manera exponencial durante los últimos 30 años. La irrupción de las tecnologías de Genética Molecular al final de la década de los 70, cuando se publicaron los primeros ensayos isoenzimáticos en poblaciones naturales de fresa, posibilitó la aparición de diferentes tipos de marcadores moleculares que permitieron la caracterización del genoma diploide primero, y el octoploide después, así como la asociación de algunos caracteres cualitativos y cuantitativos a marcadores moleculares e isoenzimas. La naturaleza de la resistencia a las principales plagas de la fresa, causadas por los hongos *Colletotrichum acutatum* y *Phytophthora fragariae*, la identificación de variedades estrechamente relacionadas o aspectos fisiológicos como la floración perpetua o el dimorfismo sexual, han sido analizados por diferentes autores desde una perspectiva genética. Estos estudios, junto con otros de caracterización de secuencias no codificantes, han permitido ir conociendo la estructura de un genoma corto y compacto, el diploide, y abrir las puertas al estudio de un genoma extremadamente complejo, el octoploide, que es esencialmente colinear y sinténico al diploide.

1.2 MOLECULAR MARKERS IN *FRAGARIA*

Since 1981, advances in structural genomics in *Fragaria* has been largely due to the improvement and availability of different types of molecular markers in *Fragaria* species, as well as markers developed in other *Rosaceae* species and transferred to *Fragaria*. Molecular markers arise from DNA mutations generated spontaneously, usually located in non-coding regions, and their pattern of inheritance is valuable for applications such as breeding, fingerprinting and genome annotation. Substitutions on a single position, larger rearrangements (insertions or deletions) or number variations on tandem-repeated DNA sequences generate marks or labels that can be easily identified and tracked by DNA visualization assays.

To select an appropriate marker, their different features should be taken into account. The stability of the mutation, as the euchromatic/heterochromatic position within the genome, critically determines the polymorphism rates. As EST- (Expressed Sequence Tags-) derived markers are usually well conserved in different taxa, they are a good choice for syntenic or phylogenetic studies, while genomic-derived markers offer more polymorphic loci between closely related individuals within a species. Genomic data or the *in silico* genomic resources of the species, accessing high technology, are crucial for the election of a marker, and while no sequence knowledge is necessary for arbitrarily-primed markers, it is needed for SSR, SNP, CAPS and STS markers.

1.2.1 ISOENZYMES

Isoenzymes are different molecular forms of a particular enzyme. They can be electrophoretically separated due to the specific migration pattern generated by the different electric charge and molecular weight. Hancock and Bringham (1978, 1979) published the first studies of variation in phosphoglucoisomerase (PGI) and peroxidase (PX) enzymes in natural populations of *F. vesca*, *F. chiloensis* and *F. virginiana*. PGI, leucine aminopeptidase (LAP) and phosphoglucomutase (PGM) variation were first employed in the characterization of diploid and cultivated strawberry cultivars (Arulsekhar *et al.* 1981, Bringham *et al.* 1981). In their study on the inheritance pattern of PGI and LAP enzymes in cultivated strawberry, Arulsekhar *et al.* (1981) showed for the first time that four active *Pgi* loci segregate following a disomic Mendelian pattern. Apart from some controversial studies, the diploidized nature of the octoploid *Fragaria* genome has been accepted by the scientific community (Ahmadi *et al.* 1990, Lerceteau-Köhler *et al.* 2003, Sargent *et al.* 2009b).

The association of a particular trait and an isoenzyme locus has never been observed in cultivated strawberry. Williamson *et al.* (1995) reported the inheritance of the yellow fruit colour and its linkage to the *Sdh* (shikimate dehydrogenase) locus in an intraspecific F₂ population of *F. vesca* 'Baron Solemancher' x *F. Vesca* 'Yellow Wonder', where SDH segregated in a 1:2:1 ratio, and fruit colour segregated 3 red/1 yellow. A few months later, Yu and Davis (1995) identified an isozyme locus weakly linked to the non-runnering trait using the same diploid varieties as Williamson *et al.* (1995) as parental lines for several F₂ and F₃ offspring. Both traits were mapped by Davis and Yu (1997) in the first linkage map of the diploid strawberry genome.

1.2.2 RFLP

With the advent of restriction enzyme technology (Smith, 1970), nucleotide sequence specific systems were developed worldwide. The first report on genetic linkage of restriction fragments in plants was done by Polans *et al.* in 1985. A robust, reliable and transferable technique, restriction fragment length polymorphism (RFLP) provides mostly co-dominant polymorphisms. It is not commonly used today, as it is time consuming, involves expensive, radioactive or toxic reagents and requires large quantity of high quality genomic DNA (Agarwal *et al.* 2008). Hybridization-based markers such as RFLP, have the advantage of being highly transferable between distantly related species.

No RFLP probes have been developed in the *Fragaria* nuclear genome so far, but *Prunus* derived probes have been shown to be a valuable tool for mapping in *Fragaria* (Viruel *et al.* 2002). Vilanova *et al.* (2008) mapped 40 polymorphic *Prunus* EST-derived RFLP markers in the diploid *Fragaria* reference population when comparing strawberry and peach genomic structure. This comparison comprised 71 common markers and revealed that only 24% of the probes gave poor or no hybridization in the *Fragaria* genome, suggesting that *Prunus* coding sequences are essentially conserved in the strawberry genome.

1.2.3 ARBITRARILY-PRIMED MARKERS

With the establishment of PCR procedures in the late 1980s, PCR-derived techniques dramatically increased the availability of molecular markers, largely due to the simplicity of the procedures; nanogram (rather than microgram) quantities of DNA are required. Arbitrarily-primed marker systems such as random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphisms (AFLP) normally generate a large number of dominant markers. They are based on

different amplification patterns of small PCR fragments generated by short primer pair combinations. No knowledge of the DNA sequence for the targeted gene or genomic region is required, as the primers bind somewhere in the sequence, and therefore markers can be quickly amplified and detected in any organism. However, reproducibility is sometimes a problem with these markers, in particular RAPDs (Jones *et al.* 1997), and they are extremely difficult to transfer between different populations. DNA amplification fingerprinting (DAF) (Bentley & Bassam, 1996) and randomly amplified DNA fingerprinting (RAF) (Waldron *et al.* 2002) techniques, with higher annealing temperatures used in the PCR, seem to be more reproducible and transferable than RAPDs.

Despite these difficulties, RAPD markers were the method of choice for the identification of strawberry varieties (Congiu *et al.* 2000), germplasm characterization (Kuras *et al.* 2004) and construction of linkage maps in diploid and octoploid species of *Fragaria* (Davis and Yu 1997; Sugimoto *et al.* 2005; Sargent *et al.* 2009b). Linkage between RAPD markers and a trait in the cultivated strawberry was first reported using bulked-segregant analysis (BSA) of individuals resistant and susceptible to *Phytophthora fragariae* (Haymes *et al.* 1997). The authors identified seven RAPD markers linked to the *Rpf1* locus, which is one of the major sources of genetic resistance to red stele root rot, caused by the soil-born fungus *P. fragariae*. Two RAPD markers were also found to be linked to the everbearing locus in octoploid strawberry in the first map-based association study in *F x ananassa* (Sugimoto *et al.* 2005).

The major advantage of AFLP over the earlier arbitrarily-primed PCR protocols is that they generate a large number of segregating markers (Vos *et al.* 1995) mostly corresponding to unique positions on the genome so they can be exploited as landmarks in genetic and physical mapping. AFLP analysis is usually preferred for increasing marker density or to identify additional markers linked to specific chromosomal regions by BSA (Kamei *et al.* 2010; Capomaccio *et al.* 2009). This technique was applied for the first time in *Fragaria* by Lerceteau-Köhler *et al.* in 2003, during the construction of the first public linkage map on the cultivated strawberry *Fragaria x ananassa*, and subsequently in related studies (Weebadde *et al.* 2008; Rousseau-Gueutin *et al.* 2008; Sargent *et al.* 2009b).

ISSR (inter-simple sequence repeat) is a general term for a genome region between two adjacent microsatellite loci. These regions are often polymorphic between different species or varieties, and have been used to generate arbitrary primers. These primers are designed containing different SSR motifs, and PCR products result from the variable amplified region. The result is a mix of a variety of amplified DNA strands which are generally short but variable in length. The small genome size of the

diploid *Fragaria* has led to the identification of variable ISSR regions (Korbin *et al.* 2002, Albani *et al.* 2004).

1.2.4 SCAR

Arbitrary markers closely linked to agronomically interesting traits can be converted to sequence characterised amplified regions (SCAR) for large scale use (Paran and Michaelmore, 1993). To date, only three economically important traits have been successfully linked to a SCAR marker in *Fragaria*; SCAR markers are difficult to develop. However, published functional SCAR markers designed on gene sequences coding for different enzymes in other species are being transferred to *Fragaria* (Palmieri *et al.* 2008). Haymes *et al.* (2000) successfully converted a RAPD marker into two SCARs, both linked in coupling phase to the *Rpf1* gene. This association was shown to be highly conserved among 133 *Fragaria* European and North American cultivars. Lerceteau-Köhler *et al.* (2005) studied the inheritance of resistance to one of the infectious agents causing anthracnose, *Colletotrichum acutatum* in 43 strawberry cultivars. They obtained two AFLP-derived SCAR markers, closely linked to the *Rca2* gene, which correctly predicted most of the resistant/susceptible genotypes (81.4% of the cultivars for STS-Rca2_417). In diploid *Fragaria*, three SCARs linked to the *SFL* locus (seasonal flowering locus) have been reported in *F. vesca* by Albani *et al.* (2004) using a backcross progeny of 1,049 individuals. This locus is considered to be a floral repressor that is inactivated by short days and cool temperatures in the autumn but reactivated by winter cold (Battey *et al.* 1998).

1.2.5 CAPS

Cleaved amplified polymorphic sequences (CAPS) are usually amplified from known gene sequences and digested with an appropriate restriction enzyme (RE). CAPS reveal the restriction fragment length polymorphisms caused by single base changes of SNPs, insertions/deletions, which modify restriction endonuclease recognition sites in PCR amplicons (Konieczny and Ausubel, 1993). The technique is limited by mutations, which create or disrupt a restriction enzyme recognition site. To overcome this limitation, Michaels and Amasino (1998) proposed a variant of the CAPS method called dCAPS (derived cleaved amplified polymorphic sequence). In dCAPS analysis, a restriction enzyme recognition site which includes the SNP is introduced into the PCR product by a primer containing one or more mismatches to template DNA (Neff *et al.* 1998). The modified PCR product is then digested with a restriction enzyme, with the resulting restriction pattern showing the presence or absence of the mutation.

CAPS and dCAPS analyses are fast and simple for SNP genotyping compared to more expensive techniques such as resequencing or high resolution melting analysis (HRM). HRM analysis discriminates two homozygous PCR amplicons containing a nucleotide variation, as well as their heterozygote, using fluorescence dyes which can bind to the double stranded DNA without inhibiting PCR (Herrmann *et al.* 2006). This approach has been used in other *Rosaceae* crops as almond (Wu *et al.* 2008, 2009) and apple (Chagné *et al.* 2008), but no studies have been reported in strawberry.

Kunihisa *et al.* (2005, 2009) used cluster-specific amplification with a set of 34 CAPS markers in *F. x ananassa*, corresponding to alleles of single diploid loci, and correctly identified 63 different cultivars using a minimum of nine CAPS markers. These markers are especially valuable, as markers detecting only a single locus are rare in *F. x ananassa*. These CAPS are cluster-specific markers, as they were developed by allele-specific amplification based on the nucleotide sequence specificity, allowing the amplification of only one of the four clusters or subgenomes in the whole *F. x ananassa* genome.

A set of new CAPS and dCAPS have recently been reported in the diploid strawberry *F. vesca* Hawaii 4 (Ruiz-Rojas *et al.* 2010). Seventy four SNPs were identified and mapped on the *Fragaria* reference map, in order to locate the T-DNA insertions generated by *Agrobacterium*-mediated mutagenesis. When a single dose transgenic plant was identified, thermal asymmetric interlaced PCR (hiTAIL-PCR) was used to obtain the insertion flanking regions. These regions were used to design primers for the identification of polymorphic SNPs in the parents of the *Fragaria* reference population and subsequently mapped in the population using a CAPS-dCAPS approach.

1.2.6 MICROSATELLITES

Simple sequence repeats (SSR) or microsatellites are tandemly arranged short sequence motifs, which occur as interspersed repetitive elements in all eukaryotic genomes (Tautz & Renz, 1984). Microsatellites are among the most variable types of DNA sequence in the genome (Weber, 1990), as the polymorphisms derive mainly from variability in length rather than in the primary sequence. Given their extensive polymorphisms, transferability between populations and species and their technical simplicity, they became the marker of choice in *Fragaria* mapping during the last decade and subsequently in population genetics studies and cultivar identification.

A set of more than 200 primer pairs amplifying polymorphic SSRs have been developed for *Fragaria* species (Table 2). Most of them have been reported in the diploid species such as *F. vesca*, *F. viridis* and *F. bucharica* (formerly *F. nubicola* (Sargent *et al.* 2006)) and also in the octoploid species *F.*

virginiana, the cultivated strawberry *F. x ananassa*, and in other rosaceous species. Whilst SSRs are excellent markers for providing a linkage framework, they are primarily developed from non-coding regions of the genome and are generally not tightly linked to genes of known function or to traits of economic importance.

In contrast to genomic SSRs, EST-derived microsatellites potentially represent functional markers (Varshney *et al.* 2005; Zorrilla-Fontanesi *et al.* 2010a) as they are present in expressed regions of the genome. Around 30% of the reported SSRs in *Fragaria* belongs to expressed regions, being especially valuable for germplasm characterization (Varshney *et al.* 2005) taking into account that EST-SSRs are not as efficient as genomic SSRs for distinguishing closely related genotypes (Gupta & Varshney, 2000). Besides genetic mapping and diversity studies, *Fragaria* microsatellites have been used for transferability and comparative mapping in the *Rosaceae* (Lewers *et al.* 2005), characterization of a high molecular weight DNA BAC library (Bonet *et al.* 2009) and the ongoing scaffold anchoring for the *Fragaria* genome sequence annotation project (Shulaev *et al.* personal communication).

Table 2. Microsatellite markers developed in *Fragaria* species

<i>Fragaria</i> species	Name	Number of SSRs	Reference
<i>F. vesca</i>	EMFv	31	James <i>et al.</i> 2003, Hadonou <i>et al.</i> 2004
<i>F. vesca</i>	UDF	68	Cipriani & Testolin 2004, Cipriani <i>et al.</i> 2006
<i>F. vesca</i>	CFVCT	35	Monfort <i>et al.</i> 2006
<i>F. vesca</i>	ChFvM	22	Zorrilla-Fontanesi <i>et al.</i> 2010
<i>F. viridis</i>	EMFvi	22	Sargent <i>et al.</i> 2003
<i>F. bucharica</i>	EMFn	31	Sargent <i>et al.</i> 2006
<i>F. virginiana</i>	Fvi	7	Ashley <i>et al.</i> 2003
<i>F x ananassa</i>	ARSFL, FAC	14,4	Lewers <i>et al.</i> 2005
<i>F x ananassa</i>	UFFa	14	Sargent <i>et al.</i> 2006
<i>F x ananassa</i>	CFACT, CFA, EMFxa	24,5, 8	Sargent <i>et al.</i> 2008
<i>F x ananassa</i>	UAFv	14	Bassil <i>et al.</i> 2006
<i>F x ananassa</i>	ChFaM	108	Gil-Ariza <i>et al.</i> 2006, Zorrilla-Fontanesi <i>et al.</i> 2010
<i>F x ananassa</i>	Fa	4	Shimomura & Hirashima 2006
<i>F x ananassa</i>	PBCESSRFXA	14	Keniry <i>et al.</i> 2006
TOTAL		425	

1.2.7 GENE-SPECIFIC MARKERS

In spite of the availability of expressed sequences in public databases, no gene-specific markers have yet been mapped in cultivated *Fragaria*. This is probably due to the octoploid nature of the genome, as the gene dose can interfere during the characterization of an expressed, and subsequently, highly conserved locus. For this reason, the intron-exon border structure, based on sequence similarity with the most highly conserved homologue gene in a model species, can be taken into account. Coding regions flanked by introns are a good choice for primer design of polymorphic sequence-specific tags

(STS), as the mutation rates are higher in the introns than in the exons. Because cultivated strawberries are octoploid, the fragments amplified by one primer pair can contain various sequences derived from up to eight chromosomes (Kim *et al.* 2001). Therefore, it is necessary to consider several sequences from each target gene to correctly predict polymorphisms from the sequence data (Kunihisa *et al.* 2005).

Most studies on gene discovery have used the diploid species *Fragaria vesca*. The first gene specific marker in *Fragaria*, *ADH*, was developed by Davis and Yu (1997) from the alcohol dehydrogenase gene. They used an intron length polymorphism segregating in their mapping population. Deng and Davis (2001) developed a candidate-gene approach to determine the molecular identity of the *F. vesca* *c* (yellow fruit) locus. They mapped intron-length polymorphic markers for 5 candidate genes and a transcription factor involved in anthocyanin biosynthesis pathways, using two diploid *Fragaria* populations. Sargent *et al.* (2007) located 29 loci for 24 genes in the diploid *Fragaria* genome using public EST databases, predominantly from *Fragaria* mRNA sequences, as a source for primer design. Many of their primer pairs were shown to be highly transferable, amplifying orthologous genes in other genera (*Malus* and *Prunus*). EST-derived markers have also been useful in synteny analysis between *Prunus* and *Fragaria*, where 21 new markers have been developed from *Prunus* EST sequences, and mapped in the diploid reference *Fragaria* linkage map (Vilanova *et al.* 2008).

Four more STS markers were mapped by Sargent *et al.* (2008) during the development of a bin mapping strategy. With the addition of two unlinked RNase loci (loci *S* and *T*) to the linkage groups FG1 and FG6, both involved in the *Fragaria* RNase-based self-incompatibility system (Bošković *et al.* 2010), the number of genes of known function in the *Fragaria* genome was brought to 38 (table 3).

Following a variant methodology, Davis *et al.* (2006) developed a new molecular marker type in *Fragaria*, called gene pair marker. As the genome size of the diploid strawberry is small (about 200 Mbp), genes are closely spaced: about 1 gene per 5.7kb (Pontaroli *et al.* 2009). As a consequence of the small intergenic distances, PCR primers targeted to conserved exon sequences in two adjacent genes (the gene pair) can be exploited as a source of polymorphisms, by amplifying an entire intergenic region flanked by gene-defining exon sequences. The authors used this strategy to amplify the *CDPK1-BHLH1* strawberry gene pair locus in other taxa, suggesting that gene pair markers can be especially useful in intergenic synteny studies or comparative linkage mapping in the *Rosaceae* family (Sargent 2009a).

Table 3. Locus name, gene annotation and map position of mapped genes in the *Fragaria* genome. Gene locations are given in cM with respect to the origin of the LG for FV x FB, and LG or flanking markers are reported for the octoploid *Fragaria* genes. * The reported position belongs to a previous version of the map.

¹ Sargent *et al.* 2007. ² Davis and Yu 1997. ³ Deng and Davis 2001. ⁴ Sargent *et al.* 2004a. ⁵ Rousseau-Gueutin *et al.* 2009. ⁶ Boškovic *et al.* 2010. ⁷ Sugimoto *et al.* 2005. ⁸ Haymes *et al.* 1997. ⁹ Sargent *et al.* 2009b.

MAP	LOCUS	GENE ANNOTATION	POSITION
	ABP ¹	Auxin-binding protein	LG2 21,5
	ACO ¹	1-Aminocyclopropane-1-Carboxylate oxidase	LG6 58,2
	ADH ²	alcohol dehydrogenase	LG2 16,9
	AKR ¹	Aldo-keto reductase superfamily member	LG4 61
	ANS ^{1,3}	Anthocyanidin synthase	LG5 9,2
	APX ¹	Cytosolic ascorbate peroxidase	LG3 42,3
	ARP ¹	Auxin-repressed mRNA	LG3 28
	BG-1 ¹	Beta-1,3-glucanase-1	LG4 60,4
	BG-2 ¹	Beta-1,3-glucanase-2	LG3 39,8
	CAD-1 ¹	Cinnamyl alcohol dehydrogenase-1	LG7 17
	CAD-2 ¹	Cinnamyl alcohol dehydrogenase-2	LG1 31,1
	CAD-3 ¹	Cinnamyl alcohol dehydrogenase-3	LG1 54,2
	CEL-1 ¹	Cellulase-1	LG4 75,3
	CEL-2 ¹	Cellulase-2	LG5 20,3
	CHI ³	chalcone isomerase	LG7 25,7
	CHS ³	chalcone synthase	LG7 17,3
	DFR ^{1,3}	Dihydroflavonol 4-reductase	LG2 59,1
	DHAR ⁵	DeHydro Ascorbate Reductase	LG7 0*
	EKO ¹	Ent-kaurene oxidase	LG2 14,8
FV x FB (Ruiz-Rojas <i>et al.</i> 2010)	EXP-1 ¹	Expansin-1	LG4 19,0*
	EXP-2 ¹	Expansin-2	LG3 54,5*
	EXP-3 ¹	Expansin-3	LG7 33,3*
	F3H ³	flavanone-3-hydroxylase	LG1 43,1
	FvNES1 ⁴	Nerolidol Synthase 1	LG3 12,5
	GAST ¹	GAST-like gene product	LG2 52,5*
	LEAFY ¹	Leafy protein	LG3 16,1*
	LOX ¹	Lipoxygenase	LG4 27,8
	MET ¹	Metallothionein-like protein	LG6 42
	MSR ¹	Methionine sulfoxide reductase	LG5 27,5*
	MYB ¹	Transcription factor	LG5 38
	PDC ¹	Pyruvate decarboxylase	LG6 43,8
	PES ¹	Pectinesterase	LG1 37,3*
	PGLM ¹	Phosphoglyceromutase	LG6 28,4
	QR ¹	Quinone oxidoreductase	LG6 76,1
	RAN ³	Del-like regulatory gene	LG5 14,8
	S Locus ⁶	S-RNase	LG1 36
	T Locus ⁶	T-Rnase	LG6 (7,3-15,2)
	ZIP ¹	Zinc transporter protein	LG6 52,4
Ever Berry x Toyonoka (Sugimoto <i>et al.</i> 2005)	EV ⁷	Everbearing	(OPE07-1 - OPB05-1)
Md683 x Senga Sengana (Haymes <i>et al.</i> 1997)	Rpf1 ⁸	<i>Phytophthora fragariae</i> resistance	(OPO-08A - OPO-16A)
	ACO ⁹	1-Aminocyclopropane-1-Carboxylate oxidase	RG6-D 22, HA6-D
	ACP ⁹	Acyl carrier protein	HA6-D
	DFR ⁹	Dihydroflavonol 4-reductase	RG2-B
Redgauntlet x Hapil (Sargent <i>et al.</i> 2009)	bEXP-204 ⁹	Expansin	HA3-A
	EXP-185 ⁹	Expansin	RG4-D
	aEXP-177 ⁹	Expansin	RG5-A, HA5-A
	b-EXP-235 ⁹	Expansin	HA7-C
	EXP-248 ⁹	Expansin	HA7-D, RG7-D
	aEXP-245 ⁹	Expansin	RG7-C
	aLOX-374/385 ⁹	Lipoxygenase	RG4-B
	bLOX-376/379 ⁹	Lipoxygenase	HA4-B

1.3 LINKAGE MAPPING IN *FRAGARIA*

The construction of linkage maps in polyploids has several limitations: There are a large number of possible genotypes for each polymorphism expected in a segregating population, and these genotypes cannot always be identified readily by their banding phenotypes due to possible co-migration of fragments on gel electrophoresis. Furthermore, the genome constitutions (allopolyploidy versus autopolyploidy) in many polyploids are not clearly understood, making it difficult to determine the patterns of inheritance. The development of genetic maps in octoploid *Fragaria* species has lagged behind that of diploid species, where meiotic recombination events can be easily characterized and, consequently, the order of segregating loci can be clearly established. In the framework of the assessment of the synteny and colinearity between both diploid and octoploid genomes, more genetic maps have been developed using diploid species, since Davis and Yu (1997) developed the first genetic map for *Fragaria*. This study, using an F₂ mapping population derived from a cross between an alpine (*F. vesca* ssp. *vesca* 'Baron Solemacher') and a non-alpine (*F. vesca* ssp. *americana* 'WC6') parent, provided the position of 80 loci, most of them RAPD markers, in a 445 cM genetic map. In the map there is also one morphological marker (the runner locus, *r*), one STS marker for the *ADH* gene and the isoenzymes *Pgi-2* and *Sdh-1*. Some of the markers developed in this work were used by Deng and Davis (2001) to anchor five genes and one transcription factor involved in anthocyanin biosynthesis pathways to their hosting linkage group, using two F₂ populations of 40 individuals each, derived from two crosses of *F. vesca* ssp. *vesca*, var. 'Yellow Wonder' with *F. nubicola*, and *F. vesca* ssp. *bracteata* with *F. vesca* ssp. *vesca*, var. 'Yellow Wonder'.

1.3.1 THE *FRAGARIA* REFERENCE MAP FVxFB

With the advances in the production of transferable molecular markers for map-based studies of the genus, such as SSR, in both diploid and octoploid species, Sargent *et al.* (2004a) constructed a reference linkage map for the genus, from a cross between *F. vesca* ssp. *vesca* f. *semperflorens* (FDP815) and *F. bucharica* (FDP601). The latter accession, commonly referred to as *F. nubicola* FDP601 (Lin & Davis, 2000; Potter *et al.* 2000; Deng & Davis, 2001; Sargent *et al.* 2004a, 2004b, 2006, 2007, 2008; Davis *et al.* 2006, Monfort *et al.* 2006, Vilanova *et al.* 2008), was recently reclassified as *F. bucharica*, a closely related species, probably of hybrid origin (Staudt, 2006). The map, previously designated as FVxFN, was renamed FVxFB. In its initial version, this map had 68 SSR, one SCAR, six STS markers and three morphological traits: seasonal/perpetual flowering, runner development and pale/dark-green leaf colour.

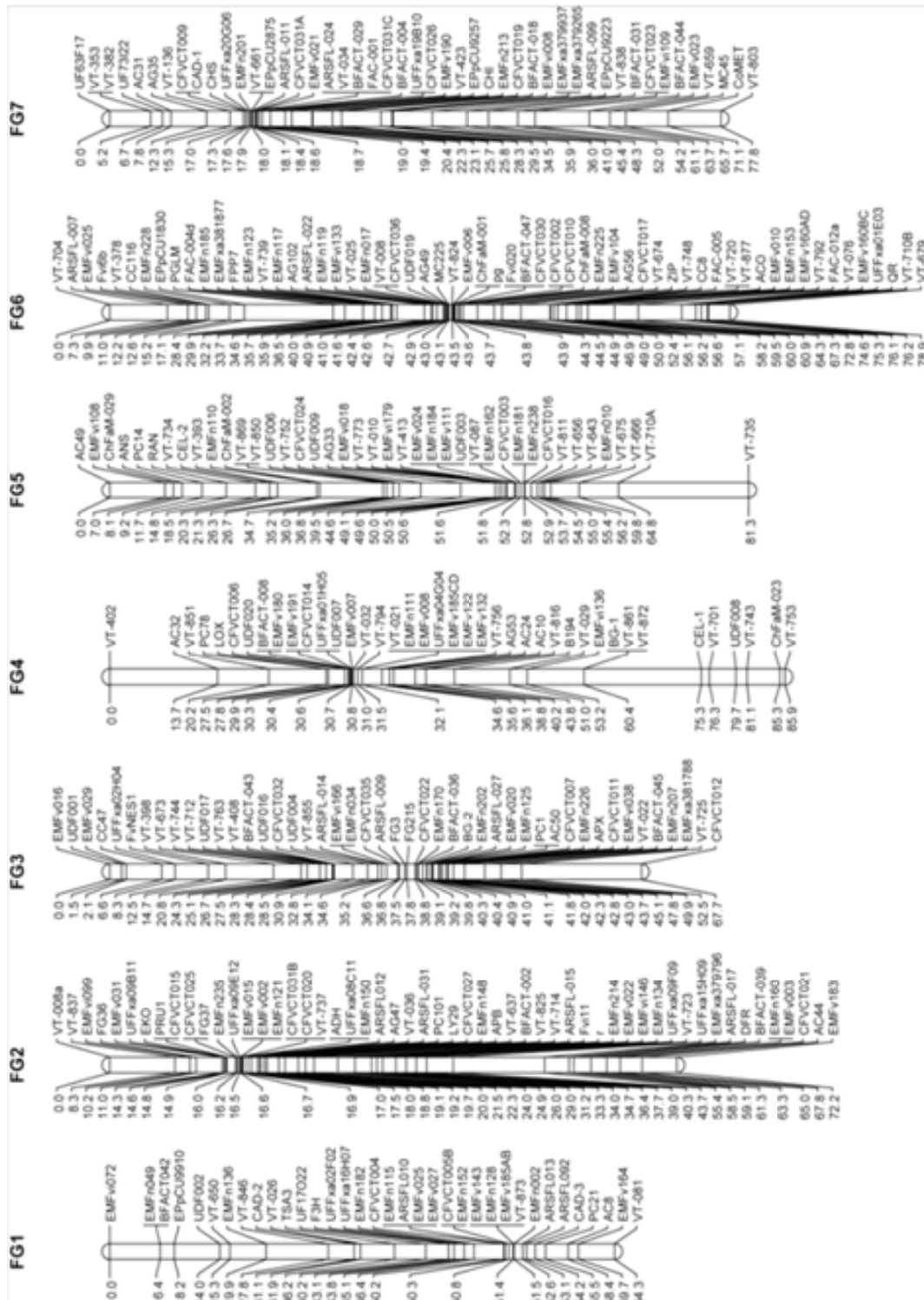


Figure 5 The diploid *Fragaria* reference map **FV x FB** in its last version (Ruiz-Rojas et al. 2010), derived from an F_2 population obtained in an interspecific *F. vesca* 815 x *F. bucharica* cross.

Even considering the high level of distorted segregation ratios (54%, where 80% of skewed loci were biased towards an excess of alleles from the paternal crossing parent, *F. bucharica*), the map was adopted as a framework for additional marker incorporation for map coverage of the diploid *Fragaria* genome. In a collaboration of several laboratories, a new set of 109 genomic and EST-derived SSR markers, developed in both diploid and cultivated strawberries, were located in this map by Sargent *et al.* in 2006, increasing the number of molecular markers to 182 (175 SSR, 6 STS and one SCAR) to provide a comprehensive coverage of the diploid *Fragaria* genome with an average density of one marker every 2.3 cM. In 2007, Sargent *et al.* developed and mapped 29 functional markers from 24 gene sequences, greatly increasing the number of genes of known function mapped in the genus. Seventeen STS were mapped using the FVxFB population, and the remaining 12 were mapped in a BC1 population derived from a cross between *F. vesca* 815 and *F. viridis* 903 (Nier *et al.* 2006).

In 2008, 40 RFLP and 29 EST-derived markers were located in the FVxFB map, in a comparative study of *Prunus* and *Fragaria* genomes (Vilanova *et al.* 2008) and a new SSR, located in an uncovered region of linkage group four, was added to the reference map during the development of a 'bin mapping' strategy for *Fragaria* (Sargent *et al.* 2008). In a phylogenetic study on the relationships among *Fragaria* species, Rousseau-Gueutin *et al.* (2008) described a new STS marker for the *DHAR* gene. Rousseau-Gueutin *et al.* (2009) added four new SSR loci to the reference map and included some rearrangements in the relative position of several markers from four linkage groups during the assessment of the synteny relationships between diploid and octoploid genomes.

Two loci encoding RNases (*S* and *T*) involved in the self-incompatibility mechanism have been described and mapped in the FVxFB map (Bošković *et al.* 2010). This mechanism includes at least two RNases and a complementary SFB pollen-expressed component. The authors showed that the presence of one of the active alleles at either RNase loci is sufficient to confer self-incompatibility in an appropriate genetic background and that a non-RNase factor, which segregates independently, is necessary for self-incompatibility.

Recently, a set of 74 CAPS and dCAPS markers derived from a T-DNA mutant insertion collection has been developed and incorporated in the FVxFB map (Ruiz-Rojas *et al.* 2010). With the inclusion of these markers, the *Fragaria* reference map (fig. 5) coverage has been brought to 529 cM, 18% more than that of Davis and Yu (1997), with 310 molecular markers (172 SSR, 16 gene-specific STS, 40 RFLP, 74 CAPS-dCAPS, seven EST and one SCAR) and three morphological traits distributed in seven linkage groups.

1.3.2 BIN MAPPING IN *FRAGARIA*

Normally, linkage mapping populations consist of randomly selected individuals of a large progeny, where the distribution of the meiotic crossover sites (breakpoints) in the genome of the different individuals of the population provides a framework for the assessment of the molecular marker positions. However, when a typical mapping population is well characterized, it is possible to add new markers by selecting a few plants of the population (the most informative set of breakpoints), minimizing time and effort over conventional mapping strategy (Vision *et al.* 2000; Howad *et al.* 2005). This 'selective' or 'bin mapping' approach was recently adapted to the FVxFB map by Sargent *et al.* (2008). The defined bin set, with 6 F₂ individuals plus one parental line and the F₁ individual, was shown to be sufficiently robust to locate 103 new markers (99 genomic and EST-derived SSRs and four gene-specific markers originally described as *Prunus* ripening genes) in one of the 46 mapping bins spanning the *Fragaria* genome, with an average length of 12.6 cM per bin.

1.3.3 OCTOPLOID MAPPING

Breeding for many traits in the cultivated strawberry could be enhanced through the use of molecular markers, but the octoploid nature of the genome provides a challenge to the development of traditional molecular breeding tools. Until 2008, the absence of a public consensus linkage map meant marker assisted selection in strawberry was delayed with respect to other crops, with only reports of partial maps or markers linked to single traits. With meiotic behaviour clearly established and multi-allelic transferrable markers anchored to the octoploid and diploid *Fragaria* genome (Rousseau-Gueutin *et al.* 2008; Spigler *et al.* 2008; Sargent *et al.* 2009b), the assessment of accurate trait-marker linkage relationships should be feasible in the near future.

The first map-based association study in cultivated strawberry was carried out to identify markers for the everbearing locus in the octoploid strawberry (Sugimoto *et al.* 2005). An F₁ progeny from a cross between 'Ever Berry' (a Japanese everbearing variety) and 'Toyonoka' (a Junebearing variety) was used to identify the locus governing the flowering pattern. The flowering test gave the expected 1:1 ratio for everbears to Junebears, suggesting that the inheritance of the trait is controlled by a monogenic dominant gene. In order to locate this gene, the authors used 175 RAPD primer pairs. With five out of 89 polymorphic fragments relating to the everbearing locus, a 39.7 cM long linkage group was constructed, with two markers flanking the everbearing gene, at 11.8 and 15.8 cM.

In the first linkage map of the cultivated strawberry, derived from an F_1 population of 113 individuals from the cross of the variety 'Capitola' and the clone CF1116 ['Pajaro' x ('Earliglow' x 'Chandler')], using AFLP analyses, Lerceteau-Köller *et al.* (2003) applied a two-step mapping procedure, previously reported by Grivet *et al.* (1996), to assign the linkage between the coupling- and repulsion-phase markers. The authors identified a large cluster of markers segregating only in the coupling phase, and deduced that the meiotic behaviour of the octoploid *Fragaria* genome may not be as completely diploidized as assumed by several authors (Arulsekhar *et al.* 1981b; Ahmadi *et al.* 1990; Ashley *et al.* 2003). Their maternal map had 235 markers distributed among 30 linkage groups covering 1,604 cM and the paternal map was built with 280 markers assigned to 28 linkage groups, yielding a map size of 1,496 cM.

Transferable markers and new individuals were later added to this map (Rousseau-Gueutin *et al.* 2008), allowing the integration of most of the paternal and maternal linkage groups of Lerceteau-Köller *et al.* (2003) and demonstrating that the meiotic behaviour of the strawberry genome is mainly disomic. However, the absence of repulsion-phase markers in some linkage groups can only be explained by the existence of residual polysomic pairing that may still occur within the *F x ananassa* genome (Sargent *et al.* 2009a). In this version, this map covers a genetic distance of 5,017 cM, defined by 162 transferable loci across both maps, 23 of which, shared between the two maps, were used for map integration.

To identify the genetic control of day-neutrality in the commercial strawberry, an AFLP-based map was constructed from a cross between 'Tribute', a day-neutral, and 'Honeoye', a short-day variety. In an F_1 population of 127 individuals, 429 single dose restriction fragments (SDRFs) were located on an integrated map of 1,541 cM with 43 linkage groups. SDRFs are particularly useful for constructing genetic maps of polyploids when meiosis behaviour is unclear (Wu *et al.* 1992). Eight significant QTLs were identified, none of which explained more than 36% of the variation, as expected for a polygenic-inherited trait (Weebadde *et al.* 2008).

The genetic mechanisms involved in sex determination were also investigated by a map-based association approach in the wild octoploid *F. virginiana* (Spigler *et al.* 2008). Using SSRs, the authors mapped sex determination as two qualitative traits, male and female function, both located in linkage group 41, separated by 6 cM. The existence of recombination events between the sex-determining loci, an important hallmark of incipient sex chromosomes, suggested *F. virginiana* as a novel model system for the study of sex chromosome evolution. Their map covered a total distance

of 2,373 cM on 42 linkage groups, much higher than the 28 expected linkage groups, including 212 transferable markers, 39 of which segregate in both the maternal and paternal maps, and were used for map integration. Skewed segregation was apparent in almost a third of the mapped markers, and many of the primer pairs used in this study produced alleles that mapped to more than one locus on the same linkage group. The authors suggested that this may have been due to changes within and between genomes of the highly heterozygous, polyploid genome of *F. virginiana*, so representing alleles from the same genetic locus. This would reduce the number of transferable markers of the map to 161 (Sargent *et al.* 2009b).

A new integrated octoploid-diploid *Fragaria* genome map was developed by Sargent *et al.* (2009b) from an F₁ population of 174 seedlings derived from a cross between two *F. x ananassa* cultivars, 'Redgauntlet' x 'Hapil', segregating for interesting agronomic traits. The resultant map has 315 molecular markers (218 SSR, 11 gene-specific markers and 86 AFLP and RAPD markers) spanning 3,116 cM in 69 different linkage groups that could be associated to one of the 56 theoretical scaffolds by at least one anchor marker. It contains 230 transferable markers, 28 of which are shared between the two parental maps and were used for map integration.

1.4 COMPARATIVE MAPPING

Genome comparisons based on the map position of homologous sequences between closely related species can be used to construct robust phylogenies (Rokas *et al.* 2003) or to identify changes in genome structure (Deleu *et al.* 2007). When well-characterized sequence data are available, the conservation of synteny between different plant species can be detected even comparing members of different families. Genome comparisons based on the map position of homologous markers between distantly related species is quite limited (Rong *et al.* 2005), but this approach has been used to identify genome rearrangements and syntenic and extensive colinear chromosomal regions in different plant species within a family (McClellan *et al.* 2010; Wu & Tanksley, 2010). These comparisons can be very useful for breeding purposes.

Diploid *Fragaria* species seem to be essentially colinear at the macrosynteny level when comparing the available maps constructed by interspecific crosses, such as *F. vesca* x *F. bucharica* or *F. vesca* x *F. viridis* (Nier *et al.* 2006). Sargent *et al.* (2007) used PCR-based methods for comparative mapping between *Fragaria*, *Malus* and *Prunus*, indicating that *DFR* and *EKO* are orthologous genes and that their hosting linkage group (LG2) could be homeologous to *Malus* LG15 and *Prunus* LG1. This

relationship has been confirmed for *Fragaria* and *Prunus* genus by hybridization methods. *Prunus* and *Fragaria* reference maps were compared by analyzing the position of 71 anchor markers covering both genomes, and it was found that, in spite of the conservation of large chromosomal fragments as expected for confamilial species, many chromosomal rearrangements, occurring since divergence of the two genera, separate the *Fragaria* and *Prunus* genomes, supporting their distant position within the *Rosaceae* family. However, there is sufficient synteny between the genomes to allow the information on marker, gene or quantitative trait locus (QTL) position from one species to be used in studies in the other (Vilanova *et al.* 2008). Further comparisons between different *Rosaceae* genus is being carried out (Illa E, personal communication), and new DNA sequence data from high-throughput sequencing will permit a next generation of microsynteny studies in the *Rosaceae* very soon.

Using transferable markers for mapping purposes enables the identification of homologous linkage groups between maps, allowing the selection of further markers from any other study to enrich the existing maps in areas not covered. Recently, the synteny relationship between the diploid and octoploid *Fragaria* genome has been established by Rousseau-Gueutin *et al.* (2008) and Sargent *et al.* (2009b). Anchoring syntenic regions with transferrable markers (most of them SSR), all the octoploid *Fragaria* linkage groups were identified on the diploid *Fragaria* reference map, according to their homologous groups, and the homeologous groups within the octoploid genome were identified by mapping marker loci derived from the same primer pair to different linkage groups on each of the parental maps. The authors observed almost complete conservation of marker order between the octoploid and diploid genetic maps. Except for an apparent duplication of the *Fvi6b* locus on homologues of diploid *Fragaria* linkage groups one and six, which may indicate an ancient chromosomal duplication or translocation event in *Fragaria* (Sargent *et al.* 2009b), no evidence of any chromosomal rearrangements between the diploid and octoploid maps has been reported.

1.5 FUTURE PERSPECTIVES

Although the architecture of the strawberry genome has been investigated over the last thirty years, the amount of genomic data has dramatically increased since 2005, with the advent of high throughput sequencing technology. It has been used for the construction of the first draft of the strawberry genome, which will be released soon (Shulaev *et al.* submitted), marking another milestone in the extraordinarily fast moving field of plant genomics. 'Despite all the sequencing advances, very little is known about how to read the *book of life* that is opening before us', says Michael Egholm, one of the people responsible for next generation sequencing technology (Wadman, 2008).

New genomic resources in *Fragaria* are being constructed worldwide: a collection of near isogenic lines (NILs) in diploid strawberry is coming out soon (Bonet *et al.* unpublished data); a collection of T-DNA insertional mutants is available (Ruiz-Rojas *et al.* 2010); transgenic lines for many interesting traits are being investigated in different laboratories (Qin *et al.* 2008). The new resources are unexplored sources of phenotypic variation. Information on genome-wide and transcriptome-wide variation in diverse germplasms can be used in different studies, ranging from mapping interesting mutations to studying genome structure and function. Identification of alleles (including QTLs and eQTLs (expression QTLs) controlling phenotypes, genetic mechanisms such as genome-wide patterns of recombination, will benefit from the availability of dense marker maps, to give a sound basis for marker assisted selection procedures in strawberry production.

**2. DESARROLLO Y CARACTERIZACIÓN
DE MARCADORES MOLECULARES Y
MEJORA DEL MAPA DE REFERENCIA
DE *FRAGARIA* DIPLOIDE**

2.1 INTRODUCCIÓN

Los marcadores RFLP, desarrollados a principios de la década de los 80 (Botstein *et al.* 1980; Wyman y White, 1980), fueron los primeros marcadores moleculares basados en la variabilidad del ADN. Este avance supuso un extraordinario progreso en el análisis del ligamiento genético dado que, anteriormente, y a pesar de que este concepto había sido claramente establecido en la Genética clásica, su aplicación práctica quedaba limitada a un reducido grupo de especies con un escaso número de marcadores disponibles. En los últimos años, la tecnología RFLP, aun siendo considerada como de gran utilidad, ha ido perdiendo terreno frente a otros métodos basados en la reacción de PCR. La mayor parte de marcadores moleculares empleados en especies del género *Fragaria* han sido desarrollados durante la última década y por tanto forman parte de este último grupo, principalmente microsatélites o SSRs. Los microsatélites, en su mayor parte, pertenecen a regiones heterocromáticas del genoma. La mejora asistida por marcadores, así como otras aplicaciones de la Genómica estructural como las que se desarrollan en esta Tesis Doctoral, precisan de herramientas de bajo grado de complejidad metodológica y alta resolución, pues suponen trabajar con un gran número de individuos, lo que hace inviable el uso de técnicas demasiado complejas o poco resolutorias. La tecnología RFLP, que implica trabajar con material radioactivo, es además compleja en su ejecución, por lo que no parece indicada para un uso a gran escala.

En *Fragaria* diploide, tan sólo el trabajo publicado por Vilanova *et al.* (2008) se basa en marcadores tipo RFLP, si bien las sondas empleadas para la hibridación fueron desarrolladas a partir de secuencias codificantes del genoma de *Prunus*. Estos ESTs fueron localizados en el mapa FV x FB (fig. 6). La distancia media observada entre un marcador RFLP y un marcador de tipo PCR es de 5.3 cM, frente a los 2.5 cM de media que se dan en el el conjunto del mapa (fig. 6). Otros autores han atribuido este fenómeno a la naturaleza genómica de los marcadores, en función de su localización en regiones codificantes y no codificantes del genoma (Dawe, 1998). Sin embargo, no podemos asumir que los marcadores RFLP localizados por Vilanova *et al.* (2008) procedan de regiones codificantes del genoma de *Fragaria*, dado que su estudio no analiza la expresión de estos loci en fresa.

Se da la circunstancia de que muchos marcadores RFLP del mapa ocupan posiciones aisladas, generando *gaps* o regiones no cubiertas por marcadores a lo largo de los grupos de ligamiento imposibles de analizar. Es por ello que existe la necesidad de caracterizar estas regiones de FV x FB no cubiertas por marcadores. Una manera de afrontar este problema es mediante la colocalización de estos loci con otros más polimórficos, como los SSRs, que son, además, de fácil detección. Este

trabajo de colocación física entre marcadores, que tradicionalmente ha venido realizándose mediante técnicas de paseo cromosómico o *chromosome walking* empleando BACs, puede ahora realizarse *in silico*, gracias a la gran cantidad de datos genómicos de fresa de los que se dispone actualmente. En 2008, el VBI (*Virginia Bioinformatics Institute*) de la Universidad de Virginia (*Virginia Tech*) creó el Consorcio de Secuenciación del Genoma de la Fresa, liderado por el doctor Vladimir Shulaev y formado por investigadores de Genómica de fresa todo el mundo. VBI puso a disposición de este Consorcio la mayor base de datos genómicos de fresa creada hasta la fecha, generada únicamente a partir de tecnología de secuenciación de segunda generación (sistema 454 GS FLX).

En este capítulo, dicha información ha sido utilizada para realizar un estudio de colocación de secuencias con motivos SSR próximas a marcadores RFLP localizados en el mapa de referencia de *Fragaria* diploide, con objeto de generar marcadores de tipo PCR que puedan ser empleados a gran escala en posteriores experimentos en esta Tesis Doctoral y, además, ahondar en la identificación de las causas de la distribución irregular de los marcadores RFLP respecto a otros marcadores del mapa. Así mismo, hemos incluido en este estudio otros marcadores SSR que, por causas desconocidas, no se pueden resolver con normalidad en nuestro laboratorio y que, por su posición en el genoma, son de vital importancia para el desarrollo de metodologías de análisis genómico de poblaciones segregantes de alto tamaño poblacional. Para ello, se han identificado motivos microsatélite cercanos a las regiones de máxima homología con dichas secuencias. Estos marcadores han sido amplificados en el ADN de la población FV x FB, se ha analizado su segregación y han sido localizados en el mapa de referencia, junto con la de los marcadores publicados y nuevos marcadores desarrollados por otros grupos de investigación.

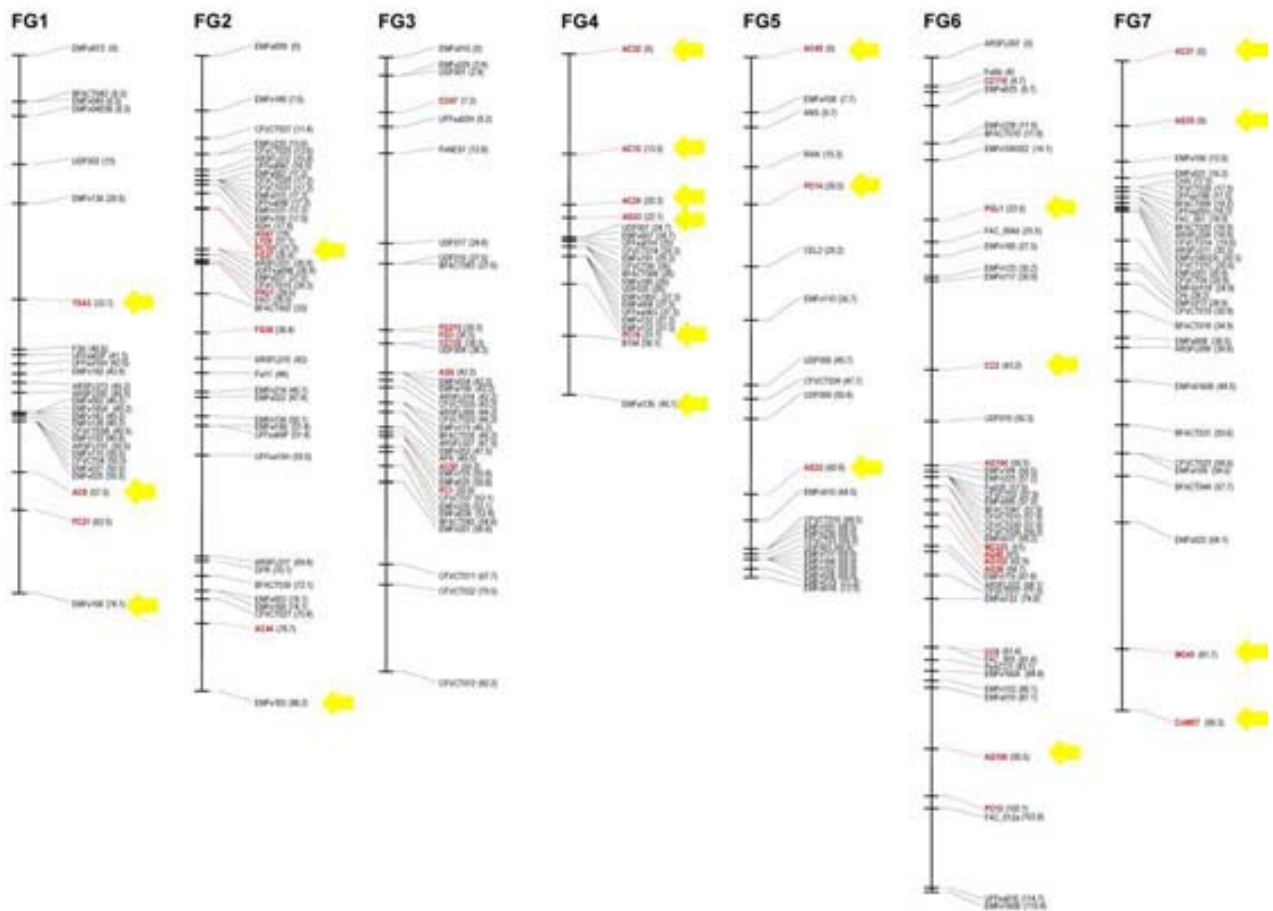


Figura 6. Mapa de ligamiento FV x FB generado por Villanova *et al.* (2008). Las líneas verticales representan los grupos de ligamiento (FG), y las marcas horizontales, la posición de los marcadores moleculares que conforman el mapa. Los marcadores que aparecen en rojo son los RFLPs que Vilanova *et al.* localizaron en su estudio. Los marcadores señalados con una flecha amarilla son los seleccionados en este estudio para su optimización.

2.2 MATERIAL Y MÉTODOS

2.2.1 CONSERVACIÓN DE LA POBLACIÓN FV x FB

Las plantas que forman la población FV x FB, así como las líneas parentales, *Fragaria vesca* FDP815, *Fragaria bucharica* FDP601 y el híbrido F_1 , fueron amablemente cedidas por *East Malling Research Center* (East Malling, Inglaterra) para su conservación en el Programa de Genómica y Biotecnología del IRTA en Cabrils, en el año 2004. La población, formada por 76 individuos F_2 , ha sido mantenida desde entonces en condiciones de invernadero de malla de sombreo (umbráculo), en macetas de 5 litros de capacidad, en un sustrato formado al 25% por turba rubia neutralizada TKS1 (Floragard), 25% de perlita de grano grueso A13 y 50% de corteza de pino. El riego y abonado de las plantas se realiza de manera mecanizada mediante un sistema de goteo. Las plagas y enfermedades son

controladas permanentemente por personal especializado mediante control integrado y tratamientos fitosanitarios de tipo químico. Flores, frutos y estolones son retirados periódicamente para mantener la pureza de las líneas.

2.2.2 RECOLECCIÓN DE LAS MUESTRAS Y OBTENCIÓN DE ADN GENÓMICO

La extracción del ADN se realizó empleando material vegetal fresco. Se recolectaron uno o dos foliolos (en función del tamaño) de hoja joven, procedente de la roseta central de la planta. Cada muestra se introdujo en un tubo de 1.5 ml de capacidad que fue inmediatamente sumergido en N₂ líquido. El material vegetal fue triturado con la ayuda de unas mazas de plástico hasta conseguir un polvo muy fino. En cada tubo se transfirieron 500 µl de tampón de extracción (Doyle y Doyle, 1990) modificado para fresa con la adición de PVP-40 al 2%, previamente calentado hasta los 65°C. Los tubos se incubaron durante 45 minutos en un baño en agitación continua a 65°C. Después, se añadió a cada tubo 500 µl de una mezcla de cloroformo-alcohol isoamílico (24:1), se agitó fuertemente cada tubo con la ayuda de un *vortex* y se centrifugó la mezcla a 14.000 rpm durante 15 minutos a temperatura ambiente. Se traspasaron los sobrenadantes a tubos limpios, donde se añadió etanol absoluto a -20°C. Tras invertir los tubos en 5 ocasiones, se centrifugaron a 14.000 rpm durante 15 minutos a 4°C y se descartó la fracción líquida de cada tubo. El precipitado se lavó con etanol 70° frío y se volvió a centrifugar durante 3 minutos a 14.000 rpm a 4°C. Una vez decantada de nuevo la fracción líquida, los pellets obtenidos se secaron a temperatura ambiente durante dos horas bajo una campana de extracción de gases y fueron resuspendidos en 50 µl de agua estéril ultrapura (grado HPLC, Panreac) e incubados durante un mínimo de doce horas a 4°C. La concentración de ADN de cada una de las muestras se midió en un espectrofotómetro (NanoDrop, Thermo scientific, DE, USA) y se prepararon alícuotas a una concentración final de 10 ng/µl.

2.2.3 SELECCIÓN DE MARCADORES

Se seleccionaron un total de 20 marcadores RFLP previamente localizados en el mapa de referencia (fig. 6) de *Fragaria* (Vilanova *et al.* 2008). La selección se realizó en función de su posición en el mapa, con objeto de alcanzar la mayor cobertura posible, descartando aquellos que distan menos de 10 cM de un marcador de tipo PCR (Tabla 4). Además, atendiendo al criterio de posición en el mapa de referencia y por ser de difícil resolución, se seleccionaron 3 marcadores de tipo SSR, EMFvi136, EMFv164 y EMFv183 (Sargent *et al.* 2003, 2004a; Hadonou *et al.* 2004) para ser sustituidos por otros más resolutivos.

2.2.4 LOCALIZACIÓN DE SECUENCIAS *IN SILICO*

Las secuencias empleadas en el desarrollo de nuevos marcadores próximos a marcadores tipo RFLP fueron obtenidas a partir del alineamiento de secuencias EST del genoma de *P. persica* empleadas por Vilanova *et al.* (2008) en el desarrollo de sondas de hibridación con secuencias genómicas de fresa diploide, mediante el algoritmo BLASTN (Altschul *et al.* 1997). Para la localización de motivos SSR próximos a los marcadores EMFvi136, EMFv 164 y EMFv183, se alinearon las secuencias correspondientes a dichos marcadores. Los alineamientos fueron llevados a cabo mediante la aplicación BLAST disponible en la base de datos de acceso restringido '*Strawberry Genome Assembly version 8*' (<http://strawberry.vbi.vt.edu/blast/>). Esta base de datos recoge una versión preliminar del genoma de *Fragaria* compuesta de contigs y scaffolds formados a partir de secuencias cortas obtenidas a partir de tecnología de secuenciación 454. Los alineamientos identificaron contigs en los que la estimación de la similitud de las secuencias alineadas fue estadísticamente significativa (valor $E \leq 0.05$). En los casos en los que el alineamiento arrojó más de un resultado significativamente consistente, sólo se tuvo en cuenta la secuencia con un valor E más bajo (tabla 4).

2.2.5 DISEÑO DE CEBADORES ESPECÍFICOS Y REACCIÓN DE PCR

Las secuencias seleccionadas para el desarrollo de cebadores específicos fueron descargadas desde la base de datos para la localización de motivos microsatélite en regiones próximas a los loci de interés mediante el uso de la herramienta 'Websat' (Martins *et al.* 2009; <http://wsmartins.net/websat/>). Una vez localizados los SSR, el diseño de los cebadores se realizó de forma rutinaria mediante la herramienta Primer3 (Rozen y Skaletsky, 2000; <http://frodo.wi.mit.edu/primer3/>), añadiendo en el extremo 5' de los cebadores F (Forward), una secuencia estándar formada por 17 nucleótidos (5' GTAAAACGACGGCCAGT 3') que denominamos 'cola M13' por ser homóloga a una región del genoma de dicho virus. La denominación de los cebadores comprende las letras CFV (Cabrls *Fragaria vesca*) seguidas de cuatro números que corresponden a las cuatro últimas cifras del número de accesión del EST empleado para su desarrollo.

Cada reacción de PCR tuvo lugar en un volumen final de 10 μ l, que contiene 20 ng de ADN molde, 1x tampón Lab (KCl 50mM, Tris-HCl 10 mM pH 8.3, gelatina 0.001%), 2.5 mM MgCl₂, 0.2 mM de cada desoxinucleótido (dATP, dCTP, dGTP, dTTP), 2 U de DNA polimerasa *AmpliTaq* (Perkin-Elmer, IL, USA), 0.2 μ M de cebador R (*Reverse*), 0.05 μ M de cebador F y 0.2 μ M de cebador M13, cuya secuencia es idéntica a la cola M13 presente en el cebador F y que incorpora un marcaje fluorescente en uno de

los cuatro posibles colores o *dyes* (VIC para el verde, FAM para el azul, PET para el rojo o NED para el amarillo). Las diferencias cromáticas permiten realizar cargas múltiples de PCR en la electroforesis capilar posterior a la reacción de PCR. La elección del color se hizo en función de los tamaños esperados para los diferentes amplicones en cada reacción de PCR (Boutin-Ganache *et al.* 2001). Los termocicladores empleados fueron los modelos PE9600 y PE9700 de Perkin Elmer, PTC-200 de MJ Research y PE GenAmp PCR System 9700 de Applied Biosystems, (CA, USA). Las condiciones de amplificación empleadas fueron: Una desnaturalización inicial de 2 minutos a 94°C, seguida de 10 ciclos compuestos de una desnaturalización (30 segundos a 94°C), una hibridación (45 segundos a 55°C) y una extensión (45 segundos a 72°C) para favorecer la hibridación de los cebadores específicos, seguidos de 25 ciclos compuestos de una desnaturalización (30 segundos a 94°C) una hibridación (45 segundos a 50°C) y una extensión (45 segundos a 72°C) para favorecer la hibridación del cebador marcado (Fig. 7), seguidos de una extensión final de 7 minutos a 72°C.

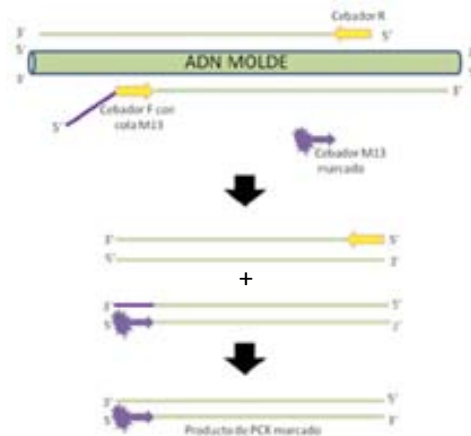


Figura 7. Amplificación de ADN genómico con marcaje fluorescente. Tras los primeros ciclos de PCR, los amplicones incorporan la secuencia M13 a través de la hibridación del cebador F con cola M13. En los 25 ciclos finales de la reacción se favorece la incorporación del cebador M13, generando amplicones que incorporan la señal fluorescente

2.2.6 VISUALIZACIÓN DE PRODUCTOS DE PCR

Los amplicones generados durante la PCR se separaron mediante electroforesis capilar en un analizador de fragmentos ABI3130 (Applied Biosystems) siguiendo las instrucciones del fabricante. En cada electroforesis se analizaron simultáneamente diferentes amplicones hasta alcanzar un máximo de 3 fluorocromos distintos por capilar. Para ello se tomaron 2 µl de la mezcla de productos de PCR y se mezclaron con 10 µl de formamida desionizada y 0.3 µl de marcador de peso molecular LIZ-500 (Applied Biosystems), marcado con fluorocromo naranja. La mezcla se sometió a una desnaturalización de 3 minutos a 94°C, se incubó unos minutos en hielo y, a continuación, se transfirió al analizador de fragmentos.

Los archivos de salida generados por el analizador de fragmentos fueron transferidos a la aplicación informática GeneMapper (Applied Biosystems) que, a partir los datos generados durante la electroforesis, genera un cromatograma correspondiente a los diferentes fragmentos de ADN separados. En el caso de no observar ningún polimorfismo en el tamaño de los amplicones, se procedió a la secuenciación y búsqueda de marcadores de tipo SNP (Single Nucleotide Polymorphism).

2.2.7 SECUENCIACIÓN DE ADN GENÓMICO

2.2.7.1 PREPARACIÓN DE LOS AMPLICONES A SECUENCIAR

La reacción de PCR tuvo lugar en un volumen final de 30 μ l, que contiene 40 ng de ADN molde, 1x tampón Lab, 2,5 mM $MgCl_2$, 0,4 mM de cada desoxinucleótido, 2 U de DNA polimerasa *AmpliTaq* (Perkin-Elmer, IL, USA), 0,6 μ M de cebador F y 0,6 μ M de cebador R. Las condiciones de amplificación empleadas fueron: Una desnaturalización inicial de 2 minutos a 94°C, seguida de 35 ciclos de desnaturalización (30 segundos a 94°C), hibridación (45 segundos a 58- 60°C) y extensión (45 segundos a 72°C). Para comprobar la calidad de la amplificación y la existencia de un único amplicón, se tomaron alícuotas de 10 μ l de cada uno de los productos de PCR, se añadió tampón de carga (azul de bromofenol 0.6% (p/v), glicerol 70% (v/v), SDS 0.2% (v/v), NEB 0.5X, EDTA 20mM) hasta alcanzar una concentración 1x y la mezcla se sometió a una electroforesis horizontal en gel de agarosa al 1% (p/v) fundida en tampón NEB 0.25x (Tris 25mM, EDTA 0.25mM, AcONa (30mM) y ajustado a pH 8.1 con ácido acético). Una vez finalizada la electroforesis, el gel se tiñó con bromuro de etidio y se expuso a luz UV para la observación de los amplicones. Cuando el amplicón obtenido fue de buena calidad, los 20 μ l restantes se sometieron a una purificación mediante columnas de sefarosa Sephadex G-50 Superfine (Amersham Biosciences, NJ, USA) siguiendo las instrucciones del fabricante, se midió su concentración en un espectrofotómetro (NanoDrop) y se prepararon alícuotas a 10ng/ μ l.

2.2.7.2 REACCIÓN DE SECUENCIACIÓN

Para la secuenciación de los productos de PCR purificados se empleó el método Sanger (Sanger *et al.* 1977). Para ello se usó el kit 'Big Dye™ Terminator Cycle DNA Sequencing' (Applied Biosystems) siguiendo las instrucciones del fabricante, empleando 25 ng de producto de PCR purificado como ADN molde. La reacción de secuenciación comprende 25 ciclos de desnaturalización (10 segundos a 96°C), hibridación de cebadores (5 segundos a 50°C) y extensión (4 minutos a 60°C).

Tabla 4. Desarrollo de cebadores específicos para la optimización de marcadores RFLP y SSR. El número de copias y la posición en FV x FB hacen referencia a los datos obtenidos por Vilanova *et al.* (2008) mediante hibridación con sondas RFLP (*southern blot*).

MARCADOR DE ORIGEN	N° DE ACCESIÓN EMBL	TIPO DE MARCADOR	POSICIÓN (FV x FB)	N° DE COPIAS	MARCADOR ASOCIADO	N° DE RESULTADOS SIGNIFICATIVOS PARA EL ALINEAMIENTO	VALOR E	MOTIVO SSR SELECCIONADO	SECUENCIA DE CEBADORES (5' - 3')	DISTANCIA ENTRE MARCADORES (RFLP/SSR - SSR1/SSR2)
TSA3	T14021	RFLP	LG1 39.3	1	CFV4021	1	0.016	[TAA] ₁₆	F: CTGGAATGGTGAGAAAACG R: AAAGCCCATCTAAACGTGAGAG	8.840 pb
AC8	BI203052	RFLP	LG1 61.2	2	CFV3052	1	3,00E-29	[TG] ₅	F: GGGGTTAGGGTTTTGTCTCC R: AGTGGTTCAAACCCATTCA	4.051 pb
EMFv164	8IEMFv164	SSR	LG1 74.1	1	CFV164	1	1,00E-63	[GA] ₁₂	F: AAGGCCACCTGATTTGTCC R: AGTTCGGTCAACGGTTTTG	6.180 pb
PC101	BI203099	RFLP	LG2 19.7	1	CFV3099	1	2,00E-08	[CT] ₆ + [CT] ₃₀	F: CCTACAAAATTAAGGGGATGGA R: ATCACCGATGGAGGAGTTAGA	16.912 pb
EMFv183	8IEMFv183	SSR	LG2 88.2	1	CFV183	1	5,00E-19	[GT] ₈	F: GCGTCCTTGCCATGTATTTT R: GCACTTCTCGAATCACACGA	8.640 pb
AC32	BI203135	RFLP	LG4 0	2	CFV3135	1	7,00E-42	[CT] ₃₈	F: GCTTTAGCGGTTGACACATACA R: GTCCATGTGGTTACTGGATTAA	43.275 pb
AC24	BI203138	RFLP	LG4 20.3	1	CFV3138	2	2,00E-11	[AG] ₆	F: ACCCAGCCAACATTACAA R: CTCTCGCCTCGAGTATCCTG	130 pb
AG53	BH023829	RFLP	LG4 22.1	2	CFV3819	2	3,00E-35	[CT] ₁₂	F: CTCTCATGGCTTCATCGTCA R: ATTGTCGGTACCGTTTCATC	15 pb
PC78	BI203148	RFLP	LG4 31.1	1	CFV3148	1	8,00E-14	[GA] ₁₇	F: GGCCAAGACAACATCCTAAAAC R: TGTACGCTGGTGTATCTTTGACA	15.140 pb
AC49	BI203072	RFLP	LG5 0	1	CFV3072	1	2,00E-10	[TAAAAA] ₅	F: TTTGCCTTGGTTGAGCGTAT R: TTACGTTGTGGTCGTTCCAGG	137 pb
PC14	BI203132	RFLP	LG5 19.1	1	CFV3132	1	1,00E-14	[CA] ₂₂	F: TTGTTTGCCGCATATCTCTC R: GGACACTCAAAGCACTGCA	8.640 pb
AG33	BH023821	RFLP	LG5 56.4	1	CFV3821	1	5,00E-46	[TAA] ₁₃	F: TATGCGGTTGATTAGGAGTTCA R: GCTAGTTTTGGTGAATGTCAA	7.280 pb
AC31	BI203096	RFLP	LG7 0	1	CFV3096	1	7,00E-57	[AT] ₂₁	F: TTCCGCTGATGGTTAGGATG R: GCAAATGGCAGAAACATGAAG	107.330 pb
AG35	BH023896	RFLP	LG7 9	1	CFV3896	1	2,00E-49	[AT] ₁₃ + [TTTTA] ₆	F: CCAGGCAATCAGGTGAAGAT R: ATGGCCACTATGATTGAGCA	1.180 pb
MC045	BI203117	RFLP	LG7 81	2	CFV3117	2	6,00E-18	[AG] ₅₁	F: CTCACGGACGGCAAGAAA R: CCTCAATATGCCACTCCATCTC	16.062 pb
CoMET	X83217	RFLP	LG7 89.6	1	CFV3217	1	7,00E-91	[TC] ₁₈	F: CGACATCAACGATGGATTTG R: TGCCATATGACTCAAGTAGCAG	22.850 pb
AC10	BI203087	RFLP	LG4 13.5	2	CFV3087	2	7,00E-42	[AT] ₈	F: CGGCTCATGTGCATTCTCT R: CCCAAGTCGATGAAAACCTCAA	7.975 pb
EMFvi136	AJ564350	SSR	LG4 46.1	1	CFV136	1	1,00E-28	[TA] ₁₁	F: TTGAAGTTTTGGGTGATCTGG R: CCCCCTTAGAAAGAACTCTGTG	14.170 pb
Pgl1	X75020	RFLP	LG6 22.8	1	CFV5020	1	E-129	[GA] ₁₁	F: GTTCCACTAATTTGGGTATGGG R: GTCTGGGCTTGTGGGTGTAT	5.110 pb
CC2	BI203070	RFLP	LG6 43.7	2	CFV3070	2	9,00E-26	[TCT] ₉	F: AGTGGAATATCGTTGGCCTTG R: GAAACAAGCAGGAAACTGGC	808 pb
AG106	BH023814	RFLP	LG6 96	2	CFV3814	2	2,00E-30	[TAA] ₁₀	F: GCTTCAAGCTTGTGTGGTCA R: CTGCTGGCTGATGAAAACCTG	1.513 pb

2.2.7.3 PURIFICACIÓN DE LOS PRODUCTOS DE LA REACCIÓN DE SECUENCIACIÓN

Los productos de PCR obtenidos durante la reacción de secuenciación fueron precipitados para su purificación en la misma placa en la que fueron generados. A los 10 μ l de producto de PCR de secuenciación se añadieron 10 μ l de agua, 2 μ l de EDTA 125mM, 2 μ l de CH₃COONa 3M (pH 4.6) y 50 μ l de etanol absoluto. La mezcla se incubó 15 minutos a temperatura ambiente en oscuridad y se centrifugó durante 45 minutos a 3000 rpm a 4°C. Una vez eliminada la fase líquida mediante un pulso de centrifugación de la placa invertida sobre papel secante, los pellets se lavaron con etanol 70° y se volvió a centrifugar 15 minutos a 3000 rpm a 4°C. Se descartó de nuevo la fase líquida y los pellets se resuspendieron en 12 μ l de formamida desionizada. La solución se incubó a 4°C un mínimo de 30 minutos en condiciones de oscuridad y se mantuvo en dichas condiciones hasta su carga en el secuenciador.

2.2.7.4 ELECTROFORESIS CAPILAR Y TRATAMIENTO DE LOS DATOS

Una vez resuspendidos, los productos de la reacción de secuenciación fueron transferidos a una placa y cargados en un secuenciador automático ABI-Prism 3130 (Applied Biosystems) siguiendo las instrucciones del fabricante. Una vez finalizado el proceso de electroforesis capilar, los datos fueron analizados mediante el software Sequencher (Gene Codes Co. MI, USA) y se procedió a interpretar visualmente el cromatograma obtenido a partir de la lectura del secuenciador.

2.2.8 ANÁLISIS GENÉTICO DE LOS NUEVOS MARCADORES Y CONSTRUCCIÓN DE UN MAPA DE LIGAMIENTO

Los datos obtenidos a partir de la segregación de polimorfismos de los nuevos marcadores fueron analizados mediante la prueba Chi-cuadrado (χ^2) para la bondad del ajuste a un ratio de segregación 1:2:1 para los marcadores codominantes y 3:1 para marcadores dominantes. Los análisis de ligamiento fueron llevados a cabo mediante el programa MAPMAKER 3.0 para MS-DOS (Lander *et al.* 1987), incorporando la función de mapeo de Kosambi (Kosambi, 1944). La construcción de los grupos de ligamiento, así como el orden de los marcadores fue determinado empleando un LOD mínimo de 3.0 y una distancia máxima entre marcadores de 50 cM. Los datos de segregación de los nuevos marcadores CFV fueron analizados en conjunto con datos de segregación de marcadores previamente publicados (Sargent *et al.* 2004a, 2006, 2007, 2008; Vilanova *et al.* 2008) y datos de marcadores no publicados en ese momento, amablemente cedidos por los doctores Dan Sargent, Iraida Amaya y Juan Jairo Ruiz Rojas. Quedan excluidos del estudio los marcadores de tipo RFLP. Las pruebas de bondad del ajuste, así como las medidas de distorsión se tomaron mediante el uso del

programa JoinMap® 4 (Van Ooijen, 2006). A partir de los datos de ligamiento arrojados por el análisis anterior se construyó un mapa de *bins*, usando la nomenclatura de Sargent *et al.* (2008). Para ello se analizó la segregación de todos los marcadores estudiados en las plantas 32, 46, 48, 59, 62 y 83, que conforman el *bin set* de *Fragaria* y se definieron *bins* para las regiones genómicas delimitadas por marcadores moleculares con genotipo común.

2.3 RESULTADOS

2.3.1 DISEÑO DE CEBADORES E IDENTIFICACIÓN DE POLIMORFISMOS

Con objeto de evaluar la utilidad de los pares de cebadores diseñados, se desarrollaron reacciones de PCR empleando ADN de los parentales de la población FVxFB, y el individuo híbrido F₁ como ADN molde. De los 21 pares de cebadores diseñados para la amplificación de marcadores tipo SSR, 19 amplificaron al menos un producto de PCR (Tabla 5). Tan sólo CFV3070 y CFV3814 no amplificaron ningún locus. Los tamaños de los productos de PCR difirieron de los tamaños esperados en un rango que varía entre 3 pb (alelo 161 pb del marcador CFV3821 en parental *F. bucharica*) y 164 pb (alelo 316 pb del marcador CFV183 en parental *F. vesca*). De los 19 loci amplificados, tres (CFV3217, CFV136 y CFV5020) resultaron ser monomórficos para el tamaño de los fragmentos generados en la amplificación del ADN de los parentales de la población. El parental *F. vesca* resultó ser heterocigótico para 3 de los 19 loci analizados (CFV3819, CFV3096 y CFV3087) mientras que el parental *F. bucharica* mostró heterocigosis para 8 de los 19 loci (CFV3135, CFV3819, CFV3148, CFV3072, CFV3132, CFV3821, CFV3117 y CFV3087).

La secuenciación del amplicón generado por el par de cebadores CFV3217 (monomórfico para el tamaño del amplicón obtenido) reveló la existencia de un polimorfismo de tipo SNP G/C (alelos *F. vesca* y *F. bucharica* respectivamente) segregando en la población, mientras que las secuencias obtenidas con los pares de cebadores CFV5020 y CFV136 resultaron monomórficas para todas sus bases nucleotídicas. Sólo dos de los 16 polimorfismos encontrados se heredaron de manera dominante en la población (CFV4021 y CFV3896), en ambos casos por ausencia de señal en el parental *F. bucharica* y por tanto de dominancia materna.

2.3.2 LOCALIZACIÓN DE MARCADORES MEDIANTE BIN MAPPING

Una vez identificados los polimorfismos, se realizaron reacciones de PCR empleando ADN de las 6 plantas F_2 que forman el *Bin Set* de *Fragaria*, con objeto de comparar la posición de los polimorfismos obtenidos con la de los marcadores originales en el mapa de *bins* (Sargent *et al.* 2008). La posición de los polimorfismos generados mediante PCR coincidió con la de los marcadores a sustituir en 13 de los 15 casos estudiados a través de la estrategia de *Bin Mapping* (Tabla 5). La pareja de cebadores CFV183 amplificó un locus situado en el *bin* II:75 (tabla 5), mientras se esperaba en el *bin* II:88. Al tratarse de *bins* contiguos cuyo *bin*-genotipo sólo difiere en un individuo, se consideró como resultado correcto a la espera de un estudio más detallado. El par de cebadores CFV3087 amplificó un locus cuya segregación corresponde con un *bin*-genotipo doble que designa dos *bins*, en el GL2 (II:50) y en el GL3 (III:53). Dado que tanto los datos de *southern blot* como el alineamiento revelaron la existencia de secuencias parálogas para este locus, y puesto que se esperaba haber amplificado un locus del GL4, el marcador fue desechado para los análisis posteriores. Los marcadores MC45 y CoMET no fueron localizados en el mapa de *bins* de Sargent *et al.* (2008), pero sus homólogos CFV3117 y CFV3217 aparecen en los *bins* VII.40 y VII.64, alejados de la posición esperada, dada su localización en el mapa de Vilanova *et al.* (2008) (fig. 6).

Tabla 5. Datos obtenidos a partir del genotipado de los parentales y *Bin set* de la población *FV x FB*. El tamaño de los amplicones esperados se calculó en base a los datos genómicos disponibles en la base de datos 'Strawberry Genome Sequencing Project', que hacen referencia a la variedad 'Hawaii 4' de *Fragaria vesca*. Los alelos marcados en negrita son los que han sido heredados en la generación F_1 . La letra A denota el genotipo heredado del parental femenino (*F. vesca*), B hace referencia al parental masculino (*F. bucharica*), H representa el genotipo híbrido (AB). En el caso de marcadores con herencia dominante, D representa genotipo A o H. La denominación de los *bins* responde a la empleada por Sargent *et al.* (2008). nd: No descrito

LOCUS ORIGINAL	NUEVO LOCUS	AMPLIFICACIÓN EN LOS PARENTALES	PESO MOLECULAR ESPERADO	ALELO(S) <i>F. vesca</i> (pb)	ALELO(S) <i>F. bucharica</i> (pb)	SEGREGACIÓN EN <i>FV x FB</i>	SEGREGACIÓN ESPERADA EN <i>BIN SET</i> <i>BIN</i>	SEGREGACIÓN OBSERVADA EN <i>BIN SET</i> <i>BIN</i>	SEGREGACIÓN OBSERVADA EN <i>BIN SET</i> GENOTIPO	
TSA3	CFV4021		233	249	0		I.41	BHHBBA	I.41	BDDDBD
AC8	CFV3052		259	278	267		I.74	HBHHHH	I.74	HBHHHH
EMFv164	CFV164		217	245	267		I.74	HBHHHH	I.74	HBHHHH
PC101	CFV3099		209	179	153		II.8	HBHBBH	II.119	HBHBBH
EMFv183	CFV183		152	316	307		II.88	HAAAAA	II.88	HAABAA
AC32	CFV3135		258	232	236-254		IV.26	HBHHAH	IV.26	HBHHAH
AC24	CFV3138		281	300	308		IV.26	HBHHAH	IV.26	HBHHAH
AG53	CFV3819		245	234-256	234-251		IV.26	HBHHAH	IV.26	HBHHAH
PC78	CFV3148	SI	153	174	157-174	polimórfica	IV.46	HBHHAH	IV.46	HBHHAH
AC49	CFV3072		191	212	204-210		V.15	ABBAAAB	V.15	ABBAAAB
PC14	CFV3132		152	170	130-139		V.29	ABBAAH	V.29	ABBAAH
AG33	CFV3821		158	175	168-161		V.73	HBHHAH	V.73	HBHHAH
AC31	CFV3096		242	174-267	226		VII.21	BHBAHH	VII.21	BHBAHH
AG35	CFV3896		189	206	0		VII.21	BHBAHH	VII.21	BDDDDD
MC45	CFV3117		230	207	228-240		nd	nd	VII.40	BHBHHH
CoMET	CFV3217		216	203	203		nd	nd	VII.64	BHBHHH
AC10	CFV3087		293	149-310	308-323		IV.26	HBHHAH	II.50/III:53	HHHBBH
EMFv136	CFV136	SI	277	238	238	monomórfica	*	*	*	*
Pgl1	CFV5020		268	283	283		*	*	*	*
CC2	CFV3070	NO	256	*	*	*	*	*	*	*
AG106	CFV3814		265	*	*	*	*	*	*	*

2.3.3 ANÁLISIS DE LA SEGREGACIÓN DE LOS NUEVOS MARCADORES EN FV x FB

Los datos de la segregación de los nuevos marcadores en todos los individuos de la población FV x FB se muestran en la tabla 6. La posición encontrada para el 87,5% de los marcadores evaluados varió respecto a la descrita para el marcador original, a partir del cual fueron diseñados los nuevos marcadores, si bien su posición en el mapa de bins sólo varió para el marcador CFV183 (bin II:75) respecto de EMFv183 (bin II:88). El 12.5% de los marcadores evaluados mantiene su posición, al encontrarse en el extremo de un grupo de ligamiento. Sin embargo, la distancia respecto al marcador más cercano sí varía; el marcador CFV-3072 dista 3.5 cM de EMFvi108, que dista 7,7 cM de AC49, lo que revela una diferencia de 4,2 cM entre CFV-3072 y AC-49. El marcador CFV-3096 dista 7,5 cM de EMFv190, que dista 13,9 cM de AC-31, por lo que la distancia entre CFV-3096 y AC-31 es de 6 cM. La mayor diferencia encontrada la registran los marcadores CFV-3117 y CFV-3217 respecto a MC-45 y CoMET, de 65,6 y 38,6 cM respectivamente.

La tabla 7 muestra el análisis de la segregación de los polimorfismos encontrados tras la amplificación del ADN de los 76 individuos F_2 que conforman la población FV x FB. En ella se detallan los datos referentes a la segregación y los valores de Chi-cuadrado para la bondad del ajuste de los datos observados respecto a los ratios esperados siguiendo el modelo mendeliano. Nueve marcadores mostraron desviaciones significativas respecto al ratio de segregación esperado.

Tabla 6. Segregación de los marcadores desarrollados en este trabajo en la población FV x FB y comparación con la segregación obtenida para los marcadores RFLP y SSR originales (Vilanova *et al.* 2008; Sargent *et al.* 2006). Los datos (A, genotipo *F. vesca*; B, genotipo *F. bucharica*; H, genotipo híbrido; D, genotipo A o H en marcadores dominantes; - Dato no disponible), corresponden la segregación de los marcadores, distribuidos por pares, separados por una línea horizontal, en los 76 individuos F₂ que componen la población. La columna N hace referencia al número de datos obtenidos para el mapeo de los marcadores. GL y posición indican el grupo de ligamiento y posición de los marcadores, en centiMorgans (cM).

NOMBRE	2	3	4	6	7	8	9	11	12	13	14	15	17	18	19	20	21	22	24	25	26	27	29	30	31	32	33	34	36	37	39	41	43	44	45	46	47	48	49	50		
TSA3	H	-	H	H	-	H	-	H	-	-	A	-	-	-	-	A	A	H	-	H	H	H	-	A	H	-	H	H	H	H	H	A	-	A	-	-	A	H	A	-		
CFV-4021	B	B	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	
AC8	B	H	-	-	-	B	-	H	-	A	A	-	B	B	-	-	-	-	-	-	A	A	-	-	B	-	-	H	-	-	H	H	-	-	-	-	-	H	A	-		
CFV-3052	H	B	H	H	H	H	A	H	A	A	A	A	B	B	A	A	A	H	B	H	A	A	H	A	B	H	H	H	H	-	H	H	H	A	H	B	H	H	A	A		
EMFv164	B	B	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D		
CFV-164	H	H	-	H	H	H	H	H	A	A	A	A	B	B	A	A	A	H	B	H	A	A	H	A	B	H	H	H	-	A	H	H	H	A	H	B	H	H	A	A		
PC101	H	H	B	B	B	H	-	H	-	B	-	-	-	B	-	H	B	B	-	-	B	H	-	H	-	H	-	B	B	H	H	H	-	B	-	-	-	H	H	-		
CFV-3099	H	B	B	B	B	H	B	H	B	B	H	B	B	B	B	B	B	B	H	B	B	H	B	H	B	H	B	B	B	B	H	H	H	H	B	B	H	H	H	H	H	
EMFv183	A	A	B	H	A	A	A	H	A	A	H	B	-	B	H	H	B	H	A	A	A	H	H	B	H	H	A	A	H	H	H	B	H	H	B	A	B	A	B	A		
CFV-183	A	A	H	H	A	H	A	H	H	B	H	B	H	B	H	H	B	H	A	A	A	H	H	B	H	H	A	A	H	H	-	B	H	H	B	A	B	A	B	A	A	
AC32	H	H	H	B	A	H	-	A	-	H	H	-	B	H	-	-	A	A	-	B	H	H	-	A	H	H	H	H	H	H	H	B	H	H	-	-	-	H	H	H	-	
CFV-3135	H	B	H	B	A	-	H	A	H	H	H	H	B	H	H	H	A	A	B	B	H	B	B	A	H	H	B	B	B	-	B	H	B	A	H	B	H	H	B	H		
AC24	-	H	H	H	-	H	-	A	-	H	-	-	-	H	-	H	-	A	-	B	H	B	-	A	H	-	H	B	B	H	B	-	B	-	-	-	H	H	B	-		
CFV-3138	H	B	H	B	A	H	H	A	H	H	H	H	B	H	H	H	A	A	B	B	H	B	B	A	H	H	B	B	B	-	B	H	B	A	H	B	H	H	B	H		
AG53	H	H	H	B	A	H	-	A	-	H	H	-	B	H	-	B	A	A	-	B	H	B	-	A	H	H	H	B	B	H	B	H	B	H	-	A	-	-	H	H	B	-
CFV-3819	H	B	H	B	A	H	H	A	H	H	H	H	B	H	H	B	-	A	B	B	H	B	B	A	H	H	B	B	B	H	B	H	B	A	H	B	H	H	B	H		
NOMBRE	51	52	53	54	55	57	58	59	60	61	62	63	64	67	68	69	70	71	72	73	75	76	78	79	80	81	83	86	87	88	89	90	91	92	93	94	N	GL	POSICIÓN			
TSA3	-	H	H	B	H	-	B	-	-	H	B	B	A	H	-	-	H	-	-	B	H	A	A	A	-	H	A	B	-	A	H	-	-	B	-	B	47	1	33,7			
CFV-4021	D	-	D	-	D	D	B	B	D	D	B	B	D	D	-	B	D	D	D	B	D	D	-	D	D	D	D	B	D	D	D	D	D	D	B	B	B	72	1	33,4		
AC8	-	-	-	-	B	-	H	-	-	H	H	B	A	H	-	-	H	-	-	B	A	A	A	H	-	H	H	B	A	B	-	-	B	-	-	36	1	57,5				
CFV-3052	H	H	H	B	H	A	H	H	H	H	H	B	A	H	H	A	H	H	H	B	A	A	A	H	H	H	H	B	B	A	H	H	B	B	B	A	75	1	57,8			
EMFv164	D	D	D	B	D	B	D	D	B	D	D	B	D	D	D	D	D	D	D	B	D	D	D	D	D	D	D	B	B	D	-	B	B	B	B	D	75	1	74,1			
CFV-164	H	-	H	-	H	A	H	H	H	H	H	B	H	H	-	A	-	H	H	B	A	A	A	H	H	H	-	B	A	H	H	B	B	B	-	68	1	61,2				
PC101	-	-	B	B	-	-	B	-	-	H	-	B	B	B	-	-	H	-	-	H	B	-	H	H	-	H	H	H	B	-	H	-	-	H	-	-	42	2	21,3			
CFV-3099	B	H	B	B	H	B	B	B	A	H	B	B	B	B	A	H	H	B	H	H	B	H	H	H	B	H	H	H	B	H	B	H	B	H	H	B	76	2	16,4			
EMFv183	A	B	B	A	A	B	H	A	H	H	A	A	A	A	H	A	B	B	H	H	H	H	H	H	A	H	A	H	A	H	A	H	H	H	B	H	75	2	88,2			
CFV-183	A	B	B	-	A	B	H	B	H	H	A	A	H	A	H	A	B	B	H	H	H	-	H	H	-	H	A	H	A	-	H	H	H	-	B	H	69	2	72,2			
AC32	-	-	B	H	H	-	H	-	-	H	A	H	H	-	B	-	B	B	-	H	H	H	H	H	-	H	H	H	H	A	B	-	-	H	-	H	53	4	0			
CFV-3135	A	H	B	B	H	-	H	H	B	B	A	H	H	H	H	B	B	B	H	B	B	B	B	B	B	B	B	H	H	H	A	B	A	H	H	B	B	73	4	12,1		
AC24	-	-	B	B	-	-	H	-	-	B	A	H	B	H	-	-	B	-	-	B	B	B	B	B	-	B	H	H	H	A	B	-	-	H	-	B	45	4	20,3			
CFV-3138	A	H	B	B	H	H	H	H	B	B	A	H	B	H	H	B	B	B	H	-	-	-	-	-	-	B	H	H	H	A	B	A	H	-	B	B	69	4	10,1			
AG53	-	-	B	B	H	-	H	-	-	B	A	H	B	-	B	-	-	-	-	B	B	B	B	B	-	B	H	H	H	A	B	-	-	H	-	B	52	4	22,1			
CFV-3819	A	H	B	B	H	H	H	H	B	B	A	H	B	H	H	B	B	B	H	B	B	B	B	B	B	B	B	H	H	H	A	B	A	H	H	B	B	75	4	9,4		

Tabla 6 (continuación).

NOMBRE	2	3	4	6	7	8	9	11	12	13	14	15	17	18	19	20	21	22	24	25	26	27	29	30	31	32	33	34	36	37	39	41	43	44	45	46	47	48	49	50		
PC78	-	H	H	B	-	B	-	-	-	H	H	-	B	H	-	-	A	A	-	B	H	B	-	A	H	H	B	-	B	H	B	H	-	A	-	-	H	B	B	-		
CFV-3148	H	B	H	B	A	H	H	A	-	H	H	B	B	H	H	H	A	A	B	B	H	B	B	A	H	H	B	B	B	H	B	H	B	A	H	B	H	B	B	H		
AC49	-	A	A	H	-	A	-	B	-	A	B	-	H	B	-	B	H	B	-	A	-	H	-	H	H	-	A	H	H	B	H	H	-	-	-	-	-	B	B	H	-	
CFV-3072	H	A	B	-	H	A	B	B	H	A	B	B	H	B	B	B	-	B	A	H	A	H	A	B	H	H	A	H	H	B	H	H	H	H	H	H	H	B	B	B	H	H
PC14	-	H	H	H	-	A	-	B	-	A	B	-	H	B	-	B	A	B	-	A	-	A	-	H	H	A	A	H	H	B	H	B	-	-	-	-	-	B	B	H	-	
CFV-3132	H	H	A	-	H	A	H	B	H	A	B	B	H	B	H	B	H	B	B	A	B	A	B	H	H	A	H	H	H	B	H	B	A	B	B	B	B	B	B	H	H	
AG33	H	H	H	A	H	A	-	A	-	H	H	-	B	B	-	H	H	H	-	A	B	A	-	A	A	H	H	H	H	B	H	B	H	A	-	-	B	B	H	-		
CFV-3821	H	H	H	A	H	A	H	A	B	H	H	H	B	B	A	H	B	H	B	A	H	A	H	A	H	B	-	H	B	H	B	B	A	B	B	B	B	B	B	H	B	
AG35	H	B	B	H	A	B	-	H	-	A	B	-	H	A	-	H	H	B	-	B	H	H	-	B	B	B	B	B	B	H	B	H	H	H	B	-	-	B	B	H	-	
CFV-3896	D	D	B	D	D	B	B	D	D	D	B	B	D	D	B	D	D	D	D	B	D	D	D	B	B	B	D	B	D	B	D	D	D	B	D	D	B	B	D	B		
AC31	-	B	B	-	-	B	-	B	-	A	B	-	B	A	-	H	H	B	-	B	A	H	-	-	B	B	-	H	H	H	H	-	-	-	-	-	B	B	H	-		
CFV-3096	H	A	B	B	H	B	B	B	A	A	B	B	H	A	B	H	B	B	B	B	B	H	H	H	B	B	A	B	H	B	H	H	A	B	H	H	B	B	H	A		
MC45	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	-	
CFV-3117	H	A	B	A	-	-	-	H	A	A	B	B	H	-	B	H	H	H	-	B	H	H	H	B	B	B	A	H	H	H	A	H	A	B	H	H	B	B	H	B		
CoMET	H	A	A	H	B	H	-	A	-	A	A	-	H	H	-	H	B	A	-	H	B	B	-	A	B	B	H	A	A	A	H	H	H	-	-	-	-	H	H	H	-	
CFV-3217	-	B	A	H	-	H	H	A	H	A	A	H	H	H	H	H	H	A	B	H	B	B	B	A	B	-	A	H	A	A	A	H	A	H	B	A	B	H	H	B		
NOMBRE	51	52	53	54	55	57	58	59	60	61	62	63	64	67	68	69	70	71	72	73	75	76	78	79	80	81	83	86	87	88	89	90	91	92	93	94	N	GL	POSICIÓN			
PC78	-	H	B	B	H	H	H	H	-	H	A	H	H	H	H	-	B	B	-	B	B	B	B	H	-	B	H	H	H	A	B	-	-	H	-	-	61	4	31,1			
CFV-3148	A	H	B	B	H	-	H	H	B	H	A	H	H	H	H	B	B	B	H	B	B	B	B	B	B	B	H	H	H	A	B	A	H	-	-	B	75	4	19,6			
AC49	-	A	H	A	A	-	A	-	-	B	A	B	H	H	-	-	H	-	-	H	H	-	A	H	-	A	B	H	H	H	H	-	-	A	-	-	61	5	0			
CFV-3072	B	H	H	A	H	H	A	A	H	B	A	B	H	A	H	B	A	H	-	H	H	-	A	H	A	H	H	A	H	H	H	B	B	A	H	B	74	5	0			
PC14	-	-	H	A	A	-	A	-	-	B	A	B	H	A	-	-	A	-	-	B	H	A	A	H	-	A	B	H	A	H	-	-	A	-	-	62	5	20,5				
CFV-3132	B	H	H	A	H	H	A	A	H	B	A	B	H	A	A	B	A	H	B	B	H	H	A	H	H	A	B	H	A	H	-	B	B	A	B	B	75	5	10,8			
AG33	-	B	H	H	H	-	B	-	-	H	H	H	H	H	-	-	A	B	-	B	B	H	B	A	-	H	A	H	H	A	B	-	-	H	-	B	67	5	60,9			
CFV-3821	H	B	H	H	H	H	B	H	B	H	H	H	H	H	H	B	A	B	B	B	B	H	B	A	B	-	A	H	H	A	B	B	H	-	-	-	75	5	49,5			
AG35	-	H	B	A	-	H	B	A	-	A	H	B	H	B	B	-	B	H	-	B	A	B	B	B	-	H	H	H	B	B	A	-	-	A	-	B	67	7	9			
CFV-3896	D	D	B	D	B	D	B	D	D	D	D	B	D	B	B	B	D	B	B	D	B	B	B	B	D	D	D	B	B	D	D	B	D	D	B	D	D	B				
AC31	-	H	B	-	A	-	B	-	-	A	-	H	B	B	-	B	-	-	B	A	-	-	B	-	H	H	H	B	B	-	-	-	-	-	-	B	60	7	0			
CFV-3096	A	H	B	A	B	H	B	A	B	A	H	H	B	B	B	B	B	H	B	B	A	B	B	B	B	H	H	H	B	B	A	H	B	A	H	B	76	7	0			
MC45	-	-	-	A	-	A	H	H	-	H	H	B	H	H	-	-	B	H	-	H	A	H	H	H	-	H	H	H	H	A	-	-	A	-	-	H	38	7	81,7			
CFV-3117	A	H	H	A	H	H	B	H	H	H	B	H	B	B	B	H	H	H	B	A	-	-	-	B	H	-	-	-	B	A	H	B	-	-	-	71	7	16,1				
CoMET	-	H	H	A	-	-	H	-	-	H	H	B	H	H	-	-	H	A	-	H	H	-	B	H	H	H	A	-	-	H	A	-	-	-	-	67	7	90,3				
CFV-3217	H	H	H	A	-	B	H	H	H	H	B	B	B	B	B	H	A	H	H	H	B	H	H	B	B	H	H	B	H	A	B	B	-	B	H	73	7	51,7				

Tabla 7. Análisis de la segregación en FV x FB de los marcadores obtenidos. Valores de chi cuadrado para la bondad del ajuste respecto a la segregación mendeliana esperada. *, **, ***, significativamente distinto respecto a la segregación esperada con $P < 0.05$, 0.01 y 0.001 respectivamente; - no se aprecian diferencias significativas respecto al valor esperado

Locus	Clasificación	Ratio esperado	Ratio observado	χ^2	Distorsión	Grados de libertad	Posición (cM)	
CFV4021	a_.bb	3:1	57:15	0,67	-	1	GL1	33,40
CFV3052	aa:ab:bb	1:2:1	22:39:14	1,83	-	2	GL1	57,80
CFV164	aa:ab:bb	1:2:1	20:37:11	2,91	-	2	GL1	61,20
CFV3099	aa:ab:bb	1:2:1	1:35:39	38,84	***	2	GL2	16,40
CFV183	aa:ab:bb	1:2:1	20:33:16	0,59	-	2	GL2	72,20
CFV3135	aa:ab:bb	1:2:1	10:32:31	13,19	***	2	GL4	12,10
CFV3138	aa:ab:bb	1:2:1	10:32:26	7,76	**	2	GL4	10,10
CFV3819	aa:ab:bb	1:2:1	9:33:33	16,44	***	2	GL4	9,40
CFV3148	aa:ab:bb	1:2:1	10:31:31	13,64	***	2	GL4	10,10
CFV3072	aa:ab:bb	1:2:1	16:32:24	2,67	-	2	GL5	0,00
CFV3132	aa:ab:bb	1:2:1	18:29:27	5,65	*	2	GL5	10,80
CFV3821	aa:ab:bb	1:2:1	13:31:27	6,66	**	2	GL5	49,50
CFV3096	aa:ab:bb	1:2:1	14:23:39	28,29	***	2	GL7	0,00
CFV3896	a_.bb	3:1	42:33	14,44	***	1	GL7	10,80
CFV3117	aa:ab:bb	1:2:1	11:30:21	3,29	-	2	GL7	16,10
CFV3217	aa:ab:bb	1:2:1	15:35:22	1,42	-	2	GL7	51,70

2.3.4 CONSTRUCCIÓN DEL MAPA DE LIGAMIENTO FV x FB

El nuevo mapa de ligamiento obtenido (fig. 8) cubre una distancia total de 541,3 cM distribuidos en 7 grupos de ligamiento. El mapa contiene 294 marcadores moleculares de tipo PCR, de los cuales 210 son SSRs, 53 son CAPS/dCAPS y 22 son STS. Además, hay un SCAR, un SNP y un marcador morfológico. La cobertura media es de un marcador cada 1,87cM y la región menos cubierta, en el GL4, abarca 13,8 cM sin marcadores. Para su construcción, se han empleado los datos de segregación de 321 marcadores moleculares, 197 de los cuales proceden de anteriores trabajos publicados sobre esta población, 108 son nuevos marcadores cedidos por otros grupos de investigación y 16 son nuevos marcadores desarrollados aquí, y que sustituyen a otros que han sido eliminados. Veintisiete marcadores moleculares parecen no estar ligados a ninguno de los grupos de ligamiento.

El genotipado de la población se muestra en el anexo I. El 48,8% de los marcadores del mapa presentan valores distorsionados ($P \leq 0.05$) respecto a la segregación mendeliana esperada (fig. 9), si bien esta distorsión no está homogéneamente distribuida; sólo dos de los marcadores del GL1 están distorsionados ($P < 0,1$), mientras que el 84% de los marcadores del GL2 presentan una fuerte distorsión, en la mayoría de los casos para valores de $P \leq 0,0001$, que va disminuyendo hasta desaparecer al final del grupo. Los marcadores del GL3 presentan bajos niveles de distorsión, mientras que el 74,3% de los marcadores del GL4 están distorsionados ($P < 0,05$). El 71,4% de los marcadores del GL5 están distorsionados ($P \leq 0.05$), si bien la distorsión es menor que en otros grupos y está uniformemente distribuida a lo largo de todo el grupo, aunque se incrementa al final.

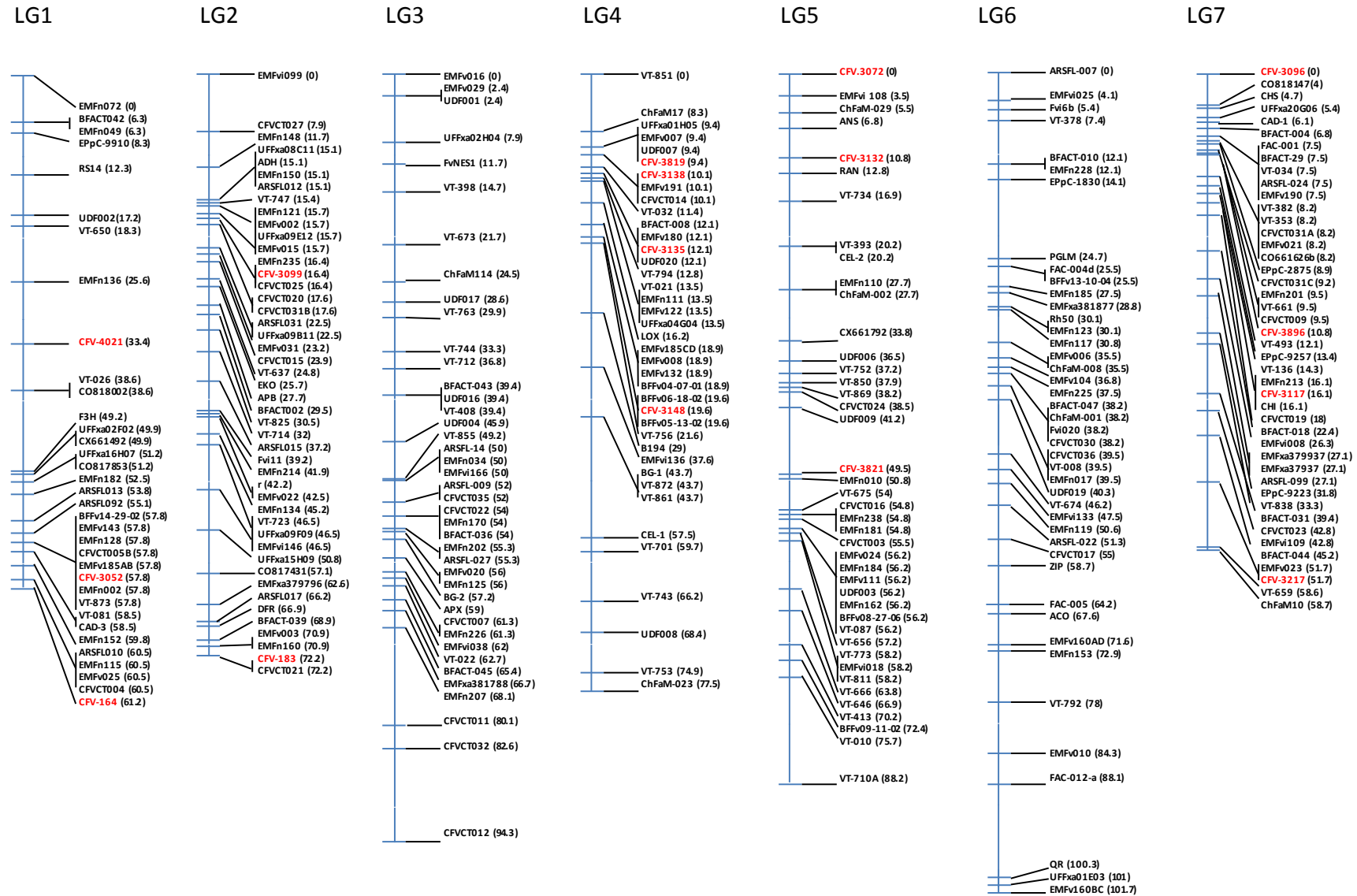


Figura 8. Mapa de ligamiento de fresa diploide (FV x FB) tras la localización de los marcadores desarrollados en esta investigación. Cada grupo de ligamiento (GL) está representado por una línea vertical de color azul, acompañados de los marcadores que los conforman. Los marcadores desarrollados en este trabajo aparecen marcados en rojo. Las distancias genéticas, en centiMorgans (cM), están indicadas a la derecha de cada marcador.

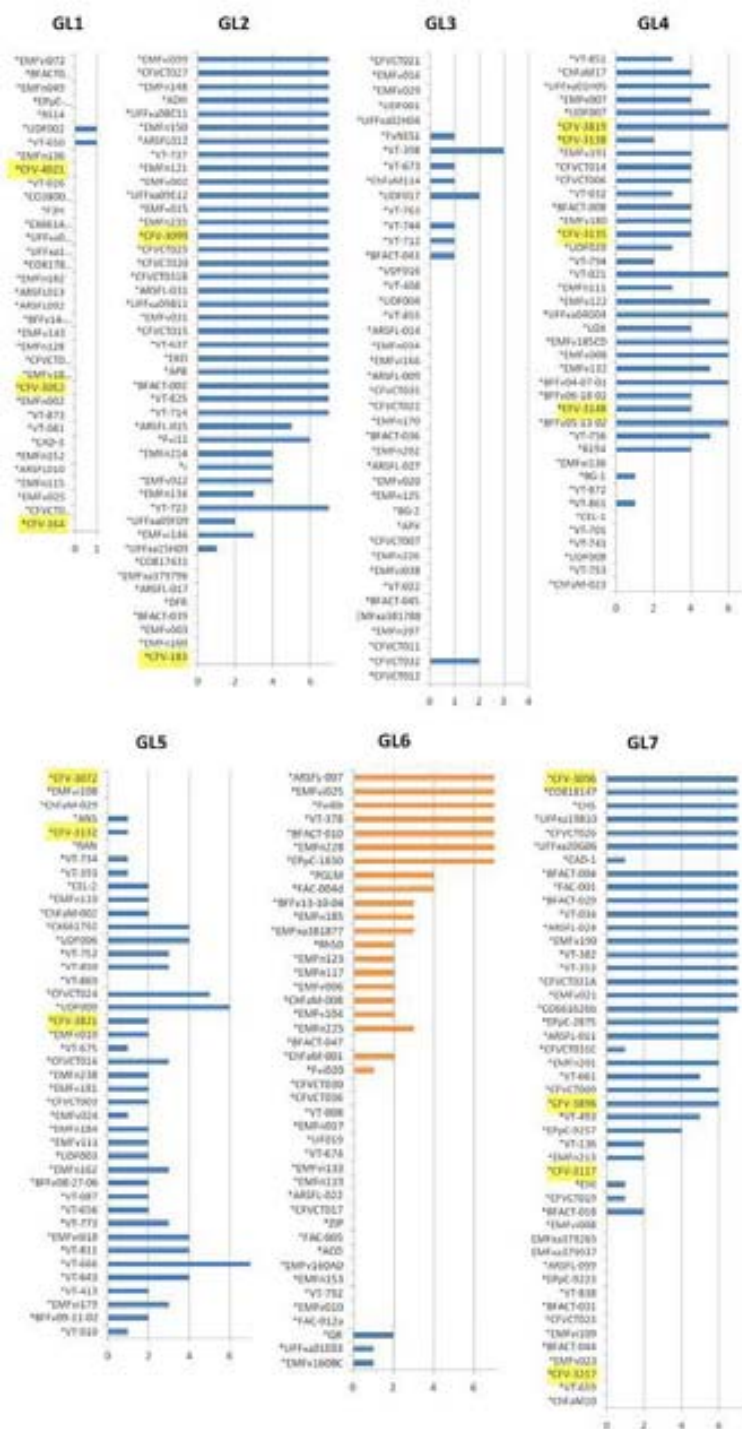


Figura 9. Valores de distorsión de los marcadores moleculares localizados en el mapa FV x FB en esta investigación. En vertical aparecen los marcadores moleculares que conforman cada grupo de ligamiento (GL) con su correspondiente valor de distorsión respecto de la segregación esperada. En horizontal, los niveles de significación para diferentes valores de P : 0, no significativo 1: $P < 0,1$. 2: $P < 0,05$. 3: $P < 0,01$. 4: $P < 0,005$. 5: $P < 0,001$. 6: $P < 0,0005$; 7: $P < 0,0001$. Las barras de color azul designan la prevalencia de alelos paternos (*F. bucharica*), mientras que las barras de color naranja indican prevalencia de los alelos maternos (*F. vesca*). Los marcadores desarrollados en esta investigación aparecen marcados en color amarillo.

El 50% de los marcadores del GL6 presentan distorsión ($P \leq 0.05$). Los marcadores más distorsionados se sitúan a la cabeza del grupo, donde los alelos maternos predominan sobre los paternos. Este fenómeno va disminuyendo hasta desaparecer en los marcadores situados aguas abajo de la posición 40 cM, y vuelve a aparecer al final del grupo. El 60% de los marcadores del GL7 presenta valores de distorsión muy fuertes ($P \leq 0,0001$), con prevalencia de los alelos paternos, que va disminuyendo hasta desaparecer a partir de la posición 22 cM.

La figura 8 representa la localización de los marcadores analizados en el mapa genético de referencia. La longitud del GL1 queda reducida en un 17,4% (12,9 cM) respecto a la versión de Vilanova *et al.* (2008) y el GL3 se reduce en un 18,1% (16 cM). Los marcadores AC-32 (CFV3135), AC-24 (3138) y AG-53 (3819), que ocupaban el extremo superior del GL4 quedan invertidos y pasan a una región más central, formando parte del *cluster* de marcadores que ya había sido descrito por Sargent *et al.* (2004), por lo que el grupo pierde 15,3 cM en esta región. Sin embargo, la localización en el otro extremo del grupo de nuevos marcadores moleculares procedentes de otras investigaciones provoca que el grupo sólo pierda un 5,4% (4,5 cM). Los marcadores MC-45 (3117) y Co-MET (3217), abandonarían el extremo inferior del GL7 pasando a formar parte de las regiones interiores, por lo que el grupo queda reducido en un 34,9% (31,6 cM).

2.3.5 REVISIÓN DEL MAPA DE *BINS* DE FV x FB

A la vista de los resultados de segregación obtenidos, se analizaron las modificaciones generadas en el mapa de *bins* desarrollado por Sargent *et al.* (2008) para esta población. El mapa queda ahora definido por 49 fragmentos de un tamaño medio de 8,9 cM, tras la adición de 4 nuevos bins (II:62, III:21, IV:66 y VII:58) y la eliminación de uno (bin II:88), con un promedio de 6 marcadores por *bin* (tabla 8). El mayor *bin*, localizado en el grupo 5, contiene 24 marcadores moleculares y abarca una región de 35 cM (6,5% del genoma).

Quince *bins* (30,6%) están definidos por un solo marcador o por varios marcadores que ocupan la misma posición, por lo que su extensión no puede ser estimada y no se han tenido en cuenta a efectos de cobertura (fig.10). La transición de unos *bins* a otros se definió mediante la identificación de un punto de recombinación en el genoma de una de las seis plantas. Cuando se identifica más de un evento de recombinación en un solo punto, es decir, más de una planta con genotipo recombinante para dos marcadores moleculares contiguos, se deduce que existe un *bin* intermedio que no ha sido detectado. De este modo, identificamos 8 *bins* intermedios a lo largo de todos los grupos de ligamiento a excepción de los grupos 2 y 5.

El mapa contiene 4 *bins* dobles (los *bins* dobles son aquellos en los que las 6 plantas del *bin set* comparten el mismo genotipo) que ya fueron descritos por Sargent *et al.* (2008). Tres de ellos implican *bins* de diferentes grupos de ligamiento: I:55 - IV:57, II:45 - III:54 y III:28 - VI:64. Para la identificación inequívoca de un locus localizado en alguno de estos *bins* es necesario genotipar el individuo 31 de FV x FB. El cuarto *bin* doble se localiza en el GL2 (II:11 - II:17), y se resuelve mediante el genotipado del individuo 14 de FV x FB (anexo I).

Tabla 8. Características de los 49 *bins* de FV x FB descritos en esta investigación. Posición inicial, nombre, genotipo, marcadores flanqueantes, intervalo cubierto y número de marcadores de cada *bin*. Cuando la segregación en el *bin set* es igual para dos o más *bins*, estos se señalan con un color en la casilla '*bin*'. Dos *bins* señalados con el mismo color comparten genotipo. A, genotipo *F. vesca*; B, Genotipo *F. bucharica*; H, Genotipo híbrido; D, genotipo A o H (marcador dominante)

	BIN	GENOTIPO	INICIO	FINAL	INTERVALO (cM)	N° DE MARCADORES
	GL1	32 46 48 59 62 83				
1	0	I:0 B H A B B H	EMFvi072	RS14	12,3	6
2	25,6	I:25 B H A B B A	EMFn136	EMFn136	0,0	1
3	38,6	I:38 B H H B B A	CO380002	UFFxa02F02	11,3	4
4	51,2	I:51 B B H B B A	UFFxa16H07	EMFn182	1,3	3
5	53,8	I:54 B B H B H A	ARSFL013	ARSFL013	0,0	1
6	55,1	I:55 B B H H H H	ARSFL092	ARSFL092	0,0	1
7	57,8	I:58 H B H H H H	EMFv143	CFV-164	3,4	15
	GL2	32 46 48 59 62 83				
8	0	II:0 H B B B B H	EMFvi099	EMFvi099	0,0	1
9	11,7	II:11 H B H B B H	EMFn148	EMFv015	4,0	10
10	16,4	II:16 H H H B B H	EMFn235	EMFn235	0,0	3
11	17,6	II:17 H B H B B H	CFVCT020	APB	10,1	9
12	29,5	II:29 H B H B H H	BFACT-002	EMFv022	13,0	8
13	45,2	II:45 H H H B H H	EMFn134	EMFn134	0,0	1
14	46,5	II:46 H A H B H H	UFFxa09F09	UFFxa09F09	0,0	1
15	50,8	II:51 H A H B A H	UFFxa15H09	CO817431	10,6	2
16	62,6	II:62 H A A B A H	EMFxa379796	EMFxa379796	0,0	1
17	66,2	II:66 H A A B A A	ARSFL-017	CFVCT021	6,0	7
	GL3	32 46 48 59 62 83				
18	0	III:0 A H H H H B	EMFv016	FvNES1	14,1	5
19	21,7	III:21 H B H H H B	VT-673	ChFaM114	2,8	2
20	28,6	III:28 H H H H H B	UDF017	VT-408	3,9	6
21	49,2	III:49 H H H H H H	VT-855	CFVCT035	2,8	6
22	54	III:54 H H H B H H	CFVCT022	VT-022	8,7	13
23	65,4	III:65 H H B B H H	BFACT-045	EMFn207	2,7	3
24	82,6	III:80 H H H B A H	CFVCT011	CFVCT032	2,5	2
25	94,3	III:94 B D D B D D	CFVCT012	CFVCT012	0,0	1
	GL4	32 46 48 59 62 83				
26	0	IV:0 H B H H A H	VT-851	UDF020	12,1	15
27	13,5	IV:13 H B B H A H	VT-021	EMFvi136	24,1	15
28	43,7	IV:43 B B B H H H	BG-1	BG-1	0,0	3
29	57,5	IV:57 B B H H H H	CEL-1	CEL-1	0,0	1
30	66,2	IV:66 H B H B H H	VT-743	VT-743	0,0	1
31	74,9	IV:75 H B H B H A	VT-753	ChFaM-023	2,6	2
	GL5	32 46 48 59 62 83				
32	0	V:0 A B B A A B	CFV-3072	RAN	12,8	6
33	20,2	V:20 A B B A A H	VT-393	VT-393	0,0	2
34	27,7	V:27 A B B H A H	EMFn110	EMFn110	0,0	2
35	33,8	V:34 A B B H H H	CX661792	VT-752	3,4	3
36	37,9	V:38 A B B H H A	VT-850	CFVCT024	0,6	3
37	41,2	V:41 H B B H H A	UDF009	VT-010	35,0	24
	GL6	32 46 48 59 62 83				
38	0	Vt0 A A A H B H	ARSFL-007	VT-378	7,4	4
39	12,1	Vt:12 A A A H A H	EMFn228	EPpC-1830	2,0	3
40	24,7	Vt:24 H A A H A H	PGLM	EMFn117	6,1	8
41	35,5	Vt:35 H A A H A B	EMFv006	ARSFL-022	15,8	16
42	55	Vt:55 H H H H A B	CFVCT017	ZIP	3,7	2
43	64,2	Vt:64 H H H H B B	FAC-005	EMFn153	8,7	4
44	88,1	Vt:88 B H H A H H	FAC-012a	EMFv160BC	13,6	3
	GL7	32 46 48 59 62 83				
45	0	VII:0 B H B A H H	CFV-3096	VT-661	9,5	23
46	13,4	VII:13 B H B H H H	EPpC-9257	ARSFL-099	13,7	11
47	31,8	VII:31 B H H H H H	EPpC-9223	VT-838	1,5	1
48	39,4	VII:39 B A H H B H	BFACT-031	VT-659	19,2	7
49	58,7	VII:58 B A H H H H	ChFaM10	ChFaM10	0,0	1

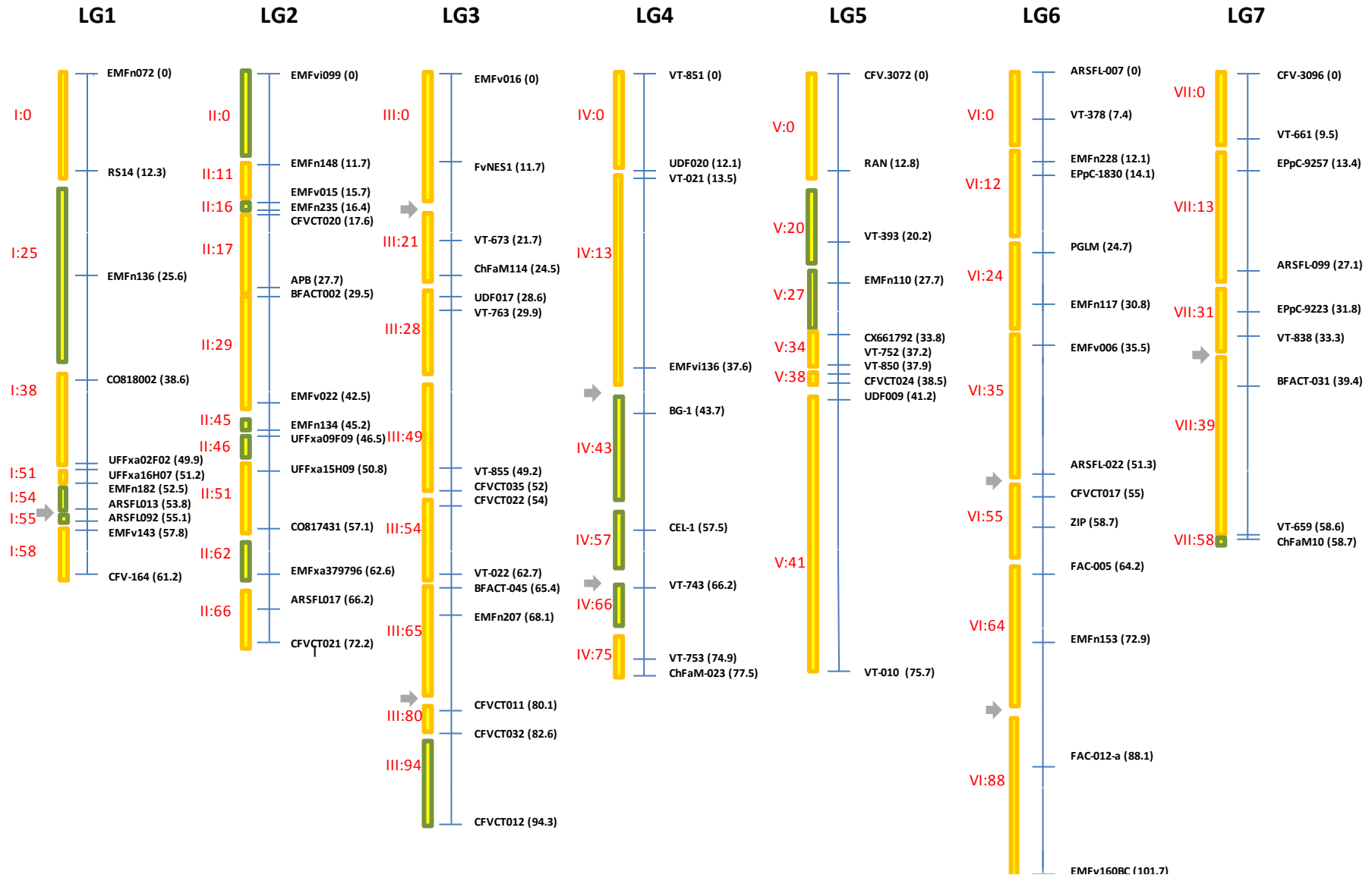


Figura 10. Mapa de bins en FV x FB. Distribución de los bins (en amarillo) en los grupos de ligamiento, representados con líneas verticales azules sobre las que figuran los marcadores que definen cada bin. Los bins delimitados por borde verde son aquellos definidos por una sola posición y cuyos límites no pueden ser definidos aquí. Las flechas grises indican la probable existencia de otros bins no detectados.

2.4 DISCUSIÓN

2.4.1 OPTIMIZACIÓN DE MARCADORES MOLECULARES MEDIANTE EL USO DE BASES DE DATOS

La estrategia empleada para la localización de nuevos marcadores próximos a RFLPs y SSRs descritos por otros autores ha funcionado correctamente en el 76,2% de los casos estudiados. Nuestros resultados son sensiblemente mejores que los obtenidos por otros autores mediante aproximaciones similares en otras especies como el triticale (35,3%), la cebada (61%) o la manzana (15,5%) (Bushman *et al.* 2008; Hackauf *et al.* 2009; Chagné *et al.* 2008). El uso de bases de datos genómicos de *Fragaria* se ha revelado como una aproximación útil para la localización y el estudio de las secuencias cercanas a marcadores moleculares sin necesidad de hacer uso de un mapa físico. A pesar de que se han detectado secuencias parálogas para el 28.5% de los ESTs de *P. persica* empleados en la generación de marcadores, la selección de las secuencias que mostraron un menor valor *E* para el alineamiento ha funcionado correctamente en la mayoría de los casos.

Trece de los 15 polimorfismos estudiados (87%) cosegregaron con los marcadores originales empleando la estrategia de *bin mapping*. En manzano, la tasa de identificación de secuencias parálogas, empleando una estrategia similar, fue del 50% (Chagné *et al.* 2008). Sólo en el caso del marcador CFV-3087 se amplificó un locus parálogo al EST de interés. En el caso del marcador CFV-183, localizado en un *bin* contiguo al de EMFv183, consideramos que su nueva posición es correcta, dado que no se han detectado secuencias parálogas *in silico* y los valores de Chi-cuadrado observados son más próximos al modelo mendeliano para el nuevo SSR ($\chi^2=0,59$ para CFV183 vs. $\chi^2=6,12$ para EMFv183), con valores de distorsión muy similares a los detectados para los marcadores moleculares localizados en la misma región.

Se ha detectado distorsión ($P \leq 0.05$) para los valores de segregación esperados en un 56% de los marcadores moleculares desarrollados aquí, y para el 48,8% de los marcadores que conforman el mapa. Pueden encontrarse segregaciones distorsionadas en prácticamente todos los mapas de ligamiento de todas las especies vegetales estudiadas, debido a factores como la estructura poblacional, la selección gamética y zigótica, la recombinación no homóloga, los elementos transponibles o los agentes ambientales (Jing *et al.* 2007; Knox y Ellis 2002; Zhao *et al.* 2006a; Joobeur *et al.* 1998; Foolad *et al.* 1995). Los valores de distorsión obtenidos en este trabajo son altos, pero semejantes a los encontrados en FV x FB por Sargent *et al.* (2004a) y Ruiz-Rojas *et al.* (2010), con un 54% y 42% de marcadores con ratios de segregación distorsionados respectivamente, así como a otros cruzamientos intraespecíficos publicados en *Fragaria* diploide (Davis y Yu, 1997). En

otras especies, la tasa de distorsión de marcadores moleculares localizados empleando poblaciones F_2 varía considerablemente; con valores que van del 11% en girasol (Lai *et al.* 2005), 33% en arroz (Zhao *et al.* 2006b) y hasta un 54% en *Avena sativa* (Tanhuanpää *et al.* 2008).

La posición descrita para los nuevos marcadores moleculares difiere, en mayor o menor medida, de la posición que ocupan los marcadores a optimizar. Este fenómeno responde a varias causas; la primera de ellas es que hemos empleado un set de marcadores moleculares más amplio que en versiones anteriores del mapa, que ha permitido identificar un mayor número de puntos de recombinación a lo largo de los grupos de ligamiento, y por tanto, las distancias genéticas han quedado modificadas. Sin embargo, algunos de los marcadores que se han desarrollado en este capítulo han ocupado posiciones distintas a las esperadas sin mediación de nuevos marcadores moleculares y empleando un número de individuos similar en su genotipado, como en el caso de marcadores CFV-164 o CFV-183, en los extremos de los grupos de ligamiento uno y dos. La caracterización de estos nuevos loci ha eliminado la amplia distancia genética que separaba los marcadores EMFv164 y EMFv183 del resto de marcadores de sus respectivos grupos de ligamiento (figuras 6 y 8). Puesto que no se han detectado secuencias parálogas en el genoma para estos loci, y dado que las distancias físicas entre los marcadores no superan las 9 Mbp (distancia física insuficiente como para generar una distancia genética tan grande), existe la posibilidad de que algunos de los datos obtenidos por otros autores para la segregación en estos loci incluyeran errores de genotipado.

2.4.2 MAPEO GENÉTICO DE NUEVOS MARCADORES EN EL MAPA DE REFERENCIA FV X FB

El desarrollo de nuevos marcadores presentado aquí, así como el análisis de datos de segregación de otros marcadores desarrollados por otros laboratorios, ha permitido la construcción de un nuevo mapa formado casi en su totalidad por marcadores de tipo PCR (294, de los cuales 210 son SSRs) y un marcador morfológico (*r*). La densidad del mapa es de un marcador cada 1,87 centiMorgans. Este valor es un 19,7% menor al obtenido por Sargent *et al.* (2006), un 25,8% menor al de Vilanova *et al.* (2008) y un 16,8% mayor al obtenido por Ruiz-Rojas *et al.* (2010).

La simplicidad técnica del genotipado con marcadores SSR ha sido esencial para conseguir un amplio set de datos de segregación con un mayor número de individuos analizados, mejorando considerablemente el poder de resolución del mapeo de los nuevos marcadores con respecto al mapeo de RFLPs de Vilanova *et al.* (2008). Las amplias distancias genéticas obtenidas por estos autores podrían ser atribuidas al origen eucromático de las sondas empleadas en el desarrollo de los

marcadores, ya que el 70% de SSRs de FV x FB son de origen genómico. En ocasiones, los marcadores de origen genómico se localizan en regiones centroméricas o teloméricas (Edmé *et al.* 2006) generando *clusters* de marcadores (Liebhard *et al.* 2002; Sargent *et al.* 2004a), ya que la recombinación en estos loci no es aleatoria, sino que responde al control ejercido por centrómeros y telómeros que, normalmente, suprimen la recombinación durante la meiosis de los gametos (Simchen y Hugerat, 1993; Dawe, 1998). En las regiones eucromáticas, en cambio, la tasa de recombinación es mayor y, en consecuencia, el ligamiento es menor. Sin embargo, el trabajo de Vilanova *et al.* (2008) no constata la expresión de los ESTs de *P. persica* en fresa, por lo que las distancias genéticas descritas pudieron deberse a artefactos instrumentales asociados a la tecnología empleada. Al sustituir estos marcadores por otros de tipo PCR, el orden que ocupan los marcadores AC32, AC24 y AG53 situados en uno de los extremos del GL4 del mapa de Vilanova *et al.* (2008) ha quedado invertido en esta investigación, una vez definida la posición de sus marcadores asociados, CFV-3135, CFV-3138 y CFV-3819, que pasan a formar parte de regiones más centrales del grupo. Igualmente, la posición relativa de los marcadores situados en uno de los extremos del GL7 de Vilanova *et al.* (2008), MC45 y CoMET se mantiene, pero sus marcadores asociados, un SSR y un SNP, cambian su posición; el marcador CFV-3117, que mantiene una distancia física de 16 kpb con MC45, pasa de ocupar la posición 81.7 cM del mapa de Vilanova *et al.* (2008) a la posición 16.1 cM en el mapa desarrollado aquí. MC45, del que se han detectado secuencias parálogas tanto mediante *southern blot* (Vilanova *et al.* 2008) como *in silico* aquí (tabla 4), fue mapeado empleando sólo 26 individuos, por lo que es probable que se corresponda con el locus CFV-3117, aunque no podemos concluir que así sea. No se han encontrado duplicaciones del marcador CoMET, que ocupa la posición 90,3cM del grupo en el mapa de Vilanova *et al.* (2008). Sin embargo, el SNP segregante detectado en el amplicón generado por el par de cebadores CFV-3217 ocupó la posición 51,7 cM de nuestro mapa, cosegregando con EMFv023, reduciendo la longitud del grupo de ligamiento en 26 cM.

La identificación de un SSR de herencia codominante cosegregando con EMFv164, de herencia dominante, y por tanto, menos informativo, ha permitido cerrar el *gap* que Sargent *et al.* (2004a, 2006, 2008) y Vilanova *et al.* (2008) definieron al final del GL1. Igualmente, la caracterización del locus CFV183 ha permitido demostrar que la distancia genética que separaba al marcador EMFv183 del resto de marcadores del grupo es de origen instrumental y no genético.

Aunque la presencia de *clusters* de marcadores estrechamente ligados es menor aquí que en otras versiones del mapa, pueden identificarse grupos de hasta ocho marcadores ocupando la misma posición (posición 57.8 cM del GL1, por ejemplo). Este fenómeno ha sido descrito en numerosas

ocasiones en otras especies como el tomate (Arenshchenkova y Ganal, 1999), donde los autores localizan estas regiones en los centrómeros de los cromosomas, ya que la recombinación está generalmente reprimida, dando una errónea impresión de proximidad entre marcadores. Este patrón ha sido también descrito en diversos cultivos como *Phaseolus* (Blair *et al.* 2003), *Helianthus* (Tang *et al.* 2002), *Pinus* (Elsik y Williams 2001) o *Hordeum vulgare* (Li *et al.* 2003). Ramsay *et al.* (2000) señalaron que la supresión de la recombinación en las regiones centroméricas en cebada puede estar acentuada en cruzamientos entre especies alejadas filogenéticamente. Dado que la población FV x FB deriva de un cruzamiento interespecífico, la lejanía filogenética podría ser uno de los factores implicados en la presencia de *clusters* de marcadores en el mapa de referencia de *Fragaria* (Sargent *et al.* 2006).

El *bin* set creado por Sargent *et al.* (2008) contenía 46 *bins* y fue generado a partir de los datos de segregación de 200 marcadores moleculares y 3 marcadores morfológicos. Tras la adición de 94 nuevos marcadores al mapa de referencia de *Fragaria*, el mapa de *bins* de Sargent *et al.* (2008) se ha visto sensiblemente modificado y queda ahora formado por 49 *bins*, tras la aparición de los *bins* II:62, III:21, IV: 66 y VII:58 y la desaparición del *bin* II:88, localizado en el extremo del GL2 y definido únicamente por el marcador EMFv183. El marcador CFV183, que sustituye a EMFv183, ha sido incluido en el *bin* contiguo, II:66. El número de *bins* dobles se ha mantenido respecto a la versión anterior. Sin embargo, hemos reducido el número de *bins* no descritos (aquellos que teóricamente deben existir pero que no han sido observados hasta el momento, por ausencia de marcadores ligados a ellos), que ha pasado de 9 a 8 en esta investigación.

El mapa genético generado aquí proporciona una nueva plataforma compuesta exclusivamente por marcadores de tipo PCR para el desarrollo de nuevos recursos genómicos en *Fragaria*, abriendo las puertas al uso en gran escala de sus marcadores moleculares para el avance en la mejora asistida por marcadores en fresa. Su naturaleza interespecífica garantiza un alto nivel de polimorfismos, esencial para la localización de genes de función conocida, que son menos polimórficos en poblaciones intraespecíficas (Deng y Davis, 2001). Si tenemos en cuenta las relaciones de sintenia y colinearidad identificadas por otros autores entre los genomas diploide y octoploide (Sargent *et al.* 2009b, Rousseau-Gueutin *et al.* 2008), los marcadores desarrollados y/o analizados en este trabajo podrán ser analizados en otras poblaciones de *Fragaria*, tanto silvestres como cultivadas, siendo de gran utilidad para saturar mapas en regiones sinténicas poco caracterizadas, como sucede en la gran mayoría de mapas de *Fragaria x ananassa*. Además, la mejora en la caracterización de los *bins* del mapa de *Fragaria* supone un avance para el mapeo selectivo de nuevos loci en el mapa de

referencia, que pueden ser transferidos a otras poblaciones en función de su localización en el mapa de *bins*. Dicho mapa queda ahora reforzado por un mayor número de marcadores al empleado por Sargent *et al.* (2008), que ha resultado ser de gran utilidad para el mapeo selectivo de nuevos marcadores desarrollados en este trabajo.

3. DESARROLLO Y CARACTERIZACIÓN DE UNA GENOTECA DE BACs EN *FRAGARIA VESCA*

El contenido de esta sección ha sido publicado en la revista

BMC Research Notes

bajo el título

'The development and characterisation of a bacterial artificial chromosome library for *Fragaria vesca*.'

Bonet J, Lopez-Girona E, Sargent DJ, Muñoz-Torres MC, Monfort A, Abbott AG, Arús P, Simpson DW, Davik J
(2009)

BMC Research Notes 2:188

doi:10.1186/1756-0500-2-188

Acceso web al artículo:

<http://www.biomedcentral.com/content/pdf/1756-0500-2-188.pdf>

<http://www.biomedcentral.com/content/supplementary/1756-0500-2-188-S1.DOC>

Este trabajo se enmarca dentro de una colaboración entre la Universidad de Clemson (USA), el *East Malling Research Center* (UK) y el CRAG (España) para el desarrollo de un mapa físico en *Fragaria* diploide. La contribución del autor de esta Tesis Doctoral a este artículo se limita a la generación de un sistema de análisis y caracterización de la genoteca y redacción del manuscrito, no habiendo participado en su construcción. Los coautores de esta publicación están informados y de acuerdo con su inclusión en esta Tesis Doctoral.

La bibliografía citada en el artículo está englobada en la bibliografía general de la Tesis. La numeración de tablas y figuras no se corresponde con la versión pública, sino que responde al orden establecido en esta Tesis. Debido a las restricciones de la revista, no pudieron publicarse ciertas figuras que sí se incluyen en esta versión, para proporcionar un mayor entendimiento del trabajo realizado.

3.1 RESUMEN

La fresa cultivada, *Fragaria x ananassa*, es una de las frutas de mayor importancia económica en nuestro país. Sin embargo, existen pocos recursos genómicos estructurales disponibles para el género y hay una necesidad urgente de desarrollar herramientas para el mapeo físico de secuencias específicas del genoma de *Fragaria*. El primer paso en el desarrollo de un mapa físico es la construcción y caracterización de una genoteca de BACs de alto peso molecular. Por ello, hemos construido una genoteca formada por 18.432 clones a partir de la accession 'Ali Baba' de *Fragaria vesca* f. *semperflorens*.

Los BACs que forman parte de la genoteca contienen un inserto de un tamaño medio de 85 kb, que representa una cobertura genómica aproximada de siete equivalentes. Para su caracterización, el ADN de cada clon ha sido incluido en una serie de mezclas o *pools* de ADN, con objeto de crear un ensayo basado en la identificación de BACs mediante amplificación por PCR, donde cualquier clon puede ser localizado en sólo 40 reacciones de PCR desarrolladas en tres fases: 1) Identificación de *superpools*, 2) Identificación de fila, columna y placa, 3) deconvolución del *pool*. Estos *pools* han sido empleados para localizar e identificar clones con secuencias SSR caracterizadas y distribuidas de manera homogénea por todo el el mapa de referencia del genoma de *Fragaria* diploide (FVxFB).

Los *pools* y *superpools* de ADN han permitido la identificación de un conjunto de BACs que contienen 70 marcadores moleculares previamente mapeados en FVxFB. El número de colonias identificadas para cada marcador sugiere que la genoteca abarca entre 4 y 10 veces el genoma completo de *Fragaria* diploide. Este dato está en concordancia con las estimaciones de cobertura basadas en el tamaño medio de los insertos.

Short Report

Open Access

The development and characterisation of a bacterial artificial chromosome library for *Fragaria vesca*

Julio Bonet¹, Elena Lopez Girona², Daniel J Sargent², Monica C Muñoz-Torres^{3,5}, Amparo Monfort¹, Albert G Abbott³, Pere Arús¹, David W Simpson² and Jahn Davik^{*4}

Address: ¹IRTA, Centre de Recerca en Agrigenòmica CSIC-IRTA-UAB, 38348 Cabells, Spain, ²East Malling Research (EMR), New Road, East Malling, Kent ME19 6BJ, UK, ³Clemson University, Department of Genetics and Biochemistry, Clemson, SC 29634 USA, ⁴Bioforsk Midt-Norge, Kvithamar, N-7500 Stjørdal, Norway and ⁵Current address: Georgetown University, Department of Biology, Washington, DC 20057 USA

Email: Julio Bonet - julio.bonet@irta.cat; Elena Lopez Girona - elenalopezgirona@hotmail.com; Daniel J Sargent - dan.sargent@emr.ac.uk; Monica C Muñoz-Torres - mcm262@georgetown.edu; Amparo Monfort - amparo.monfort@irta.cat; Albert G Abbott - aalbert@clemson.edu; Pere Arús - pere.arus@irta.cat; David W Simpson - david.simpson@emr.ac.uk; Jahn Davik* - jahn.davik@bioforsk.no

* Corresponding author

Published: 23 September 2009

Received: 23 June 2009

BMC Research Notes 2009, 2:188 doi:10.1186/1756-0500-2-188

Accepted: 23 September 2009

This article is available from: <http://www.biomedcentral.com/1756-0500/2/188>

© 2009 Davik 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.

Abstract

Background: The cultivated strawberry *Fragaria ×ananassa* is one of the most economically-important soft-fruit species. Few structural genomic resources have been reported for *Fragaria* and there exists an urgent need for the development of physical mapping resources for the genus. The first stage in the development of a physical map for *Fragaria* is the construction and characterisation of a high molecular weight bacterial artificial chromosome (BAC) library.

Methods: A BAC library, consisting of 18,432 clones was constructed from *Fragaria vesca* f. *semperflorens* accession 'Ali Baba'. BAC DNA from individual library clones was pooled to create a PCR-based screening assay for the library, whereby individual clones could be identified with just 34 PCR reactions. These pools were used to screen the BAC library and anchor individual clones to the diploid *Fragaria* reference map (FV×FN).

Findings: Clones from the BAC library developed contained an average insert size of 85 kb, representing over seven genome equivalents. The pools and superpools developed were used to identify a set of BAC clones containing 70 molecular markers previously mapped to the diploid *Fragaria* FV×FN reference map. The number of positive colonies identified for each marker suggests the library represents between 4× and 10× coverage of the diploid *Fragaria* genome, which is in accordance with the estimate of library coverage based on average insert size.

Conclusion: This BAC library will be used for the construction of a physical map for *F. vesca* and the superpools will permit physical anchoring of molecular markers using PCR.

Research Hypothesis

The strawberry species *Fragaria ×ananassa* is a member of the Rosaceae, and is one of the most economically-import-

tant soft-fruit species. However, it is an allo-octoploid species ($2n = 8x = 56$), which grants it genetically complex status. *Fragaria vesca* is a diploid ($2n = 2x = 14$) relative of

3.2 INTRODUCTION

The strawberry species *Fragaria × ananassa* is a member of the *Rosaceae*, and is one of the most economically-important soft-fruit species. However, it is an allo-octoploid species ($2n = 8x = 56$), which grants it genetically complex status. *Fragaria vesca* is a diploid ($2n = 2x = 14$) relative of the cultivated strawberry, that has been demonstrated to be the closest extant diploid to the ancestor of the cultivated octoploid strawberry (Bringhurst, 1990; Rousseau-Gueutin *et al.* 2009). *F. vesca* has a genome size of approximately 206 Mb/C (Folta & Davis 2006), a short reproductive cycle and facile vegetative and seed propagation, as well as amenability to genetic transformation. These features make it an ideal model organism for forward and reverse genetics studies in *Fragaria*.

There currently exists a well characterised, saturated genetic linkage map for diploid *Fragaria* developed from an interspecific cross between *F. vesca* and another closely-related diploid species, *F. bucharica* (FV×FB). Comparisons between this map and maps of the cultivated strawberry have shown that the diploid genome of *F. vesca* is essentially completely collinear with the four genomes that make up the allooctoploid genome of the cultivated strawberry (Rousseau-Gueutin *et al.* 2008). Despite a growing body of molecular genetics resources, to date, very few structural genomic resources have been reported for the genus *Fragaria* (Davis *et al.* 2006). High molecular weight DNA libraries are an essential resource for genomic investigations, including positional cloning of genes of economic importance, such as the *fw2.2* QTL in tomato (Frary *et al.* 2000); for the determination of gene structure and function (Xu *et al.* 2001); for region-targeted marker development (Georgi *et al.* 2002); comparative genome analysis (Jung *et al.* 2006); genome sequencing (Zhang & Wu 2001); and the construction of physical maps (Zhebentyayeva *et al.* 2008). Due to their stability, ability to maintain relatively high molecular weight genomic DNA inserts (over 150 kb), and their ease of handling and propagation, bacterial artificial chromosome (BAC) libraries have become the vector of choice for the development of DNA libraries in many plant species, including *Malus*, *Prunus* and *Rosa* in the *Rosaceae* (Xu *et al.* 2001, Kaufmann *et al.* 2003, Horn *et al.* 2005).

In this paper, we report the development of a high molecular weight DNA BAC library from a diploid strawberry, *F. vesca* f. *semperflorens* cv. 'Ali Baba', as a resource for the construction and characterisation of a physical map for diploid *Fragaria* and for positional cloning efforts. We have calculated that the library contains an average insert size of 85 kb and thus represents an estimated 7.6× coverage of the *Fragaria* genome. To allow rapid characterisation of this library, we developed a set of PCR pools and superpools from the library and identified 102 BAC clones containing 70

genetically mapped molecular markers (ten per linkage group), distributed throughout the seven diploid *Fragaria* linkage groups. These results demonstrate the efficiency of our PCR pools and superpools and provide a foundation for the development of physical mapping resources for the genus.

3.3 METHODS

3.3.1 HIGH MOLECULAR WEIGHT DNA PREPARATION

A total of 20 g of young leaf tissue from the *Fragaria vesca* f. *semperflorens* cultivar 'Ali Baba' that had been given 72 h dark treatment at 16°C, was snap-frozen in liquid nitrogen and ground to a fine powder. The powdered tissue was then transferred to a 1 litre flask with 200 ml of nucleic isolation buffer ((NIB) 10 mM Tris, 10 mM EDTA, 100 mM KCl, 500 mM sucrose, 4 mM spermidine, 1 mM spermine, 2% w/v PVP MW 40000, 0.13% DIECA, 0.1% ascorbic acid, and 0.2% β -mercaptoethanol (BME)). The homogenate was shaken gently for 30 minutes at room temperature, before filtering through 2 layers of cheesecloth and 2 layers of Miracloth into a clean flask. A 1/20th volume of lysis buffer (10% Triton-X 100 in NIB) was added and the flask was incubated on ice for 10 min, gently swirling every other minute.

Aliquots were transferred to 50 ml falcon tubes and the nuclei spun down at 1800 g for 15 min at 4°C. The supernatant was discarded and the nuclei resuspended in approximately 1 ml of NIB. Nuclei from two tubes were then combined, NIB was added to fill the tube and it was centrifuged at 1800 g for a further 15 min at 4°C. This procedure was repeated until all nuclei had been combined in a single tube. The supernatant was removed and 2 ml of NIB (without BME) was added along with an equal volume of 1.5% low-melting temperature agarose. The mix was then allowed to solidify in plug moulds at room temperature.

3.3.2 TREATMENT OF THE PLUGS

After solidification, the plugs were treated twice with EPS buffer (0.5 M EDTA pH9.2, 1.0% sarcosyl, 0.1% BME, 1 mg/ml proteinase K) in 50 ml falcon tubes and incubated at 50°C for 24 h. Subsequently, the plugs were rinsed twice in sterile water with 0.1% BME before washing twice in 1 mM PMSF, 10 mM Tris, 10 mM EDTA at room temperature. The plugs were then rinsed twice with 0.1% BME followed by a double wash for 60 min each time with 10 mM Tris, 1 mM EDTA.

3.3.3 PARTIAL DIGESTION AND SIZE SELECTION OF THE HIGH MOLECULAR WEIGHT DNA

Three units of *Hind*III was used to digest eight DNA plugs. Size selection of *Fragaria* DNA was performed with two consecutive rounds of pulse field gel electrophoresis (PFGE), conducted at 12°C, with 1 - 40 sec switch time, 6 V/cm, 120 deg angle for 18 hrs on a CHEF-DR-II system (Bio-Rad, Hercules, CA, USA). Fragments between 100 and 300 Kb in size were cut out and divided into three equal sections. A second size selection was performed on the three sections but the DNA was allowed to electrophorese into low melting point agarose with 3 - 5 sec switch time. Subsequently, fragments in the 100 to 300 Kb size range were excised from the gel and cut into cubes for electro elution using a Bio-Rad electro eluter (Bio-Rad, Hercules, CA, USA), following the manufacturer's protocol.

3.3.4 LIGATION AND TRANSFORMATION

The eluted DNA was ligated into the pIndigoBAC-536 vector (fig. 11 (Luo & Wing 2003)) using a target DNA to vector ratio of 10:1 and transformed into ElectroMAX™ DH10B™ *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA) by electroporation following manufacturer's suggestions. Transformations were plated on Q-trays with LB-agar medium containing X-gal, IPTG and chloramphenicol. The colonies were grown overnight and picked with a Genetix robot (Genetix, New Milton, UK) following the manufacturer's operating guidelines. A total of 18,432 colonies were picked and arrayed into forty-eight 384-well plates.

3.3.5 ESTIMATION OF BAC CLONE INSERT SIZES

Thirty-eight randomly-selected BAC clones were grown for 16 h at 37°C in 5 ml of LB containing 12.5 µg/mL chloramphenicol. BAC DNA was extracted from 38 clones following a standard alkaline lysis miniprep technique (Sambrook *et al* 1989), digested with *Not*I and insert sizes were estimated against the λ size standard using PFGE (Fig. 12) in a CHEF system (Bio-Rad, Hercules, CA, ISA).

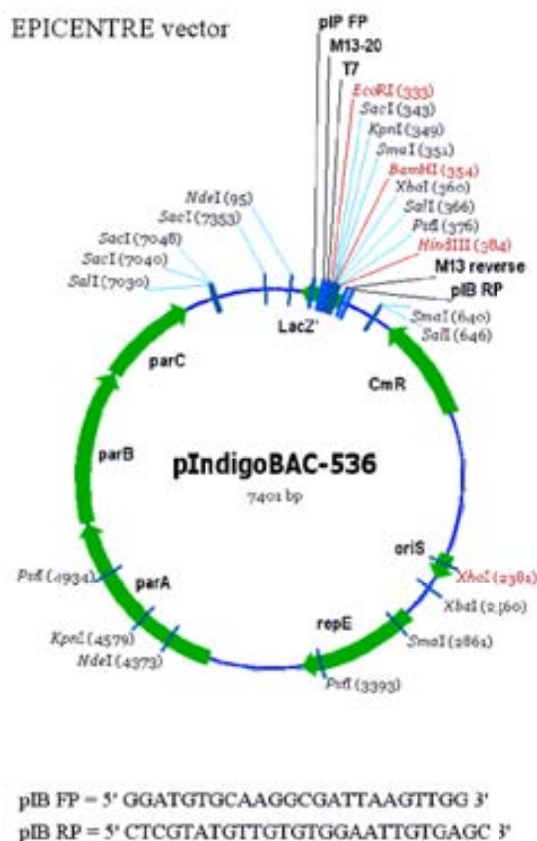


Figure 11. Schematic outline of the plasmid pIndigo-536. The BAC vector has the *lacZ* reporter gene and a multiple cloning site region. Source: <http://cnrgv.toulouse.inra.fr/fr>. This figure is not included in the public version of this paper.

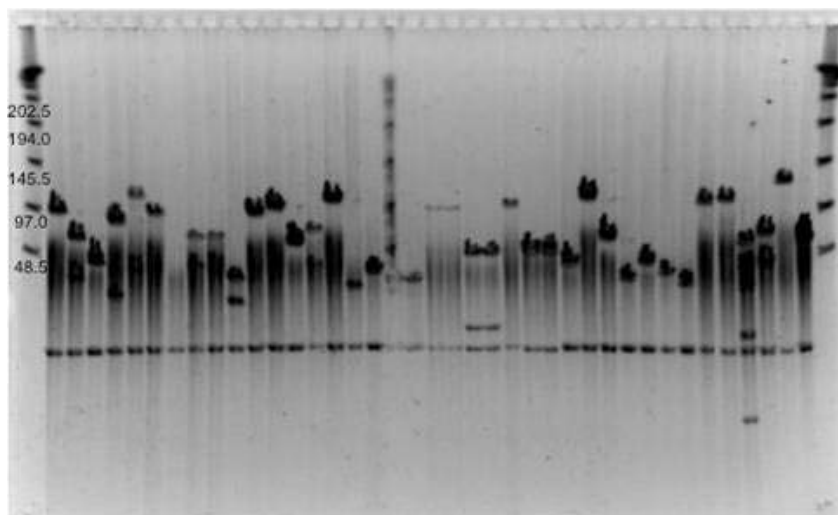


Figure 12. Analysis of BAC clones with PFGE. Analysis of BAC clones using PFGE. Randomly selected clones from the *F. vesca* 'Ali Baba' BAC library were digested with *NotI*. The clones were electrophoresed through a 1% agarose CHEF gel (0.5× TBE). The size standard loaded at each end is Lambda Ladder (NEB) and fragment sizes are indicated in Kb on the left. The size marker in lane 19 is Midrange II (NEB). This figure is not included in the public version of this paper.

3.3.6 POOL AND SUPERPOOL DEVELOPMENT

The 18,432 clone library, contained in forty-eight 384-well plates, was condensed into forty-eight 96-well plates (4,608 cultures), with each culture containing four adjacent clones from the original 384-well plates. BAC colony cultures were then pooled in three separate orientations (fig. 13) from the 4,608 culture pools (48 plate pools, 48 column pools and 32 row pools containing the entire library). The DNA was isolated from the pooled cultures following a standard alkaline lysis miniprep technique (Sambrook *et al.* 1989). Subsequently, DNA from each of six plates (2,304 BAC clones each) was pooled to create eight superpools for prescreening of the library.

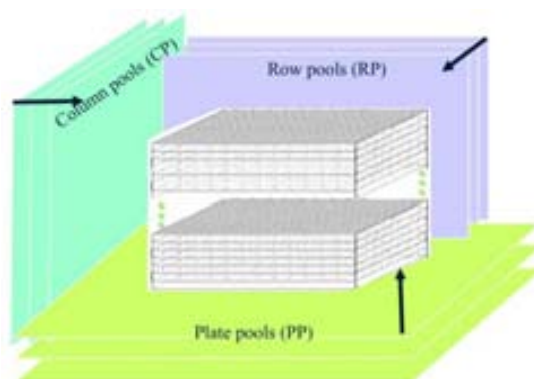


Figure 13. Three-fold BAC DNA pooling strategy. Forty-eight 384-well plates containing 18,432 individual BAC clones were arranged in four three-dimensional stacks consisting of 48 layers (or plates), 48 columns and 32 rows. Each stack was pooled on the three unique coordinate axes as shown to generate a total of 128 DNA pools. This figure is not included in the public version of this paper.

3.3.7 LIBRARY SCREENING AND IDENTIFICATION OF BAC CLONES CONTAINING MARKERS FROM THE DIPLOID *FRAGARIA* REFERENCE MAP

A set of 70 molecular markers (47 SSRs, one SNP, two EST-SSRs, nine ESTs, and 11 gene specific STS markers) evenly spaced throughout the seven linkage groups of the diploid *Fragaria* reference map (ten per linkage group) (Sargent *et al.* 2008; Vilanova *et al.* 2008) were selected for PCR analysis of pooled BAC DNA (see Table 9). PCR reactions were performed using the DNA superpools as template, followed by a second round of PCR with the DNA plate, row and column pools corresponding to the positive DNA superpools. Reactions were performed following touch-down conditions from Sargent *et al.* (2003) from 55-50°C and visualised following agarose gel electrophoresis. The four clones identified by each positive PCR were then grown overnight and used as template for PCR following the procedures described above, and single positive BAC clones were identified following electrophoresis and visualization over UV light (fig 14).

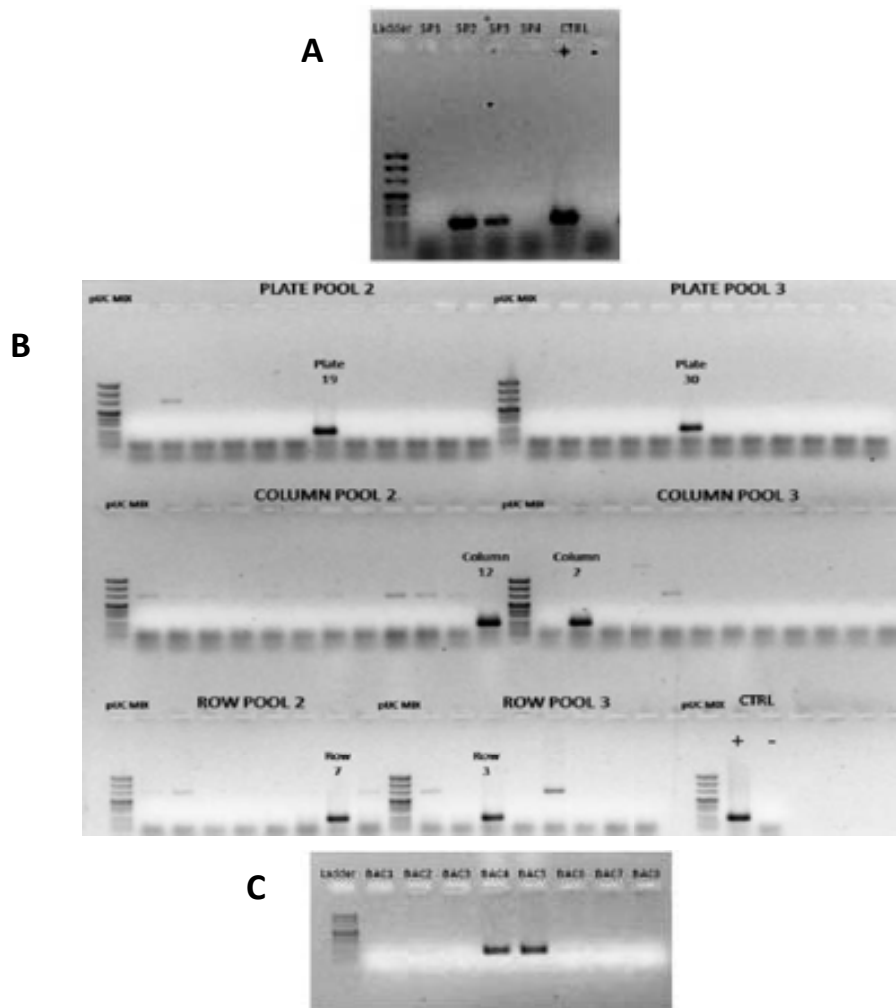


Figure 14. BAC screening of an SSR locus, ARSFL007, in two different DNA bulked superpools. A: Superpool screening. **B:** Plate, column and row screening. **C:** Deconvolution of individual BACs. Positive controls contain strawberry genomic DNA. This figure is not included in the public version of this paper.

3.4 RESULTS AND DISCUSSION

3.4.1 BAC CLONE INSERT SIZES AND LIBRARY GENOME COVERAGE

The *Fragaria* BAC library created is composed of 18,432 BAC clones arrayed into forty-eight 384-well microplates. Of the 38 randomly-selected BAC clones analysed to assess the average insert size of the library, all but one (97%) contained inserts. The insert sizes of the clones ranged from 35 kb to 145 kb, with an average insert size of 85 kb, which is comparable with BAC libraries developed for other Rosaceous crops such as peach (70 kb; (Georgi *et al.* 2002)), rose (102 kb; (Kaufmann *et al.* 2003)) and apple (110 kb; (Xu *et al.* 2001)). The *F. vesca* genome size has been estimated to be in the region of 206 Mb (Folta & Davis 2006), and thus, a total of 18,432 clones with an average insert size of 85 kb represents coverage of approximately 7.6× the *F. vesca* genome. The library has been denoted

CUFvAB (Clemson University *F. vesca* 'Ali Baba') for ease of reference, and each clone is identified by its plate number and 384-well cell reference i.e. CUFvAB04A01 refers to clone A01 on plate 4.

3.4.2 PCR ANALYSIS OF BAC LIBRARY POOLS AND SUPERPOOLS

All markers screened on the superpools of the *F. vesca* 'Ali Baba' BAC library amplified a PCR product of the expected size in relation to amplicons generated from *F. vesca* genomic DNA, and when single positive colonies were identified, all contained the expected PCR marker, indicating no cross-contamination in the library. The results of the pool and superpool screens suggests the BAC library represents between 4× and 10× coverage of the diploid *Fragaria* genome, which is in accordance with the estimate of the library coverage based on average insert size. The BACs to which the markers are anchored are detailed in table 9. In total, 102 BAC clones anchored to 70 genetically-mapped molecular markers (47 SSRs, one SNP, two EST-SSRs, nine ESTs, and 11 gene specific STS markers) that were evenly distributed (ten per linkage group) across the seven linkage groups of the diploid *Fragaria* reference map were identified and verified in this investigation (fig 15).

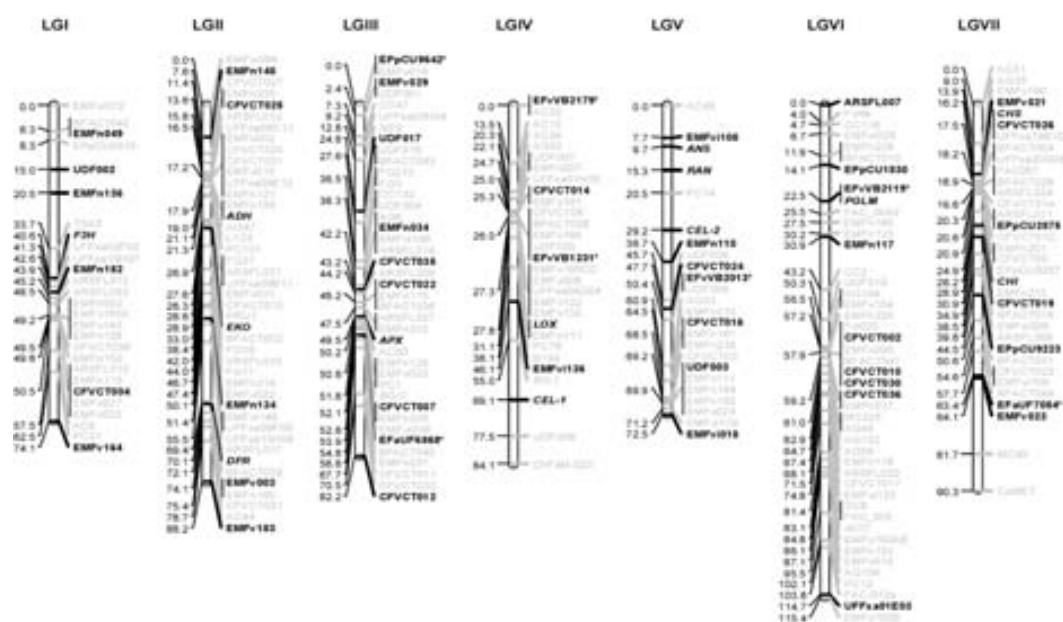


Figure 15. Genetically-anchored BAC clones. The *Fragaria* reference map shows the map locations of the 60 genetic markers used to anchor 89 BAC clones in this investigation. Anchor markers are given in bold. Loci names followed by an asterisk (*) denote loci bin-mapped in the FV×FB population. This figure is not included in the public version of this paper.

3.5 CONCLUSION

We have developed a BAC library consisting of 18,432 clones from DNA of the *F. vesca* f. *semperflorens* accession 'Ali Baba' with an average insert size of 85 Kb. At an effective 7.6× coverage of the *Fragaria* haploid genome, this library provides a good basis for the development of further genomics resources for *Fragaria*. This is the first high molecular weight genomic library reported for the genus and as such, is a valuable resource for genomics studies in *Fragaria*. Despite the initial time investment in the development of the library pools and superpools being relatively high, the availability of a PCR-based library screen means large numbers of markers, including SSRs, can now be rapidly associated with individual BAC clones in an efficient and cost-effective manner as has been demonstrated here. To construct physical maps, BAC libraries consisting of high molecular weight genomic DNA inserts are an essential prerequisite. The library presented here will enable the development of a physical map resource for *Fragaria* and the associated PCR pools and superpools will permit the rapid anchoring of genetic markers to a physical map framework. The BAC library will also prove to be a valuable tool for the continued development of linkage maps for the cultivated strawberry (*F. × ananassa*). The identification of BAC clones anchored to regions of the diploid *Fragaria* reference map in which there is a current paucity of markers on the maps of the cultivated strawberry (Rousseau-Gueutin *et al.* 2008; Spigler *et al.* 2008), will permit the targeted development of polymorphic markers in those regions, as has been previously successfully demonstrated for other species such as peach (Georgi *et al.* 2002).

The library reported here and the associated PCR pools and superpools are available on a cost recovery basis from the communicating author.

Table 9. *Fragaria* molecular markers and associated BAC clones. A list of the 70 genetic markers and the 102 BAC clones they identified through PCR screening is provided. The table also lists the marker type, position on the diploid *Fragaria* reference map and the primer sequences. *Markers bin mapped in the FV×FB population and mapping interval of the bin to which they were located is given.†References: 1: Deng & Davis (2001), 2: Ashley *et al.* (2003), 3: Sargent *et al.* (2003), 4: Cipriani & Testolin (2004), 5: Hadonou *et al.* (2004), 6: Sargent *et al.* (2004a), 7: Sargent *et al.* (2006), 8: Monfort *et al.* (2006), 9: Sargent *et al.* (2007), 10: Vilanova *et al.* (2008). This table is included in the additional files linked to this paper.

Locus name	Marker type	Primer sequences (5' – 3')	Linkage group	Map position	BAC clones	Ref†
EMFn049	SSR	F: GCTAGCGTTCAAATTAAGACCA R: GGTATCGGATTGAGGAAATCAA	FG1	6.3	CUFvAB19G11	1
UDF002	SSR	F: TATGGCCAGGATTGTTTCTGCT R: TAGGAGGAGGCGTTGAAATG	FG1	15	CUFvAB01I20	2
EMFn136	SSR	F: TTTCTCTTTTGCTCCATAGTTCC R: TTCATCAGGATCCAGAAGTCC	FG1	20.5	CUFvAB28D19	3
F3H	gene STS	F: GGTTGAAGGCACTGTCCATTA R: GTCACAATCTCGGCCAATCCT	FG1	40.6	CUFvAB07G13	4
EMFn182	SSR	F: GCAACAAAGGAGGTTAGAGTCCG R: TGGTGAGTGCTCATTGTTC	FG1	43.9	CUFvAB31E17 CUFvAB31E23	3
EMFv185	SSR	F: TCCAACAATGATTTCTCTCTGG R: ATCGTGGTTTACTCCTGTGG	FG1	49.2	CUFvAB35B13	5
EMFv143	SSR	F: ACAATTGGAGCGACACCTTC R: CCGCGGTTTGATGAGTAAAA	FG1	49.2	CUFvAB02A09	5
EMFn128	SSR	F: CATCAACATTCACATGAATTTACC R: CGGCGGATCTAGTTTTGAGG	FG1	49.2	CUFvAB42L24 CUFvAB42O04	3
CFVCT004	SSR	F: ATGAGTGGTGCTGGGGTAGT R: CCCATCTTCGCTTCACTCA	FG1	50.5	CUFvAB26C14 CUFvAB32F19	6
EMFv164	SSR	F: AAGTTGGTGCTATGGTGT R: GGTGACTGACAGGGACCACT	FG1	74.1	CUFvAB14K20	5
EMFn148	SSR	F: TTACTCTCACAGAAACAACG R: CAACTTCTCTCACTCACC	FG2	7.6	CUFvAB44G19	3
CFVCT025	SSR	F: CCTTCACTATGCTGCTCCAA R: GAGGTTCCCTGTGCATTGT	FG2	13.6	CUFvAB17M15	5
ARSFL012	SSR	F: GCGGAACCAAGCCAATAAGATG R: GCGACCACGACAGTTTCTCACTCT	FG2	15.8	CUFvAB08P13 CUFvAB31A05	7
ADH	SNP	F: CCAAGGTACACATCTTTTTTTC R: CTCTCCACAATCCTGAATTTTAG	FG2	17.9	CUFvAB04F12	8
EMFv031	SSR	F: AGGTTTTGTTTGCTTTTT R: GTGTTTTGCTACTACTTGAT	FG2	27.6	UCFvAB21O16	5
EKO	gene STS	F: AGAAGAAACGAATTGCTTCAGG R: TTGGTCCATTATCAGTGTC	FG2	28.9	CUFvAB03P21	6
EMFn134	SSR	F: TGATTCTTTGAAAGGCTTTGG R: AAAACAACCCCTCTCATCC	FG2	50.1	CUFvAB40B03 CUFvAB40K09	3
DFR	gene STS	F: CACCGGAGTGTTTCATGTCCG R: AACCTCCGAACGTCTTTGC	FG2	70.5	CUFvAB06L09	4
EMFv003	SSR	F: CTCTGATTCTTCTCGTCCACCAT R: TCCCCAGAGAATTAACAGTCGTA	FG2	74.1	CUFvAB31I23 CUFvAB42L09 CUFvAB44A05	3
EMFv183	SSR	F: TAGCCTTGCCAATCAAATCC R: CCACCTTTGGAAGACCAGTGC	FG2	88.2	CUFvAB07P21	3
EMFv029	SSR	F: TACTATTGAAGAACTCCTACTGA R: TCTTTGATCTGCTTCCACCTT	FG3	2.4	CUFvAB45I03	5
UDF017	SSR	F: GGACGTTCCACATCCGTAGA R: CGGTGGAGATGTGATTTTATGA	FG3	24.8	CUFvAB19B09	2
EMFn034	SSR	F: GCCTCAAAGATCACTCATTTC R: TCTTCATCTTTCAACCTCAAA	FG3	42.2	CUFvAB22O08	1
CFVCT035	SSR	F: CCGGTCAAACACCAAAACT R: CTGGAAAGGAAACGATTGGA	FG3	43.2	CUFvAB40P19 CUFvAB38P20	6
CFVCT022	SSR	F: ATAATCCCCTTCCCATCACC R: GACATTTTTAGCGGGTGTC	FG3	46.2	CUFvAB1M19 CUFvAB5O19 CUFvAB10M19	6
APX	gene STS	F: GCATATCTGGTCTCGATTCTGC R: CCCATAGAAACAACCTACTGTC	FG3	49.5	CUFvAB13L21	9
CFVCT007	SSR	F: AAACAACACTCAAAGTACCGATCT R: TTCAACATGCAAGCCTGATA	FG3	52.1	CUFvAB05I20 CUFvAB44P15	6

Table 9 (cont)

Locus name	Marker type	Primer sequences (5' – 3')	Linkage group	Map position	BAC clones	Reft
CFVCT012	SSR	F: GGAAAGGTTGGTACGACAGG R: CCCCTCCACAATATCTCTT	FG3	82.2	CUFvAB34M13 CUFvAB36J2	6
EPpCU9642	EST	F: TTCAGTTGGCAGATCCTGTG R: TGCTGAGACCCCTCCAATTT	FG3	0,0-26,7*	CUFvAB33F02 CUFvAB21C08	8
EFaUF6868	EST	F: GCTCTCCAGGTCGAGTACG R: GTTCCACTTGGGCAGTTGT	FG3	53,9-73,8*	CUFvAB15C01 CUFvAB15M01 CUFvAB38A09	8
UDF007	SSR	F: TGAGTAAATGATGCAACCCAGA R: GCTTGAGTATGTATTTGAGTGTATGTG	FG4	24.7	CUFvAB14K12	2
UFFxa01H05	EST-SSR	F: GGGAGCTTGCTAGCTAGATTTG R: AGATCCAAGTGTGGAAGATGCT	FG4	25	CUFvAB15P07	3
CFVCT014	SSR	F: GGCACCACGGATTCAAGTA R: TGTTGCGTTTTCAAGCTCAC	FG4	25.3	CUFvAB15P07 CUFvAB32L14	6
EMFv180	SSR	F: AAGTTGGCTGCCTATGGTGT R: GGTGACTGACAGGGACCACT	FG4	26	CUFvAB14K20 CUFvAB33E08	3
EMFvi136	SSR	F: GAGCCTGCTACGCTTTTCTATG R: CCTCTGATTGATGATTTGCT	FG4	46.1	CUFvAB14I04	10
CEL1	gene STS	F: TACATCATGGTTGTGTCTGTCC R: TCCAGAAAGCCTGATTCTAAGG	FG4	68.3	CUFvAB23O15	9
UDF020	SSR	F: ATGTCCATTTGCCGACATTT R: GGGTTTATTGGGTTGGGTTT	FG4	26	CUFvAB32F24	2
UDF008	SSR	F: TGTTTGCCTGCCGATTATTA R: TTAGCTCGCGTAAACTTCAGA	FG4	77.5	CUFvAB25P04	2
EFvVB2179	EST	F: ATCTGCTGACAATGCAAAG R: AAGAGCCTTCAGTTGCTCCA	FG4	0,0-26,0*	CUFvAB06A13 CUFvAB40C09	8
EFvVB1231	EST	F: CCAACTGTGACATCCACGAC R: GCTGTCACGCAGAAAATCAA	FG4	26,0-46,1*	CUFvAB41P20 CUFvAB08A14	8
EMFvi108	SSR	F: GGACCCCAAACATTGAATAAA R: GAAGAGGGGAGGAGCAATAAAT	FG5	7.7	CUFvAB03H03 CUFvAB30G01 CUFvAB41E14	10
ANS	gene STS	F: AAGGAGAAGTATGCCAATGACC R: CTCCCTTCTTAATCCCAAGC	FG5	9.7	CUFvAB13A02	4
RAN	gene STS	F: TTCACAAAACAACCCAG R: GTACCCATCACCCCACTC	FG5	15.3	CUFvAB21J01	4
CEL2	gene STS	F: GGTGTTTCCAGCCCTTGTGTC R: ATGCTTACCAGCTTTACCTGTC	FG5	29.2	CUFvAB06C16	9
EMFn110	SSR	F: GACGCTTCGGAGACTGAGG R: CCCCTTAAAATAATTAATCTCC	FG5	36.4	CUFvAB44G07	3
CFVCT024	SSR	F: CCCCAGAAAGCTGGTAAATA R: TCTTCAAATTGTGGCTGCAT	FG5	47.7	CUFvAB07E15 CUFvAB14P15 CUFvAB43B11	6
CFVCT016	SSR	F: CACAACGGAGTACACGGCTA R: TCAGATTCACCTCGGACCTC	FG5	68.5	CUFvAB20O01 CUFvAB20O11	6
CFVCT003	SSR	F: TCGTCGATCTCTAAAACAAGGTC R: ACCACCACCCATTCTGAGTT	FG5	69.2	CUFvAB42K22	6
EMFvi018	SSR	F: CAAACATGGAAGGAAAAGAAGG R: GTCAGAGAGACCCATCTGAACC	FG5	72.5	CUFvAB32N02	10
EFvVB2013	EST	F: GTGCAGTTGCCAAAGGAGC R: AGTGGGTTTGTGCTGCT	FG5	50,4-72,5*	CUFvAB08J20	8
ARSFL-007	SSR	F: GCGCGCATAAGGCAACAAG R: GCGAATGGCAATGACATCTTCTCT	FG6	0	CUFvAB19N24 CUFvAB30F04	7
EPpCU1830	EST	F: TGATGCAATTGGCACAAGC R: CCTATCACCACTTACTTCACTGC	FG6	14.1	CUFvAB14H08	8
PGLM	gene STS	F: AAGAGGACCTTGCAAAGTTACG R: GTGTCTTGATTGCTTCAACAGC	FG6	24.5	CUFvAB14N12	9
EMFn117	SSR	F: ATCGGATCAACAAGCAAAGC R: ATGGATGAGGGGAGAAGAGG	FG6	30.9	CUFvAB31015 CUFvAB33M21	3
CFVCT002	SSR	F: GATGCTAAGTAGCCGCACCT R: TGGACCTTTTGGGTATGGT	FG6	57.9	CUFvAB44H23	6
CFVCT010	SSR	F: TGACAGAGACAAATGCATCACA R: CGACGTTTCCCTCTTTCTA	FG6	57.9	CUFvAB42K10	6
CFVCT036	SSR	F: AAGCTTTTCCACCGAGAAT R: CCAACAATCCGATGCCTAAG	FG6	59.2	CUFvAB22L24 CUFvAB34C03	6
CFVCT017	SSR	F: GCATCTCCAAGCTCTCACG R: GCCTAAACCAACCCAAAATC	FG6	71.5	CUFvAB16L06 CUFvAB16K18	6
UFFxa01E03	EST-SSR	F: ACCCCATCTTCTCAAATCTCA R: GACAAGGCCAGAGCTAGAGAAG	FG6	114.7	CUFvAB33P23 CUFvAB33H10	3

Table 9 (cont)

Locus name	Marker type	Primer sequences (5' – 3')	Linkage group	Map position	BAC clones	Ref†
EFvVB2119	EST	F: GCTCGAGCTGATTACGATTACC R: TAAAGGACCCATCAGAGAAACG	FG6	22,5-43,2*	CUFvAB29F20	8
EMFv021	SSR	F: TCATTTTTCAGGGCCACGGGTAGA R: GTGGTGGTTGAGGCAGTGGAGGAT	FG7	16.2	CUFvAB06A15 CUFvAB24E18	5
CFVCT026	SSR	F: CGCGTCAAAGATAACACGAGA R: GGCTTCCGATCGATATTCAA	FG7	17.5	CUFvAB45M14	6
CHS	gene STS	F: TATCCCCTACTACTCTC R: GCTGGCCCCATTCTTAA	FG7	17.5	CUFvAB01G08	4
EPpCU2875	EST	F: AACTCAGAGACATATCTGCACAGG R: AAGTTGAAGCGGTCTTCATAGG	FG7	20.3	CUFvAB24G05	8
EMFn201	SSR	F: CAGCTCAGAAAAGCTCACAGC R: TAGAACGCCAATCACAAACC	FG7	20.6	CUFvAB07B11	3
CHI	gene STS	F: AAGACGGCCGAGGAGTTG R: CGTCAGCGGTAGTATCATTGTC	FG7	28.2	CUFvAB40D22	4
CFVCT019	SSR	F: ATGGCGTCATACCATCATCA R: CTGCGGTTTGAGAGGACTTC	FG7	30.9	CUFvAB05H08 CUFvAB13K05 CUFvAB25K16	6
EPpCU9223	SSR	F: AACAGAGCCAAGCTTATGCGAG R: TTTCTGCGCAACCGCATC	FG7	44.5	CUFvAB13D14	8
EMFv023	SSR	F: AATTACCGAGCTCCCACTA R: CAGCGCTAAAGCGGTTGC	FG7	64.1	CUFvAB24I04	5
EFaUF7084	EST	F: CAGAAGAGGTTCAAGTTCC R: ACACCATAGCAAGCCCTG	FG7	63,4-81,00*	CUFvAB18K01	8

3.6 AUTHORS' CONTRIBUTIONS

JB developed the pools and superpools, screened the library for genetic markers and drafted the manuscript; DJS developed the pools and superpools, screened the library for genetic markers and drafted the manuscript; MCMT developed the BAC library and critically reviewed the manuscript; AM developed the pools and superpools and critically reviewed the manuscript; AGA conceived the idea for a library, planned the experiments, managed the project at Clemson and critically reviewed the manuscript; PA planned the experiments, managed the project at Cabrils and drafted the manuscript; DWS managed the experiments at East Malling and critically reviewed the manuscript; JD developed the BAC library, screened the library for cpDNA contamination and drafted the manuscript. All authors read and approved the final manuscript.

4. LOCALIZACIÓN DE GENES CANDIDATOS PARA LA CALIDAD DEL FRUTO EN EL GENOMA DE *FRAGARIA* DIPLOIDE

El contenido de esta sección ha sido publicado en la revista *Tree Genetics & Genomes* bajo el título
'Development and mapping of peach candidate genes
involved in fruit quality and their transferability and
potential use in other Rosaceae species.'

Le Dantec L, Cardinet G, Bonet J, Fouché M, Boudehri K, Monfort A, Poëssel JL, Moing A, Dirlewanger E.
(2010)

Tree Genetics & Genomes

doi: 10.1007/s11295-010-0308-8

Acceso web al artículo:

<http://www.springerlink.com/content/h316m3u45351057k/>

<http://www.springerlink.com/content/h316m3u45351057k/supplementals/>

Este trabajo se enmarca dentro de una colaboración entre el INRA (Francia) y el CRAG (España), y pertenece al proyecto europeo de investigación integrada ISAFRUIT, cuyo objetivo es la mejora de la calidad de la fruta en Europa. La contribución del autor de esta Tesis Doctoral a este artículo se limita a la localización de los genes candidatos en el genoma de fresa y el estudio de la sintenia entre el genoma de *Fragaria diploide* y de *Prunus persica*. Los coautores de esta publicación están informados y de acuerdo con su inclusión en esta Tesis Doctoral.

La bibliografía citada en el artículo está englobada en la bibliografía general de la Tesis. La numeración de tablas y figuras no se corresponde con la versión pública, sino que responde al orden establecido en esta Tesis. Debido a las restricciones de la revista, no pudieron publicarse ciertas figuras que sí se incluyen en esta versión, para proporcionar un mayor entendimiento del trabajo realizado.

4.1 RESUMEN

La calidad de la fruta determina su valor económico, y por ello, a lo largo de las dos últimas décadas se han identificado y estudiado una serie de caracteres mendelianos relacionados con características agronómicas interesantes en muchas especies de frutales. En la familia *Rosaceae*, el melocotonero es una buena especie modelo, ya que tiene una fase juvenil corta y un genoma bien caracterizado. El objetivo de este estudio es identificar genes candidatos (CGs) relacionados con la calidad del fruto del melocotonero que puedan ser transferidos a otras especies de la familia tales como la fresa diploide.

Para ello, se han construido dos genotecas de cDNA de fruto del cultivar 'Fantasia', elaboradas en dos estados de desarrollo diferentes y se ha generado una colección de ESTs correspondientes a 1730 unigenes. Las secuencias y sus correspondientes anotaciones han sido almacenadas en una base de datos relacional y están disponibles a través de una interfaz de tipo web. Cincuenta y nueve genes relacionados con procesos de crecimiento, desarrollo y maduración del fruto, potencialmente involucrados en caracteres como el dulzor, la acidez y el contenido en compuestos fenólicos, han sido seleccionados de acuerdo a su anotación. Se han diseñado cincuenta y cinco pares de cebadores a partir de las secuencias obtenidas y se ha comprobado su amplificación en el genoma de *Fragaria* diploide.

De los 55 CGs seleccionados, ocho han sido mapeados en el genoma del melocotonero y 14 en el genoma de fresa. Cuatro de estos genes han podido ser localizados en ambas especies, permitiendo analizar las relaciones de sintenia entre ambos genomas. Se ha confirmado el patrón de sintenia propuesto por otros autores y se han localizado nuevas regiones sinténicas mediante mapeo comparativo. Los CGs seleccionados están localizados en tres grupos de ligamiento del melocotonero (PG3, PG5 y PG7) y en los 7 grupos de ligamiento de *Fragaria*. Asimismo, se ha analizado la colocalización de algunos de estos genes con QTLs para caracteres de calidad de fruto en melocotonero que deben ser confirmados en futuros análisis.

Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species

Loïck Le Dantec · Gaele Cardinet · Julio Bonet · Mathieu Fouché · Karima Boudehri · Amparo Monfort · Jean-Luc Poëssel · Annie Moing · Elisabeth Dirlewanger

Received: 8 December 2009 / Revised: 30 March 2010 / Accepted: 13 May 2010
© Springer-Verlag 2010

Abstract The goal of the present study was to identify candidate genes (CGs) involved in fruit quality in peach that can be transferred to other Rosaceae species. Two cDNA libraries from fruit of the "Fantasia" peach cultivar, constructed at two stages of development, were used to generate a set of expressed sequence tag sequences. A total of 1,730 peach unigenes were obtained after clustering. Sequences and corresponding annotations were stored in a relational database and are available through a web interface. Fifty-nine CGs involved in fruit growth and development or fruit quality at maturity, focusing on sweetness, acidity, and phenolic compound content, were

selected according to their annotation. Fifty-five primer pairs, designed from peach CG sequences and giving PCR products in peach, were tested in strawberry and 36 gave amplified products. Eight CGs were mapped in peach, 14 in strawberry, four in both species and confirmed the pattern of synteny already proposed using comparative mapping. In peach, the CGs are located in three linkage groups (3, 5, 7), and in strawberry they are distributed in all seven *Fragaria* linkage groups. Colocalization between some of these CGs and quantitative trait loci for fruit quality traits were identified and are awaiting confirmation in further analyses.

Keywords Candidate gene · EST · Fruit quality · Rosaceae · Peach · Strawberry

Communicated by A. Abbott

Electronic supplementary material The online version of this article (doi:10.1007/s11295-010-0308-8) contains supplementary material, which is available to authorized users.

L. Le Dantec · G. Cardinet · M. Fouché · K. Boudehri · E. Dirlewanger (✉)
INRA, UR419, Unité de Recherches sur les Espèces Fruitières,
BP 81, F-33140 Villenave d'Ornon, France
e-mail: dirlewanger@bordeaux.inra.fr

J. Bonet · A. Monfort
IRTA, Centre de Recerca en Agrogenòmica CSIC-IRTA-UAB,
Carretera de Cabrils Km 2,
08348 Cabrils, Barcelona, Spain

J.-L. Poëssel
INRA, UR1052, Unité de Génétique et d'Amélioration
des Fruits et Légumes,
F-84000 Avignon, France

A. Moing
INRA, UMR619 Fruit Biology,
BP 81, F-33140 Villenave d'Ornon, France

Introduction

The Rosaceae family encompasses a large number (3,000) of diploid and polyploid species (Shulaev et al. 2008) including economically important crops such as those belonging to the genera *Prunus* (stone fruits: peach, almond, apricot, cherry, and plum), *Malus* (apple), *Pyrus* (pear), *Rosa* (rose), *Rubus* (raspberry), and *Fragaria* (strawberry). In this family, the basic chromosome numbers are $x=7$, 8, and 9, and many species have a compact genome size, e.g., 200 Mb in *Fragaria* ($2n=2x=14$) (Akiyama et al. 2001; Folta and Davis 2006) and 290 Mb in *Prunus* ($2n=2x=16$) (Baird et al. 1994).

Peach is a good model species for the Rosaceae family. Compared to other tree species, it has a short juvenile phase and is genetically the best-characterized *Prunus* species with a number of genes controlling important traits and showing Mendelian behavior (Hesse 1975; Monet et al.

4.2 INTRODUCTION

The *Rosaceae* family encompasses a large number (3000) of diploid and polyploid species (Shulaev *et al.* 2008) including economically important crops such as those belonging to the genera *Prunus* (stone fruits: peach, almond, apricot, cherry, and plum), *Malus* (apple), *Pyrus* (pear), *Rosa* (rose), *Rubus* (raspberry), and *Fragaria* (strawberry). In this family, the basic chromosome numbers are $x=7$, 8, and 9, and many species have a compact genome size, e.g., 200 Mb in *Fragaria* ($2n=2x=14$) (Akiyama *et al.* 2001; Folta & Davis 2006) and 290 Mb in *Prunus* ($2n=2x=16$) (Baird *et al.* 1994).

Peach is a good model species for the *Rosaceae* family. Compared to other tree species, it has a short juvenile phase and is genetically the best-characterized *Prunus* species with a number of genes controlling important traits and showing Mendelian behavior (Hesse 1975; Monet *et al.* 1996), and a framework physical map of its genome is available (Abbott *et al.* 2002; Zhebentyayeva *et al.* 2008). Details of genetic and genomics resources are described in the Genomic Database for *Rosaceae* (<http://www.bioinfo.wsu.edu/gdr>, Jung *et al.* 2008).

Based on the comparison of linkage maps of rosaceous crop species constructed with transferable markers, synteny studies within this family have revealed that the genomes of *Prunus* species are largely collinear (Dirlewanger *et al.* 2004) and that there is also a high degree of synteny between the subfamilies, *Malus* and *Prunus* (Dirlewanger *et al.* 2004; Sargent *et al.* 2009c). Vilanova *et al.* (2008) studied the synteny between *Prunus* and diploid *Fragaria* reference maps using 71 anchor markers. Their comparisons revealed a clear pattern of synteny between *Fragaria* and *Prunus* genomes, with most markers that mapped to the same linkage group in one species mapping to one or two linkage groups in the other species. The *Fragaria* bin set [from *Fragaria vesca* 815 × *Fragaria bucharica* 601 (FV × FB) progeny] that detects 46 fragments of its genome, with an average length of 12.6 cM, is a valuable resource for the rapid addition of a large number of markers to the *Fragaria* genome (Sargent *et al.* 2008). Using this *Fragaria* bin set, it is possible to directly sequence and therefore genotype and map single nucleotide polymorphisms (SNPs) in conserved coding regions of the genome, facilitating comparative mapping in rosaceous genera (Vilanova *et al.* 2008).

Fruit quality determines the economic value of most fruited *Rosaceae*. In peach, many Mendelian characters involved in fruit quality have been studied and mapped, such as peach/nectarine (Bliss *et al.* 2002; Chaparro *et al.* 1994; Rajapakse *et al.* 1995), melting flesh/stony hard flesh (Warburton *et al.* 1996), freestone/clingstone (Dettori *et al.* 2001; Warburton *et al.* 1996), polycarpel (Bliss *et al.*

2002), and flesh color (Bliss *et al.* 2002; Rajapakse *et al.* 1995; Warburton *et al.* 1996). Quantitative trait loci (QTLs) for traits such as soluble sugar (sucrose, fructose, glucose, and sorbitol) and organic acid (malic, citric, and quinic acid) content, which determine fruit quality in peach, have also been mapped (Dirlewanger *et al.* 1999; Etienne *et al.* 2002b). The candidate gene (CG) approach has been successfully developed for peach characters such as melting flesh (*M*) and freestone (*F*) (Peace *et al.* 2005) and cold storage browning (Ogundiwin *et al.* 2008a); however, no colocalization between CG and QTL has yet been found for fruit acidity and fruit development (Etienne *et al.* 2002b; Rothan *et al.* 1999).

In octoploid strawberry (*Fragaria × ananassa* Duch.; $2n = 8x = 56$), QTLs controlling fruit quality traits have been detected and characterized, for malate, citrate, total soluble solids, sucrose, hexose/sucrose, and xylose, with the percentage of variation of each QTL ranging from 6.5% to 16% (Lerceteau-Köhler *et al.* 2006).

In peach, the sugar/acid ratio is a major organoleptic quality trait of the mature fruit. Among the main sugars (sucrose, glucose, fructose, and sorbitol), sucrose is predominant at maturity and can reach 60% of dry matter, while sorbitol represents only 8% of dry matter (Moing, 2000). Phenolic compounds including phenolic acids and flavonoids are bioactive compounds with antioxidant properties and have potential health benefits. Total phenolics vary notably in peach cultivars, and their contribution to antioxidant activity has been demonstrated to be greater than those of vitamin C and carotenoids (Gil *et al.* 2002). Several peach phenolic compounds have even been reported to affect growth and differentiation in human cancer cells (Lea *et al.* 2008). In strawberry, the major components of fruit quality are sweetness, acidity, and flavor, and the sugar/organic acid ratio is a major parameter of fruit taste (Perkins-Veazie 1995). The major soluble sugars in strawberries are glucose, fructose, and sucrose, and the major organic acid is citrate. Anthocyanins, flavonoids responsible for the fruit's red color, are a major contributor to fruit phytonutrient composition. Among phenolic acid derivatives, ellagitannins are potentially interesting for human health (Clifford & Scalbert 2000). In peach and strawberry, the major soluble compounds such as soluble sugars and organic acids, together with potassium and water transporters/channels and cell wall enzymes, contribute to the cell expansion of the fruit flesh and consequently to the fruit size and juiciness.

In the present study, we developed additional peach expressed sequence tags (ESTs) from cDNA libraries constructed from fruit of the 'Fantasia' peach cultivar at two early stages of fruit development, corresponding to the first rapid growth stage and the growth plateau (Etienne *et al.* 2002a), in order to obtain genomic tools to understand the different processes involved in fruit

development, during which many physicochemical changes occur. These changes determine the final organoleptic and nutritional properties of the fruit at maturity. From the two peach cDNA libraries, EST sequences homologous to genes encoding key proteins in organic acid, sugar, phenolic acid and flavonoid metabolism, cell expansion, cell cycle, signal transduction, and transport of dicarboxylates, water, potassium, and proton were isolated in order to undertake a candidate gene approach. The following steps are described: (1) EST processing and annotation, (2) primer design and intron/exon identification, (3) SNP detection, and (4) candidate gene mapping in peach and diploid strawberry.

The primary aim of the present study was the identification and mapping of new candidate genes potentially involved in fruit organoleptic quality in *Rosaceae* species. The final goal was the generation of genomic tools for molecular-assisted selection for peach and for other *Rosaceae* fruit-producing species. The selection of these candidate genes and their mapping in peach are part of the ISAFRUIT European Integrated Project (<http://www.isafruit.org/Portal/index.php>). Colocalization between these CGs and QTLs for fruit quality traits was assessed and is awaiting confirmation in further analyses.

4.3 MATERIAL AND METHODS

4.3.1 PLANT MATERIAL

The cDNA libraries were constructed from the peach variety 'Fantasia'. 'Fantasia' is a round, normally acid nectarine and is one of the parents of a peach F_2 mapping progeny (J×F). The other parent of this progeny is 'Ferjalou-Jalousia[®]', a flat, nonacid peach (Dirlewanger *et al.* 2006). These two varieties have closely related ancestors, both from 'Le Grand' variety, which is why the polymorphism observed in the F_2 progeny from these varieties is very low (Dirlewanger *et al.* 2006). However, this progeny segregates for six Mendelian characters and various quantitative characters involved in fruit quality (Dirlewanger *et al.* 2006).

For mapping, the J×F mapping progeny of 208 F_2 individuals (Dirlewanger *et al.* 2006) and the diploid *Fragaria* bin mapping set, including six individuals from the *F. vesca* 815 × *F. bucharica* 601 reference F_2 interspecific population, FV×FB, (Sargent *et al.* 2008) were used.

4.3.2 FRUIT cDNA LIBRARIES AND EST SEQUENCING

Two cDNA libraries (L1 and L2) were cloned from fruit of 'Fantasia' at two early stages of development. L1 was obtained at 30 days after bloom (DAB) corresponding to the first rapid growth stage with cell division (stage I) and L2 was obtained at 80 DAB corresponding to the growth plateau (stage II) before the second rapid growth stage (stage III) at 100–105 DAB (Etienne *et al.* 2002a).

Total RNAs were extracted from fruit mesocarp plus epidermis following the method described previously (Etienne *et al.* 2002a). PolyA⁺ RNAs were purified from total RNAs using magnetic oligo-dT beads (Dynabeads, Dynal, Oslo, Norway) following the manufacturer's instructions. Double-stranded cDNAs were synthesized from 5 µg polyA⁺RNA measured by absorbance at 260 nm, ligated into the Uni-ZAP™ XR vector using the ZAP-cDNA synthesis kit (Stratagene, La Jolla, CA), and packaged into bacteriophage using packaging extracts (Gigapack II, Stratagene) to produce the cDNA library stock according to the manufacturer's instructions. Recombinant Uni-ZAP™ XR package phages were plated on *Escherichia coli* XL1-Blue cells. cDNA clones were randomly selected for single-pass 5' sequencing. EST sequencing was done using a MegaBace 1000 automatic sequencer (Amersham Biosciences) at the UMR-BIOGECO (Genotyping-Sequencing platform, Agro site, INRA Cestas-Pierroton, http://www.pierroton.inra.fr/biogeco/site_pole_agro/genoseq.html). All the ESTs generated from the two cDNA libraries (L1 and L2) were submitted to GenBank.

4.3.3 EST PROCESSING AND ANNOTATION

EST sequences were analyzed (from base calling to functional annotation) using the EPA bioinformatic pipeline (EST Processing and Annotation, V. García, personal communication) maintained at the Bordeaux Bioinformatics Center (<http://www.cbib.u-bordeaux2.fr/EPA/>). EST sequences were preprocessed using PHRED for base calling and trimming (Ewing & Green 1998; Ewing *et al.* 1998), cross match for vector-derived sequence screening and local PERL scripts (V. García, personal communication) for masked sequences removal, giving a set of 2,579 peach EST sequences.

To minimize redundancy, all ESTs were clustered using stackPACK v2.2 (Miller *et al.* 1999). Sequences and quality files generated in the preprocessing step were submitted to the clustering module of stackPACK. Clustering and assembly were done using the following workflow: d2cluster (Burke *et al.* 1999) for the clustering step, PHRAP (www.phrap.org) for assembly and contig generation, and STACK_ANALYSIS and CRAW (Burke *et al.* 1998) for alignment and consensus sequence creation. A gene index (unigene) comprising the consensus (contigs) and singletons (ESTs which could not be

joined with other sequences during the clustering process) was then transferred to the annotation module of the pipeline. Each contig and singleton were first annotated based on sequential searches for homologous sequences in Swiss-Prot and TrEMBL databases using BLASTX program with an E value threshold of 10^{-8} . Details of the procedure are available at <http://www.cbib.u-bordeaux2.fr/EPA/help/help.php>.

Singletons and contigs homologous to genes putatively involved in sugar, organic acid, and phenolic metabolism, and fruit growth or development were selected (table 10). They were named CGPPBxxxx, where CG stands for candidate gene, PP *Prunus persica*, and B Bordeaux, and xxxx corresponds to the last four digits of the EST accession number in GenBank for singletons and to the last four digits of one EST in the contig (table 11) for contigs. CGs were classified in functional categories using the Gene Ontology Annotation database through the UniProtKnowledgeBase (www.uniprot.org). The MapMan Ontology adapted for tomato was also used (Urbanczyk-Wochniak *et al.* 2006). The principle of MapMan ontology is a hierarchical “BIN” - base structure. To avoid redundancy, each MapMan BIN includes a similar biological function, e.g., glycolysis or transport, and is further split into subBINs corresponding to submodes of the corresponding biological function (<http://gabi.rzpd.de/projects/MapMan/>).

Table 10. Peach candidate genes potentially involved in fruit quality. ¹SA: Sweetness or acidity, ²P: Phenolics, ³GD: Fruit growth or development. This table is included at the electronic supplementary material of the published version of this paper.

Candidate gene	Sequence size of the consensus (C) or singleton (S)	Annotation based on singleton or consensus sequence : Homologous protein. <i>Species.</i>	Potential role for fruit quality ¹	Accession number	NCBI Allignment E- value	% Identity (% positives)
CGPPB2290	C-1370	Aquaporin TIP1-1 (Tonoplast intrinsic protein 1-1) (AtTIP1;1) (Gamma-tonoplast intrinsic protein) (Gamma-TIP) (Aquaporin TIP) (Tonoplastintrinsic protein, root-specific RB7). - <i>A. thaliana</i>	GD/SA	P25818	2,00E-93	69/74
CGPPB2455	C-1057	Aquaporin TIP2-1 (Tonoplast intrinsic protein 2-1) (AtTIP2;1) (Delta-tonoplast intrinsic protein) (Delta-TIP). - <i>Arabidopsis thaliana</i>	GD/SA	Q41951	1,00E-101	77/86
CGPPB2693	C- 569	Probable aquaporin PIP2-5 (AtPIP2;5) (Plasma membrane intrinsicprotein 2d) (PIP2d). - <i>Arabidopsis thaliana</i>	GD	Q9SV31	1,00E-38	90/94
CGPPB2895	C- 687	Aquaporin PIP2-7 (AtPIP2;7) (Plasma membrane intrinsic protein 3)(Salt stress-induced major intrinsic protein) - <i>Arabidopsis thaliana</i>	GD	P93004	2,00E-76	88/93
CGPPB3133	S- 646	Aquaporin SIP1-1 (Small basic intrinsic protein 1-1) (OsSIP1;1). - <i>Oryza sativa</i> subsp. <i>Japonica</i>	GD	Q5VR89	1,00E-53	59/73
CGPPB3143	C- 724	Aquaporin PIP1-1 (AtPIP1;1) (Plasma membrane intrinsic protein 1a)(PIP1a) (Plasma membrane aquaporin-1). - <i>Arabidopsis thaliana</i>	GD	P61837	1,00E-62	94/98
CGPPB3303	S- 593	Probable aquaporin PIP2-5 (AtPIP2;5) (Plasma membrane intrinsicprotein 2d) (PIP2d). - <i>Arabidopsis thaliana</i>	GD	Q9SV31	7,00E-56	73/78, 68/81
CGPPB5502	C- 591	Aquaporin PIP1-2 (AtPIP1;2) (Plasma membrane intrinsic protein 1b)(PIP1b) (Transmembrane protein A) (TMP-A) (AthH2). - <i>Arabidopsis thaliana</i>	GD	Q06611	2,00E-55	95/97
CGPPB5532	S- 526	Aquaporin PIP1-2 (Plasma membrane intrinsic protein 1-2) (ZmPIP1-2)(ZmPIP1;2) (ZmPIP1b). - <i>Zea mays</i>	GD	Q9XF59	2,00E-69	87/92
CGPPB5536	C- 752	Aquaporin TIP2-1 (Tonoplast intrinsic protein 2-1) (AtTIP2;1) (Delta-tonoplast intrinsic protein) (Delta-TIP). - <i>Arabidopsis thaliana</i>	GD/SA	Q41951	1,00E-101	77/86
CGPPB2761	S- 545	Expansin-A4 precursor (AtEXPA4) (Alpha-expansin-4) (At-EXP4) (AtEx4)(Ath-ExpAlpha-1.6). - <i>Arabidopsis thaliana</i>	GD	O48818	3,00E-35	84/92
CGPPB3043	C- 671	Expansin-A6 precursor (AtEXPA6) (Alpha-expansin-6) (At-EXP6) (AtEx6)(Ath-ExpAlpha-1.8). - <i>Arabidopsis thaliana</i>	GD	Q38865	7,00E-92	83/91, 59/70
CGPPB6033	S- 572	Expansin-A10 precursor (AtEXPA10) (Alpha-expansin-10) (At-EXP10) (AtEx10) (Ath-ExpAlpha-1.1). - <i>Arabidopsis thaliana</i>	GD	Q9LDR9	6,00E-42	83/90
CGPPB2741	C- 803	(P42652) 14-3-3 protein 4 (PBLT4). - <i>Solanum lycopersicum</i> (<i>Lycopersicon esculentum</i>).	GD	P42652	3,00E-71	92/93, 85/96
CGPPB5971	S- 622	14-3-3-like protein B (SGF14B) (Fragment). - <i>Glycine max</i>	GD	Q96451	7,00E-51	81/85
CGPPB2791	S- 659	Cyclin-D3-1 (G1/S-specific cyclin-D3-1) (CycD3;1) (Cyclin-delta-3)(Cyclin-d3). - <i>Arabidopsis thaliana</i>	GD	P42753	1,00E-23	55/70
CGPPB3501	C- 623	UDP-glucose 6-dehydrogenase (EC 1.1.1.22) (UDP-Glc dehydrogenase)(UDP-GlcDH) (UDPGDH). - <i>Glycine max</i>	GD	Q96558	3,00E-95	91/95
CGPPB6189	S- 647	Sucrose synthase (EC 2.4.1.13) (Sucrose-UDP glucosyltransferase). - <i>Phaseolus aureus</i> (<i>Vigna radiata</i>)	GD/SA	Q01390	6,00E-69	90/94
CGPPB3705	C- 597	Potassium transporter 11 (AtPOT11). - <i>Arabidopsis thaliana</i>	GD/SA	O64769	6,00E-40	88/97, 47/60, 76/87, 100/100
CGPPB2585	S- 475	Phosphoenolpyruvate carboxykinase - <i>Urochloa panicoides</i>	SA	P49292	2,00E-26	70/81
CGPPB3070	C- 708	Probable fructokinase-2 (EC 2.7.1.4). - <i>Arabidopsis thaliana</i>	SA	Q9LNE3	8,00E-51	71/79
CGPPB3190	S- 416	Probable fructokinase-1 (EC 2.7.1.4). - <i>Arabidopsis thaliana</i>	SA	Q9SID0	3,00E-58	83/91
CGPPB3212	S- 574	Enolase 1 (EC 4.2.1.11) (2-phosphoglycerate dehydratase 1) (2-phospho-D-glycerate hydro-lyase 1). - <i>Zea mays</i>	SA	P26301	7,00E-46	93/95
CGPPB3258	C- 1008	Triosephosphate isomerase, cytosolic (EC 5.3.1.1) (TIM) (Triose-phosphate isomerase). - <i>Petunia hybrida</i>	SA	P48495	1,00E-106	76/81
CGPPB3373	S- 639	Pyruvate dehydrogenase E1 component subunit alpha (EC 1.2.4.1). - <i>Porphyra purpurea</i>	SA	P51267	9,00E-46	75/86
CGPPB3547	S- 461	Pyruvate kinase, cytosolic isozyme (EC 2.7.1.40) (PK). - <i>Glycine max</i>	SA	Q42806	2,00E-35	60/69, 54/70
CGPPB3890	S- 637	Pyruvate kinase, cytosolic isozyme (EC 2.7.1.40) (PK). - <i>Nicotiana tabacum</i>	SA	Q42954	3,00E-17	45/61
CGPPB5800	S- 585	Pyruvate kinase, cytosolic isozyme (EC 2.7.1.40) (PK). - <i>Glycine max</i>	SA	Q42806	1,00E-19	44/64
CGPPB3575	C- 659	Vacuolar proton pump subunit C (EC 3.6.3.14) (V-ATPase subunit C). - <i>Arabidopsis thaliana</i>	SA	Q9SDS7	8,00E-30	73/83, 100/100

Table 10 (cont.)

Candidate gene	Sequence size of the consensus (C) or singleton (S)	Annotation based on singleton or consensus sequence : Homologous protein. <i>Species</i> .	Potential role for fruit quality ¹	Accession number	NCBI Alignment E-value	% Identity (% positives)
CGPPB5670	C- 1054	Vacuolar proton pump subunit D (EC 3.6.3.14) (V-ATPase subunit D). - <i>Arabidopsis thaliana</i>	SA	Q9XGM1	1,00E-115	79/91
CGPPB3699	C- 679	Fructose-bisphosphate aldolase cytoplasmic isozyme (EC 4.1.2.13)(Gravity-specific protein GSC 233). - <i>Oryza sativa</i> subsp. <i>Japonica</i>	SA	P17784	7,00E-50	83/87
CGPPB5676	S- 633	Aconitate hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase)(Fragment). - <i>Cucumis melo</i> var. <i>conomon</i>	SA	Q42669	9,00E-61	90/93, 86/86
CGPPB5756	S- 469	Hexokinase-1 (EC 2.7.1.1) (SoHxK1). - <i>Spinacia oleracea</i>	SA	Q9SEK3	1,00E-27	73/83
CGPPB5473	S- 547	Tonoplast dicarboxylate transporter (AtTDAT) (Vacuolar malatetransporter) (Sodium-dicarboxylate cotransporter-like) (AtSDAT). - <i>Arabidopsis thaliana</i>	SA	Q8LG88	2,00E-20	82/88
CGPPB5514	C- 566	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic (EC 1.2.1.12). - <i>Magnolia liliiflora</i> .	SA	P26518	2,00E-70	83/87
CGPPB5575	S- 679	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic (EC 1.2.1.12). - <i>Ginkgo biloba</i>	SA	Q39769	1,00E-46	77/91
CGPPB5569	S- 652	Phosphoenolpyruvate carboxykinase - <i>Arabidopsis thaliana</i>	SA	Q9T074	5,00E-26	84/91
CGPPB5795	C- 684	Fructose-bisphosphate aldolase, cytoplasmic isozyme (EC 4.1.2.13). - <i>Spinacia oleracea</i>	SA	P29356	6,00E-92	88/91
CGPPB6050	S- 607	NADP-dependent malic enzyme (EC 1.1.1.40) (NADP-ME). - <i>Populus trichocarpa</i> (<i>Populus balsamifera</i> subsp. <i>trichocarpa</i>)	SA	P34105	7,00E-77	78/83
CGPPB6190	S- 587	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 1 (EC 1.2.1.12). - <i>Zea mays</i>	SA	P08735	9,00E-79	75/83
CGPPB6194	S- 493	Probable pleiotropic drug resistance protein 4. - <i>Arabidopsis thaliana</i>	-	O81016	5,00E-32	76/88
CGPPB6212	S- 443	Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucoseisomerase) (PGI) (Phosphohexose isomerase) (PHI). - <i>Synechocystis sp.</i> (strain PCC 6803).	SA	P52983	3,00E-15	50/67
CGPPB6260	S- 520	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic (EC 1.2.1.12). - <i>Antirrhinum majus</i>	SA	P25861	3,00E-32	83/88
CGPPB6347	S- 630	Hexokinase-2, chloroplast precursor (EC 2.7.1.1) (NtHxK2). - <i>Nicotiana tabacum</i>	SA	Q6Q8A5	1,00E-84	77/85
CGPPB2799	S- 246	Dihydroflavonol-4-reductase (EC 1.1.1.219) (DFR) (Dihydrokaempferol 4-reductase). - <i>Vitis vinifera</i>	P	P51110	3,00E-28	90/91,76/91
CGPPB2805	C- 1011	Leucoanthocyanidin dioxygenase (EC 1.14.11.19) (LDOX) (Leucocyanidin dioxygenase) (Leucoanthocyanidin hydroxylase) (Anthocyanidin synthase). - <i>Malus domestica</i> , <i>Malus sylvestris</i> .	P	P51091	1,00E-103	86/88, 95/96
CGPPB3316	C- 451	Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1). - <i>Camellia sinensis</i>	P	P48386	7,00E-72	88/94
CGPPB3369	S- 644	4-coumarate--CoA ligase-like 9 (EC 6.2.1.-) (4-coumarate--CoA ligase isoform 4) (At4CL4). - <i>Arabidopsis thaliana</i>	P	Q84P23	4,00E-60	74/89
CGPPB6322	S- 686	4-coumarate--CoA ligase-like 5 (EC 6.2.1.-) (Peroxisomal OPC-8:0-CoAligase 1) (4-coumarate--CoA ligase isoform 9) (At4CL9). - <i>Arabidopsis thaliana</i>	P	Q84P21	3,00E-67	81/93
CGPPB3378	S- 697	Transcription factor MYB39 (Myb-related protein 39) (AtMYB39). - <i>Arabidopsis thaliana</i>	P	Q8GWPO	6,00E-29	71/90, 69/87
CGPPB3491	S- 620	Flavonoid 3'-monooxygenase (EC 1.14.13.21) (Flavonoid 3'-hydroxylase) (Cytochrome P450 75B2). - <i>Petunia hybrida</i>	P	Q9SBQ9	2,00E-75	69/75
CGPPB3861	S- 649	Leucoanthocyanidin reductase (EC 1.3.1.77) (LAR) (Protein BANYULS)(Anthocyanin spotted testa) (ast). - <i>Arabidopsis thaliana</i>	P	Q9SEV0	3,00E-11	26/51
CGPPB5654	C- 836	Serine carboxypeptidase-like 8 precursor (EC3.4.16.) (Sinapoylglucose--malate O-sinapoyltransferase) (EC 2.3.1.92) (SMT)(Protein SINAPOYLGLUCOSE ACCUMULATOR 1). - <i>A. thaliana</i>	P	Q8RUW5	3,00E-40	43/61
CGPPB5806	S- 707	Dihydroflavonol-4-reductase (EC 1.1.1.219) (DFR) (Dihydrokaempferol 4-reductase). - <i>Dianthus caryophyllus</i>	P	P51104	1,00E-15	35/56
CGPPB5921	S- 612	Dihydroflavonol-4-reductase (EC 1.1.1.219) (DFR) (Dihydrokaempferol 4-reductase). - <i>Gerbera hybrida</i> .	P	P51105	2,00E-23	32/55
CGPPB5769	S- 604	Serine carboxypeptidase-like 18 precursor (EC 3.4.16.-). - <i>Arabidopsis thaliana</i>	-	Q9C7Z9	2,00E-57	51/70
CGPPB5792	S- 534	Serine carboxypeptidase-like 20 precursor (EC 3.4.16.-). - <i>Arabidopsis thaliana</i>	-	Q8L7B2	2,00E-51	81/89

4.3.4 *P. PERSICA* CANDIDATE GENE ANALYSES

Primer pairs for each selected CG were designed so that the amplification product had at least one intron when possible, as introns are more variable than exons due to generally much lower selective pressure in intronic than in exonic regions (Feltus *et al.* 2006; Yang *et al.* 2007). Moreover, based on the hypothesis that intron sites are conserved between *Arabidopsis thaliana* and *P. persica* (Georgi *et al.* 2003; Jung *et al.* 2006), peach EST sequences were compared with orthologous *Arabidopsis* genomic DNA sequences to locate the probable intron/exon junctions in peach, using the Intron Finder from the tools available in the Sol Genomics Network website (<http://www.sgn.cornell.edu/>). Primer pairs were then designed, using the Primer3 software (Rozen & Skaletsky 2000), within conserved exon regions flanking putative introns in the peach gene sequences. Primers were chosen to amplify fragments from 300 bp if predicted introns were included in the amplified region, or 600 bp if no intron was predicted, with a melting temperature of 57°C for all primer pairs.

The polymerase chain reaction (PCR) amplifications were performed with the DNA of the two parents of the F₂ J×F mapping population, 'Ferjalou-Jalousia[®]' and 'Fantasia', the F₁ hybrid and 12 F₂ individuals, for peach; with the parents of the reference FV×FB map, *F. vesca* 815 and *F. bucharica* 601, for strawberry; and with the six F₂ individuals used for bin mapping as template. For peach, PCR amplification was prepared in a final volume of 10 µl including 0.2 µM of each primer, 200 µM of each dNTP, 1.5 mM MgCl₂, 10 ng of genomic DNA, and 0.4 U Taq polymerase (Invitrogen). After a 5-min initial denaturation at 94°C, DNA amplification was performed as follows: 35 cycles of 45 s at 94°C, 45 s at 57°C, and 45 s at 72°C, with a final extension for 5 min at 72°C. For strawberry, DNA amplification was performed in a final volume of 20 µl containing 20 ng DNA, 1× PCR buffer, 2.5 mM MgCl₂, 0.2 mM each dNTP, 0.2 µM each primer, and 0.5 U Taq polymerase (Promega). PCR was as follows: 2-min denaturation at 94°C, 35 cycles of 30 s at 94°C, 30 s at annealing temperature from 55°C to 50°C, after temperature gradient determination, and 45 s at 72°C and a final extension for 7 min at 72°C.

PCR products were electrophoresed on agarose gel (2% agarose for peach and 2.5% for strawberry) to control the quality of amplification, estimate the size of the amplification product, and detect possible intron length polymorphisms. For peach, acrylamide/bisacrylamide (19/1) 4.5% gel electrophoreses were also performed to detect small-size polymorphism.

For peach, amplified products were sent to GENOME Express (<https://www.gexbyweb.com>) for sequencing. For strawberry, amplicons were purified using the QiaQuick PCR purification kit (Qiagen) according to the manufacturer's instructions, resuspended in 50 µl sterile distilled water, and analyzed on an ABI PRISM 3130 genetic analyzer. The genomic sequences were compared to the EST sequences, using the Blast 2 sequence function, to confirm that each amplified fragment corresponded to the expected one.

4.3.5 SINGLE NUCLEOTIDE POLYMORPHISMS ANALYSES

For peach, no size polymorphism was detected even on acrylamide gel, so SNPs were detected in CGs by comparing the sequences of PCR amplification products of 'Ferjalou-Jalousia'® (J), 'Fantasia' (F), F₁ hybrid (H), and 12 F₂ individuals. For each CG, sequences were aligned and chromatograms were analyzed using two methods in order to avoid false-negative or false-positive SNP detection. The first method used GenalysWin2.8 software (Takahashi *et al.* 2003) and the second a local pipeline (Le Dantec *et al.* 2004), combining PHRED for base calling and quality assessment, PHRAP for assembly of the sequences, PolyBayes (Marth *et al.* 1999) and PolyPhred (Nickerson *et al.* 1997) for SNP detection, and Consed (Gordon *et al.* 1998) for visual inspection of aligned sequences and chromatograms.

For strawberry, when no size polymorphism was found, SNPs were detected by sequence alignment using the Sequencher (GENECODE) software and verified by visual inspection of the DNA chromatograms.

4.3.6 CANDIDATE GENE MAPPING

In peach, when SNPs were identified, Rebase software (<http://rebase.neb.com/rebase/rebase.html>) was used to search for a restriction enzyme specific to one allele and the restriction polymorphism revealed on 2% agarose gel following digestion of the PCR product of all 208 F₂ individuals of the mapping population. When no enzyme was found, the amplification products of the 208 individuals of the population were sequenced for SNP genotyping. CGs were then located on the linkage groups of the linkage map available (Dirlewanger *et al.* 2006) using MAPMAKER/EXP V3.0 software (Lander *et al.* 1987) with the Kosambi function.

In strawberry, the CGs found to be polymorphic in the parents of the mapping population were bin-mapped after sequencing the *Fragaria* bin set that includes FV, hybrid, and six selected individuals (Sargent *et al.* 2008).

Table 11. Origin of the peach candidate genes selected. ESTs in bold are those used to refer to the candidate gene when it was identified on a contig.

Candidate gene	cDNA library L1/L2	Consensus ID / Singleton S	EST Genbank ID
CGPPB2290	L1 - L2	CN56	AJ872290 - AJ872546 - AJ872584 - AJ872635 - AJ873712 - AJ873670 - AJ873695 - AJ873709 - AJ872999 - AJ875743 - AJ875648 - AJ875686 - AJ873023 - AJ873094 - AJ873051 - AJ873159 - AJ873192 - AJ873193 - AJ873229 - AJ873272 - AJ873327 - AJ873313 - AJ875831 - AJ875899 - AJ876176 - AJ876228 - AJ876285
CGPPB2294	L1-L2	CN279	AJ872294 - AJ873253 - AJ875851
CGPPB2300	L1	S	AJ872300
CGPPB2455	L1	CN208	AJ872455 - AJ873694 - AJ873247 - AJ873470
CGPPB2473	L1	S	AJ872473
CGPPB2585	L1	S	AJ872585
CGPPB2693	L1	CN19	AJ872693 - AJ872555
CGPPB2741	L1-L2	CN41	AJ872741 - AJ872833 - AJ872626 - AJ875725
CGPPB2761	L1	S	AJ872761
CGPPB2791	L1	S	AJ872791
CGPPB2799	L1	S	AJ872799
CGPPB2805	L1-L2	CN13	AJ872805 - AJ873290 - AJ873902 - AJ876045 - AJ876167
CGPPB2895	L1	CN193	AJ872895 - AJ873376
CGPPB3043	L1	CN145	AJ873043 - AJ873044
CGPPB3070	L1	CN250	AJ873070 - AJ873870
CGPPB3133	L1	S	AJ873133
CGPPB3143	L1-L2	CN361	AJ873143 - AJ875844 - AJ875871 - AJ876059 - AJ876066 - AJ876268
CGPPB3190	L1	S	AJ873190
CGPPB3212	L1	S	AJ873212
CGPPB3258	L1-L2	CN233	AJ873258 - AJ873754 - AJ876133
CGPPB3303	L1	S	AJ873303
CGPPB3316	L1	CN316	AJ873316 - AJ873463
CGPPB3369	L1	S	AJ873369
CGPPB3373	L1	S	AJ873373
CGPPB3378	L1	S	AJ873378
CGPPB3491	L1	S	AJ873491
CGPPB3501	L1- L2	CN204	AJ873501 - AJ876106
CGPPB3547	L1	S	AJ873547
CGPPB3575	L1	CN334	AJ873575 - AJ876204
CGPPB3699	L1-L2	CN360	AJ873699 - AJ876246
CGPPB3705	L1-L2	CN395	AJ873705 - AJ876361
CGPPB3861	L1	S	AJ873861
CGPPB3890	L1	S	AJ873890
CGPPB5473	L2	S	AJ875473
CGPPB5502	L1- L2	CN387	AJ875502 - AJ873530 - AJ876271 - AJ876326
CGPPB5514	L1- L2	CN25	AJ875514 - AJ872298
CGPPB5532	L2	S	AJ875532
CGPPB5536	L2	CN286	AJ875536 - AJ875960 - AJ875809
CGPPB5569	L2	S	AJ875569
CGPPB5575	L2	S	AJ875575
CGPPB5654	L2	CN303	AJ875654 - AJ876095
CGPPB5670	L1-L2	CN112	AJ875670 - AJ875567 - AJ873506
CGPPB5676	L2	S	AJ875676
CGPPB5756	L2	S	AJ875756
CGPPB5769	L2	S	AJ875769
CGPPB5792	L2	S	AJ875792
CGPPB5795	L2	CN388	AJ875795 - AJ875669 - AJ875877 - AJ876328
CGPPB5800	L2	S	AJ875800
CGPPB5806	L2	S	AJ875806
CGPPB5921	L2	S	AJ875921
CGPPB5971	L2	S	AJ875971
CGPPB6033	L2	S	AJ876033
CGPPB6050	L2	S	AJ876050
CGPPB6189	L2	S	AJ876189
CGPPB6190	L2	S	AJ876190
CGPPB6194	L2	S	AJ876194
CGPPB6212	L2	S	AJ876212
CGPPB6260	L2	S	AJ876260
CGPPB6322	L2	S	AJ876322
CGPPB6347	L2	S	AJ876347

Table 12. Primer pairs designed from peach CG sequences.

Candidate gene	Forward primer (5'-3')	Reverse primer (5'-3')
CGPPB2290	ATGGATGCTCCATCGAAAG	CGGCAATATCTCCCTCT
CGPPB2294	GGAACACCAGATATTGCAAAC	TGGTGAGAGTCAGATGCTTG
CGPPB2300	CTTGTGCAGCATAGGCAATA	ACCTTTCCTGGGATTGAAG
CGPPB2455	CGAGCTGGGTTTCATTGAT	CCTTCATCTCAAGTTCGTC
CGPPB2473	CACTTTCCACCCATAACTC	GCGAGTTCGTTGAGTTTCC
CGPPB2585	GAGCATCCACCTCCTTGAT	CGCGAAGTACAAGAAGACTG
CGPPB2693	ATAGTTCACCTGGCCACAAT	AAAGGCAAGTGTGGAGAA
CGPPB2741	GCTTTGATGCTTCCTTGATT	CTTGGGCTTGCACTTAACCT
CGPPB2761	GACGTGAAAGGATCCAAGAC	CACTGTCCACACACTCACAA
CGPPB2791	GCTGCAGGTCTAGCTGTA	GATTGCAGTGAGAGCAGTGA
CGPPB2799	CAAATATGGGGCCAGAGTC	ATCCAAGCTTCCCTCGTC
CGPPB2805	ACATCAGTGTGAGCTTCGAC	CCTCAAACACTGCTGATTA
CGPPB2895	TAGCTGCTGCTGAGAATG	CACAGTTTTCTCAGCCACTG
CGPPB3043	CCAACTTCTGCTCCAAC	CGACGTTGGTGATTAGAACC
CGPPB3070	GGTGAAAGCTCAGACTCG	CGGTTGCAGATACTACACCA
CGPPB3133	GCCAATCCAATAAACCAAGA	CACCACTGCAAGTTTCTACG
CGPPB3143	GTGGTACAAGCTGCAAGTG	GCTACACCAAGGGATCAGG
CGPPB3190	CTATGGCTAACAATGGCGTA	TTTGATGAGCTCCAAGTTGA
CGPPB3212	CCCATAAAGACTCGAGCAAT	AGTCACCGAAGTGGTGAAC
CGPPB3258	CCTACTCTCCTCGGGTTCA	TCAAGCAAACGTAAGTCTCTG
CGPPB3303	TCACTGTGCTTACCGTGATT	GCCACCATGTACAATACAGC
CGPPB3316	GTGGAGGAAGTTCGCAAG	ATGTCCACACCACTGTTGG
CGPPB3369	CTGCAGCCTCAGTGATATTG	CCTTGATGATGAGGGTTTCC
CGPPB3373	CCCAAGACCTCATCTTTGTC	CAGTTGACACAGCTTCTTGG
CGPPB3378	GTCCAAGATGTGGGAAGAG	GCTATAGCCGACCACCTAAG
CGPPB3491	AGAAGCTGAGCCAAGTTCAC	CACGACACCAACTTCTCAAG
CGPPB3501	TGGTGAAGATCTGTTGCATT	ACCCTCAGCAAGAACTCTG
CGPPB3547	GTGAATGGCGAACATAGACA	CTGCCCTTCTTAAGTTGAA
CGPPB3575	ACGTCGGTTTCTTGCTCTAC	TACGGAGAGGTTTTAGCTC
CGPPB3699	CAATCGTCCTCAAGCTTCTT	CAACCTTAATGCCATGAACA
CGPPB3705	GACTCTCATGAGGAACATTG	CGTTGCAAGATACGGTTACA
CGPPB3861	AATTACAAGGCCTTCTTGGGA	TTCTGTCATTACACCCCTTG
CGPPB3890	CTCGTTTGTTCAGCTGGTAG	CCATTAAGGTGAAGGCATCT
CGPPB5473	CTTCCAACAGGAACACCTTC	CAAATGGCTGCAATCATAGA
CGPPB5502	AAGAGCAGCACCAATGAATG	CCAGAGACTCTCATGTCCT
CGPPB5514	GGTCGTTTGGTTGCTAGAGT	CTTCTCATTGACACCCACAA
CGPPB5532	TGACGTATGGTCGTTTTAC	CCTGTCAGCGACAGCTT
CGPPB5536	AGAATCTGCAGCTGTCATCA	TTCATAAAGAGCTCGTGCA
CGPPB5569	GTCTAGCCCTCGATACCAAA	TCAGCTAGGTGGTCTGTTCA
CGPPB5575	CATGTGCTCTATGAGGTCCA	ATCAGGTATGCTGCTGATGG
CGPPB5654	TTCACCTTTGGAGCAATAC	GCTCCTGGTATCGATGATG
CGPPB5670	TAGAGCATTGACACGACGAT	CGCTTGACGGTAGTTCTTA
CGPPB5676	ATGGCAAATCCTTCGTATGT	AGCTCTGGGATTGAAACAAC
CGPPB5756	CGAATTGCGCAATTACAAG	TGTATGAGCACTACACCGAAT
CGPPB5769	GCATCCAGTAGAAGTGTGTTG	CGAGTACGAATCACCAGAAA
CGPPB5792	CATAGGTAGAGGCCATGGAT	TCAGCCACAGAACTACTGGA
CGPPB5795	CTCTTCGTGAGCTCCTCTTC	CTGGCACCAGCTTCATAGTA
CGPPB5800	TGAGACAATATCTGCGTGCT	TGTTAGGACAACCAGGGTCT
CGPPB5806	GCCCTTCTCCTGTAAGCTCT	AGGCTCAGCCAAGACTTATG
CGPPB5921	CTTCAACCAGACCTCCAACA	AGCTCCTGTCCATTGCTAAG
CGPPB5971	AGTGAGGTTGTCCCTCAA	AGGCTGCTGATGAGAGAAAG
CGPPB6033	ATGGACGATTCTGCGAACTA	GGGAGGAATAAGATTACCA
CGPPB6050	CCAGTACAATAAGGGGCTTG	GTGTAGACAATTGGGAGCAA
CGPPB6189	ACTCCGGTGAGAGTCAAGAG	ACCTGTGGACTTCCAACATT
CGPPB6190	GTGACAAGCCCGTCTTATT	CTTCTCATTGACACCCACAA
CGPPB6194	TGAATTCTGACTGTGCTGCT	GAAACCCAATGGCACATATT
CGPPB6212	ACAATGGGAAGGAAGATCAA	TATCATCAACCTGGTGTGGA
CGPPB6260	AACTACCACCATTCGCTGTT	AAGATGACGTGGTTTCCACT
CGPPB6322	TCGGATACTGTCCAACCTTT	TGCACAAGGATGGTTAAGAA
CGPPB6347	GGGAGTGACCTCAAGATGAT	GCCCAGAAGCAATGAAATC

4.4 RESULTS

4.4.1 PEACH cDNA LIBRARY AND EST SEQUENCES

Two thousand five hundred and seventy-nine individual 5' ESTs (1650 from the L1 cDNA library and 929 from the L2 cDNA library) were retained after elimination of poor quality sequence data. All the sequences are available in GenBank (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>). From the 2579 ESTs, 1204 were assembled into 355 contigs and 1375 were classified as singletons, giving a peach EST set of 1730 unigenes. Contigs and singletons were annotated as described in "Materials and methods." The annotated peach EST unigene set is accessible through a web-integrated database, also containing all EST sequences, contigs, and tools for analysis, at <http://www.cbib.u-bordeaux2.fr/SAM2/COMPLETE/> under the PeachESTdb project name. Distribution of contig annotation is as follows: 57 contigs were classified as similar to proteins of unknown functions (15.5%), 42 contigs had no significant similarities with any protein in the databases (11.8%), and 256 contigs had similarities with classified proteins (72.7%). Distribution of singleton annotation was slightly different, with 251 singletons having similarities with proteins of unknown function (18.3%), 468 singletons having no significant similarities with any protein in the databases (34%), and 656 singletons having similarities with classified proteins (47.7%).

4.4.2 CANDIDATE GENE SELECTION

From the 1730 unigenes, a total of 59 CGs were manually selected based on their annotation according to their potential role in (1) fruit primary metabolism (in relation to sweetness or acidity), (2) secondary metabolism including phenolic acid or flavonoid content (in relation to fruit nutritional quality), and (3) transport and cell division and growth (in relation to fruit growth and development). Thirty-seven CGs were selected from singletons (17 ESTs from L1 and 20 from L2 library) and 22 CGs from contigs (seven contigs include only ESTs from L1, three include only ESTs from L2 and 12 contigs from both L1 and L2 libraries) (table 11). Each contig includes from two to 27 ESTs (11 contigs with two ESTs, four with three ESTs, four with four ESTs, one with five ESTs, one with six ESTs and one with 27 ESTs). The names of the CGs from contigs referred to one EST of the contig. Primer pairs were designed for all 59 CGs selected (table 12) with 55 CGs giving PCR amplification products (table 13). For these CGs, functional annotation was performed following the Gene Ontology scheme and the MapMan Ontology BIN structure, based on singleton or contig sequences ('Electronic supplementary material'). Sixteen CGs involved in glycolysis, five involved in carbohydrate metabolism, and two involved in proton transport were considered as potentially involved in fruit

sweetness or acidity. Seven CGs involved in flavonoid biosynthesis and five involved in phenylpropanoid metabolism were considered as potentially involved in the control of phenolic compound content. Among the latter genes, CGPPB5654, CGPPB5769, and CGPPB5792 were selected for their homology with the *A. thaliana* serine carboxypeptidase-like gene family clade I which is thought to act as the acyltransferase, sinapoylglucose:malate sinapoyltransferase (Fraser *et al.* 2005). Four CGs involved in cell wall behavior, 12 involved in membrane water transport, and one involved in cell cycle were considered as potentially involved in fruit growth and development. In addition, two *14-3-3-like* proteins and one *MYB* factor were also selected in relation to the possible involvement of several members of their families in signaling for primary metabolism and proton transport (Roberts, 2003) and regulation of transcription for secondary metabolism (Deluc *et al.* 2006), respectively. The four CGs that did not produce a PCR product using the primer pairs designed in this study (table 13) were not included in further analyses.

4.4.3 PRIMER DESIGN TO OPTIMIZE POLYMORPHISM IDENTIFICATION

Of the 59 primer pairs designed (table 12), 53 amplified a single fragment, two amplified two fragments, and four gave no amplification in peach (table 13). The sequence lengths ranged from 246 to 697 bp for singletons and from 451 to 1,370 bp for contigs. For the two primer pairs that gave two fragments (CGPPB3316 and CGPPB5769) for all the tested individuals (J, F, H, and 12 F₂ individuals), the PCR products including both fragments were sequenced. For CGPPB3316, sequences of good quality were obtained for all individuals tested. By comparing them with the sequence of the corresponding EST, two introns were detected, one of 217 bp and the other of 181 bp (Fig. 16). Three exons were amplified, the second and the third being repeated sequences including the reverse primer sequence in their 3' end. The longest amplified fragment of 1000 bp corresponded to the three exons separated by the two introns, while the smallest fragment (650 bp) corresponded to the two first exons separated by the first intron. As the CGPPB5769 sequences were not readable, the two fragments were extracted from acrylamide gel and sequenced separately. The longest fragment of 550 bp, had a 129-bp intron, an exon of 124 bp, a second intron of 227 bp, and an exon of 63 bp (Fig. 17). In the smallest fragment, of 416 bp, the second intron was 109 bp. This may be due to indel polymorphism within the second intron or to amplification of two paralogous sequences.

Among the 55 primer pairs amplifying PCR products, 13 did not include an intron, 21 included one intron, 20 included two introns, and one included three introns (table 13). The intron number was the same in the homologous *A. thaliana* sequence for 40 of the 55 CGs giving PCR amplification

products (73%). This was the case for all CGs without an intron (observed size corresponding to the expected) except for the CGPPB3378 homolog. It was also true for 12 of the 21 CGs containing one intron, while eight of the homologous *A. thaliana* sequences did not contain an intron and one contained two introns. Of the 20 CGs that contained two introns, 15 of the homologous *A. thaliana* sequences also contained two introns, and for CGPPB5514, containing three introns, there were also three introns in the homologous *A. thaliana* sequence. For the 28 CGs that contained the same number of introns in peach and *A. thaliana*, the position of the intron in 26 of them (93%) was as predicted from the intron position of the homologous *A. thaliana* sequence. For the four primer pairs that failed to amplify any PCR product, no intron was predicted by comparison with the homologous *A. thaliana* sequence. Among the 55 primer pairs designed from peach EST sequences, 36 (65%) amplified products in strawberry (table 13). However, for five peach primer pairs, the amplification was very low, and new primers were designed based on *Fragaria* sequences from the *Fragaria* database (table 14). For four of them, the same number of introns was detected in *Fragaria* and *P. persica* (table 15).

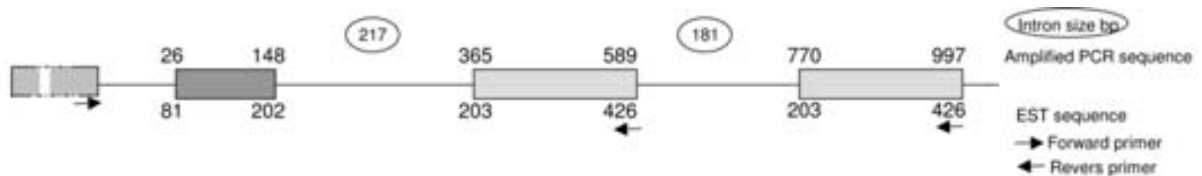


Figure 16. Comparison of EST and PCR product sequences for CGPPB3316 in peach.

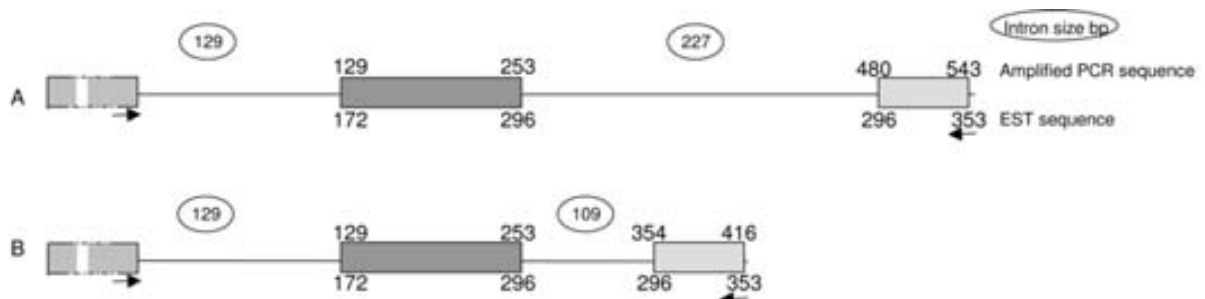


Figure 17. Comparison of EST and PCR product sequences for CGPPB5769 in peach. a: Sequence of the longest fragment (550 bp), b sequence of the smallest fragment (420 bp).

4.4.4 SNP DETECTION AND CANDIDATE GENE MAPPING

4.4.4.1 MAPPING IN PEACH

For the 55 CGs having amplification products, no indel polymorphism was observed. SNP detection in peach was done by comparing the sequences of the two parents (J, F), the hybrid and 12 F₂ individuals using two methods, GenalysWin2.8 and PolyBayes, which gave exactly the same results. SNPs were detected for eight CGs (table 16). The number of SNPs per CG ranged from one (for CGPPB2805 and CGPPB 5806) to nine (for CGPPB5502). Twenty-five SNPs were detected in analyzed sequences of a total length of 30 kb (1 SNP per 1,2 kb on average). A restriction enzyme site including the SNP was searched for to reveal polymorphism on agarose gel (fig. 20). A site was identified for seven of the CGs including SNPs: *Bfal* for CGPPB2805, *HpyCH4V* for CGPPB3369, *MseI* for CGPPB3501 and CGPPB6189, *BfuCI* for CGPPB3890, *Tsp509I* for CGPPB5502, and *HpaII* for CGPPB5806 (table 16, Fig. 18). The seven amplified polymorphic markers (cleaved amplified polymorphic sequence, CAPS) resulting from cleavage with the respective restriction enzyme were all codominant markers. For CGPPB3491, as no restriction enzyme was identified including the SNPs, PCR products were sequenced for all the progeny for genotyping.

One CG was localized on peach linkage group (PG) 3 (CGPPB5806), five on PG5 (CGPPB2805, CGPPB5502, CGPPB3369, CGPPB3491, CGPPB3501), and two on PG7 (CGPPB3890 and CGPPB6189) of the peach J×F linkage map (table 16).

4.4.4.2 BIN MAPPING IN *FRAGARIA*

The 55 CGs giving amplification products in peach were tested in *Fragaria*, and among the 36 CGs (65.5%) that were amplified in strawberry, 14 (25.5%) were polymorphic between the two parents: half showed indel polymorphism and half showed SNP polymorphism (table 16). The indel length polymorphism was from 4 to 800 bp (fig. 21). The number of SNPs per CG ranged from 1 to 4. CGs were bin-mapped in all the seven *Fragaria* linkage groups (FGs; Fig. 19). CGPPB6189 was bin-mapped to FG1 (bin I:41) and CGPPB2895 and CGPPB6050 to FG2 (bin II:08 and II:14, respectively). Five CGs were bin-mapped to FG3: CGPPB5971 (bin III:13), CGPPB3705 (III:28), CGPPB2741, and CGPPB3378 to the same bin (III:44) and CGPPB6347 to bin III:82. CGPPB3547 was bin-mapped to FG4 (bin IV:82); CGPPB2805, CGPPB3699, and CGPPB3491 to FG5 (bin V:15, V:37, V:48, respectively); CGPPB5795 to FG6 (bin VI:68); and CGPPB5502 to FG7 (bin VII:21). The bin nomenclature referring to *Fragaria* follows Sargent *et al.* (2008).

Table 13. Characterization of PCR products obtained with the CG primer pairs in peach and strawberry. The intron size was obtained using the blast 2 seq available in NCBI. When the intron is near one of the primers, the size (in italics) was deduced from the difference between the observed and expected size of the fragment.

Candidate Gene	Predicted size (bp)	Observed size (bp)	N° of predicted introns in homologous <i>A. thaliana</i> sequence, predicted position on	N° of introns in peach sequence, position on the peach EST (size bp)	Amplification in strawberry
CGPPB2290	312	420	1, 655	1, 655 (111)	no
CGPPB2294	350	470	N	1, 441 (108)	no
CGPPB2300	265	550	2, 216 - 262	2, 216 (132) - 262 (166)	no
CGPPB2455	253	800	1, 619	1, 983 (547)	yes
CGPPB2473	273	273	N	N	yes
CGPPB2585	404	500	N	1, 365 (90)	no
CGPPB2693	513	600	N	1, 120 (109)	no
CGPPB2741	235	1200	2, 358 - 475	2, 358 (230) - 475 (812)	yes
CGPPB2761	514	1000	N	1, 461(486)	yes
CGPPB2791	548	548	N	N	yes
CGPPB2799	220	310	N	1, 138 (93)	no
CGPPB2805	231	450	1, 823	1, 791 (219)	yes
CGPPB2895	275	520	2, 324 - 465	2, 324 (113) - 465 (151)	yes
CGPPB3043	228	900	1, 513	1, 513 (652)	no
CGPPB3070	250	1200	2, 410 - 527	2, 410 (127) - 530 (823)	yes
CGPPB3133	421	421	N	N	yes
CGPPB3143	308	680	2, 451 - 592	2, 194 (260) - 592 (113)	yes
CGPPB3190	410	410	N	N	no
CGPPB3212	438	No amplification	-	-	-
CGPPB3258	212	1000	2, 276 - 359	2, 276 (101) - 362 (687)	yes
CGPPB3303	232	450	1, 429	1, 429 (194)	yes
CGPPB3316	401	650 - 1000	1, 202	2, 202 (217) - 426 (181)	yes
CGPPB3369	218	700	2, 419 - 487	2, 419 (402) - 487 (116)	no
CGPPB3373	299	No amplification	-	-	-
CGPPB3378	234	234	1, 159	N	yes
CGPPB3491	245	550	1, 379	1, 379 (291)	yes
CGPPB3501	485	485	N	N	yes
CGPPB3547	308	550	N	1, 200 (293)	yes
CGPPB3575	579	900	N	2, 471 (205) - 550 (108)	yes
CGPPB3699	420	420	N	N	yes
CGPPB3705	500	500	N	N	yes
CGPPB3861	438	550	N	1, 371 (98)	no
CGPPB3890	246	1000	3, 417 - 480 - 577	2, 417 (400) - 480 (300)	yes
CGPPB5473	376	470	N	1, 153 (87)	no
CGPPB5502	199	+1000	2, 355 - 496	2, 355 (104) - 496 (697)	yes
CGPPB5514	379	+1000	3, 226 - 342 - 442	3, 226 (107) - 342 (126) - 442 (388)	no
CGPPB5532	260	390	1, 424	1, 424 (125)	yes
CGPPB5536	515	515	N	N	no
CGPPB5569	477	477	N	N	yes
CGPPB5575	210	730	2, 370 - 463	2, 370 (105) - 463 (417)	yes
CGPPB5654	600	1000	N	2, 358 (110) - 412 (290)	no
CGPPB5670	504	504	N	N	no
CGPPB5676	226	320	1, 349	1, 349 (94)	no
CGPPB5756	354	354	N	N	no
CGPPB5769	229	420-550	2, 173 - 296	420, 2, 173 (129) - 296 (109) 550, 2, 173 (129) - 296 (227)	yes
CGPPB5792	279	380	1, 278	1, 278 (85)	yes
CGPPB5795	256	600	1, 306	1, 306 (338)	yes
CGPPB5800	229	900	2, 204 - 302	2, 204 (401) - 302 (270)	yes
CGPPB5806	325	750	1, 435	1, 435 (434)	yes
CGPPB5921	533	No amplification	N	-	-
CGPPB5971	281	1150	2, 372 - 495	1, 372 (1,100)	yes
CGPPB6033	515	515	N	N	yes
CGPPB6050	243	600	2, 344 - 419	2, 344 (202) - 419 (176)	yes
CGPPB6189	248	330	1, 385	1, 385 (82)	yes
CGPPB6190	206	850	2, 285 - 385	2, 285 (119) - 385 (525)	no
CGPPB6212	366	No amplification	N	-	-
CGPPB6260	317	700	N	2, 276 (149) - 369 (383)	yes
CGPPB6322	206	520	2, 448 - 516	2, 448 (98) - 516 (216)	no
CGPPB6347	229	600	2, 252 - 403	2, 252 (252) - 403 (119)	yes

Tabla 14. *Fragaria*-specific primers for a subser of CGs.

Candidate gene	<i>Fragaria</i> - specific primer	Homology in VBI database contigs (<i>Prunus</i> EST)	Forward primer (5'-3')	Reverse primer (5'-3')
CGPPB2805	CGFVC2805	825	CCAACTTGAGTGGGAGGACT	TGCTCTGTACTTGCCGTTG
CGPPB2895	CGFVC2895	206189	GGCGGCTCTCAGAATGTACT	TATCGGCACCTTTGTGCTTG
CGPPB3491	CGFVC3491	204985	CGCTAGCACAGCTACCTCCT	ATGACCTTAAACACGTCGG
CGPPB3699	CGFVC3369	138954	TGCCGAATCTTCTGTATGG	ATTGATAAAGTACAAGGCTTATCAGG
CGPPB6189	CGFVC6189	159448	CATTCGATGTAGCGACGAC	CTTGACTGTTGTTGAGGCCA

Table 15. PCR products obtained with *Fragaria*-specific primers for a subset of CGs.

<i>Fragaria</i> primers	Fragment size (bp)	Intron size in <i>Fragaria</i> (bp)	Intron size in peach (bp)
CGFVC2805	524	137	219
CGFVC2895	462	98-114	113-151
CGFVC3491	718	712	291
CGFVC3369	471	81	82
CGFVC6189	471	81	82

Table 16. CG polymorphism and mapping in peach and diploid *Fragaria*

Name	Size of analysed sequence	<i>Prunus</i>			<i>Fragaria</i>			Bin mapping
		SNP position	SNP mapping method	Mapping LG, position from the top	Primer pair used	SNP position/Indel length		
CGPPB2741	1200	0	-	-	Peach	100 bp INDEL	III:44	
CGPPB2805	360	A/G 301	CAPS marker	5, 38,6 cM	<i>Fragaria</i>	4bp INDEL	V:15	
CGPPB2895	436	0	-	-	<i>Fragaria</i>	5 bp INDEL	II:08	
CGPPB3369	670	T/C 178, A/G 242	CAPS marker	5, 64,4 cM	Peach	No amplification	-	
CGPPB3378	167	0	-	-	Peach	G/A 145F, C/A 272F, A/G 355F	III:44	
CGPPB3491	520	A/T 129, T/A 133	Direct SNP detection	5, 77,2 cM	<i>Fragaria</i>	A/T 189F, T/A 241F	V:48	
CGPPB3501	428	A/G 278, A/T 281	CAPS marker	5: 76,6 cM	Peach	0	0	
CGPPB3547	518	0	-	-	Peach	600 pb INDEL	IV:82	
CGPPB3699	361	0	-	-	<i>Fragaria</i>	A/T 251F, C/G 264F, AG 267F,	V:37	
CGPPB3705	438	0	-	-	Peach	C/T 115R	III:28	
CGPPB3890	800	G/C 90, A/T 262, C/T 612, C/T 647, A/G 234, A/G 256, C/G 284,	CAPS marker	7, 42 cM	Peach	0	0	
CGPPB5502	780	G/A 306, C/A 321, A/G 397, G/A 426, G/C 472	CAPS marker	5, 44,5 cM	Peach	500 bp INDEL	Bin VII:21	
CGPPB5795	548	0	-	-	Peach	G/A 256R	VI:68	
CGPPB5806	750	T/G 100	CAPS marker	3, 0 cM	Peach	0	0	
CGPPB5971	1,15	0	-	-	Peach	800 bp INDEL	III:13	
CGPPB6050	571	0	-	-	Peach	50 bp INDEL	II:14	
CGPPB6189	284	A/T 63, T/C 75, G/A 105, T/G 186	CAPS marker	7, 59,2 cM	<i>Fragaria</i>	C/T 174R, C/T 221R	I:41	
CGPPB6347	545	0	-	-	Peach	G/T 398F	III:82	

F (forward) and R (reverse) indicate the primer from which the SNP position is referred in *Fragaria* sequences.

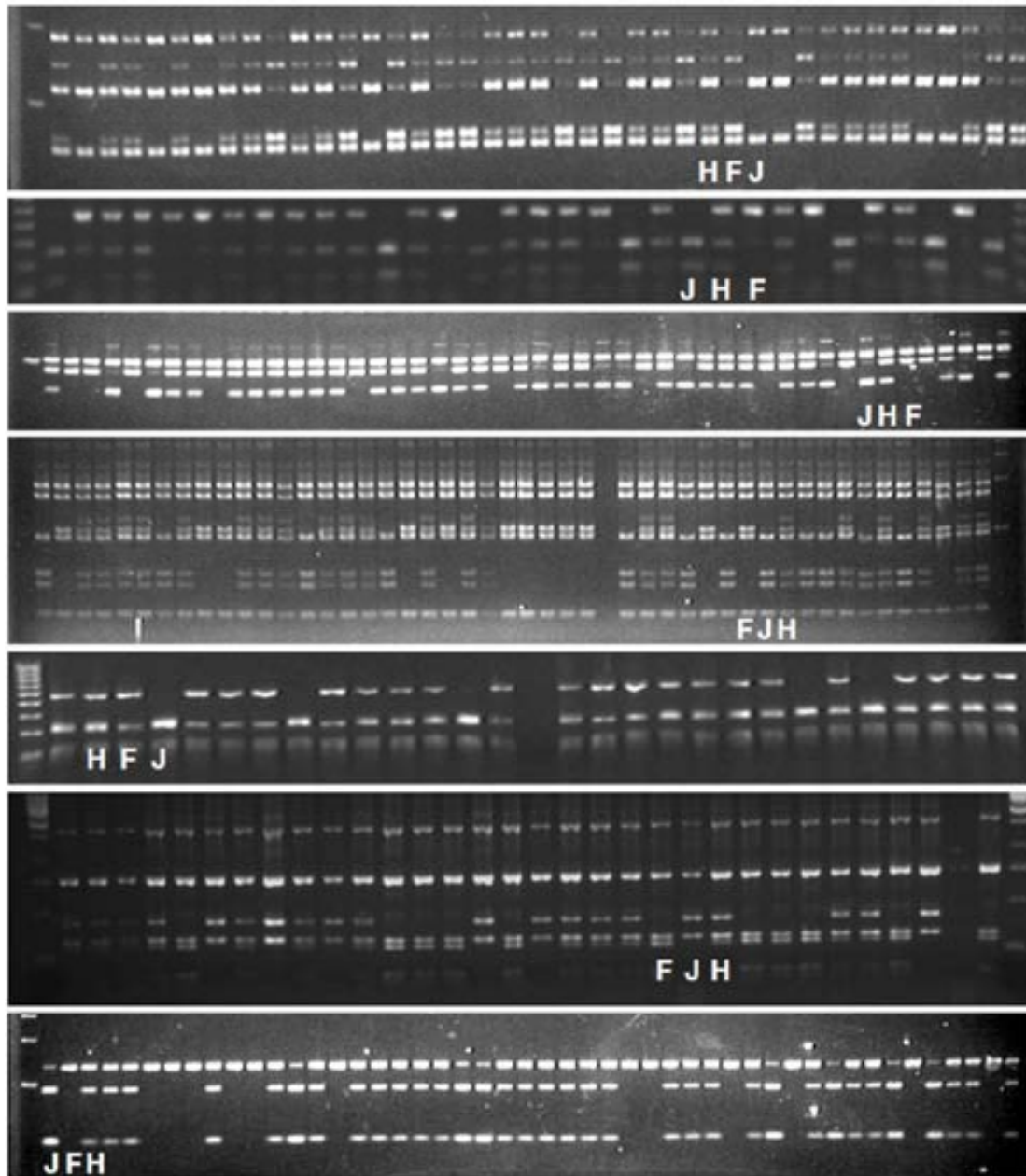


Figure 20. Agarose gels of the seven CAPS markers derived from CGs in peach. The patterns of the two parents and of the hybrid of the JxF progeny are indicated (*J* Ferjalou-Jalousia®, *F* Fantasia, *H* F₁ hybrid).

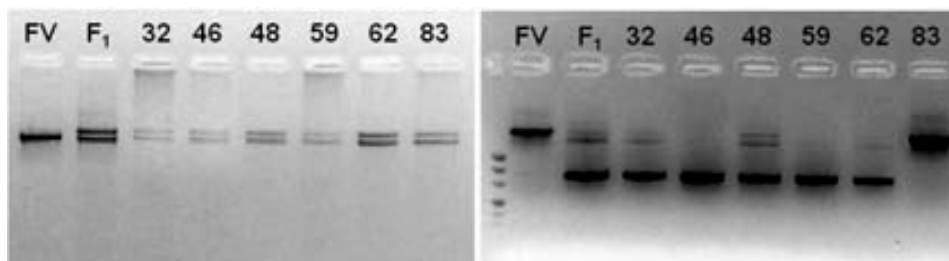


Figure 21. Agarose gels of two INDEL markers derived from CGs in strawberry. The patterns of one of the parents (FV, *F. vesca*) and the F₁ hybrid of the FVxFB bin set is followed by the corresponding mapping individuals of the diploid *Fragaria* bin set.

4.5 DISCUSSION

4.5.1 PEACH cDNA LIBRARY AND EST PROCESSING

Our preselection of peach ESTs of interest was based on the current knowledge of peach and strawberry fruit development stages, and peach, strawberry, and other fleshy fruit biochemical composition and metabolism and related membrane transport. Briefly, in peach, fruit acidity results from the synthesis and accumulation of the major organic acids, malic, citric, and quinic acids, that occur during rapid growth stages I and III, the latter corresponding to the enlargement of fruit cells (Etienne *et al.* 2002a). The increase in malic acid occurs during stage I, while the increase in citric acid is concomitant with the beginning of stage III (Etienne *et al.* 2002a; Moing *et al.* 1998). Later, during ripening, the concentration of organic acids decreases due to both dilution by sugars, which accumulate to high levels at this stage of fruit development, and reduction of synthesis/enhanced catabolism of organic acids (Tucker, 1993). In peach, as in other fleshy fruit species, the rapid growth stages I and III are both associated with the enlargement of mesocarp cells involving expansins (Hiwasa *et al.* 2003) and plasma membrane aquaporins (Hu *et al.* 2003). Organic acids are stored in mesocarp cell vacuoles before ripening using tonoplast transporters and proton pumps (Gaxiola *et al.* 2007; Masia *et al.* 1992; Shiratake & Martinoia 2007).

The major phenolic compounds in peach are hydroxycinnamic derivatives, mainly chlorogenic acid (5-*O*-caffeoyl-quinic acid) and its isomer neochlorogenic acid (3-*O*-caffeoyl-quinic acid), and flavonoids including the main anthocyanin cyanidin-3-glucoside, flavonol glycosides (quercetin and kaempferol derivatives), and the flavan-3-ol isomers catechin and epicatechin, which can polymerize into proanthocyanidins, known as condensed tannins (Senter & Callahan 1990; Tomas-Barberan *et al.* 2001). In the whole peach fruit, the phenolic level, expressed on a fresh weight basis, decreases during stage I, increases rapidly, and reaches a maximum in the middle of stage II and then decreases during stage III due to dilution by sugar and water accumulation (Kubota *et al.* 2000). Phenylalanine ammonia lyase (a key enzyme of phenolic compound biosynthesis) activity is high during stage I and the beginning of stage II, preceding the accumulation of phenolic compounds (Kubota *et al.* 2001). The cyanidin-3 glucoside anthocyanin accumulates in red parts of the peach skin (Tomas-Barberan *et al.* 2001; Tsuda *et al.* 2004). In deep red cultivars, there is a small peak of anthocyanin concentration at stage I which increases sharply at stage III of fruit development (Tsuda *et al.* 2004). This change is closely related to the increase in expression of the anthocyanin biosynthesis genes, including chalcone synthase (a key gene in flavonoid biosynthesis), flavonoid 3'-monooxygenase,

diydroflavonol-4-reductase, and leucoanthocyanidin dioxygenase genes (Tsuda *et al.* 2004). To identify CGs involved in these biological processes, cDNA libraries were constructed with mRNA extracted from fruit mesocarp plus epidermis of 'Fantasia' cultivar at two stages of fruit development: L1 was obtained at 30 DAB corresponding to the first rapid growth stage with cell division and enlargement of cells (stage I) and L2 was obtained at 80 DAB corresponding to the growth plateau (stage II) (Etienne *et al.* 2002a). This library construction strategy served to identify and map new candidate genes involved in fruit organoleptic quality in *Rosaceae* species.

From the cDNA libraries, 2579 ESTs were obtained. After clustering, the percentage singletons (ratio singletons/contigs, 53.4%) observed was within the range described for other small libraries of plant EST (Lewers *et al.* 2008).

4.5.2 SELECTION OF THE CANDIDATE GENES

Fifty-nine CGs were selected according to their functional annotation in relation to their potential role in fruit sweetness or acidity, phenolic content, and fruit growth or development. Of these, 55 were classified using the Gene Ontology scheme and MapMan structure as follows: 22 CGs corresponded to primary metabolism, five to cell wall fate, 12 to phenolic metabolism, and 14 to membrane transport. Two CGs potentially involved in regulation were also selected, two *14-3-3-like* proteins, one *MYB* factor, and one gene involved in cell cycle. Therefore, most of the candidate genes identified in the present work shared high sequence similarity with housekeeping genes functioning in primary or secondary plant metabolism and membrane transport.

In model fruit species such as tomato, randomly selected EST markers or EST markers chosen for their involvement in fruit size and composition have been used to construct genetic maps (Causse *et al.* 2004; Tanksley *et al.* 1992).

4.5.3 PRIMER DESIGN

Among the 55 primer pairs designed from peach EST sequences that amplified in peach, 36 (65.5%) also amplified in diploid strawberry. Sequence comparison between each PCR product and corresponding EST indicates that primers and PCR conditions were sufficiently selective to amplify the desired fragment. The transportability of markers from *Prunus* to *Fragaria* reported here is similar to what has been reported between *Malus* and *Prunus* (69%) and between *Malus* and *Fragaria* (61%) (Sargent *et al.* 2009c) and higher than the 50% transferability reported for primer

pairs designed arbitrarily from mRNA sequences without taking into account conserved exon regions (Sargent *et al.* 2007). The strategy of designing the primers in the exons appears to be very efficient for developing universal CGs.

By comparing genomic and EST sequences, it was possible to identify intron position and length. For 75% of the CGs, the number of introns in peach was identical to the number of introns in *A. thaliana*. Moreover, the predicted position was exactly as predicted from homologous *A. thaliana* sequences. This is in agreement with the high conservation of gene intron/exon junction reported in *Prunus* and *Arabidopsis* (Georgi *et al.* 2003; Jung *et al.* 2006).

4.5.4 SNP DETECTION AND CANDIDATE GENE MAPPING

4.5.4.1 MAPPING IN PEACH

As expected from previous studies (Dirlewanger *et al.* 2006), a very low level of polymorphism was observed in the peach J×F progeny, even taking into account intron sequences: only eight CGs of the 55 tested (14%) were polymorphic.

CGPPB5806 (flavonoid biosynthesis) is located at the distal end of PG3. In a previous study, no QTL was detected in this linkage group for phenolic concentration (Dirlewanger *et al.* 2005). Five CGs were located on PG5. Three of them are involved in phenolic metabolism: CGPPB2805 (anthocyanin biosynthesis), CGPPB3491 (flavonoid biosynthesis), and CGPPB3369 (phenylpropanoid metabolism). The other two CGs have a potential role in fruit growth or development: CGPPB5502 (aquaporin) and CGPPB3501 (oxidation reduction process). In the same region, QTLs were detected for dry flesh mass content, fresh flesh mass of fruit, fruit suture diameter, and fruit polar diameter (Quilot *et al.* 2004). Two CGs are located on PG7: CGPPB3890 (cytosolic pyruvate kinase) and CGPPB6189 (sucrose synthase) and are putatively involved in sweetness. QTLs for sucrose (Dirlewanger *et al.* 2005; Quilot *et al.* 2004), glucose, and fructose (Quilot *et al.* 2004) and for total sugar content (Dirlewanger *et al.* 2005) have been reported in this region. These colocalizations will be further investigated in other *Prunus* progenies in the European ISAFRUIT Integrated Project (FP6-FOOD-CT-2006-016279).

4.5.4.2 BIN MAPPING IN STRAWBERRY

The polymorphism detected in the FV×FB F₂ interspecific *Fragaria* progeny was much higher than in the J×F peach mapping progeny (39%). This polymorphism was half due to indels and half to SNPs in contrast to peach for which polymorphism was only due to SNPs. Fourteen CGs were polymorphic

and bin-mapped on all the seven FGs of *Fragaria*, in 13 of the 46 bins that cover the *Fragaria* genome. CGPPB2805 (anthocyanin biosynthesis) and CGPPB3491 (flavonoid biosynthesis) show syntenic positions in the *Prunus* map. CGPPB2741 (*PBL-T4*, signal transduction) and CGPPB3378 (*MYB* factor, transcription regulation) are located in the same bin (III:44). Of all the CGs involved in glycolysis, only CGPPB5795 (fructosebiphosphate aldolase) mapped in FG6, in a homoeologous region of the octoploid *Fragaria* × *ananassa* (Rousseau-Gueutin *et al.* 2008), where a QTL of sugars has been described (Lerceteau-Köhler *et al.* 2006). A QTL analysis in diploid *Fragaria* progeny is necessary to verify CG-QTL colocalization.

4.5.5 SYNTENY ANALYSIS BETWEEN *PRUNUS* AND *FRAGARIA*

Four CGs were polymorphic in peach and *Fragaria*, two of them (CGPPB2805 and CGPB3491) were located on PG5 and FG5, CGPPB5502 on PG5 and FG7, and CGPPB6189 on PG7 and FG1 (table 15). This is partially in agreement with what has been reported by Vilanova *et al.* (2008) where linkage group 5 of *Prunus* and *Fragaria* were reported to be the most collinear. PG7 has been reported to include markers of two *Fragaria* linkage groups (FG1 and FG6). CGPPB5502 (Aquaporin PIP1-2), which was bin mapped out of syntenic region (FG7 instead of FG5), gave three amplicons in *Fragaria*, but only one was polymorphic and was mapped. The mapped *Fragaria* sequence may not be orthologous to the *Prunus* CGPPB5502 but rather a paralogous sequence. In order to complete this pattern, the CGs will be bin mapped in the T×E *Prunus* reference map of Howad *et al.* (2005).

4.6 CONCLUSION

We have developed a peach EST database (PeachESTdb) including ESTs from two cDNA libraries corresponding to two fruit development stages (stage I and stage II). The annotations of the ESTs are available at <http://www.cbib.ubordeaux2.fr/SAM2/COMPLETE/> in project list: Peach- ESTdb. The 59 CGs selected according to their functional annotation are potentially involved in fruit sweetness or acidity, phenolic content, and fruit growth and development. Primer pairs were designed and used for PCR amplification in peach and diploid strawberry. A high level of marker transportability between peach and strawberry was obtained, indicating that the strategy of designing primers in exons is efficient, providing a source of CGs that could be of use within the *Rosaceae* species. Eighteen CGs were mapped: eight in peach, 14 in strawberry, and four in both species. In peach, they are located on three linkage groups and in strawberry on all the seven linkage groups. Several CG/QTL colocalizations were identified and will be further investigated for possible roles of CGs in the corresponding QTL trait.

SEGUNDA PARTE

**Desarrollo y caracterización preliminar de una colección de
líneas de introgresión en *Fragaria* diploide**

1. INTRODUCCIÓN

1.1 CONOCIMIENTO ACTUAL DE LAS BASES GENÉTICAS Y MOLECULARES DEL DESARROLLO Y MADURACIÓN DEL FRUTO DE FRESA

Los avances en Biología Molecular en los últimos 15 años han posibilitado el conocimiento del control molecular del desarrollo y maduración del fruto. La mayor parte de estos estudios se han desarrollado en tomate (Davuluri *et al.* 2005; Saladié *et al.* 2007) donde se ha identificado la gran complejidad del control de caracteres individuales como el tamaño, la forma, la firmeza o el contenido en azúcares y ácidos orgánicos (Causee *et al.* 2007). El fruto de fresa difiere significativamente del tomate, ya que éste deriva del tejido ovárico, mientras que el fruto de fresa deriva del receptáculo. Además, la fresa es un fruto no climatérico. La fresa diploide es un sistema modelo atractivo para el estudio del desarrollo y maduración en frutos no climatéricos (Mezzetti, 2009).

El desarrollo de cualquier fruto comprende varias etapas que comprenden la antesis, el desarrollo del ovario, la fertilización, la división celular, la formación de las semillas, el desarrollo de los embriones y la expansión celular (Gillaspy *et al.* 1993). El fruto de fresa difiere del resto de las bayas en que está compuesto por un receptáculo embebido en el verdadero fruto o aquenio, y por ello su desarrollo es distinto en algunas fases: Los aquenios se originan a partir de los carpelos tras la fecundación y el receptáculo se origina por el engrosamiento del eje floral (Hancock, 1999). Durante el crecimiento aumenta el número y volumen de células del fruto y, en la maduración, se producen una serie de cambios en la textura, color, tamaño, composición de azúcares y ácidos orgánicos, además de una alteración en la síntesis de pigmentos y síntesis y acumulación de compuestos aromáticos (Giovanonni, 2001). Durante la maduración de los frutos no climatéricos hay un descenso en la actividad respiratoria.

En los últimos años se han desarrollado diversos estudios de expresión génica en fresa mediante el uso de herramientas genómicas como los *microarrays* y las genotecas de sustracción de ESTs. Dichos estudios han permitido identificar una serie de genes cuyo patrón de expresión presenta grandes diferencias durante las distintas etapas de desarrollo y maduración del fruto. Las pruninas y oleosinas se expresan preferencialmente en los aquenios, mientras que en los receptáculos destacan los genes responsables de la síntesis de antocianinas, como son la chalcona isomerasa (*CHI*) o la antocianidina sintasa (*ANS*) (Aharoni y O'Connell, 2002). Bombarely (2007) identificó una serie de genes que modifican su expresión durante el desarrollo del fruto. Estos genes codifican para peroxidadas,

proteínas de unión a clorofilas, pruninas, pectato liasa, una proteína de interacción con el oncogen 6b, una proteína similar a *ETTIN* (*ETTIN* o *ETT* es una proteína de unión a ADN ligada a elementos de respuesta a auxinas), factores de elongación alfa-1 y otros genes de función desconocida. Salentijn *et al.* (2003) estudiaron las diferencias entre variedades con distintos valores para la firmeza del fruto, identificando dos pectin-metil esterasas y una poligalacturonasa como los genes que presentan mayores diferencias, junto con la expansina *FaExp2* y otra poligalacturonasa. Remarcablemente, estos autores identificaron dos patrones de expresión claramente diferenciados para los genes relacionados con la síntesis de ligninas, especialmente para un gen homólogo de la cinnamoyl CoA reductasa (*FaCCR*), que mostró niveles 20 veces superiores en la variedad menos firme.

La compleja regulación hormonal del desarrollo y maduración del fruto de fresa es poco conocida, si bien estudios recientes ponen de manifiesto la intervención de 27 genes relacionados con la síntesis de hormonas (Bombarely *et al.* 2010); 14 de ellos están relacionados con la síntesis de auxinas, tres genes están implicados en la síntesis de brasinosteroides, dos genes intervienen en la síntesis de citoquinas, dos genes están relacionados con la síntesis de giberelinas, un gen es responsable de la síntesis de ácido abscísico y 5 genes están implicados en la síntesis de etileno.

Gran parte de los genes reguladores de la maduración identificados en fresa (principalmente expansinas) están regulados negativamente por auxinas (Bellido-Cabello, 2005) o bien escapan al control de éstas, como es el caso de la expansina *FaExp2*. Parece que las señales no climatéricas independientes de auxinas podrían coordinar el desarrollo de la maduración en fresa (Civello *et al.* 1999). El papel de las auxinas ha sido analizado en diversos estudios de desarrollo y maduración de la fresa octoploide, debido a la facilidad para eliminar de forma física su fuente endógena de auxina, los aquenios (Nitsch, 1959). Entre los 6 y 12 días tras la antesis se produce un incremento del contenido en auxinas asociado al comienzo de la fase exponencial de crecimiento del fruto. Las endo- β -1,4-glucanasas (EGasas o celulasas), codificadas por una familia multigénica, se expresan predominantemente durante la maduración. La expresión de *faEG1* y *faEG3* está regulada negativamente por auxinas (Trainotti *et al.* 1999). *Cel1* y *Cel2*, localizados en los grupos de ligamiento GL4 y GL5 de *Fragaria* diploide (Sargent *et al.* 2007), se expresan diferencialmente en las distintas etapas del desarrollo del fruto (Llop-Tous *et al.* 1999); los transcritos de *Cel1* no son detectables en el fruto verde y se detecta un leve aumento de la expresión en el fruto blanco, que va aumentando gradualmente durante la maduración, alcanzando el máximo en el fruto completamente maduro. *Cel2*, en cambio, se expresa durante la etapa de fruto verde, se acumula y permanece en niveles elevados constantes durante la maduración. Los altos niveles de expresión de ambos genes en el

fruto maduro y su patrón de expresión solapante sugieren que las celulasas juegan un papel importante durante la maduración.

Muchos de los caracteres más interesantes de cara a la mejora genética de la fresa están involucrados en el desarrollo y maduración del fruto (textura, síntesis y acumulación de azúcares y ácidos orgánicos, polifenoles y vitaminas, entre otros) y su control genético escapa aún a nuestro conocimiento. Entre las variedades comerciales de fresa podemos encontrar frutos de formas y tamaños muy variados, diferentes texturas y colores, distinto contenido en sólidos solubles, polifenoles y capacidad antioxidante (Monfort *et al.* 2008). En otras especies de la familia *Rosaceae* se han identificado QTLs con importantes efectos sobre la calidad del fruto que están siendo incorporados en diferentes programas de mejora genética de variedades comerciales, como es el caso de los géneros *Prunus* (Peace *et al.* 2005; Ogundiwin *et al.* 2008b) y *Malus* (Bink *et al.* 2007; Yao *et al.* 2007). El control sobre estos caracteres permitiría incrementar la calidad nutricional del fruto de fresa o reducir los daños producidos durante el transporte de la fruta, así como generar nuevas variedades que pudieran satisfacer las necesidades cambiantes del mercado.

1.2 MAPEO GENÉTICO DE LA DIVERSIDAD FENOTÍPICA EN PLANTAS

Los caracteres interesantes a nivel agronómico tienen, en general, una herencia cuantitativa. Los factores genéticos implicados en la mayor parte de este tipo de caracteres son genéticamente complejos, ya que presentan una distribución continua; están controlados por la acción de varios genes, las interacciones entre ellos y su expresión depende de las condiciones ambientales. El análisis de QTLs es una manera de describir las características de las distribuciones fenotípicas continuas y determinar el número de loci implicados en el control de un carácter fenotípico, así como su transmisión a las siguientes generaciones (Tanksley, 1993). El término QTL fue acuñado por Gelderman (1975) para referirse a cada uno de los 'factores efectivos' que controlan parte de la variación de un carácter cuantitativo; indica un único locus, aunque puede tratarse de varios genes estrechamente ligados. En 1923, Sax describió la asociación de un carácter poligénico, el tamaño de las semillas de la judía, con un carácter monogénico, la pigmentación de la semilla. Desde entonces, los estudios sobre el ligamiento entre caracteres monogénicos o marcadores y caracteres cuantitativos o QTLs han sido los protagonistas en las principales publicaciones especializadas en genética. Los avances en genómica estructural y análisis estadístico multivariante han propiciado la aparición de nuevas vías de aproximación al estudio de la variación fenotípica en especies de alto

interés económico como el girasol (Poormohammad Kiani *et al.* 2007), el tomate (Frary *et al.* 2000), o el trigo (Raman *et al.* 2005).

En poblaciones naturales, es muy raro encontrar un marcador asociado a un QTL, a menos que el marcador esté completamente ligado al él. Por ello, el análisis de QTLs ha de realizarse asumiendo su segregación en poblaciones de mapeo, tales como las poblaciones F_1 , poblaciones F_2 , poblaciones de líneas diplohaploides (DHLs), poblaciones derivadas de un retrocruzamiento (*backcross*) o poblaciones basadas en cruzamientos recurrentes (RILs, BILs, NILs y SLs). Con el desarrollo de un gran número de marcadores moleculares, los QTLs pueden ser estudiados en relación a sets de marcadores ligados que son analizados simultáneamente para determinar sus efectos sobre los caracteres cuantitativos. Este tipo de análisis es conocido como mapeo por intervalo o *interval mapping* (Lander y Botstein, 1986, 1989). Esta aproximación permite la compensación por recombinación entre marcadores y QTLs, incrementando la probabilidad de detección de QTLs, particularmente de aquellos con efectos menores (Tanksley, 1993). Se han identificado QTLs de interés agronómico para caracteres como la calidad de fruto (Moreno *et al.* 2008), fecha de floración (Weebadde *et al.* 2008), productividad (Semel *et al.* 2006), resistencia a enfermedades y plagas (Fukino *et al.* 2008), tolerancia a estrés (Xu *et al.* 2006) o para caracteres involucrados en la domesticación (Tian *et al.* 2009), entre otros. Paralelamente, los avances en Genómica comparada han hecho posible que QTLs descritos en unas especies puedan ser identificados y analizados en otras (Doganlar *et al.* 2002), de manera que la información genética que podemos obtener a partir del estudio de QTLs es, sin duda, una fuente de conocimiento de la variación fenotípica presente en la Naturaleza.

Las poblaciones de mapeo convencionales (F_1 , F_2 , *backcross* o DHLs) pueden inducir a error durante la identificación y mapeo de QTLs, debido a la sobreestimación del efecto de los QTLs (Melchinger *et al.* 1998), a su bajo poder de resolución (van Ooijen, 1992) y la dificultad para distinguir entre pleiotropía y ligamiento de QTLs (Lebreton *et al.* 1998). El intervalo de confianza para la localización de QTLs mayores en poblaciones de este tipo depende de varios factores; En una población tipo *backcross* formada por 300 individuos, o en una población de 150 LDHs con un mapa genético poco saturado, estaría en torno a 35 cM para QTLs de alta heredabilidad (Hyne *et al.* 1995). En poblaciones de similares características, pero de las que se dispone de mapas genéticos de alta densidad, es del orden de 20 cM (Fernández-Silva *et al.* 2010). Además, existen limitaciones para identificar QTLs con efectos menores, ya que un QTL mayor puede enmascarar los efectos de otros QTLs menores que no cosegregan con él, incrementando la variación fenotípica total, por lo que genes con efectos

menores pueden quedar por debajo del umbral de detección. Este hecho interfiere con la correcta estimación del número de QTLs, y por tanto con su posterior mapeo fino y clonaje. De igual forma, la interacción entre dos loci no ligados puede reducir la diferencia entre subgrupos para el análisis de un QTL, y por tanto, los loci pueden escapar de la detección (Eshed y Zamir 1995).

Los estudios de análisis de QTLs en especies del género *Fragaria* son muy escasos. Las bases genéticas de la floración remontante han sido estudiadas en diferentes cruzamientos de *F. x ananassa*. Wu *et al.* (1992) identificaron 8 QTLs que explicaban más del 36% de la varianza, indicando que se trata de un carácter cuantitativo. Sugimoto *et al.* (2005) identificaron dos marcadores RAPDs ligados a un único QTL para el mismo carácter. Weebaddee *et al.* (2008) también identificaron dos marcadores AFLP ligados a la floración remontante, llegando a identificar hasta 8 QTLs para este carácter. El carácter *semperflorens* de la accesión FDP815 de *F. vesca* fue localizado en la posición 29.6cM del GL6 en la primera versión del mapa de referencia de *Fragaria* (Sargent *et al.* 2004a), pero este marcador no ha vuelto a ser incluido en ninguna de las versiones posteriores de dicho mapa.

1.2.1 LÍNEAS RECOMBINANTES CONSANGUÍNEAS

Las líneas recombinantes consanguíneas, comúnmente denominadas RILs por su acrónimo en inglés (*Recombinant Inbred Lines*) se construyen mediante múltiples generaciones de cruzamientos consanguíneos a partir de individuos heterocigóticos, hasta conseguir la homocigosis de la mayor parte de sus loci (Bailey, 1971).

En plantas, la obtención de este tipo de poblaciones suele hacerse mediante el método de semilla única o SSD por sus siglas en inglés (*Single Seed Descent*). A partir de un híbrido F_1 se obtiene una población F_2 y, mediante autofecundaciones sucesivas ($F_3, F_4, F_5 \dots F_n$), se van generando plantas cuya homocigosis aumenta gradualmente hasta obtener RILs (en la práctica, a partir de la octava o novena generación). La población final estará formada por líneas homocigotas para prácticamente todos sus loci (fig. 22). La homocigosidad alcanza el 99,6% en la novena generación (Nuez, 2000). Este método es muy eficaz, tanto para el mapeo de marcadores como para la asociación de un carácter a un marcador, ya que los cruzamientos entre líneas consanguíneas generan un desequilibrio de ligamiento (LD) máximo entre un QTL y sus marcadores asociados.

Las poblaciones de RILs han sido empleadas con éxito para el mapeo de caracteres cuantitativos en cultivos como el maíz (Burr *et al.* 1988; Chung *et al.* 2010) el trigo (Carrillo *et al.* 1990), el tomate (Goldman *et al.* 1995) la cebada (Toojinda *et al.* 1998), el pepino (Fazio *et al.* 2003) o la soja (Yesudas

et al. 2010). Las RILs ofrecen algunas ventajas frente a poblaciones segregantes de mapeo convencional. Dado el alto grado de homocigosidad, los marcadores dominantes ofrecen una información equivalente a la de los codominantes (Staub *et al.* 1996). Cada línea ha de fenotiparse una sola vez y pueden fenotiparse varios individuos de cada línea, reduciendo así la variabilidad individual/ambiental. Las poblaciones de RILs pueden perpetuarse en el tiempo, por lo que pueden ser utilizadas indefinidamente para el mapeo de nuevos loci sobre un mapa de referencia, lo cual es muy atractivo de cara a especies vegetales anuales o estacionales.

Una propiedad particular de las RILs sobre otras poblaciones segregantes convencionales radica en que han tenido que atravesar varias rondas de meiosis antes de alcanzar la homocigosidad; La fracción de recombinación obtenida en las líneas consanguíneas es el doble de la tasa convencional para distancias genéticas pequeñas y aumenta, de manera no lineal, hasta alcanzar el 50% para marcadores no ligados. Esto supone algunas dificultades en las etapas iniciales del mapeo, ya que es más complicado establecer el ligamiento. Sin embargo, es más sencillo detectar recombinación entre marcadores fuertemente ligados (Taylor, 1978).

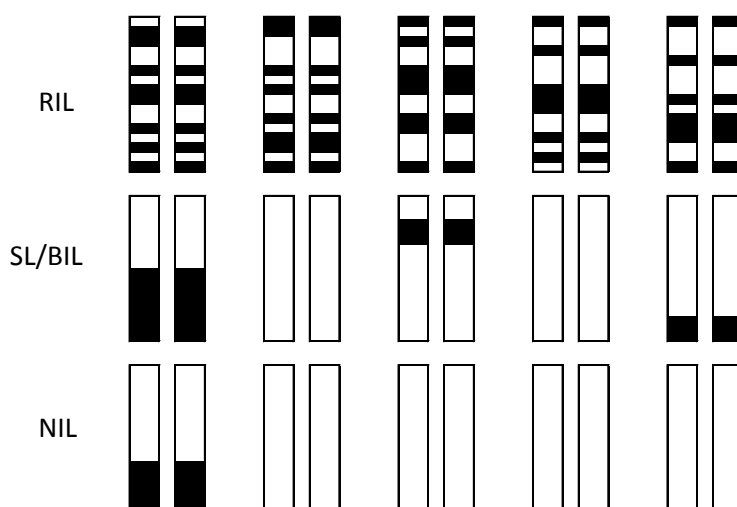


Figura 22. Representación esquemática del genoma de diferentes tipos de líneas de introgresión. Líneas de introgresión de una especie $2n=10$. Cada barra vertical representa un cromosoma. Cada par de cromosomas homólogos incluye regiones de diferentes colores que representan las regiones heredadas de cada uno de los parentales de la población.

1.2.2 LÍNEAS DE SUSTITUCIÓN

Las líneas de sustitución (SLs), a menudo referidas como CSSL (*Chromosome Segment Substitution lines*), o RCSL (*Recombinant chromosome substitution lines*) contienen varios segmentos cromosómicos de un genotipo donante en un fondo genético homogéneo de otro parental, denominado parental recurrente. Las colecciones de líneas de sustitución están formadas por líneas

que contienen, en su conjunto, el genoma completo del parental donante, donde cada línea porta una o varias sustituciones diferentes (Tanksley, 1993).

Normalmente, estas líneas proceden del retrocruzamiento recurrente de individuos que proceden, en origen, de un único individuo híbrido F_1 . Los fenotipos mostrados por este tipo de líneas pueden ser comparados con el mostrado por el parental recurrente, de manera que cualquier diferencia significativa es atribuida a la presencia de algún QTL en las regiones genómicas sustituidas.

Dado que son líneas homocigotas (las sustituciones son fijadas en una fase final de autofecundación), este tipo de poblaciones puede multiplicarse indefinidamente para mejorar la precisión en el mapeo de QTLs o eliminar el 'ruido' genético producido por otros loci segregantes no ligados.

Estas poblaciones poseen varias ventajas para el mapeo de caracteres cuantitativos sobre las poblaciones de RILs, ya que, además de reducirse el ruido producido por el fondo genético, los QTLs segregantes pueden analizarse como factores mendelianos mediante el cruzamiento de cada línea individual con el parental recurrente y la obtención de poblaciones F_2 . Además, su pequeño tamaño poblacional facilita en gran medida el mantenimiento y fenotipado de las plantas (Eshed y Zamir, 1995). Existen colecciones de líneas de sustitución en cultivos como el trigo (Sears *et al.* 1957) la colza (Burns *et al.* 2003), la cebada (Matus *et al.* 2003) o el algodón (Saha *et al.* 2006), si bien la mayor parte de la literatura reciente acerca de este tipo de poblaciones se ha desarrollado en arroz (Wan *et al.* 2004; Kwon *et al.* 2008).

1.2.3 LÍNEAS DE INTROGRESIÓN POR RETROCRUZAMIENTO

Las líneas de introgresión por retrocruzamiento o BILs, de su acrónimo en inglés (*Backcross Inbred Lines*), son una variante de las RILs. Este tipo de líneas se obtienen tras varias generaciones de retrocruzamiento con uno de los parentales (parental recurrente) que comienzan en la generación F_1 y terminan con un ciclo de autofecundación.

Las colecciones de BILs poseen características intermedias entre las poblaciones de RILs y las colecciones de líneas de sustitución, ya que cada línea posee una elevada proporción del genoma del parental recurrente (entre un 75 y un 90%, generalmente) y, en su conjunto, suelen cubrir por completo el genoma del parental donante.

La primera colección de BILs se desarrolló en trigo (Wehrhahn *et al.* 1965) y posteriormente se han desarrollado otras en cultivos como el arroz (Lin *et al.* 1998), el pimiento (Chaim *et al.* 2002), o el tomate (Eshed y Zamir, 1994), así como nuevas colecciones en trigo (Blanco *et al.* 2006).

1.2.4 LÍNEAS CASI ISOGÉNICAS

Las líneas casi isogénicas o NILs (*Near isogenic lines*) son líneas homocigóticas cuyo genoma presenta una introgresión o fragmento exógeno, procedente de un genoma donante. En plantas, una colección de NILs puede ser definida como una librería genómica de introgresiones solapantes del genoma donante sobre un genoma denominado receptor o recurrente (Eshed y Zamir 1995).

Como las BILs, las NILs se obtienen mediante el retrocruzamiento recurrente de plantas obtenidas a partir de uno o varios individuos híbridos (un individuo F_1 , individuos F_2 o LDHs) hasta obtener individuos con un único segmento cromosómico introgresado, que finalmente se autofecundan hasta conseguir la homocigosis completa del genoma (fig. 23).

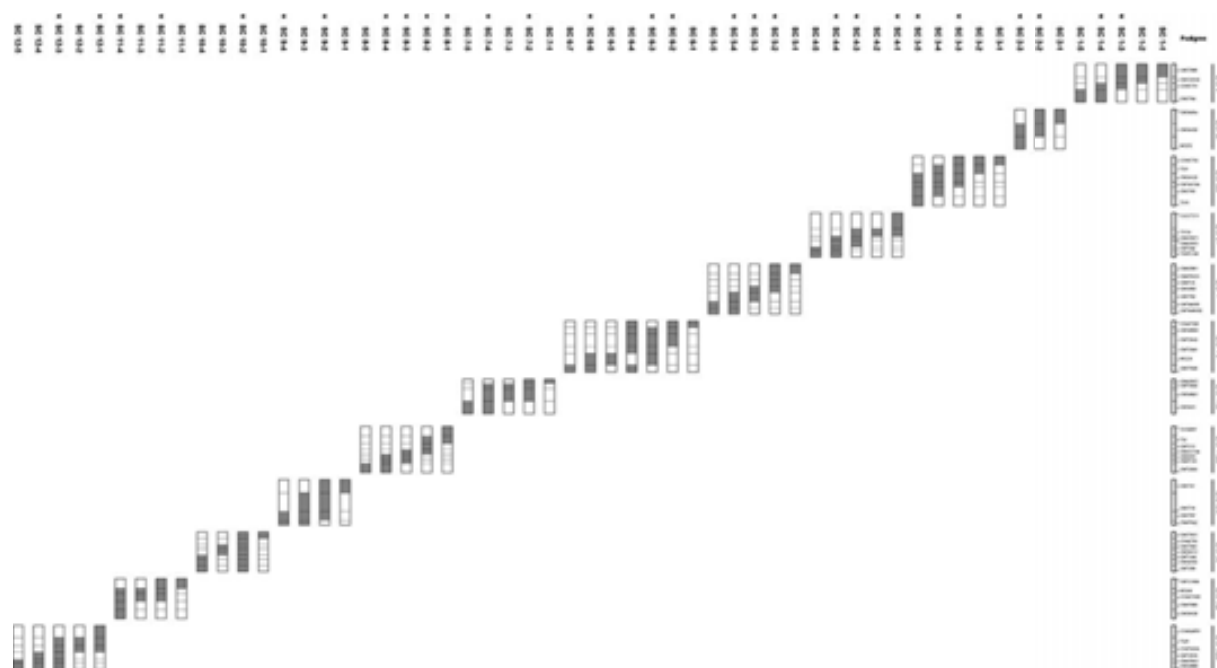


Figura 23. Representación gráfica del genoma de la colección de NILs desarrollada en melón. Arriba figura el nombre de cada línea y a la derecha, los distintos grupos de ligamiento. Cada barra vertical representa un par de cromosomas homólogos con regiones introgresadas, donde las introgresiones, delimitadas por marcadores moleculares, aparecen en gris. Extraído de Eduardo *et al.* (2005).

Dada su isogenicidad con respecto al genoma recurrente, toda la variación genética que distingue unas NILs de otras puede ser directamente asociada al segmento introgresado. Además, como cada línea contiene una única introgresión, la mayor parte de los problemas de fertilidad asociados a

poblaciones interespecíficas convencionales (en las que los individuos poseen una gran fracción del genoma una especie generalmente silvestre) son eliminados, de manera que los caracteres de interés agronómico pueden ser evaluados correctamente. Esta es la razón por la que, por ejemplo, no pudieron identificarse QTLs para el incremento de la productividad en tomate antes de la aparición de la primera colección de NILs de Eshed y Zamir (1995).

La primera colección de NILs en tomate de Eshed y Zamir (1995) constaba de 50 líneas y permitió la detección de un total de 104 QTLs: 16 para el peso de la planta, 22 para el peso del fruto inmaduro, 18 para la masa del fruto, 23 para el contenido en azúcares, 14 para el contenido en sólidos solubles y 11 para el rendimiento de la planta; una cifra nunca antes alcanzada para un estudio de caracteres cuantitativos en plantas. Esta primera colección fue seguida de otras muchas, con cobertura parcial (Bernancchi *et al.* 1998; Frary *et al.* 2003; Fridman *et al.* 2004; Brouwer y St Clair 2004; Canady *et al.* 2005; Semel *et al.* 2006) o total (Monforte y Tanksley, 2000a; Chetelat y Meglic, 2000) del genoma del tomate.

Se han desarrollado colecciones completas de NILs en especies vegetales como el trigo (Pestsova *et al.* 2001) *Arabidopsis thaliana* (Koumproglou *et al.* 2002; Keurentjes *et al.* 2007), lechuga (Jeuken y Lindhout, 2004), y melón (Eduardo *et al.* 2005) y colecciones parciales que, o bien sólo abarcan regiones específicas del genoma, o no cubren la totalidad del genoma, en cultivos como la col silvestre, *Brassica oleracea*, (Ramsay *et al.* 1996) el arroz (Wan *et al.* 2004; Zhang *et al.* 2009), el maíz (Szalma *et al.* 2007), el sorgo (Harris *et al.* 2007), el trigo (Zhou *et al.* 2005) el tabaco (Lewis *et al.* 2007), la cebada (Marcel *et al.* 2007) o la soja (Kopisch-Obuch y Diers 2006; Bolon *et al.* 2010).

El análisis de caracteres cuantitativos en poblaciones de NILs supera las limitaciones inherentes a las poblaciones F₂ DHLs o *backcross*, empleadas tradicionalmente para este fin, así como las ligadas a cualquier otro tipo de población segregante (RILs, BILs o SLs). En una NIL que alberga un QTL, éste pasa a ser la única fuente de variación genética por la ausencia de otros QTLs segregantes. El QTL es pues, considerado un carácter de herencia mendeliana simple (Alonso-Blanco y Koornneef, 2000) de manera que, mediante el desarrollo de subNILs con introgresiones solapantes sobre la región que contiene el QTL, éste puede diseccionarse a través del mapeo por sustitución (Paterson *et al.* 1990), descartando la posibilidad de que dos introgresiones solapantes con un efecto fenotípico similar porten el mismo QTL.

El uso de NILs ha transformado la tarea del clonaje de QTLs, que ahora es similar al realizado para caracteres de herencia mendeliana simple (Frary *et al.* 2000; Fridman *et al.* 2004) ya que las NILs proporcionan un alto poder de resolución en el mapeo que facilita la inclusión de un QTL en un único BAC o YAC. Empleando un número mínimo de réplicas por genotipo, pueden detectarse un gran número de QTLs en comparación con otros estudios similares en los que cada genotipo debe ser medido hasta cientos de veces.

Las colecciones de NILs son el escenario óptimo para verificar los efectos de un QTL (Tanksley *et al.* 1996; Eshed y Zamir 1995), estudiar las interacciones de un QTL con otros QTLs (Monforte y Tanksley 2000b), de un QTL con el ambiente (Warrington *et al.* 2008; Fernandez-Silva *et al.* 2010), investigar la herencia de caracteres oligogénicos (Essafi *et al.* 2009) o identificar las interacciones generadas entre un QTL con su fondo genético o con otros QTLs (Monforte *et al.* 2001; Reif *et al.* 2009). Este tipo de interacciones incluyen fenómenos genéticamente complejos como la pleiotropía, la epistasia o la estimación del ligamiento entre QTLs. Estos estudios son difíciles de llevar a cabo en poblaciones segregantes convencionales, dado que la frecuencia de algunas combinaciones genotípicas es demasiado baja como para poder establecer comparaciones significativas. Otras ventajas adicionales del uso de NILs es la facilidad para distinguir entre efectos aditivos y dominantes de un QTL, o la ventaja estadística que supone el uso de un control común para todas las líneas en el diseño experimental, lo que posibilita un análisis estadístico con un alto valor de grados de libertad. Frente a las múltiples ventajas que supone el uso de NILs en el mapeo genético, existen una serie de inconvenientes. La principal desventaja es la gran inversión de tiempo y esfuerzo requerido para su desarrollo. El fenotipado de estas líneas requiere exhaustivas mediciones (Paran y Zamir, 2003).

Además de su uso en mapeo, las colecciones de NILs han sido empleadas en la introducción de nueva variabilidad genética procedente de especies silvestres en variedades élite (Tanksley y McCouch 1997; Zamir 2001), para estudiar las bases genéticas de la heterosis (Melchinger *et al.* 2007; Fernandez-Silva *et al.* 2009) o para el estudio de la distorsión de la tasa de transmisión de marcadores (TDR) en poblaciones de mapeo (Moyle y Graham 2006).

1.3 CONSTRUCCIÓN DE LÍNEAS CASI ISOGÉNICAS

En la literatura pueden encontrarse diferentes maneras de obtener NILs. En la mayoría de los casos, estas líneas derivan de RILs, BILs o SLs que se retrocruzan con el parental recurrente y se autofecundan para obtener NILs (Zhang *et al.* 2009). Esta estrategia es la más adecuada cuando se

pretende aislar cierta región genómica sobre un fondo isogénico. Sin embargo, cuando lo que se persigue es generar una colección completa de NILs cubriendo todo el genoma del parental donante, pueden emplearse diferentes métodos, en función de una serie de parámetros que deben tenerse en cuenta.

En primer lugar, ha de tenerse en cuenta el nivel de ploidía de la especie en cuestión. Durante la generación de NILs, los ciclos de retrocruzamiento van seguidos de selección asistida por marcadores. El análisis genético en especies poliploides puede llegar a ser extremadamente complejo, ya que los alelos de un locus pueden confundirse fácilmente con alelos procedentes de otras copias parálogas de dicho locus. La poliploidía dificulta enormemente la correcta asignación del número y localización de introgresiones. A día de hoy, no existen poblaciones de NILs en ninguna especie vegetal con genoma poliploide.

Es importante tener en cuenta el tamaño del genoma a cubrir. En especies con genomas grandes y un alto número de cromosomas tales como el trigo, es interesante partir de líneas previamente retrocruzadas. De no ser así, el aislamiento de una única región introgresada puede demorarse durante muchas generaciones. Este fue el caso de la población de NILs de arroz desarrollada por Li *et al.* (2004), que deriva de cinco generaciones de retrocruzamiento seguidas de siete generaciones de autofecundación.

El ciclo de vida de la especie con la que se trabaja es crucial en el desarrollo de NILs. En el caso de los árboles, que tienen ciclos de vida largos, la generación de NILs puede demorarse durante décadas. A pesar de ello, pueden desarrollarse NILs en especies arbóreas con genoma pequeño en un espacio de tiempo corto mediante la optimización de la selección asistida por marcadores en poblaciones *backcross* formadas por un alto número de individuos. El gran tamaño poblacional maximiza la probabilidad de encontrar individuos con bajo número de introgresiones cubriendo todo el genoma, de manera que pueden obtenerse NILs heterocigóticas en un par de generaciones de retrocruzamiento (Picañol *et al.* 2007).

La disponibilidad de mapas saturados para la especie con la que se trabaja es indispensable para la correcta selección asistida por marcadores. Cuando los mapas disponibles son poco saturados o contienen grandes fracciones del genoma sin marcadores moleculares, es probable que no se detecten introgresiones segregantes y/o fijadas en la población final, que pueden sesgar los análisis posteriores de análisis de QTLs. En la primera colección de NILs de melón (Eduardo *et al.* 2005), una

serie de pequeñas introgresiones no fueron identificadas. Gracias a la publicación de nuevos mapas genéticos, los autores fueron capaces de identificarlas y eliminarlas (Eduardo *et al.* 2007).

La correcta selección de las líneas parentales a partir de las cuales se desarrollará la población de NILs es un punto crucial en el diseño experimental. Suelen emplearse líneas homocigóticas como parental recurrente, puesto que la segregación en regiones no introgresadas puede enmascarar los efectos fenotípicos de las introgresiones. Es interesante emplear una variedad comercial como parental recurrente. Esto supone una ventaja adicional al uso de NILs sobre otro tipo de poblaciones de mapeo, ya que empleando variedades comerciales, la identificación de introgresiones con un efecto positivo sobre el genoma receptor es inmediatamente aplicable a la mejora de la variedad, además de servir de base para la introducción del carácter en nuevos cultivares (Eshed y Zamir, 1995). Igualmente, es interesante el uso de una especie o variedad silvestre como parental donante. De este modo, además de garantizar un alto grado de polimorfismo en la población, se consigue introducir nueva variabilidad genética procedente de una fuente exótica en un cultivar moderno. Una alta tasa de polimorfismo entre los parentales permite seleccionar de manera eficaz. Si el polimorfismo en la población es bajo, la selección se limita a unos cuantos loci, por lo que se dejarán descubiertas algunas regiones del genoma. Además, una tasa de polimorfismo baja es un indicador de que ambas especies o variedades están emparentadas genéticamente, por lo que probablemente, las diferencias fenotípicas entre ambas serán mínimas, limitando la segregación de caracteres fenotípicos interesantes en la población.

Por último, es preciso contar con un amplio número de marcadores moleculares de herencia codominante. El uso de marcadores moleculares de herencia dominante, que suelen analizarse por la presencia o ausencia del marcador en la descendencia, queda inevitablemente limitado a aquellos marcadores ausentes en el genoma recurrente y presentes en el genoma donante, ya que la dominancia en sentido opuesto no permite la identificación de introgresiones del genoma donante en poblaciones de tipo *backcross*, de modo que no es posible diferenciar un individuo heterocigótico (que presenta introgresión) de un individuo homocigótico (genotipo recurrente).

1.4 CONSTRUCCIÓN DE UNA COLECCIÓN DE NILs EN *FRAGARIA* DIPLOIDE

El genoma de las especies diploides del género *Fragaria* es de pequeño tamaño (Folta y Davis 2006). Su ciclo de vida es muy corto (en condiciones óptimas, puede obtenerse una generación en sólo tres meses) y las plantas pueden propagarse de forma vegetativa fácilmente, sin necesidad de recurrir a

la propagación *in vitro*. Durante la última década se ha avanzado mucho en la caracterización genómica del género, con la creación de una población de mapeo extraordinariamente polimórfica y la construcción de un mapa genético de alta densidad (Sargent *et al.* 2006; Ruiz-Rojas *et al.* 2010). Además, cabe destacar que las especies silvestres de *Fragaria*, en especial *F. vesca* y *F. virginiana*, son mucho más aromáticas que las cultivadas (Hirvi y Honkanen, 1982; Ulrich *et al.* 2007) Parece pues, un género ideal para el desarrollo de líneas de introgresión.

Para la construcción de una colección de NILs en *Fragaria* se ha seleccionado la línea PI 551824 de *Fragaria vesca*, variedad conocida como 'Reine des Vallées', y, en nuestro país, comúnmente conocida como 'Reina de los Valles', como parental recurrente. La principal ventaja de esta variedad francesa es que su cultivo está bastante extendido en España, ya que se emplea en la industria de las conservas de la fresa, principalmente en la producción de mermeladas y purés (Almenar, 2005), de manera que la mejora de cualquier carácter de interés agronómico puede ser transferida directamente a la industria. 'Reine des Vallées' es una variedad que se caracteriza por el intenso aroma de sus frutos, de carne blanda y fina. Se trata de una variedad de día neutro de rendimiento elevado que no desarrolla estolones. Es por ello que la variedad se ha propagado por semillas tradicionalmente, lo que, *a priori* garantiza un alto grado de homocigosidad en su genoma. Como la mayoría de variedades de *F. vesca*, es robusta y vigorosa, con largos peciolo, grandes folíolos y flores pequeñas. Sus frutos son alargados, de color rojo escarlata muy brillante (fig. 24). Además, es una variedad muy resistente a las enfermedades comunes de la fresa, por lo que las NILs que deriven de ella serían candidatas idóneas para realizar estudios de patología (Marcel *et al.* 2007).

El parental donante escogido, la línea PI 657844, también conocida como FDP601 de *Fragaria bucharica*, ya fue empleado como parental de la población de referencia para el género (FV x FB) por Sargent *et al.* (2004a). Este hecho facilita la elección de marcadores moleculares durante la selección asistida por marcadores. A pesar de ser una especie autoincompatible (Bošković *et al.* 2010), es compatible con otras especies de *Fragaria*. Se trata pues, de una especie silvestre, de día corto, sin valor comercial, de porte pequeño, poco vigorosa, de cortos peciolo, hojas pequeñas y flores grandes. Sus frutos son redondeados, con protuberancia superior o 'cuello', planos en la parte superior y de color rojo oscuro (fig. 24). Filogenéticamente es bastante próxima a *F. vesca*, pero ambas son morfológicamente divergentes (Sargent *et al.* 2005), ya que pertenece al grupo asiático de especies de *Fragaria*, lo que garantiza un amplio rango de fenotipos en poblaciones segregantes.

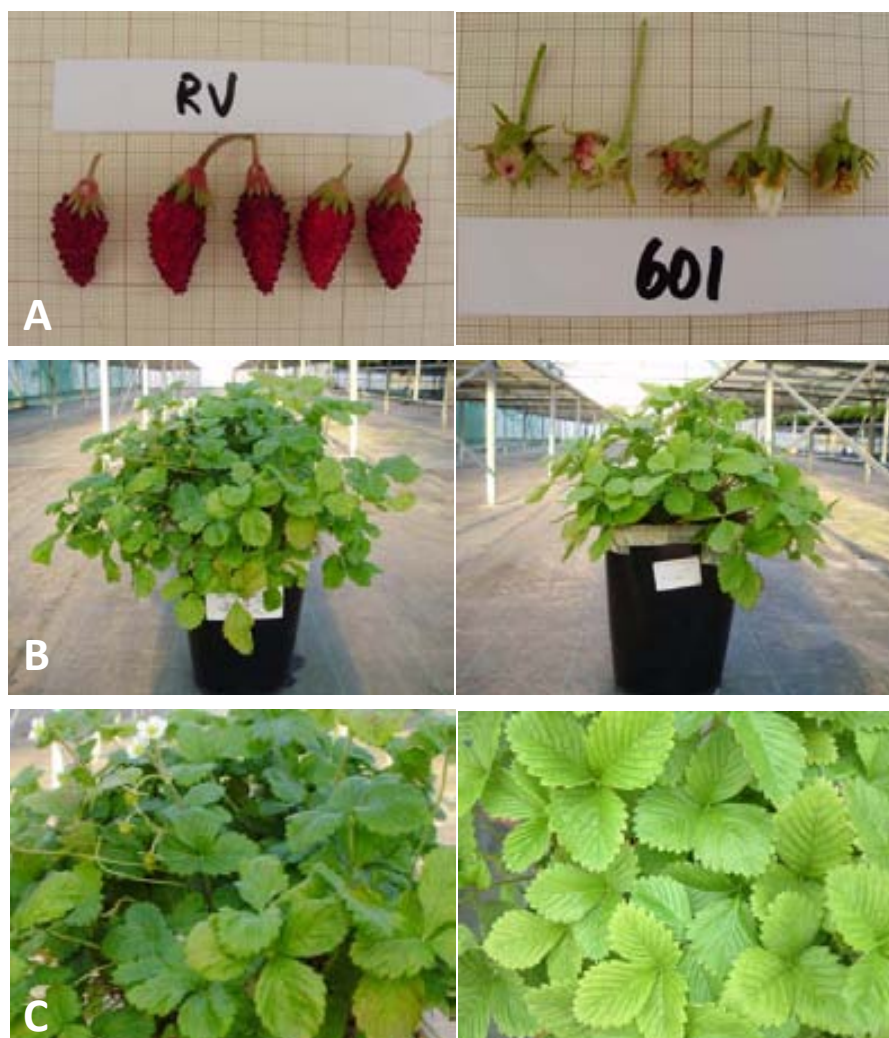


Figura 24. Parentales empleados en la creación de una colección de NILs en *Fragaria*. Fotografías de fruto(A), planta (B) y hojas (C). A la izquierda, parental recurrente, *F. vesca* 'Reine des Vallées'. A la derecha, parental donante, *F. bucharica* FDP601.

2. MATERIAL Y MÉTODOS

2.1 OBTENCIÓN Y MANTENIMIENTO DE LAS PLANTAS

Los frutos maduros fueron recogidos y preservados a 37°C durante un mínimo de 3 días para provocar su deshidratación. Después, las semillas se extrajeron manualmente y se colocaron en tubos de 1.9 ml de capacidad. Las semillas se mantuvieron a 4°C durante un periodo mínimo de dos semanas. Antes de la germinación, las semillas fueron sometidas a un tratamiento de desinfección de su superficie.

En primer lugar, en cada tubo se añadió 1.5 ml de etanol 70°. Los tubos se agitaron fuertemente durante un minuto y el etanol se retiró mediante una pipeta Pasteur. Tras un enjuague con agua

estéril, se añadió en cada tubo 1.5 ml de una solución de Captan (1,2,3,6-Tetrahidro-N-(triclorometilitio)fthalimida) 3% (p/v) que se agitó fuertemente durante 5 minutos. El antifúngico se retiró mediante una pipeta Pasteur y se realizaron tres lavados consecutivos con agua estéril para eliminar los restos de Captan. Después, las semillas permanecieron sumergidas en agua estéril durante un mínimo de doce horas. Tras un último lavado con agua estéril, y bajo una campana de flujo laminar, las semillas fueron transferidas a una placa petri estéril, cuyo fondo había sido cubierto previamente con varias capas de papel absorbente autoclavado. El fondo de la placa se humedeció con agua estéril y las placas fueron selladas y mantenidas en una cámara de cultivo a 28°C, sometidas a un fotoperiodo de 12 horas de luz durante un mínimo de dos semanas y un máximo de dos meses. Durante ese tiempo, eventualmente se añadió agua estéril en el interior de las placas para evitar la deshidratación de las plántulas.

Una vez desarrolladas las raíces y los cotiledones, las plántulas se retiraron de la placa y fueron transferidas de manera individual a tacos de turba prensada de 33 mm de diámetro hidratados (*Jiffy*, Benelux, Países Bajos). Los tacos se mantuvieron en el interior de propagadores de plástico (mini-invernaderos) de dimensiones 50 x 38 x 37 cm (*Puteaux*, Les Clayes sous Bois, Francia), que fueron mantenidos en el interior de un invernadero, bajo condiciones de riego y exposición solar mínimas.

Las plantas seleccionadas para continuar el experimento tras los análisis genéticos fueron transferidas a tiestos de 200 ml de capacidad con sustrato multiplicativo y mantenidas bajo condiciones normales de invernadero. Pasados dos o tres meses, o al alcanzar un tamaño adulto, las plantas fueron transferidas a tiestos de 5 litros de capacidad y mantenidas bajo condiciones de invernadero de malla de sombreo (ver sección 2.2.1 de la primera parte de esta Tesis, 'Conservación de la población FV x FB') en Cabrils, Barcelona (latitud Norte 41° 31' 09", longitud Este 2° 22' 32" (fuente: *Google Earth* – NOAA, *US navy*). Las condiciones climáticas de esta localidad del Maresme, situada a 90 metros sobre el nivel del mar, son las propias de un clima mediterráneo templado, con temperaturas moderadas (15°C de temperatura media anual), precipitaciones bajas (apenas alcanzan los 600 mm anuales) que se concentran durante la primavera y el otoño, con inviernos y veranos suaves.

2.2 INDUCCIÓN FLORAL Y POLINIZACIONES

Las plantas que no florecieron de manera natural fueron sometidas a un tratamiento de inducción floral descrito por Sargent (2005). Para ello, fueron transferidas a una cámara a 4°C durante cuatro

semanas en condiciones de oscuridad total. Pasado este periodo, fueron devueltas al invernadero, bajo condiciones de fotoperiodo mínimo de 12 horas.

En el caso de las polinizaciones cruzadas, 24 horas antes de la polinización se recolectaron las anteras de las flores del parental masculino recién abiertas y se mantuvieron en placas de petri selladas y refrigeradas a 4°C durante un periodo máximo de 72 horas. Dos horas antes de la polinización, las placas se abrieron y las anteras fueron incubadas a 28°C durante una hora para provocar su deshidratación. Transcurrido este tiempo, las placas se sellaron de nuevo y se agitaron suavemente para desprender el polen. Las flores del parental femenino se emascularon 24 horas antes de la polinización. Para ello, se seleccionaron los botones florales en la fase del desarrollo en que los pétalos comienzan a emerger y empiezan a visualizarse entre los sépalos, aún fusionados. Con la ayuda de una cuchilla estéril, se seccionaron los sépalos, los pétalos y finalmente se emascularon las flores seccionando las anteras y eliminando los estambres. Los receptáculos florales fueron cubiertos con pequeños sobres de papel sellados con cinta adhesiva hasta el momento de su polinización. Los receptáculos fueron polinizados empleando un pincel fino tras 24, 48 y 72 horas de la emasculación y los sobres fueron repuestos después de cada polinización y retirados 72 horas después.

Las autofecundaciones se realizaron mediante el aislamiento de los botones florales con sobres de papel sellados con cinta adhesiva. Los sobres fueron ligeramente agitados tras 24, 48 y 72 horas de su colocación, para provocar la caída del polen sobre los receptáculos florales. Trascorridas 72 horas desde la última polinización, se retiraron los sobres.

2.3 RECOLECCIÓN DE LAS MUESTRAS

Cuando las plantas hubieron desarrollado un mínimo de cuatro hojas verdaderas, se procedió a realizar las extracciones de ADN. La extracción se realizó en placas de 96 pocillos, empleando siempre material vegetal fresco. Para ello se recolectaron uno o dos folíolos (en función del tamaño) de hoja joven. Cada muestra se introdujo en uno de los pocillos de las placas de extracción, emplazadas en todo momento sobre un recipiente con hielo. Cada pocillo contiene, además, 300 µl de tampón de extracción (Doyle y Doyle, 1990) modificado para fresa con la adición de PVP-40 al 2% y una esfera de tungsteno (Quiagen, Hilden, Alemania) para la posterior homogenización del material vegetal.

2.4 EXTRACCIÓN Y PREPARACIÓN DE ADN GENÓMICO

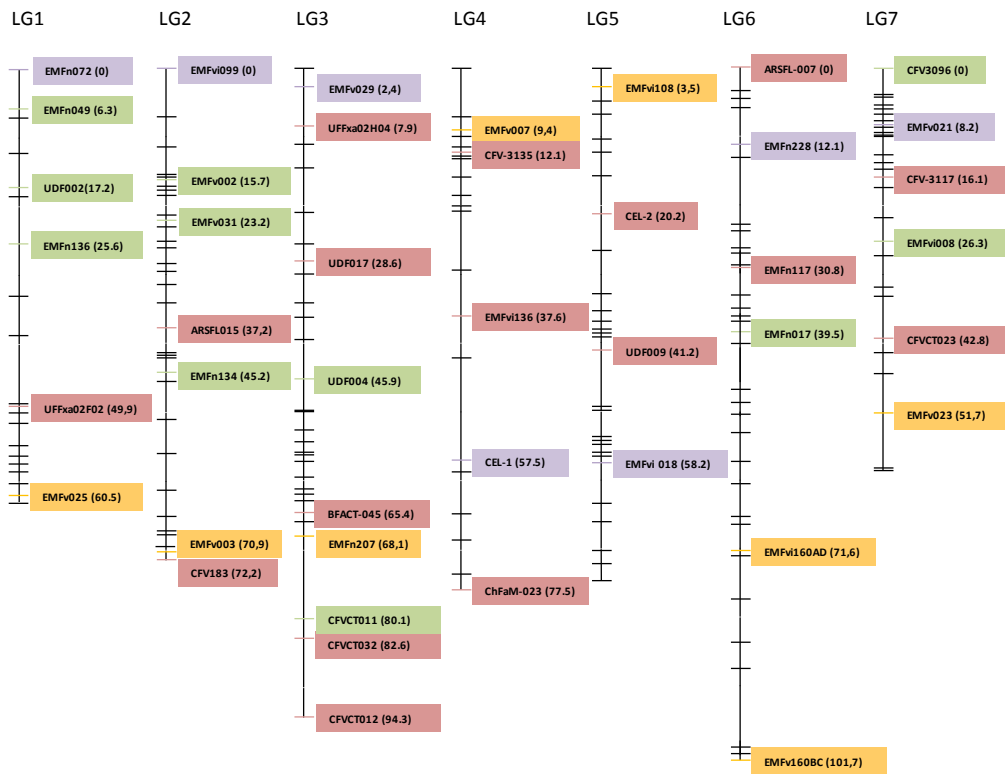
Las placas se sellaron, se instalaron en el homogeneizador de tejidos de velocidad variable Mixer Mill 300 (Retsch, Haan, Alemania) siguiendo las instrucciones del fabricante y se sometieron a dos pulsos de 30 hz y 45 segundos cada uno. A continuación, las placas se incubaron en un baño a 65°C durante 45 minutos, en agitación constante. Finalizado ese tiempo, las placas se centrifugaron durante 15 segundos a 1000 rpm en una centrífuga de placas modelo Allegra (Beckman Coulter Inc, CA, USA) y se añadió, a cada pocillo, 300 µl de una solución de cloroformo-alcohol isoamílico 24:1. Las placas se sellaron, se agitaron fuertemente y se centrifugaron a 3700 rpm durante 15 minutos. Se recogieron 200 µl de sobrenadante de cada pocillo y se transfirieron a una nueva placa, donde se mezclaron con 200 µl de etanol absoluto a -20°C. La mezcla se agitó suavemente 5 veces y se incubó a -20°C durante dos horas. A continuación, las placas se centrifugaron durante 15 minutos a 3700 rpm y 4°C, se descartó la fracción líquida y los pellets, situados en la base de cada pocillo, se lavaron con etanol 70°C frío (4°C). Las placas se centrifugaron durante 3 minutos a 3700 rpm y 4°C, y se descartó la fracción líquida de nuevo. Los pellets obtenidos se secaron a temperatura ambiente durante dos horas bajo una campana de extracción de gases y fueron resuspendidos en 50 µl de una solución de agua estéril ultrapura y 100 µg de RNasa. La solución se incubó durante 30 minutos a temperatura ambiente, seguidos de una segunda incubación de doce a dieciséis horas a 4°C. La concentración de ADN de cada una de las muestras se midió en un espectrofotómetro (NanoDrop, DE, USA) y se prepararon alícuotas a una concentración final de 10 ng/µl.

2.5 SELECCIÓN DE MARCADORES

Los marcadores moleculares empleados en la caracterización genómica de las poblaciones BC1, BC2 y BC2 \otimes se seleccionaron en función de su posición genómica, su disponibilidad y su sencillez metodológica. La figura 25 muestra los nombres y posiciones de los marcadores empleados en las diferentes fases de M.A.S. de la población BC1 y en las familias BC2. En éstas familias sólo se analizaron los marcadores localizados en regiones segregantes. Las familias BC2 \otimes y las líneas seleccionadas para formar parte de la colección final de NILs fueron analizadas con un mayor número de marcadores moleculares para un genotipado más exhaustivo. La tabla 17 muestra las características principales de los 76 marcadores moleculares empleados en esta investigación.

La estimación del tamaño de las regiones introgresadas, así como del porcentaje del genoma de *F. bucharica* presente en las poblaciones segregantes obtenidas se calculó en base a la suma de distancias entre marcadores en heterocigosis en cada planta.

A



B

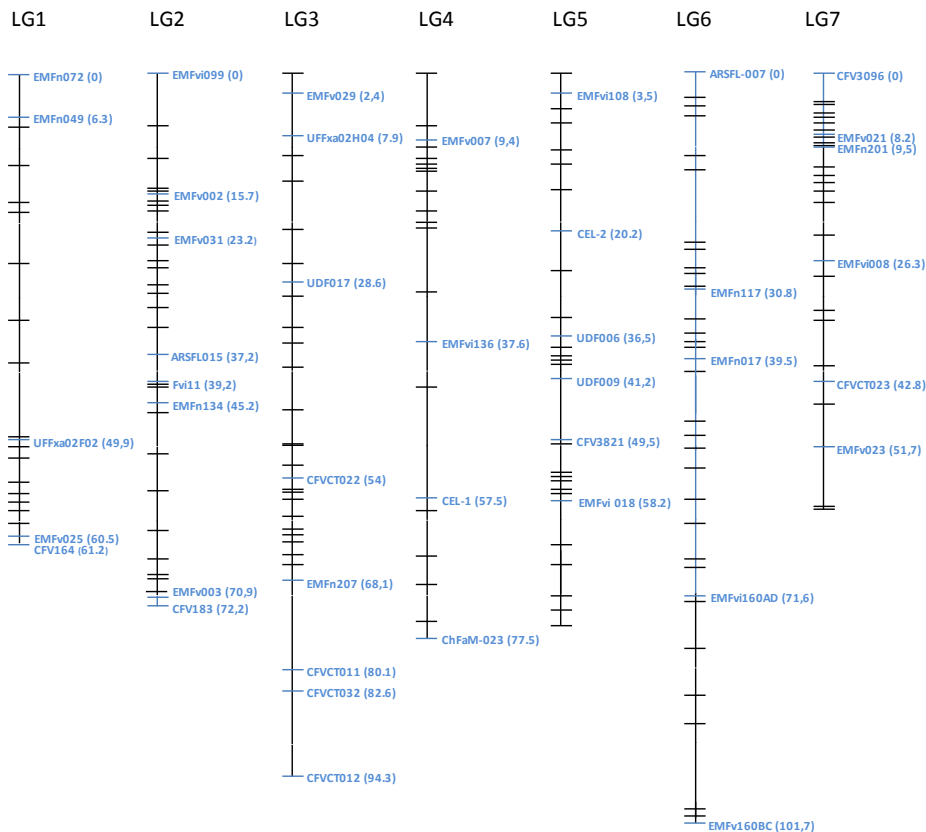


Figura 25. Distribución de los SSRs analizados en las poblaciones BC1 y BC2 . A: SSRs analizados durante las diferentes fases de M.A.S en la población BC1. Los colores indican la fase de M.A.S en la que fueron empleados; Violeta, fase uno. Amarillo, fase dos. Rojo, fase tres y finalmente, en verde, fase cuatro. **B:** SSRs analizados en los grupos de ligamiento segregantes de las familias BC2.

Tabla 17. Marcadores moleculares analizados en esta investigación. Se indica el grupo de ligamiento (GL), nombre y posición de los marcadores, métodos de detección empleados (ABI: electroforesis capilar. GEL: electroforesis en gel de agarosa), tamaño de los alelos y tipo de herencia observada en el cruzamiento *F. vesca* var. 'Reine des Vallées' x *F. bucharica* 601 (C: codominante, D: dominante) Enzima de restricción (R.E.) y porcentaje de agarosa empleado en la detección en gel. En negrita, alelos presentes en la planta F₁ que se empleó para generar la colección de NILs. *Agarosa de alta resolución Metaphor®.

GL	MARCADOR	TIPO	DETECCIÓN		POSICIÓN	ALELOS		HERENCIA	R.E.	GEL (%p/v)
			ABI	GEL		<i>F. vesca</i>	<i>F. bucharica</i>			
1	EMFv072	SSR	SI	SI	0	259	282-295	C	-	2.5
	EMFn049	SSR	SI	NO	6,3	202	200-211	C	-	-
	UDF-002	SSR	SI	NO	17	145	132	C	-	-
	EMFn136	SSR	SI	NO	25,6	177-207	167-181	C	-	-
	CFV-4021	SSR	SI	SI	33,4	240-249	0	D	-	1.2
	CO818002	SSR	SI	NO	38,6	225	229	C	-	-
	UFF02F02	SSR	SI	NO	49,9	206	208-214	C	-	-
	CFV-3052	SSR	SI	NO	57,8	277-280	269	C	-	-
	EMFv025	SSR	SI	NO	60,5	221	218	C	-	-
	CFV164	SSR	SI	SI	61,2	245	267	C	-	2.5
2	EMFv099	SSR	SI	SI	0	162	210	C	-	3.0
	EMFv148	SSR	SI	SI	11,7	300	220-230	C	-	2.5
	EMFv002	SSR	SI	SI	15,7	292	209	C	-	1.2
	CFV-3099	SSR	SI	NO	16,4	179	153-260	C	-	-
	CFVCT020	SSR	SI	NO	17,6	146	144	C	-	-
	EMFv031	SSR	SI	NO	23,2	227	213-238	C	-	-
	BFACT002	SSR	SI	NO	29,5	217-228	263-271	C	-	-
	Fv11	SSR	SI	NO	39,2	322	306	C	-	-
	EMFn134	SSR	SI	SI	45,2	214	250-263	C	-	2.5
	EMFxa379796	SSR	SI	NO	62,6	200	195	C	-	-
3	BFACT039	SSR	SI	NO	68,9	181	190	C	-	-
	EMFv003	SSR	SI	SI	70,9	260	210-225	C	-	2,5
	CFV183	SSR	SI	NO	73	316	307	C	-	-
	EMFv016	SSR	SI	SI	0	233	0	D	-	1.2
	EMFv029	SSR	SI	NO	2,4	202-225	208-218	C	-	-
	UFFxa02H04	SSR	SI	NO	7,9	224	226	C	-	-
	CFVCT020	SSR	SI	NO	17,6	146	144	C	-	-
	UDF017	SSR	SI	SI	28,6	237	190	C	-	2.0
	BFACT043	SSR	SI	NO	39,4	228	224	C	-	-
	CFVCT022	SSR	SI	NO	54	130-132	127	C	-	-
4	CFVCT007	SSR	SI	NO	61,3	99	120	C	-	-
	BFACT045	SSR	SI	NO	65,4	190	196	C	-	-
	EMFn207	SSR	SI	NO	68,1	266	254	C	-	-
	CFVCT011	SSR	SI	NO	80,1	155	170	C	-	-
	CFVCT032	SSR	SI	NO	82,6	170-180	156-180	C	-	-
	CFVCT012	SSR	SI	NO	94,3	168-201	158	C	-	-
	VT-851	CAPS	NO	SI	0	300	200	C	<i>BbvI</i>	2.0
	UDF007	SSR	SI	NO	9,4	150	141	C	-	-
	CFV-3819	SSR	SI	NO	9,4	256-260	251	C	-	-
	EMFv007	SSR	SI	NO	9,7	226	232	C	-	-
5	CFV-3135	SSR	SI	NO	12,1	232-289	254	C	-	-
	CFV-3148	SSR	SI	NO	20,3	174	157-174	C	-	-
	VT-861	CAPS	SI	SI	43,7	400	280	C	<i>MseI</i>	2.0
	CEL-1	SSR	NO	SI	57,5	400	500	C	-	1.2
	ChFaM23	SSR	SI	NO	77,5	158-164	150	C	-	-
	CFV-3072	SSR	SI	NO	0	212	204-210	C	-	-
	EMFv108	SSR	SI	NO	3,5	225	203	C	-	-
	CFV-3132	SSR	SI	SI	10,8	170	130-139	C	-	2.5
	CEL2	SSR	SI	NO	20,2	238	236	C	-	-
	EMFn110	SSR	SI	NO	27,7	238	239	C	-	-
6	UDF006	SSR	SI	NO	36,5	201	197	C	-	-
	CFVCT024	SSR	SI	NO	38,5	110	105	C	-	-
	UDF009	SSR	SI	NO	41,2	170	158 - 150	C	-	-
	CFV-3821	SSR	SI	NO	49,5	175	168-161	C	-	-
	EMFn010	SSR	SI	NO	51	224	224-204	C	-	-
	EMFv024	SSR	SI	NO	56,2	216	216-219	C	-	-
	EMFv018	SSR	SI	SI	58	262	210-240	C	-	3.0
	VT-010	CAPS	NO	SI	75,7	400	375	C	<i>HapII</i>	4.0*
	ARSFL007	SSR	SI	NO	0	232	236	C	-	-
	EMFv228	SSR	SI	SI	11,4	287	350	C	-	2.0
7	EMFn117	SSR	SI	NO	30,1	162-173	191	C	-	-
	EMFn017	SSR	SI	NO	38,8	171	181	C	-	-
	CFVCT017	SSR	SI	NO	54,3	155	149	C	-	-
	EMFv160AD	SSR	SI	NO	84,8	171	106	C	-	-
	EMFv010	SSR	SI	SI	83,6	192	0	D	-	1,2
	UFFxa01E03	SSR	SI	NO	100,3	192	185	C	-	-
	EMFv160BC	SSR	SI	NO	101	141	138	C	-	-
	CFV3096	SSR	SI	SI	0	261-267	226	C	-	2.5
	EMFv021	SSR	SI	SI	8,2	225-212	271	C	-	2,5
	EMFn201	SSR	SI	NO	9,5	223	227	C	-	-
CFV-3896	SSR	SI	NO	10,8	206	0	D	-	-	
CFV-3117	SSR	SI	SI	16,1	207	228-240	C	-	2,5	
EMFv008	SSR	SI	NO	26	306	315	C	-	-	
ARSFL099	SSR	SI	NO	27,1	189	191-195	C	-	-	
CFVCT023	SSR	SI	NO	42,8	155	130	C	-	-	
EMFv023	SSR	SI	NO	51,7	225	235	C	-	-	
ChFaM010	SSR	SI	NO	58,7	312	315	C	-	-	

2.6 REACCIÓN DE PCR

La mayor parte de las reacciones de PCR que se realizaron durante la caracterización del genoma de los individuos obtenidos para este estudio pueden agruparse en dos tipos; aquellas cuyos productos de PCR serían después visualizados en el analizador de fragmentos ABI3130 (Applied Biosystems, CA, USA), cuyas condiciones de amplificación figuran en la sección 2.2.5 de la primera parte de esta Tesis (temperatura de hibridación de cebadores para todos ellos: 55°C), y las reacciones de PCR cuyos productos serían después visualizados tras una electroforesis horizontal sobre gel de agarosa. Estas últimas tuvieron lugar en volumen final de 15 µl, que contiene 20 ng de ADN molde, 1x tampón Lab (KCl 50mM, Tris-Hcl 10 mM pH 8,3, gelatina 0.001%), 2,5 mM MgCl₂, 0,4 mM de cada desoxinucleótido, 2 U de DNA polimerasa *AmpliTaq* (Perkin-Elmer, IL, USA), 0,7 µM de cebador F y 0,7 µM de cebador R. La elección del método de detección de polimorfismos (analizador de fragmentos ABI3130 o gel de agarosa) se hizo en función del marcador molecular, del número de individuos a analizar y de la disponibilidad del equipamiento en el momento del análisis. Las condiciones de amplificación empleadas fueron: Una desnaturalización inicial de 2 minutos a 94°C seguida de 35 ciclos de desnaturalización (30 segundos a 94°C), hibridación (45 segundos a la temperatura óptima de hibridación del par de cebadores empleados, nunca inferior a 55°C) y extensión (30 segundos a 72°C) seguidos de una extensión final de 7 minutos a 72°C.

2.7 DIGESTIONES ENZIMÁTICAS

Los marcadores CAPS se analizaron mediante la separación de los fragmentos obtenidos tras la digestión con un enzima de restricción en gel de agarosa. Quince microlitros de producto de PCR fueron digeridos siguiendo las instrucciones del fabricante (fig. 26).

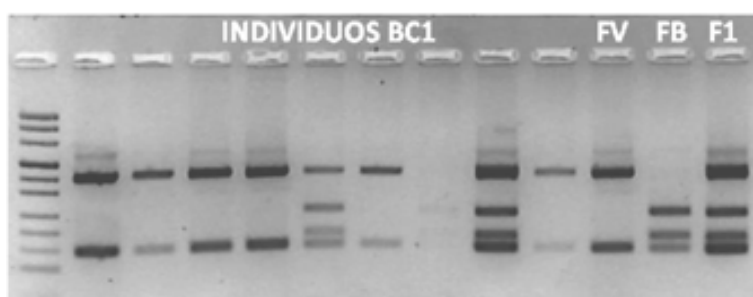


Figura 26. Detección de introgresiones en el genoma de *Fragaria vesca* con marcadores CAPS. Imagen de un gel de agarosa mostrando la separación de los fragmentos obtenidos tras la digestión enzimática de los productos de PCR obtenidos para el marcador VT-861, tras la digestión con el enzima *MseI*. A la izquierda figura el marcador de peso molecular pUC-mix. FV: *F. vesca*. FB: *F. bucharica*. F1: Híbrido F₁.

2.8 VISUALIZACIÓN DE PRODUCTOS DE PCR

La visualización de los productos de PCR se realizó mediante tinción con bromuro de etidio de geles de agarosa tras una electroforesis, o mediante el análisis de los cromatogramas obtenidos en el analizador de fragmentos ABI 3130 (Applied Biosystems, CA, USA), tal y como se especifica en el la sección 2.2.6 de la primera parte de esta Tesis (Fig. 27).

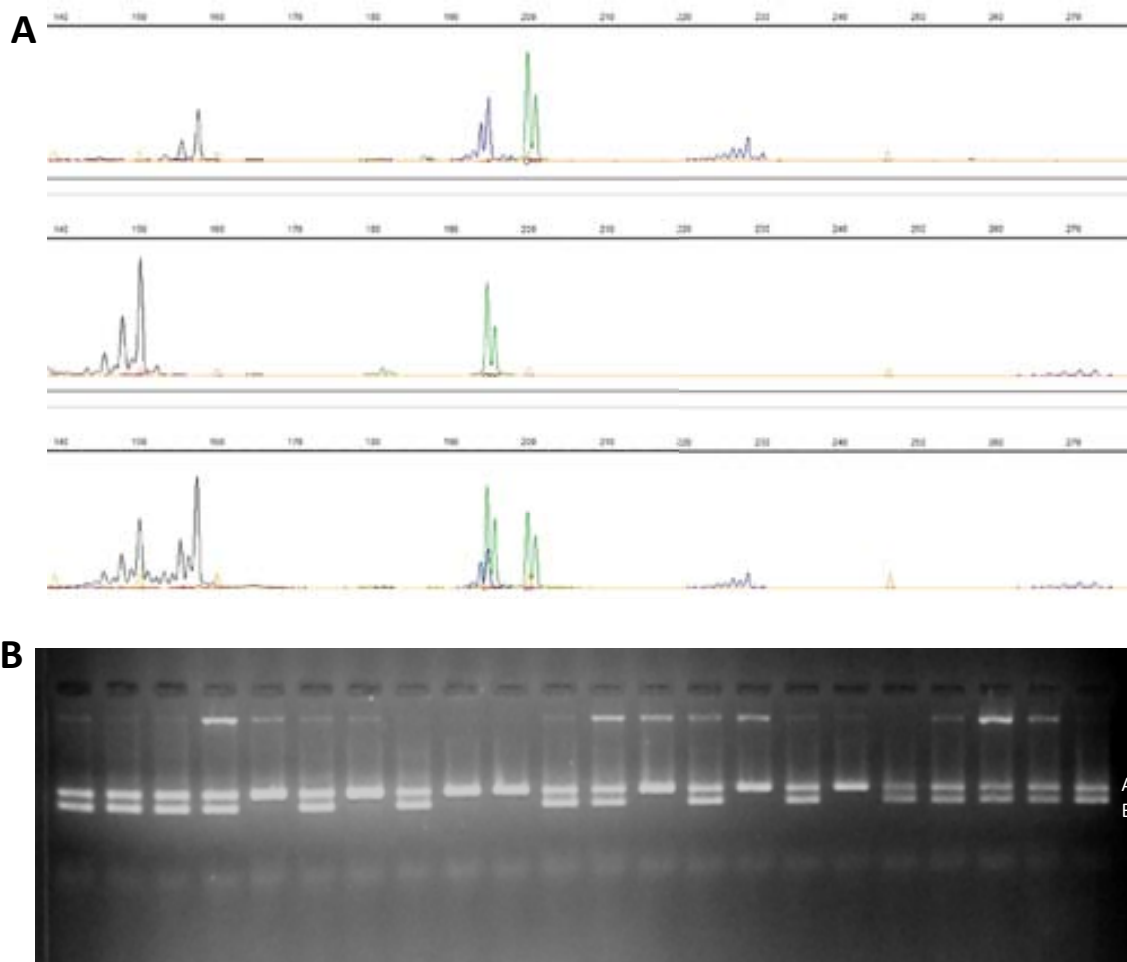


Figura 27. Detección de introgresiones en el genoma de *Fragaria vesca* con marcadores SSR. A: Electroferogramas obtenidos tras la electroforesis capilar de tres productos de PCR procedentes de los parentales de la población de NILs en modo *multiloading* (carga simultánea). El gráfico superior corresponde al parental *F. vesca*, el del medio corresponde al parental *F. bucharica* y el de la parte inferior, al híbrido F_1 . A la izquierda, en color negro, aparecen los alelos correspondientes al marcador CFVCT017 (A:158 pb, B:150 pb), marcados con el fluorocromo NED. En el centro, en color verde, los alelos correspondientes al marcador EMFxa379796 (A:200 pb, B:194 pb) marcados con el fluorocromo VIC. A la derecha de la imagen, de menor intensidad y en azul, los alelos correspondientes al marcador BFACT002 (A:228 pb, B:272 pb) marcados con el fluorocromo FAM. Los picos naranjas corresponden al marcador de peso molecular LIZ-500. **B:** Gel de agarosa mostrando los productos de PCR de un subconjunto de individuos BC1 para el marcador SSR EMFn148. El alelo A corresponde a *F. vesca*, mientras que el alelo B, corresponde a *F. bucharica*. La imagen permite diferenciar individuos AA para este locus (una sólo banda) de los individuos AB (dos bandas)

2.9 ANÁLISIS DE CARACTERES CUANTITATIVOS EN LAS LÍNEAS DE INTROGRESIÓN

Para la evaluación de las modificaciones en el tamaño de los folíolos de las líneas obtenidas, se tomaron cuatro hojas al azar y se midió la longitud de cada uno de los folíolos, obteniendo así 12 medidas para cada línea que fueron consideradas repeticiones del mismo evento a efectos estadísticos.

Para la evaluación de los caracteres 'longitud', 'anchura' y 'forma' de los frutos, se tomaron 8 frutos de cada una de las líneas evaluadas en diferentes momentos del verano de 2010, que fueron considerados repeticiones del mismo evento a efectos estadísticos. Los frutos fueron cortados de la planta, fotografiados y medidos en el momento de su recolección.

2.10 MÉTODOS PROBABILÍSTICOS Y DE ANÁLISIS ESTADÍSTICO

En los retrocruzamientos, y previo a la germinación de las semillas, se estableció el número mínimo de plantas a generar para obtener una alta probabilidad de encontrar plantas con el genotipo deseado, empleando la fórmula

$$P = 1 - (1 - 0,5^l)^N$$

Donde P es la probabilidad de obtener un genotipo determinado tras el análisis de N individuos con l introgresiones, asumiendo la ausencia de recombinación durante la meiosis.

Los tests de bondad del ajuste efectuados para la estimación de la distorsión de la segregación de los marcadores analizados fueron llevados a cabo mediante el programa JoinMap® 4 (Van Ooijen, 2006).

Los análisis estadísticos para el análisis de caracteres cuantitativos se llevaron a cabo mediante la aplicación SAS *Enterprise guide* v4.2 (SAS Institute Inc., Cary, NC, USA), a través del análisis de la varianza ANOVA de una entrada. Para el análisis de los caracteres 'tamaño de los folíolos', 'longitud', 'grosor' y 'forma' del fruto, las medias obtenidas para cada línea fueron comparadas con los valores obtenidos en el parental recurrente RV mediante el test de Dunnett (Dunnett, 1955) con un error de tipo I de $\alpha \leq 0.01$. Ante la falta de individuos control para la comparación de medias para evaluar el efecto de las introgresiones procedentes del genoma de *F. bucharica* en el carácter 'periodo vegetativo', se realizó un test de comparación de medias de rango múltiple SNK (Student-Newman-Keuls) (Keuls, 1952) con un error de tipo I de $\alpha \leq 0.05$, donde las líneas con introgresiones en regiones comunes fueron consideradas repeticiones de un mismo evento.

3. RESULTADOS

3.1 DISEÑO EXPERIMENTAL DEL DESARROLLO DE UNA COLECCIÓN DE NILs EN *FRAGARIA* DIPLOIDE

El esquema del desarrollo de la genoteca de líneas de introgresión a partir de un cruzamiento interespecífico entre *Fragaria vesca* var. 'Reine des Vallées' (PI 551824) y *Fragaria bucharica* (FDP601) se muestra en la figura 28. En el primer retrocruzamiento se empleó la planta F₁-30 como parental masculino y *F. vesca* como parental femenino. Las semillas obtenidas fueron germinadas y se extrajo el ADN de los individuos BC1 para realizar el cribado de plantas en función de su dotación mediante selección asistida por marcadores. En una primera etapa de M.A.S., se analizó uno de los extremos de cada uno de los 7 grupos de ligamiento del genoma de *Fragaria* diploide mediante el uso de marcadores moleculares de tipo SSR. Tras la primera criba, el genoma de los individuos seleccionados fue analizado para el otro extremo de cada grupo de ligamiento, empleando un nuevo set de SSRs. Una vez determinado el número de introgresiones por planta (asumiendo la hipótesis de que no se han dado eventos de doble recombinación durante la meiosis) los individuos con menos de 4 introgresiones fueron seleccionados y genotipados para 17 loci situados en regiones internas de los grupos de ligamiento. Tras esta tercera fase de M.A.S., se seleccionaron 19 individuos cuyo genoma fue analizado para 11 nuevos loci. Finalmente se seleccionaron los 8 individuos con menor número de introgresiones que, sumadas, cubren la totalidad del genoma del parental donante.

Cada uno de los individuos BC1 seleccionados fue empleado como parental femenino para la segunda fase de retrocruzamientos, estableciéndose así 8 familias de plantas BC2. El genoma de las plantas BC2 fue analizado para los loci segregantes en cada familia y se seleccionaron 10 plantas BC2 para ser los progenitores de las líneas finales.

Las plantas BC2 que desarrollaron flores fueron autofecundadas, obteniéndose un total de 1340 plantas BC2 pertenecientes a 7 familias, que fueron analizadas para loci localizados en regiones segregantes de cada familia. Tras esta última fase de M.A.S., se seleccionó un set formado por individuos con una sólo introgresión, tanto en homocigosis como en heterocigosis, e individuos con dos o tres introgresiones en regiones no cubiertas por otras líneas. Se recogieron nuevas muestras de material vegetal para una nueva extracción de ADN y se realizó un análisis genético final para 68 loci distribuidos por todo el genoma, con objeto de delimitar exhaustivamente las introgresiones seleccionadas e identificar posibles introgresiones contaminantes en las líneas seleccionadas. El análisis final condujo a la identificación de las 30 líneas con introgresiones cubriendo la mayor parte del genoma donante.

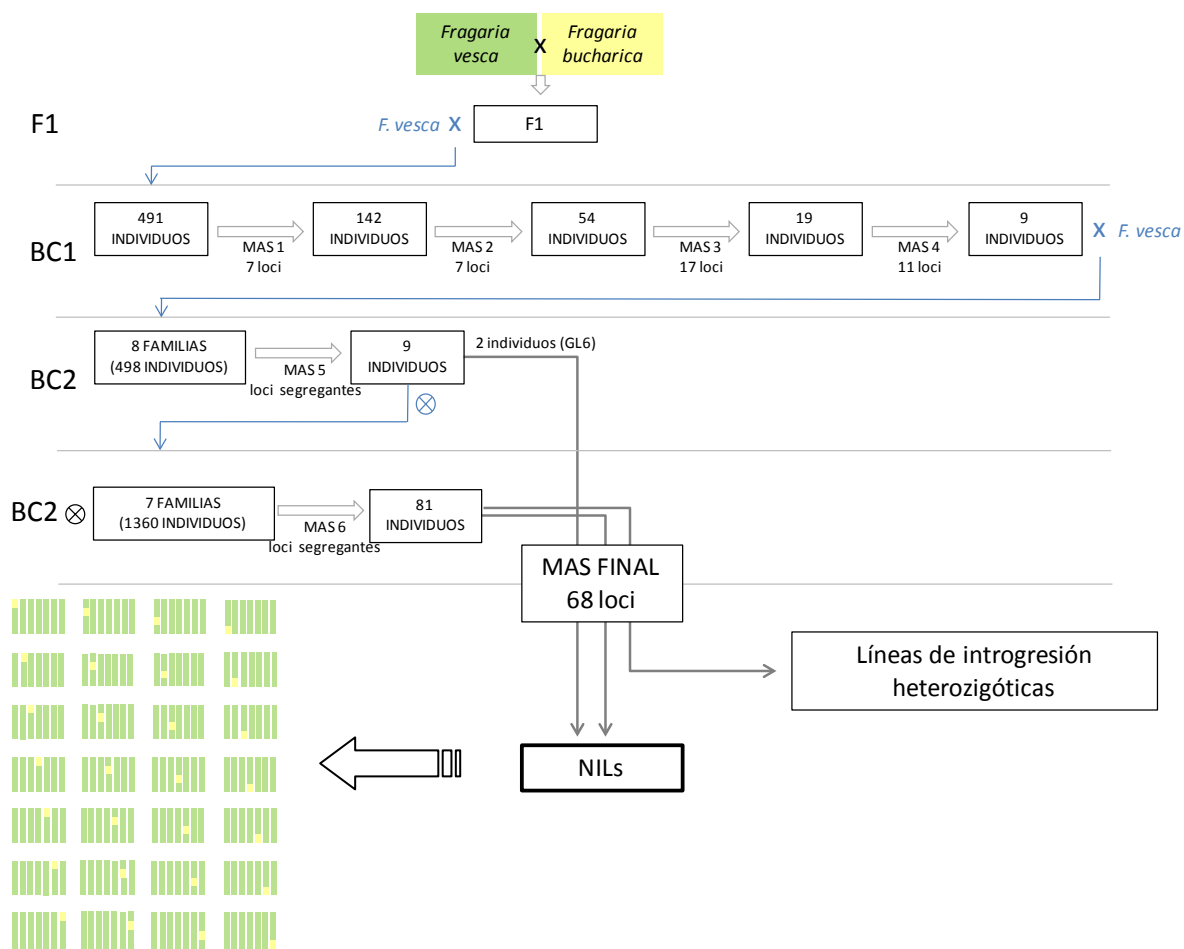


Figura 28. Diseño experimental para la obtención de una colección de líneas de introgresión en *Fragaria*. De arriba a abajo, las distintas fases cruzamiento y selección asistida por marcadores, empleadas hasta la obtención de líneas de introgresión (ILs) y NILs. En los retrocruzamientos (en azul) en primer lugar figura el parental femenino y, tras el símbolo 'X', el parental masculino. Abajo, a la izquierda, representación esquemática de una población de NILs similar a la que se pretende conseguir en esta investigación. Cada conjunto de 7 barras representa el genoma de cada línea y cada barra representa un grupo de ligamiento, donde las introgresiones del parental donante figuran como segmentos de color amarillo sobre un fondo genético homogéneo de color verde.

3.2 OBTENCIÓN DE LA GENERACIÓN F₁

El cruzamiento entre los parentales, empleando al parental *F. bucharica* como parental masculino (donador de polen) y el parental *F. vesca* como parental femenino, dio lugar a una generación de plantas F₁ entre las que se seleccionó el individuo F₁-30, por ser el primero en florecer. Los híbridos mostraron características de ambos parentales (fig. 29), como la producción de estolones, la ausencia de floración durante el primer año o la floración en día corto (caracteres paternos), y el gran porte, vigor y resistencia a enfermedades (caracteres maternos).



Figura 29. Características morfológicas de la generación F₁. La fotografía A muestra el individuo seleccionado de la generación F₁ obtenida (F₁-30). En la fotografía B se muestra la morfología de los frutos de la planta híbrida seleccionada para generar la población BC1.

3.3 CARACTERIZACIÓN Y SELECCIÓN ASISTIDA POR MARCADORES DE UNA POBLACIÓN BACKCROSS 1

El retrocruzamiento del híbrido F₁ con el parental *F. vesca* dio como resultado una población de 515 individuos BC1. El ADN de 491 fue extraído y caracterizado mediante PCR para 7 loci: EMFv072, EMFvi099, EMFv029, CEL-1, EMFvi018, EMFn228 y EMFv021, cada uno de ellos localizado en un extremo de un grupo de ligamiento del mapa de referencia del genoma de *Fragaria* diploide (fig. 25A). El porcentaje medio de regiones introgresadas para cada locus fue del 57,7%. El máximo porcentaje de introgresión se detectó para el marcador EMFvi099 (82% de individuos con genoma introgresado para este locus), localizado en la posición 0 cM del GL2 (tabla 18).

Tras este primer análisis, 373 individuos (72,4%) fueron desechados por poseer cuatro o más grupos de ligamiento introgresados total o parcialmente, y 142 individuos (27,6%) fueron seleccionados para un segundo análisis, donde su genoma fue analizado para la presencia de alelos heredados del parental *F. bucharica* en marcadores localizados en los extremos opuestos a los analizados en la fase anterior: EMFv025, EMFv003, EMFn207, EMFv007, EMFvi108, EMFv160, EMFv023 (Fig 25B). El promedio de individuos que presentaron introgresiones para estos loci fue de 52.2%, siendo EMFv160BC, localizado en el extremo inferior del GL6, el marcador que mostró mayor tasa de heterozigosis (73,6%).

Ochenta y ocho individuos seleccionados en primera fase (62%) mostraron tener más de 3 introgresiones en su genoma y por tanto fueron desechados, mientras que 54 plantas (38%) fueron seleccionadas para la siguiente fase de caracterización genómica. El número de introgresiones detectadas para los individuos seleccionados fue de una única introgresión en dos individuos (0,38%

de la población BC1 inicial), dos introgresiones en 11 individuos (2,13%) y tres introgresiones en 41 individuos (7,96%). Los individuos seleccionados mostraron tener al menos 147 introgresiones, localizadas principalmente en los grupos de ligamiento 2 y 6, con 32 introgresiones en cada uno de ellos (21,8%), seguidos de los grupos 3 y 5, con 24 introgresiones cada uno (16,3%). Los grupos 1, 4 y 7 presentaron respectivamente 10 (6,8%), 12 (8,2%) y 13 introgresiones (8,8%).

Tabla 18. Caracterización genómica de la población *Backcross* 1. Fases de la selección (M.A.S.), grupos de ligamiento (GL), loci (SSR), número de datos obtenidos (*N*), genotipos, distorsión y valores de Chi-cuadrado (χ^2) para bondad del ajuste a una segregación mendeliana. *, **, ***, valores significativamente distintos a los de la segregación esperada con $p < 0.05$, 0,01 y 0,001 respectivamente; - valor no significativamente distinto al esperado.

MAS	GL	SSR	N	GENOTIPO AA (%)	GENOTIPO AB (%)	DISTORSIÓN	Chi-cuadrado
FASE 1	1	EMFv072	448	60,5	39,5	***	19,72
	2	EMFvi099	417	18,0	82,0	***	170,96
	3	EMFv029	447	43,2	56,8	***	8,32
	4	CEL-1	450	54,9	45,1	**	4,3
	5	EMFvi018	413	30,0	70,0	***	65,92
	6	EMFn228	464	47,8	52,2	-	0,86
	7	EMFv021	462	41,6	58,4	***	13,2
FASE 2	1	EMFv025	138	52,2	48,8	-	0,18
	2	EMFv003	133	57,9	42,1	*	3,32
	3	EMFn207	132	43,9	56,1	-	1,94
	4	EMFv007	135	41,5	58,5	**	3,92
	5	EMFvi108	133	54,9	45,1	-	1,27
	6	EMFv160	140	26,4	73,6	***	31,11
	7	EMFv023	126	58,7	41,3	**	3,8
FASE 3	1	UFFO2F02	54	81,0	18,5	***	21,41
	2	ARSFL015	54	37,0	63,0	*	3,63
	2	CFV183	54	42,6	57,4	-	1,19
	3	UFFxa02H04	54	51,9	48,1	-	0,07
	3	UDF017	54	46,3	53,7	-	0,30
	3	BFACT045	54	42,6	57,4	-	1,19
	3	CFVCT032	54	46,3	53,7	-	0,3
	3	CFVCT012	53	50,0	50,0	-	0
	4	CFV3135	54	55,6	44,4	**	0,67
	4	EMFvi136	53	66,0	34,0	***	4,74
	4	ChFaM23	54	75,9	24,1	-	14,52
	5	CEL2	53	60,4	39,6	-	1,85
	5	UDF009	54	61,1	38,9	-	2,67
	6	ARSFL007	53	52,8	47,2	-	0,07
6	EMFn117	53	43,4	56,6	-	1,19	
7	CFV3117	54	66,7	33,3	**	6	
7	CFVCT023	54	59,3	40,7	-	1,9	
FASE 4	1	EMFn049	19	78,9	21,1	**	6,37
	1	UDF002	19	84,4	15,8	***	8,89
	1	EMFn136	19	78,9	21,1	**	6,37
	2	EMFv002	18	44,4	55,6	-	0,22
	2	EMFv031	19	26,3	73,7	**	4,26
	2	EMFn134	18	38,9	61,1	-	0,89
	3	UDF004	18	55,6	44,4	-	0,22
	3	CFVCT011	18	61,1	38,9	-	0,89
	6	EMFn017	19	52,6	47,4	-	0,05
	7	CFV3096	19	78,9	21,1	**	6,37
7	EMFvi008	18	72,2	27,8	*	3,6	

El ADN de las 54 plantas seleccionadas fue caracterizado para 17 nuevos loci distribuidos a lo largo de los 7 grupos de ligamiento del genoma de *Fragaria* diploide en una tercera fase de M.A.S. Este análisis permitió la detección de un total de 255 introgresiones, con una media de 4,7 introgresiones por planta que se distribuyeron de la siguiente manera: El grupo que presentó un mayor número de introgresiones fue el GL3 con 58, seguido de los grupos 2 y 6 con 43 y 40 introgresiones respectivamente, seguido de los grupos 5 y 4 con 38 y 33 introgresiones cada uno. Los grupos 1 y 7 presentaron 19 y 24 introgresiones respectivamente.

Se seleccionaron 19 individuos en función del número y localización de sus introgresiones. Se primaron introgresiones grandes, que cubriesen grupos de ligamiento completos. El número de introgresiones por individuo en la subpoblación seleccionada varió entre 1 y 5. Dichas plantas fueron caracterizadas con 11 nuevos SSR (tabla 18) en la cuarta y última fase de M.A.S. en la población BC1. Una vez analizados los datos obtenidos para los 42 loci, se detectaron un total de 88 introgresiones, con una media de 4,6 introgresiones por planta. El grupo de ligamiento que presentó un mayor número de introgresiones fue el GL3, con 25 introgresiones, seguido de los grupos 2 y 6 con 18 y 14 introgresiones respectivamente. A continuación, el GL4 con 10 y finalmente, los grupos 1, 5 y 7, con 7 introgresiones cada uno. Las introgresiones detectadas en esta subpoblación BC1 tuvieron un tamaño medio de 44,7 cM, si bien éstas pueden clasificarse en diferentes tipos en función de su origen; Identificamos 19 introgresiones pequeñas, definidas por un solo marcador molecular, cuya extensión no podemos estimar. Por otro lado, identificamos introgresiones que cubren grupos de ligamiento completos, que proceden de la unión de dos gametos con genotipo parental para todos los loci del cromosoma y cuya extensión depende exclusivamente del tamaño del grupo de ligamiento. En este caso, el tamaño medio de las introgresiones detectadas fue de 86,3 cM. También encontramos otro tipo de introgresiones más pequeñas, ocupando una parte de los distintos grupos de ligamiento, generadas por la unión de dos gametos donde al menos uno ha sufrido uno o varios eventos de recombinación durante la meiosis. En este caso, el tamaño medio de las introgresiones detectadas fue de 30,6 cM. La figura 30 muestra la distribución de las introgresiones detectadas a lo largo del genoma de la población BC1. Ocho individuos fueron seleccionados para la siguiente fase de retrocruzamiento. Cuatro de ellos poseen tres introgresiones y los cuatro restantes poseen cuatro introgresiones. Su genoma se esquematiza en la figura 31.

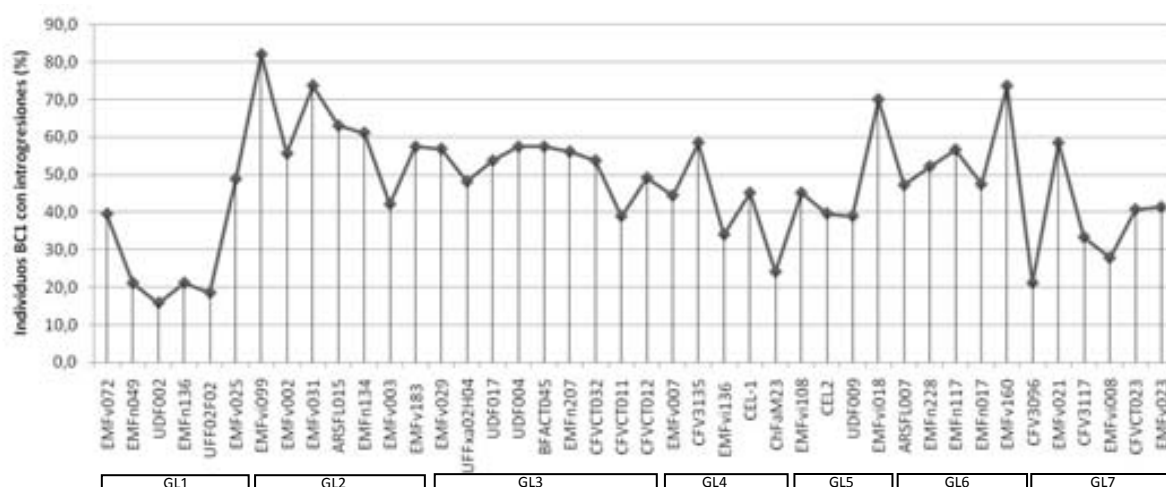


Figura 30. Análisis posicional de las de introgresiones detectadas en la población *backcross 1*. Porcentaje de individuos con introgresiones identificadas en las posiciones correspondientes a los marcadores moleculares analizados.

3.4 CARACTERIZACIÓN Y M.A.S. EN NUEVE FAMILIAS *BACKCROSS* 2 (BC2)

Tras las polinizaciones, se obtuvo una población *backcross* 2 formada por 498 individuos distribuidos en 8 familias. Estas familias BC2 vieron reducido drásticamente su tamaño poblacional debido a serios problemas agronómicos provocados por las altas temperaturas que se alcanzaron durante el verano de 2009 en el área de Barcelona, por lo que la probabilidad de encontrar individuos con una sola introgresión tras el análisis genómico disminuyó, tal y como se muestra en la tabla 19. Este fenómeno afectó especialmente a la familia 256, que desapareció por completo.

Tabla 19. Características de las familias BC2. Nombre, número de individuos, valor de la probabilidad (P) de encontrar individuos con una sola introgresión, dado el tamaño poblacional de cada familia. En la última columna figuran los nombres y los grupos de ligamiento (GL) segregantes que han sido caracterizados.

FAMILIA BC2	N	GRUPOS DE LIGAMIENTO SEGREGANTES
256	0	-
505	21	1,2,7
473	83	2,3,6
90	12	2,4,6
480	42	2,3,5
356	75	3,4,6
460	30	2,4,7
306	55	1,2,5

El ADN de las familias BC2 fue analizado para sus loci segregantes (tabla 20), seleccionando de antemano el grupo o grupos de ligamiento que se quisieron mantener introgresado(s). Las familias 306 y 505 fueron seleccionadas para mantener introgresados los extremos del GL1. La familia 480 fue seleccionada para mantener los grupos 2 y 5, la familia 473 fue seleccionada para mantener los grupos 2 y 3, la familia 90 fue seleccionada para mantener el grupo 4, la familia 356 fue seleccionada para mantener el grupo 6 y la familia 460 fue seleccionada para mantener el grupo 7. En primer lugar se analizaron los loci correspondientes a este o estos grupos. Se seleccionaron los individuos que presentaron un grupo completo introgresado y finalmente se analizó el resto de loci segregantes.

El DNA de los 21 individuos de la familia 505 fue analizado inicialmente para los marcadores localizados en el GL1. Seis plantas (28%) mantuvieron la introgresión en este grupo y fueron caracterizadas para el resto de loci segregantes. Se seleccionó el individuo 505.011 por ser el único que mantuvo la introgresión del grupo uno y perdió las introgresiones de los grupos 2 y 7 (fig 32).

Tabla 20. Segregación de marcadores obtenida en las familias BC2. N: número de datos obtenidos, porcentajes para genotipo, grado de significación de la distorsión detectada y valores de chi-cuadrado (χ^2). *, **, ***, valores significativamente distintos a los de la segregación esperada con $p < 0.05$, 0,01 y 0,001 respectivamente; - valor no significativamente distinto al esperado.

FAMILIA	GL	LOCUS	N	Genotipo		Distorsión	Chi-cuadrado	
				<i>F. vesca</i> (%)	<i>F. bucharica</i> (%)			
505	1	UFF02F02	21	66,6	33,3	-	2,33	
		EMFv025	21	66,6	33,3	-	2,33	
		CFV-164	21	57,4	42,9	-	0,43	
	2	EMFvi099	19	42,1	57,9	-	0,47	
		EMFv002	14	35,7	64,3	-	1,14	
		EMFv031	18	55,6	44,4	-	0,22	
		ARSFL015	19	52,6	47,4	-	0,05	
		EMFn134	19	52,6	47,4	-	0,05	
		EMFv003	19	47,4	52,6	-	0,05	
		CFV-183	19	52,6	47,4	-	0,05	
		7	EMFv021	13	53,8	46,2	-	0,08
			EMFn201	21	52,4	57,6	-	0,05
			EMFvi008	17	47,1	52,9	-	0,06
			CFVCT023	17	29,4	70,6	*	2,88
473	2	EMFvi099	73	46,6	53,4	-	0,34	
		EMFv002	10	70	30	-	1,60	
		EMFv031	25	36	64	-	1,96	
		ARSFL015	19	31,6	68,4	-	2,58	
		EMFn134	10	70	30	-	1,60	
		EMFv003	53	52,8	47,2	-	0,17	
	3	CFV-183	11	72,7	27,3	-	2,27	
		EMFv029	8	62,5	37,5	-	0,50	
		UFFxa02H04	58	51,7	48,3	-	0,07	
		UDF017	71	56,3	43,7	-	1,14	
		CFVCT022	81	65,4	34,6	***	7,72	
		EMFn207	10	30	70	-	1,60	
		CFVCT011	24	62,5	37,5	-	1,50	
		CFVCT012	81	55,6	44,4	-	1,00	
6	ARSFL007	25	52	48	-	0,04		
	EMFv160BC	20	50	50	-	0		
90	2	EMFvi099	3	0	100	*	3	
		EMFv002	3	0	100	*	3	
		EMFv031	3	0	100	**	3	
	4	UDF007	12	8,3	91,6	***	8,33	
		EMFvi136	3	0	100	*	3	
		CEL-1	12	41,6	58,3	-	0,33	
		ChFaM23	12	66,6	33,3	-	1,33	
	6	ARSFL007	3	33,3	66,6	-	0,33	
		EMFn117	3	33,3	66,6	-	0,33	
	480	2	EMFvi099	42	42,9	57,1	-	0,86
EMFv002			40	45,23	54,76	-	0,38	
Fvi11			41	36,6	63,4	*	2,95	
CFV-183			41	34,1	65,9	**	4,12	
3		UDF004	12	58,4	41,6	-	0,33	
		EMFvi108	42	47,6	52,4	-	0,10	
5		CEL2	39	48,7	51,3	-	0,03	
		UDF006	41	39	61	-	1,98	
		CFV3821	3	66,6	33,3	-	0,33	
		EMFvi018	42	11,9	88,1	***	24,38	
356	3	UFFxa02H04	20	45	55	-	0,20	
		UDF017	51	60,8	39,2	-	2,37	
		EMFn207	20	50	50	-	0	
		CFVCT012	20	45	55	-	0,20	
		EMFvi136	20	35	65	-	1,80	
	4	CEL-1	30	40	60	-	1,20	
		ChFaM23	14	50	50	-	0	
		ARSFL007	69	66,7	33,3	***	7,67	
	6	EMFn117	32	28,1	71,9	**	6,12	
		EMFn017	32	37,5	62,5	-	2,00	
EMFV160AD		62	43,5	56,5	-	1,03		
EMFv160BC		75	53,3	46,7	-	0,33		
460	2	EMFvi099	17	47,1	52,9	-	0,06	
		EMFv002	4	50	50	-	0	
		EMFv031	4	50	50	-	0	
		ARSFL015	4	50	50	-	0	
		EMFn134	4	50	50	-	0	
		EMFv003	17	47,1	52,9	-	0,06	
		CFV-183	4	50	50	-	0	
	4	UDF007	4	50	50	-	0	
		EMFvi136	4	50	50	-	0	
		CFV3096	16	43,8	56,3	-	0,25	
		EMFv021	29	41,4	58,6	-	9,86	
		EMFn201	4	50	50	-	0	
		EMFvi008	29	55,2	44,8	-	0,31	
		CFVCT023	4	50	50	-	0	
306	1	EMFv023	28	53,6	46,4	-	0,14	
		EMFn072	60	56,7	43,3	-	1,07	
		EMFn049	58	56,9	43,1	-	1,10	
	2	EMFvi099	60	48,3	51,7	-	0,07	
		ARSFL015	60	56,6	43,3	-	1,07	
		EMFv003	58	62,1	37,9	*	3,38	
		CFV-183	59	64,4	35,6	**	4,90	
		CEL-2	8	37,5	62,5	-	0,5	
	5	UDF009	8	12,5	87,5	**	4,5	
		EMFvi018	8	0	100	***	8	

La familia 473 fue analizada para la presencia de alelos de *F. bucharica* en los grupos de ligamiento 2, 3 y 6. Siete individuos (6,7%) perdieron la introgresión del GL2 y mantuvieron gran parte de la introgresión del grupo 3. Tres de ellos mantuvieron la introgresión del GL3 completa. Tres individuos (2,9%) perdieron la introgresión del grupo 3 y mantuvieron la introgresión del grupo 2 completa. Estos 10 individuos fueron analizados para los dos loci segregantes localizados en los extremos del GL6. El individuo 473.061 fue seleccionado por mantener la introgresión del GL3 completa y haber perdido las introgresiones de los grupos 2 y 6 (fig 32).

Los doce individuos de la familia 90 fueron analizados para su composición genética en el GL4. Sólo 3 de ellos (25%) mantuvieron la introgresión completa. Estos individuos fueron analizados para los loci segregantes de los grupos 2 y 6. Se seleccionó el individuo 90.009 que, habiendo perdido la introgresión del GL6, conservó parte de la introgresión del GL2 (fig 32).

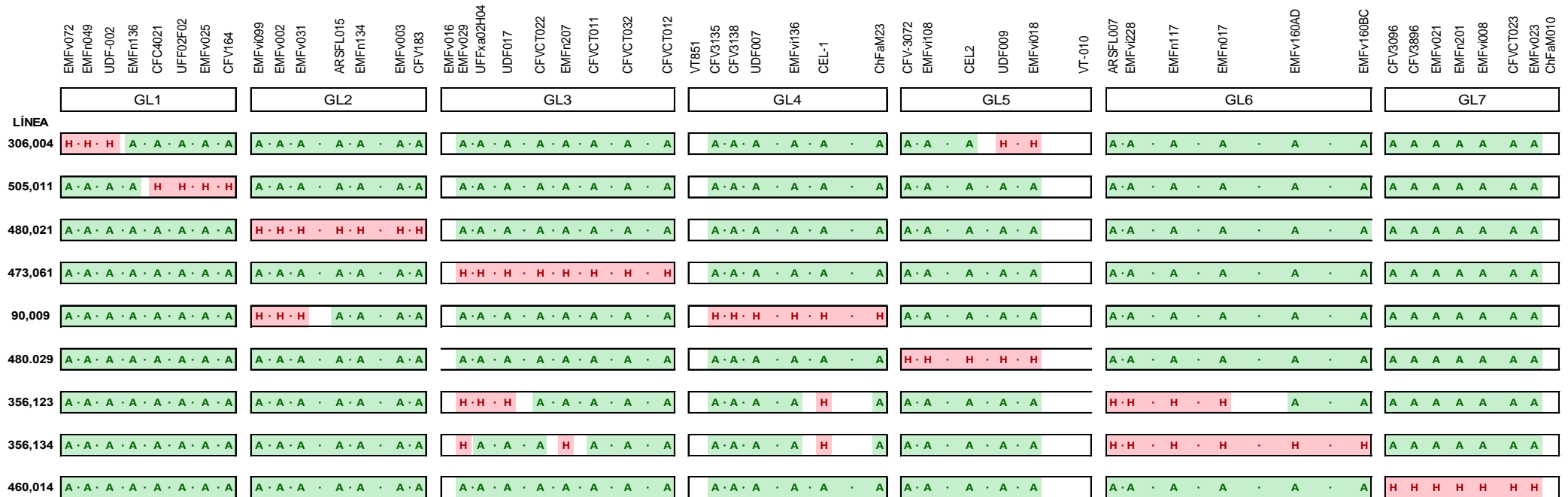
La familia 480 fue analizada para los marcadores de los grupos 2, 3 y 5. Veinte plantas (47,6%) mantuvieron la introgresión del GL5 completa. Se seleccionó el individuo 480.029, ya que había perdido las introgresiones de los grupos 2 y 3. Ocho plantas (19%) mantuvieron completa la introgresión del GL2, dos de las cuales perdieron la introgresión del GL5 completa. El individuo 480.021 fue seleccionado por perder, además, la introgresión del GL3.

Las plantas que componen la familia 356 fueron analizadas para marcadores moleculares del GL6. Nueve individuos (12%) mantuvieron la introgresión completa. De ellos, ninguno perdió el resto de introgresiones, de manera que se analizaron los marcadores segregantes de los grupos 3 y 4 de 30 individuos que mantenían gran parte del GL6 introgresado. Se seleccionaron 2 individuos. 356.123 mantuvo parte de la introgresión del GL6 (40 cM) y parte de las introgresiones de los grupos 3 y 4. La línea 356.134 mantuvo la introgresión completa del GL6 y mantuvo introgresiones en los grupos 3 y 4.

La familia 460 fue analizada para los marcadores del GL7. Nueve de las plantas (30%) mantuvieron la introgresión completa, de las cuales sólo dos perdieron la introgresión del GL2. Una de ellas, la línea 460.014 perdió también la introgresión del GL4 y fue seleccionada.

Las plantas de la familia 306 fueron analizadas para los marcadores moleculares del GL1. Veintisiete de ellas (49%) mantuvieron la introgresión. De ellas, 10 perdieron las dos introgresiones del GL2. Dos de estas 10 líneas perdieron también la introgresión del GL5. Sin embargo, ambas murieron y se hubo de seleccionar el individuo 306.004, cuyo genoma contiene parte del GL5 introgresado.

Figura 32. Representación gráfica de los genotipos seleccionados en la población BC2. Cada barra horizontal representa un grupo de ligamiento. En la parte superior figura el nombre de los marcadores moleculares empleados en la caracterización genómica. A la izquierda, en vertical, los nombres de las nueve líneas seleccionadas. ‘H’ representa genotipo híbrido (AB) y ‘A’, genotipo *F. vesca* (AA). Los segmentos de color rosa representan introgresiones, mientras que los segmentos verdes corresponden al fondo genético. Las regiones en blanco no han sido caracterizadas. En las regiones donde figura un punto (-) se asume que el genotipo es igual al de los marcadores moleculares flanqueantes



3.5 CARACTERIZACIÓN Y M.A.S EN SIETE FAMILIAS GENERADAS MEDIANTE AUTOFECONDACIÓN DE LAS SELECCIONES-BC2 (BC2 ⊗)

Las líneas seleccionadas fueron llevadas a floración de manera natural, a excepción de las líneas 356.123 y 356.134, que no florecieron de manera natural ni respondieron al tratamiento de inducción floral. El resto fueron autopolinizadas y sus semillas germinadas para generar siete familias. La tabla 21 muestra las características y loci segregantes en cada una de estas familias. En la tabla 22 se recogen los datos de segregación para todos los marcadores analizados. En función de su estructura genética se seleccionó un set de individuos para formar parte de la colección final, o por presentar genotipos interesantes para obtener líneas con introgresiones más pequeñas en el futuro.

Tabla 21. Características de las familias generadas mediante autofecundación de las líneas BC2 seleccionadas. Nombre, tamaño poblacional y loci segregantes analizados en cada una.

FAMILIA BC2 AUTOFECONDADA	N	LOCI ANALIZADOS
306.004	179	GL5: UDF009, CFV3821, EMFv024, VT-010. GL1: EMFv072, EMFn049
505.011	213	GL1: CFV4021, CO818002, UFFxa02F02, CFV164
480.021	196	GL2: EMFvi099, EMFvi148, EMFv002, EMFv031, BFACT002, Fvi011, EMFn134, BFACT039, CFV183
473.061	147	GL3: EMFv016, EMFv029, UFFxa02H04, VT-398, UDF017, CFVCT022, EMFn207, CFVCT011, CFVCT012
90.009	231	GL2: EMFvi099, EMFv002, BFACT002. GL4: VT-851, CFV3819 CFV3135, CFV3148, VT-861, CEL-1, ChFaM023
480.029	168	GL5: CFV3072, CFV3132, CEL-2, EMFn110, UDF006, UDF009, CFV3821, EMFn010, EMFv024, EMFvi018, VT-010
460.014	188	GL7: CFV3096, EMFv021, CFV3117, EMFvi008, CFVCT023, EMFv023
TOTAL	1322	

La familia 306.004 fue analizada para seis loci, 5 SSR y un CAPS. La segregación de los marcadores se ajustó al modelo mendeliano. Los genotipos seleccionados en esta familia se muestran en la figura 33. Se seleccionaron 5 individuos. De ellos, tres conservaron la introgresión del GL5 y perdieron la introgresión del GL1, y dos conservaron la introgresión del GL1 y perdieron la introgresión del GL5.

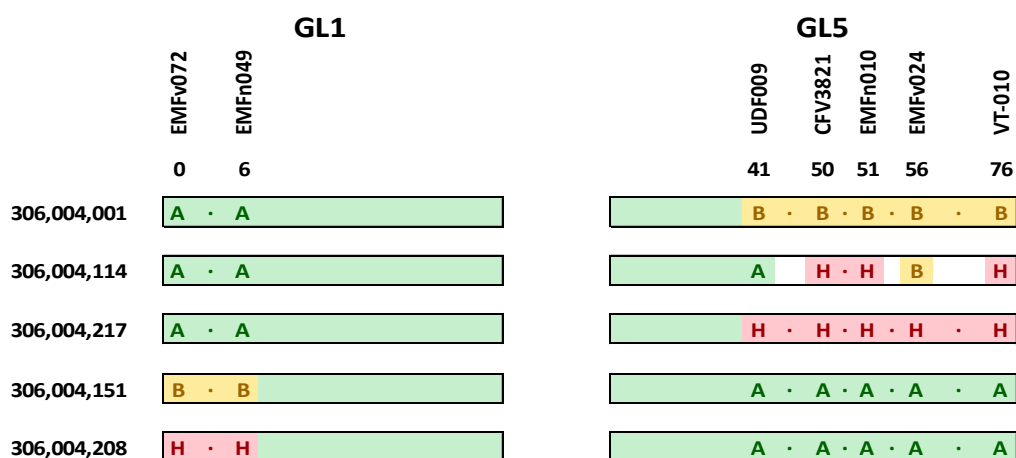


Figura 33. Esquema de los grupos de ligamiento 1 y 5 de los genotipos seleccionados en la familia 306.004. El color verde y la letra A representan el genotipo *F. vesca*. B indica genotipo *F. bucharica* y H, genotipo híbrido.

La familia 505.011 fue analizada para cuatro SSRs localizados en extremo inferior del GL1. Los marcadores segregaron siguiendo el modelo mendeliano esperado. Se seleccionaron tres individuos pertenecientes a esta familia, cuyos genotipos se muestran en la figura 34. El individuo 505,011,002 mantuvo la introgresión completa en homocigosis, mientras que 505.011.001, y 505,011,021 mantuvieron la región inferior introgresada en homocigosis.

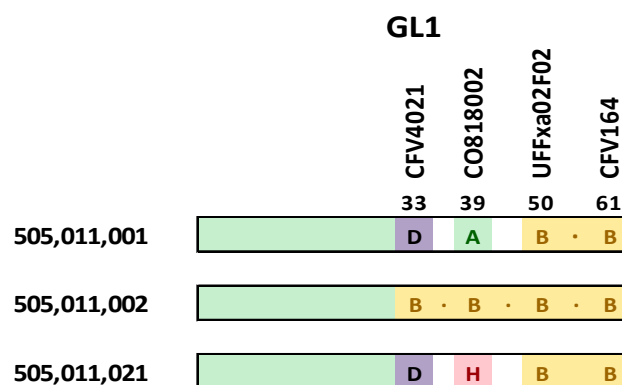


Figura 34. Esquema del grupo de ligamiento 1 de los individuos con genotipo seleccionado en la familia 505.011. El color verde representa el genotipo *F. vesca*. B indica genotipo *F. bucharica*, H, genotipo híbrido y D genotipo A o H (marcador dominante).

La familia 480.021 fue analizada para 9 SSRs del GL2. Los valores de segregación obtenidos están fuertemente distorsionados, con predominio de los genotipos híbrido y *F. bucharica* desde el inicio del grupo hasta la posición 45,2 cM (EMFn134). Los genotipos seleccionados se muestran en la figura 35. Un individuo mantiene la introgresión completa en homocigosis, el resto presentan distintas regiones del grupo introgresadas, tanto en homocigosis como en heterocigosis.

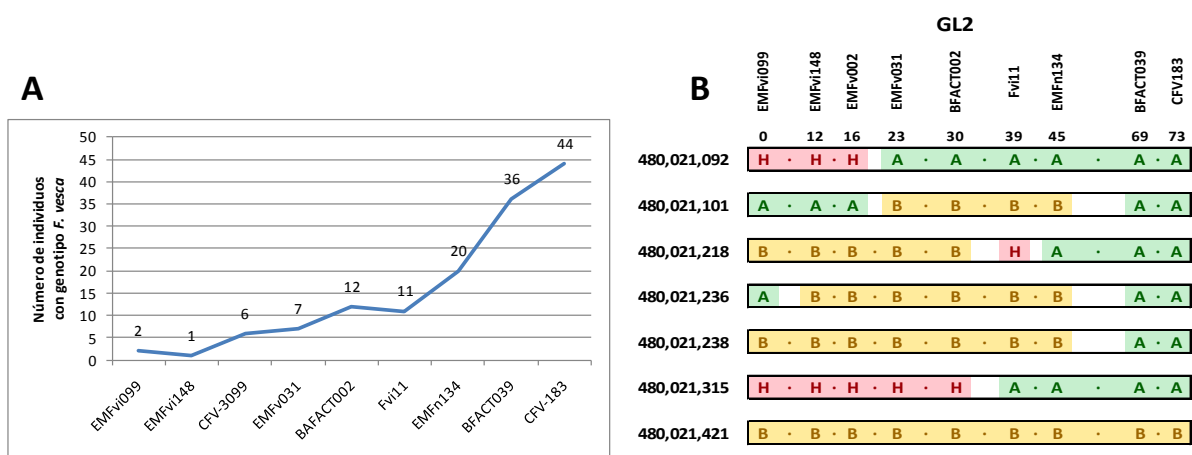


Figura 35. Análisis de los individuos pertenecientes a la familia 480.021. **A:** Número de individuos con genotipo *F. vesca* en la familia 480.021 (N=196 individuos) para cada uno de los marcadores moleculares que se indican. **B:** Esquema del grupo de ligamiento 2 de los individuos con genotipo seleccionado en la familia 480.021. La letra A representa el genotipo *F. vesca*. B indica genotipo *F. bucharica*, y H genotipo híbrido.

En la familia 473.061 se analizaron 9 SSRs. Los marcadores segregaron de manera mendeliana y 5 individuos fueron seleccionados (fig. 36). Un individuo mantuvo la introgresión del grupo completo en homocigosis. El resto mantuvieron introgresadas diferentes regiones del grupo, la mayoría en homocigosis, excepto 473,061,312, que mantiene introgresada una región de 64 cM en heterocigosis.

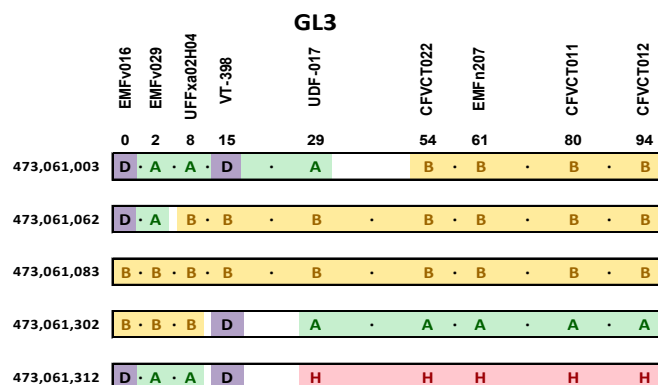


Figura 36. Esquema del grupo de ligamiento 3 de los individuos con genotipo seleccionado en la familia 473.061. A representa el genotipo *F. vesca*. B indica genotipo *F. bucharica*, H, genotipo híbrido y D genotipo A o H (marcador dominante).

La familia 90.009 fue analizada para los marcadores localizados en las regiones segregantes de los GL2 y GL4. Los marcadores del GL2 mostraron una fuerte distorsión, mientras que los del GL4 mostraron una segregación mendeliana normal (tabla 22). Se seleccionaron 8 genotipos, cuyo genoma se esquematiza en la figura 37. Uno de los individuos mantuvo la introgresión del GL2 en homocigosis y perdió la introgresión del GL4. El resto perdieron la introgresión del GL2 y mantuvieron diferentes regiones del GL4 introgresadas. Una de las líneas presentó la introgresión del GL4 completa en heterocigosis.

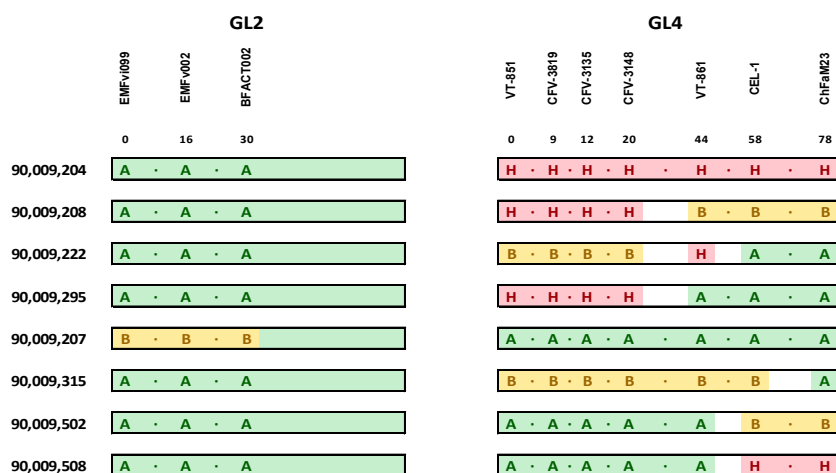


Figura 37. Esquema de los grupos de ligamiento 2 y 4 de los individuos con genotipo seleccionado en la familia 90.009. El color verde y la letra A representa el genotipo *F. vesca*. B indica genotipo *F. bucharica*, H, genotipo híbrido.

La familia 480.29 fue analizada para los marcadores moleculares localizados en el GL5. Los marcadores segregaron de manera distorsionada para todos los loci analizados, con predominio del genotipo *F. bucharica* frente al genotipo *F. vesca* (tabla 22). Se seleccionaron 8 individuos cuyos genotipos se esquematizan en la figura 38. Una línea mantuvo la introgresión completa en homocigosis, mientras que el resto fueron seleccionadas por mantener diferentes partes del grupo introgresado, tanto en homocigosis como en heterocigosis.

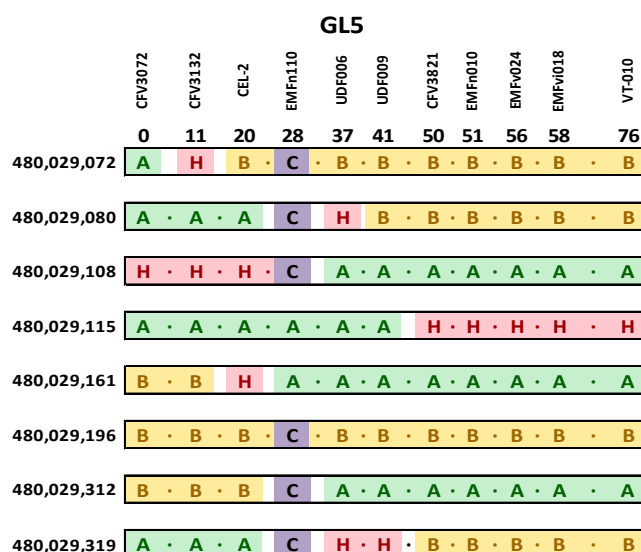


Figura 38. Esquema del grupo de ligamiento 5 de los individuos con genotipo seleccionado en la familia 480.029. A representa genotipo *F. vesca*. B indica genotipo *F. bucharica*, H indica, genotipo híbrido y C indica genotipo B o H (marcador dominante).

La familia 460.014 fue analizada para 6 marcadores del GL7. La segregación de los marcadores en este grupo estuvo muy distorsionada, con prevalencia de los alelos de *F. vesca* sobre los de *F. bucharica*, como puede apreciarse en la figura 39. De los 196 individuos analizados, sólo uno, 460.14.010, presentó introgresada la región comprendida entre los marcadores CFV3117 y EMFvi008 en homocigosis. En el resto, se observa una tendencia a la generación de puntos de recombinación flanqueando los marcadores CFV3117 y EMFvi008 cuando el genotipo, tanto aguas arriba como aguas abajo de dichos marcadores, es *F. bucharica*, de manera que, salvando la línea 460,14,010, ninguna de las introgresiones homocigotas obtenidas rebasó las posiciones centrales del grupo. La líneas 460,014,081, con una introgresión homocigótica en el extremo inferior del GL7, murió semanas después de ser genotipada.

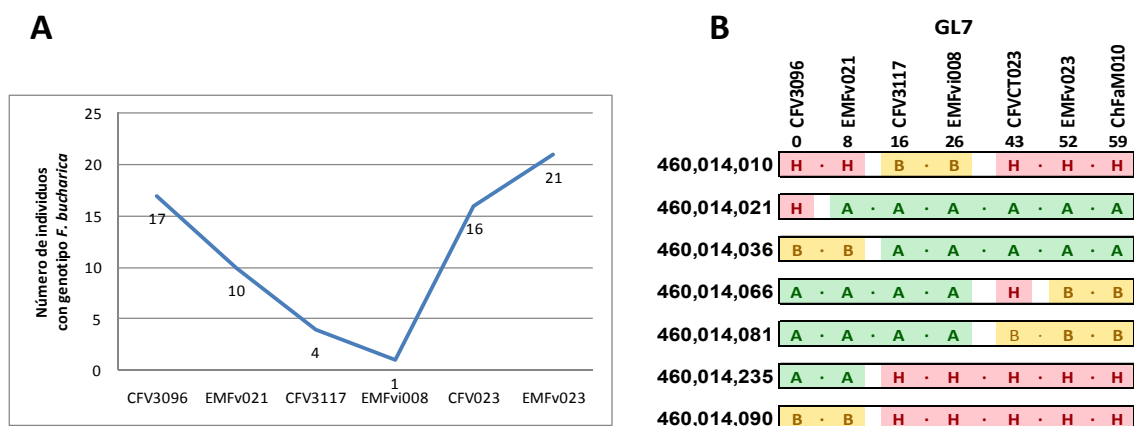


Figura 39. Análisis de los individuos pertenecientes a la familia 460.014. **A:** Número de individuos con genotipo *F. bucharica* pertenecientes a la familia 460.014 (N=188 individuos) para cada uno de los marcadores moleculares que se indican. **B:** Esquema del grupo de ligamiento 7 de los individuos con genotipo seleccionado en la familia 460.014. El locus ChFaM010 sólo se caracterizó para las líneas seleccionadas.

3.6 CARACTERIZACIÓN GENÓMICA DE LAS LÍNEAS BC2 SELECCIONADAS.

Tras la selección de 43 líneas BC2, se realizaron nuevas extracciones de ADN, tanto de éstas como de sus líneas parentales BC1 y BC2. Se analizó su composición genómica para 68 marcadores moleculares distribuidos uniformemente a lo largo de todo el genoma de *Fragaria* diploide, con objeto de detectar posibles introgresiones contaminantes y delimitar los límites de las introgresiones con mayor exactitud. Con este análisis se completó el genotipado de las líneas. Se constata que queda sin cubrir una pequeña región del GL1, menor de 16 cM localizada entre los marcadores UDF-002 y CFV-4021. El 78,2% del genoma de *F. bucharica* está introgresado de forma homocigótica en alguna de las líneas de la colección. Sólo el GL6 carece de introgresiones en homocigosis, pero queda cubierto por dos líneas que contienen, además, otras tres introgresiones segregantes en los grupos 3 y 4. Treinta y una líneas fueron seleccionadas por poseer introgresiones en homocigosis, mientras que 12 contienen introgresiones en heterocigosis. El grupo de ligamiento que presenta un mayor número de introgresiones es el GL5, con 11 líneas, seguido del 2 con 8 líneas. Siete líneas tienen introgresiones en los grupos 4 y 7. Los grupos 1 y 3 presentan introgresiones en 5 líneas. Las familias con mayor representación son la 480.029 y la 90.009, con 8 líneas, seguidas de las familias 460,014 y 480,021, con 7 líneas cada una. Las familias 306,004 y 473,061 están representadas con 5 líneas cada una, y en último lugar, la familia 505,011, con 3 líneas.

En relación al fondo genético de las introgresiones, lejos de ser una línea pura, 'Reine des Vallées' presenta heterocigosis para al menos 14 marcadores moleculares (tabla 17) localizados en todos los grupos de ligamiento. La tabla 23 muestra la segregación para algunos de estos marcadores en las líneas seleccionadas.

Tabla 22. Segregación de marcadores obtenida en las familias BC2. Nombre, grupos de ligamiento segregantes, loci y número de individuos analizados, porcentajes obtenidos para cada uno de los genotipos y grado de significación de la distorsión detectada. En el caso de marcadores dominantes, el porcentaje se sitúa entre dos columnas. *, **, ***, valores significativamente distintos a los de la segregación esperada con $p < 0.05$, 0,01 y 0,001 respectivamente; - valor no significativamente distinto al esperado.

FAMILIA	GL	LOCUS	N	Genotipo <i>F. vesca</i> (%)	Genotipo híbrido (%)	Genotipo <i>F. bucharica</i> (%)	Distorsión	χ^2
306.004	1	EMFv072	166	22,3	58,4	19,3	*	4,72
		EMFn049	164	23,2	56,7	20,1	-	3,51
	5	UDF009	173	15,0	52,0	32,9	***	11,38
		CFV-3821	164	14,6	52,4	32,9	***	11,37
		EMFv024	166	13,2	55,5	31,3	***	12,80
		VT-010	176	14,2	54,5	31,2	***	11,68
505.011	1	CFV-4021	191		80,1	19,9	-	2,65
		CO818002	168	26,3	48,7	25,0	-	0,05
		UFF02F02	180	25,0	51,1	23,9	-	1,13
		CFV164	195	28,7	49,2	22,1	-	1,78
480.021	2	EMFvi099	176	1,1	51,1	47,8	***	76,50
		EMFvi148	190	1,0	51,0	48,0	***	83,46
		CFV-3099	187	3,2	50,8	46,0	***	68,50
		EMFv031	194	3,6	53,1	43,3	***	61,87
		BAT002	186	6,4	50,0	43,5	***	51,19
		Fvi11	171	6,4	54,4	39,2	***	37,99
		EMFn134	175	11,4	48,6	40,0	***	28,71
		BFACT039	177	20,3	54,2	25,4	-	2,19
473.061	3	CFV-183	187	23,5	54,0	22,4	-	1,25
		EMFv016	126		73	27	-	0,15
		EMFv029	141	26,2	48,9	24,8	-	0,12
		UFFxa02H4	156	26,3	48,1	25,6	-	0,24
		VT-398	150		74	26	-	0,00
		UDF017	169	25,4	50,9	23,7	-	0,16
		CFVCT022	139		81,3	18,7	**	4,70
		EMFn207	161	23,6	57,8	18,6	*	4,96
		CFVCT011	164	28	54,9	17,1	*	5,51
		CFVCT012	166	24,1	50,6	25,3	-	0,07
90.009	2	EMFvi099	227	5,3	53,3	41,4	***	60,23
		EMFv002	203	4,4	54,7	40,9	***	55,73
		EMFv031	218	5,5	55,0	39,4	***	52,46
	4	VT-851	223	20,2	52,2	27,8	-	2,96
		CFV-3819	216	24,2	52,0	23,8	-	0,09
		CFV-3135	213	21,1	51,6	27,2	-	1,82
		CFV-3148	224	21,4	49,6	29,0	-	2,60
		VT-861	164	22,0	55,5	22,6	-	2,21
		CEL-1	224	24,6	53,6	21,9	-	1,46
ChFaM-023	223	25,6	54,7	19,7	-	3,49		
480.029	5	CV-3072	166	15,7	54,2	30,1	**	8,12
		CV-3132	167	13,8	50,9	35,3	***	15,57
		CEL-2	163	14,7	48,5	36,8	***	16,06
		EMFn110	141	16,3		83,7	**	5,68
		UDF006	158	18,4	41,8	39,9	***	18,91
		UDF009	138	14,5	43,5	42	***	23,28
		CV-3821	156	14,1	46,2	39,7	***	21,44
		EMFn010	150	14	47,3	38,7	***	18,68
		EMFv024	145	15,9	44,8	39,3	***	17,50
		EMFvi018	156	15,4	46,2	38,5	***	17,54
460.014	7	VT-10	148	12,2	49,3	38,5	***	20,58
		CFV3096	172	33,1	57,0	9,9	***	21,95
		EMFv021	171	37,4	56,7	5,8	***	37,20
		CFV-3117	176	40,3	57,4	2,3	***	54,85
		EMFvi008	181	40,3	59,1	0,6	***	63,30
		CFVCT023	182	29,7	61,5	8,8	***	25,56
EMFv023	127	36,2	47,2	16,5	***	10,23		

Tabla 23. Genotipo de algunos loci segregantes en el parental *F. vesca* en los individuos F1-30, BC1, BC2 y BC2 seleccionados para el análisis genómico final. Arriba se indican los tamaños de los alelos segregantes identificados *F. vesca*, a los que se designa como alelos A₁ y A₂. B indica que el individuo no posee alelos *F. vesca*. -, dato no disponible.

	EMFn136	EMFv029	CFVCT022	CFVCT012	CFV-3819	CFV-3135	UDF009	EMFn010	EMFv024	CFV3096
A1 (pb)	177	202	131	170	255	232	171	224	216	261
A2 (pb)	207	224	133	203	259	290	158	204	219	267
F1-30	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
90,009	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁
90,009,204	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁
90,009,207	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁
90,009,208	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	-	A ₁ A ₁
90,009,222	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	B B	B B	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
90,009,295	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁
90,009,315	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	B B	B B	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁
90,009,502	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁
90,009,508	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	-	A ₂ A ₂
306,004	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	-	-	A ₁ A ₂
306,004,001	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	B B	B B	B B	A ₁ A ₂
306,004,114	A ₂ A ₂	-	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	-	A ₁ A ₁	A ₁ A ₁	-	A ₁ A ₁
306,004,151	A ₁ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	-	A ₁ A ₁	A ₁ A ₂
306,004,208	A ₁ A ₂	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂
306,004,217	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂
356	A ₁ A ₂	-	A ₁ A ₁	A ₁ A ₂	A ₁ A ₁	-	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
356,123	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₂ A ₂
356,134	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂
460,014	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	-	A ₁ A ₁	B B
460,014,010	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
460,014,021	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	B B	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
460,014,036	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	B B
460,014,066	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	-	A ₁ A ₁	A ₂ A ₂
460,014,081	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
460,014,090	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	B B
460,014,235	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
473,061	A ₁ A ₂	-	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	-	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂
473,061,003	A ₁ A ₂	A ₁ A ₁	B B	B B	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂
473,061,062	A ₁ A ₂	A ₁ A ₁	B B	B B	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	-	A ₁ A ₂	A ₂ A ₂
473,061,083	-	-	B B	B B	A ₁ A ₁	-	A ₁ A ₂	A ₁ A ₂	-	A ₂ A ₂
473,061,312	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
473,061,328	A ₁ A ₂	B B	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	-	A ₁ A ₁	A ₂ A ₂
480,021	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	-	A ₁ A ₂	A ₂ A ₂
480,021,092	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂
480,021,101	A ₂ A ₂	A ₁ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂
480,021,218	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂
480,021,236	A ₂ A ₂	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂	-	A ₁ A ₁	-	A ₁ A ₂	A ₂ A ₂
480,021,238	A ₂ A ₂	A ₁ A ₂	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	-	A ₂ A ₂
480,021,315	A ₂ A ₂	A ₁ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂
480,029	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	-	A ₂ A ₂
480,029,072	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	B B	B B	B B	A ₂ A ₂
480,029,080	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	B B	B B	B B	A ₂ A ₂
480,029,108	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	-	A ₂ A ₂	A ₂ A ₂
480,029,115	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂
480,029,161	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂
480,029,196	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	B B	B B	B B	A ₂ A ₂
480,029,312	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂
480,029,319	A ₂ A ₂	-	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₁	B B	B B	A ₂ A ₂
505,011	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂
505,011,001	A ₁ A ₁	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₂	-	A ₁ A ₁	A ₂ A ₂
505,011,002	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	A ₁ A ₁	-	A ₂ A ₂
505,011,021	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₂ A ₂	A ₂ A ₂	A ₂ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂
505,011,061	A ₁ A ₂	A ₂ A ₂	A ₁ A ₁	A ₂ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₂	A ₁ A ₁	A ₁ A ₁	A ₂ A ₂

3.7 SELECCIÓN DE UNA COLECCIÓN DE LÍNEAS DE INTROGRESIÓN.

Las líneas seleccionadas para formar parte de la colección final de líneas de introgresión se eligieron en base a la posición de sus introgresiones homocigóticas. Dieciocho de las treinta líneas pueden ser consideradas NILs (fig. 41). Las introgresiones seleccionadas tienen un tamaño medio de 43 cM (7,9% del genoma). La introgresión de mayor tamaño, localizada en el GL6 de la línea 356.134, abarca 101,7 cM (18,8% del genoma), y la más pequeña, presente en la línea 505.011.001, abarca 3 cM del GL1 (0,5% del genoma). Los grupos de ligamiento con mayor representación en la colección son los GLs 2 y 5, con 6 líneas cada uno, seguido del grupo 7, con 5 líneas. Los grupos 1 y 3 están representados con cuatro líneas cada uno y, por último, los grupos 4 y 6 están representados con 3 y 2 líneas respectivamente. Del total de regiones introgresadas, que abarcan un total de 1355 cM, el 70,1% corresponde a introgresiones homocigóticas, mientras que el 29,9% restante corresponde a introgresiones heterocigóticas. El 47% de las líneas mantiene alguna región en heterocigosis, generalmente de pequeño tamaño, adyacente a las introgresiones homocigóticas. Tan sólo las líneas 356.123 y 356.134 mantienen introgresiones heterocigóticas localizadas en grupos de ligamiento distintos al GL6 (4% del genoma en el primer caso, y 6,6% en el segundo) además de poseer introgresiones heterocigóticas para el GL6.

El conjunto de introgresiones seleccionadas define un conjunto de *bins* o segmentos genómicos delimitados por dos puntos de recombinación (pertenecientes a un mismo individuo o a dos individuos con introgresiones contiguas o solapantes) o entre un punto de recombinación y un telómero (Fig. 40). El tamaño medio de los *bins* de una población de NILs define la resolución genética de este tipo de poblaciones para el mapeo de nuevos marcadores moleculares o QTLs. (Vision *et al.* 2000; Eduardo, 2005). El tamaño medio de los 38 *bins* generados es de 17,7 cM, si bien varía en un rango que va de los 2 (bin 14) a los 62 (bin 32) cM.

Para simplificar la denominación de las líneas, se ha elaborado una nueva nomenclatura. El nombre de cada línea comienza por las letras FB (*Fragaria bucharica*) seguido de un número del 1 al 7, que se corresponde con el grupo de ligamiento que contiene la introgresión principal, seguido de un punto. Después del punto hay dos números separados por un guión que se corresponden con la posición de inicio y final de la introgresión principal. En las líneas que tienen regiones segregantes, figura la letra h al final. Esta nomenclatura permite que, en el futuro, puedan introducirse nuevas líneas a la colección sin modificar los nombres de las ya existentes. La figura 42 muestra las imágenes de las 30 líneas seleccionadas.

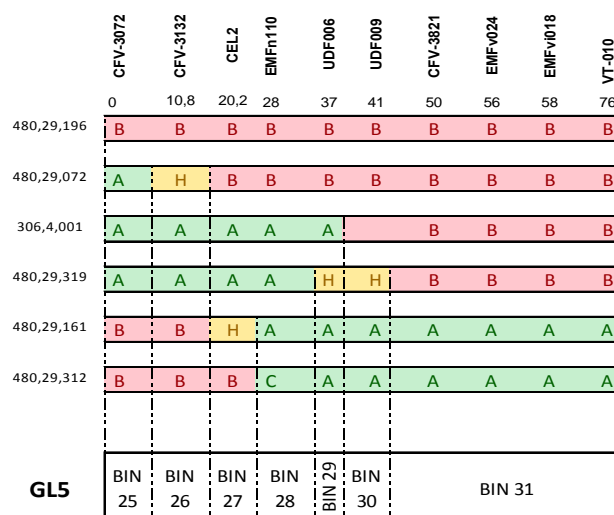


Figura 40. Representación gráfica de los bins del GL5. Las líneas discontinuas indican los límites entre bins, es decir, entre regiones genómicas definidas por la segregación única de los marcadores moleculares que lo componen en la colección.



Figura 42. Imágenes de las 30 líneas de introgresión seleccionadas en esta investigación.

3.8 MAPEO DE CARACTERES CUALITATIVOS EN LA COLECCIÓN DE LÍNEAS DE INTROGRESIÓN

Durante la construcción de la colección se han ido recogiendo datos acerca del desarrollo de las plantas obtenidas. Tal y como se preveía por la posición en el mapa del locus *r*, que codifica para la formación de estolones, y que está situado en la posición 42,2 cM del GL2, sólo los individuos FB2.0-73, FB2.0-63, FB2.12-63 y FB2.12-45 presentan estolones.

Hemos observado que algunas plantas desarrollan gruesos tallos florales, muy largos, que sobresalen en longitud del resto de la planta. (fig. 42). Este comportamiento aparece en todas las líneas seleccionadas de la familia 480.021 excepto en la 480,021,092 (que no ha sido incluida en la colección final por tener la introgresión en heterozigosis). Este fenotipo sí aparece en la línea 480.021.315, por tanto, los factores genéticos que determinan este tipo de tallos florales se localizan en el GL2, en una región de 23,5 cM definida por los marcadores EMFv002 y Fvi11 (fig. 35).

Se ha detectado que las plantas con genotipo *F. bucharica* en el GL6 (líneas FB6.0-39h y FB6.0-101h), concretamente en una región de 54,3 cM delimitada por los marcadores ARSFL007 y CFVCT017 (fig. 32), son sensibles a la vernalización.

Todas las plantas seleccionadas en la familia 505.011 (incluyendo al individuo BC2 505.011) poseen una clorosis característica en sus hojas. Parece que se trata de un carácter dominante, cuyo origen genético se localiza en una región de 12 cM del GL1, flanqueada por los marcadores UFF02F02 y CFV164 (fig. 34).

Las líneas FB5.0-76, FB5.50-76h y FB5.41-76 detuvieron el desarrollo de flores al comienzo del otoño, cuando los días comenzaron a acortarse. Mientras, en el resto de líneas la floración se intensificó en esta época (exceptuando la línea FB5.20-76h, que nunca floreció). El análisis efectuado a fecha 30/09/2010 (fotoperiodo de 11 horas y 49 minutos) mostró que FB5.0-76, FB5.50-76, y FB5.41-76 detuvieron por completo el desarrollo de flores.



Figura 43. Fenotipos mostrados por algunas líneas de introgresión. A: Clorosis de la hoja, presente en las líneas pertenecientes a la familia 505.011. B: Tallos florales engrosados en la línea FB2.0-73. C: Desarrollo de estolones en la línea FB2.12-63.

3.9 ESTUDIO PRELIMINAR DE ALGUNOS CARACTERES CUANTITATIVOS EN LA COLECCIÓN

3.9.1 TIEMPO DE FLORACIÓN

Las plantas que componen la colección de líneas de introgresión comenzaron a desarrollar flores en diferentes momentos de su desarrollo. El número de días que pasaron desde la siembra hasta el desarrollo del primer botón floral sigue una distribución normal con media $\mu_v = 78,16$ y desviación estándar $\delta_v = 10,94$. El periodo vegetativo varía en las líneas evaluadas en un rango que va de los 51 días en la línea FB1.50-61h a los 100 días en la línea FB5.41-76. Los individuos FB2.12-63 y FB1.0-6h no fueron evaluados, puesto que su desarrollo es mucho más lento que el del resto de líneas. Las líneas FB6.0-39h Y FB6.0-101h no fueron evaluadas puesto que no han desarrollado flores hasta el momento debido a su sensibilidad a la vernalización. La línea FB5.20-76h tampoco ha desarrollado flores a día de hoy. La tabla 24 muestra la distribución de la duración del periodo vegetativo en los individuos de la colección. Las líneas que precisan de un mayor periodo vegetativo son las que poseen introgresiones en la región inferior del GL5 (líneas FB5.0-76, FB5.20-76h, FB5.50-76h, FB5.41-76). El periodo vegetativo medio de las líneas con introgresiones en el GL5 es de 88,4 días. Esta cifra es un 13,1% superior al valor μ_v . El periodo vegetativo medio de las líneas FB5.0-76, FB5.50-76 y FB5.41-76 es de 96,7 días, lo que supone un aumento del 23,67% sobre la media de la colección y del 9,34% sobre la media de las líneas con introgresiones en el GL5. Si eliminamos del análisis las líneas que contienen introgresiones en el GL5, la media de la población es de 75,6 días. El promedio de la duración del periodo vegetativo en las líneas con introgresiones en la región inferior del GL5 muestra diferencias estadísticamente significativas respecto a los promedios de las líneas con introgresiones en otras regiones genómicas ($Pr > 0.0233$). Este modelo explica un 52,2% de la variabilidad total encontrada.

Tabla 24. Periodo vegetativo de los individuos que componen la colección de líneas de introgresión. Grupo de ligamiento introgresado, nombre de la línea, fechas de siembra y primera floración y periodo vegetativo, expresado en número de días desde la fecha de siembra a la aparición del primer botón floral. Las líneas señaladas en color gris no fueron evaluadas.

INTROGRESIÓN (GL)	LÍNEA	FECHA DE SIEMBRA	FECHA DE FLORACIÓN	PERIODO VEGETATIVO (n° días)	INTROGRESIÓN (GL)	LÍNEA	FECHA DE SIEMBRA	FECHA DE FLORACIÓN	PERIODO VEGETATIVO (n° días)
1	FB1.0-6h	18/03/2010	20/09/2010	186	5	FB5.0-76	13/03/2010	14/06/2010	93
	FB1.33-61	11/02/2010	07/05/2010	85		FB5.20-76h	13/03/2010	-	-
	FB1.50-61h	11/02/2010	03/04/2010	51		FB5.50-76h	30/03/2010	05/07/2010	97
	FB1.58-61	11/02/2010	07/05/2010	85		FB5.41-76	06/03/2010	14/06/2010	100
2	FB2.0-73	30/03/2010	23/06/2010	85	FB5.0-20	30/03/2010	21/06/2010	83	
	FB2.0-63	12/03/2010	07/06/2010	87	FB5.0-11h	19/03/2010	27/05/2010	69	
	FB2.12-63	12/03/2010	-	-	6	FB6.0-39h	16/08/2009	-	-
	FB2.12-45	06/03/2010	28/05/2010	83		FB6.0-101h	16/08/2009	-	-
	FB2.0-30h	12/03/2010	07/06/2010	87	7	FB7.52-59h	16/02/2010	04/05/2010	77
FB2.0-30	18/03/2010	28/05/2010	71	FB7.43-59		16/02/2010	28/04/2010	71	
FB3.0-94	23/02/2010	30/04/2010	66	FB7.26-27h		16/02/2010	27/04/2010	70	
FB3.54-94	23/02/2010	10/05/2010	76	FB7.0-10		16/02/2010	26/04/2010	69	
3	FB3.0-8	12/04/2010	28/06/2010	77	FB7.0-16h	16/02/2010	07/05/2010	80	
	FB3.8-94	16/02/2010	27/04/2010	70					
4	FB4.58-78	30/03/2010	21/06/2010	83	media				78,16
	FB4.0-20h	18/03/2010	25/05/2010	68					
	FB4.0-58	18/03/2010	28/05/2010	71					

3.9.2 RESISTENCIA A OIDIO

Los síntomas de la infección por oidio, fueron examinados en la colección en diferentes momentos durante el verano de 2010. En función de la gravedad de los síntomas, a cada planta se le asignó una puntuación siguiendo una escala del 0 al 3, donde 0 representa la ausencia de síntomas y 3 representa el máximo grado de afección por el hongo. Las líneas FB1.0-6h, FB2.12-63, FB3.0-94, FB3.0-8, FB3.54-94, FB4.58-78, FB4.0-58 y FB5.20-76 no fueron evaluadas, dado que las características de las plantas hicieron inviable el análisis. Este fenómeno se debe a diferentes causas: En la línea FB5.20-76, los graves síntomas de la infección por el virus del retorcimiento de la hoja enmascararon los síntomas de infección por oidio en algunas mediciones. En las líneas FB1.0-6h, FB2.12-63, el desarrollo vegetativo estuvo muy retrasado respecto al resto de líneas, de manera que el área foliar era mucho menor, imposibilitando el análisis. Las líneas FB3.0-94, FB3.0-8 y FB3.54-94 presentaron un fenotipo foliar que impide la identificación de los síntomas de infección por oidio, como se estudiará más adelante con mayor detenimiento. Las líneas FB4.58-78 y FB4.0-58 presentaban enanismo, de modo que el área foliar fue también mucho menor.

Las líneas FB6.0-39h y FB6.0-101h no presentaron síntomas de la infección por oidio en ninguna de las mediciones. A la vista de estos resultados, se observa que las introgresiones en la región superior del GL6 podrían estar en el origen de la resistencia a la infección por el hongo. La línea 5.41-76 presentó bajos síntomas de infección a lo largo de toda la temporada y fue la única, junto con FB6.0-39h y FB6.0-101h que no presentó síntomas en la primera medición, cuando se detectó la mayor

afectación por parte del hongo en la población. Tan sólo presentó síntomas leves en una de las mediciones. Sin embargo, la ausencia de fenotipo en la línea FB5.0-76 plantea serias dudas sobre la correlación entre la resistencia a oidio y las introgresiones del GL5.

Tabla 25. Caracterización de la colección de líneas de introgresión para los síntomas de infección por oidio. Grupo de ligamiento introgresado, nombre de la línea, y gravedad de la afección, expresada en una escala de 0 a 3 en diferentes mediciones efectuadas durante el verano de 2010. Las líneas señaladas en gris no fueron evaluadas

LÍNEA	28 jul	17 ago	30 ago	8 sep	23 sep	FENOTIPO	LÍNEA	28 jul	17 ago	30 ago	8 sep	23 sep	FENOTIPO
FB1.0-6h	-	-	-	-	1	-	FB4.0-20h	1	1	0	0	0	tolerante
FB1.33-61	2	0	0	0	0	tolerante	FB4.0-58	-	-	-	-	-	tolerante
FB1.50-61h	2	0	0	1	0	tolerante	FB5.0-76	1	2	1	0	0	tolerante
FB1.58-61	2	0	0	1	1	tolerante	FB5.20-76h	0	3	-	-	0	tolerante
FB2.0-73	2	1	0	0	0	tolerante	FB5.50-76h	1	2	0	0	0	tolerante
FB2.0-63	1	2	0	0	0	tolerante	FB5.41-76	0	1	0	0	0	tolerante
FB2.12-63	-	-	-	-	-	-	FB5.0-20	2	3	0	2	2	tolerante
FB2.12-45	2	1	0	0	1	tolerante	FB5.0-11h	2	2	0	2	0	tolerante
FB2.0-30h	2	2	0	0	0	tolerante	FB6.0-39h	0	0	0	0	0	resistente
FB2.0-30	3	1	2	1	2	tolerante	FB6.0-101h	0	0	0	0	0	resistente
FB3.0-94	-	-	-	-	-	-	FB7.52-59h	3	0	1	1	0	tolerante
FB3.8-94	-	-	-	-	0	-	FB7.43-59	3	0	2	1	1	tolerante
FB3.54-94	-	-	-	-	0	-	FB7.26-27h	3	0	1	1	0	tolerante
FB3.0-8	3	0	2	2	1	tolerante	FB7.0-10	3	0	1	1	2	tolerante
FB4.58-78	-	-	-	-	0	tolerante	FB7.0-16h	3	0	2	1	1	tolerante

3.9.3 TAMAÑO DE LOS FOLIOLOS

Veintiocho de las 30 líneas de introgresión fueron analizadas para la longitud de sus folíolos. Las líneas FB4.0-58 y FB2.12-63 no fueron evaluadas, ya que su desarrollo estuvo muy retrasado respecto al resto de líneas. El parental RV mostró un valor medio de 51 mm para la longitud de sus folíolos. En la colección, el valor medio observado fue de 42,3 mm. Catorce líneas mostraron diferencias significativas respecto al control ($\alpha = 0,01$). Todas ellas presentan tamaños menores al control. Tan sólo la línea FB6.0-101h mostró valores superiores a RV, si bien las diferencias detectadas no son estadísticamente significativas. La figura 44 muestra las diferencias detectadas para el tamaño de los folíolos respecto al control RV. Las diferencias significativas detectadas varían en un rango que va del 17,2% de la línea FB3.0-8 al 64,8% de la línea FB3.0-94.

Las líneas FB3.0-94, FB3.54-94 y FB3.8-94 mostraron los mayores valores de variación, con una disminución del 64,8%, 60,6% y 48,3% respecto al control respectivamente. Estas líneas poseen una región introgresada solapante de 40,3 cM. La línea FB3.0-8, cuya introgresión no solapa con el resto, también presenta efectos significativos, si bien no son atribuibles a este QTL, ya que su introgresión no solapa con la de la línea FB3.54-94. La disminución del tamaño de los folíolos en estas líneas va

acompañada de una gran reducción del tamaño de los peciolo, por lo que el fenotipo resultante es una planta de pequeño porte con hojas muy pequeñas (21 mm. en promedio).

Las líneas con introgresiones en el GL1 mostraron diferencias significativas y valores muy similares entre sí, si bien la línea FB1.0-6h no comparte la región introgresada con el resto de líneas. Las líneas FB7.0-10 y FB7.26-27 también son significativamente distintas al control, pero el fenotipo mostrado por FB7.0-16h, que presenta introgresiones en la misma región, no es estadísticamente significativo. Lo mismo sucede con las líneas FB5.0-20 y FB5.0-11h con respecto a la línea FB5.0-76. FB4.0-20h y FB4.58-78 presentan diferencias significativas, si bien sus introgresiones no son solapantes. Por todo ello, sólo se tendrá en cuenta el QTL localizado en el GL3.

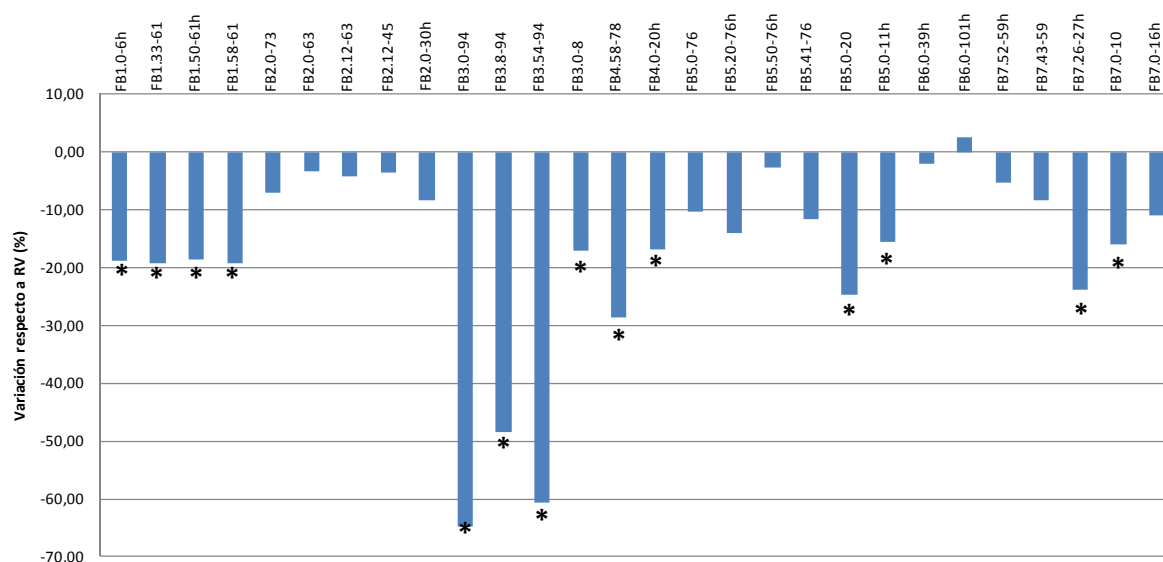


Figura 44. Variación detectada para la longitud de los folíolos de las líneas de introgresión analizadas. Porcentaje de variación observado respecto al valor control (RV) en 28 líneas. Las diferencias estadísticamente significativas se señalan con un asterisco (*).



Figura 45. Imágenes de las hojas de algunas líneas de introgresión. Imagen de las hojas de las líneas que presentan diferencias significativas para el tamaño de los folíolos frente al control, 'Reine des Vallées'. De izquierda a derecha, individuos FB3.8-94, FB3.54-94, FB3.0-94 y, a la derecha, 'Reine des Vallées'.

3.9.4 TAMAÑO DEL FRUTO

Durante los meses de julio, agosto y septiembre de 2010 se tomaron medidas de los frutos desarrollados por algunas de las líneas de introgresión obtenidas. Las líneas FB2.12-45, FB3.0-94, FB3.0-8, FB3.8-94, FB3.54-94, FB1.58-61, FB1.0-6h, FB2.12-63, FB4.58-78, FB4.0-58, FB5.0-76, FB5.20-76h, FB6.0-39h y FB6.0-101H no fueron evaluadas, bien porque no desarrollaron frutos o porque el número de frutos desarrollados fue insuficiente para realizar el estudio. La longitud media de los frutos obtenidos fue de 21,77 mm, su anchura media fue de 12 mm y el ratio medio L/A (largo/anchura) fue de 1,82 mm. Las dimensiones de los frutos del control, RV, fueron 25,6 mm de longitud, 13,20 de anchura y su ratio medio L/A fue de 1,96. Las figuras 46 y 48 muestran los valores alcanzados por las líneas analizadas.

3.9.4.1 LONGITUD DEL FRUTO

Nueve líneas presentaron diferencias significativas para la longitud del fruto. Todas ellas presentaron valores de longitud de fruto menores que el control. Las diferencias detectadas van del 14 al 41% y son distintas con un nivel de significación $\alpha = 0,01$. El número estimado de QTLs para este carácter fue de 3 ya que, según los criterios fijados para la estimación del número de QTLs, se ha considerado que cuando dos líneas que contienen regiones introgresadas solapantes presentan efectos significativos, el QTL se localiza en la región solapante.

Las líneas FB1.58-61 y FB1.50-61h, con introgresiones solapantes en el GL1, entre las posiciones 33,4 y 61,2 cM, presentaron diferencias significativas. Sus frutos fueron un 27 y un 25% menor, respectivamente, respecto al control.

Se han detectado diferencias significativas en todas las líneas con introgresiones en el GL2 analizadas. La región común en todas ellas está comprendida entre los 0 y los 39,2 cM del grupo. Las diferencias encontradas aquí son las mayores registradas en toda la población. La línea FB2.0-73 presentó frutos cuya longitud difiere un 34,1% respecto al control. La diferencia entre FB2.0-63 y el control fue del 41%. La línea FB2.0-30h desarrolló frutos cuya longitud difiere en un 20,3% del valor control y la línea FB2.30 mostró un valor medio de longitud inferior en un 36% al valor control.

Las líneas FB5.50-76h y FB5.41-76, con introgresiones solapantes en una región de 34,8 cM del GL5, presentaron valores inferiores en un 14,1% respecto al valor control.

La línea FB7.0-10h, que posee una pequeña región del GL7 introgresado en homocigosis presentó valores inferiores a la media en un 23,8%. Sin embargo, no se aprecian efectos significativos en la línea FB7.16h, cuya introgresión solapa con la de FB7.10.

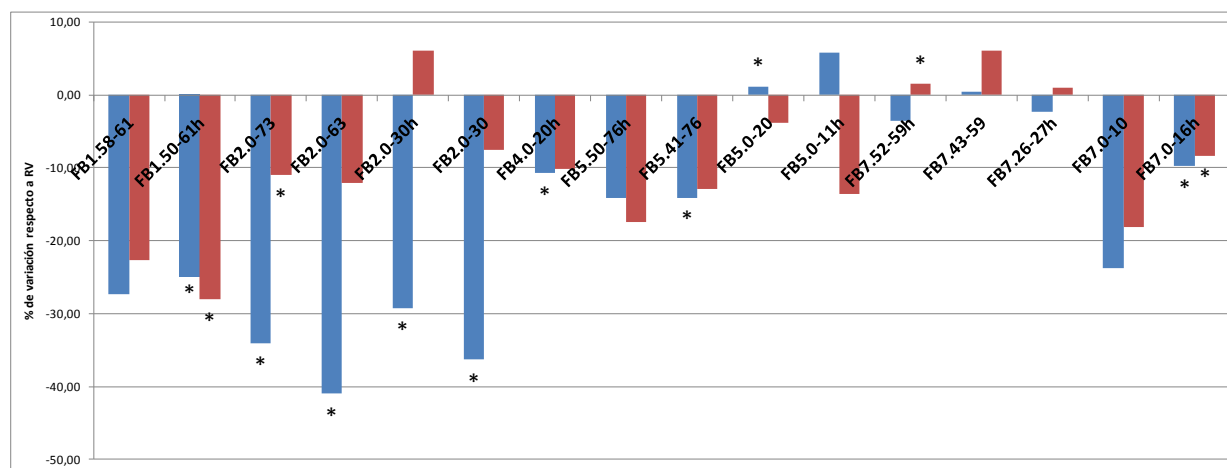


Figura 46. Datos obtenidos para el análisis de las dimensiones longitudinal y transversal de los frutos analizados. Porcentaje de variación observado respecto al valor control (RV) para la longitud (L, en azul) y la anchura (A, en rojo) en 16 líneas. Las diferencias estadísticamente significativas se señalan con un asterisco (*).

3.9.4.2 ANCHURA DEL FRUTO

Cuatro líneas presentaron diferencias significativas en la anchura del fruto respecto al control. Todas ellas presentan valores significativamente inferiores al control ($\alpha = 0,01$). FB1.58-61 y FB1.50-61h, poseen introgresiones solapantes en el GL1, entre las posiciones 33,4 y 61,2 cM. Esta región coincide con la posición de uno de los QTLs detectados para la longitud del fruto. Estas líneas presentaron valores medios de anchura de sus frutos un 22,7 y un 28% inferior, respectivamente, al control.

Se han detectado diferencias significativas en los frutos de las líneas FB5.50-76h y FB7.0-10. Sin embargo, no se han detectado efectos significativos en líneas con introgresiones comunes a las de estas líneas.

3.9.5 FORMA DEL FRUTO

Cinco líneas presentaron diferencias significativas para la forma del fruto, estimada como el ratio entre el valor L y el valor A, para un nivel de significación $\alpha = 0,01$. Todas ellas presentan frutos de aspecto redondeado. Las líneas FB2.0-73, FB2.0-63, FB2.0-30h y FB2.0-30 presentaron valores inferiores al control en un 27%, 33,3%, 33,3% y 31,3% respectivamente. La región común en todas ellas, y que albergaría este QTL para la forma del fruto, está comprendida entre los 0 y los 39,2 cM del grupo y coincide con uno de los QTLs de longitud del fruto detectados (fig. 47).

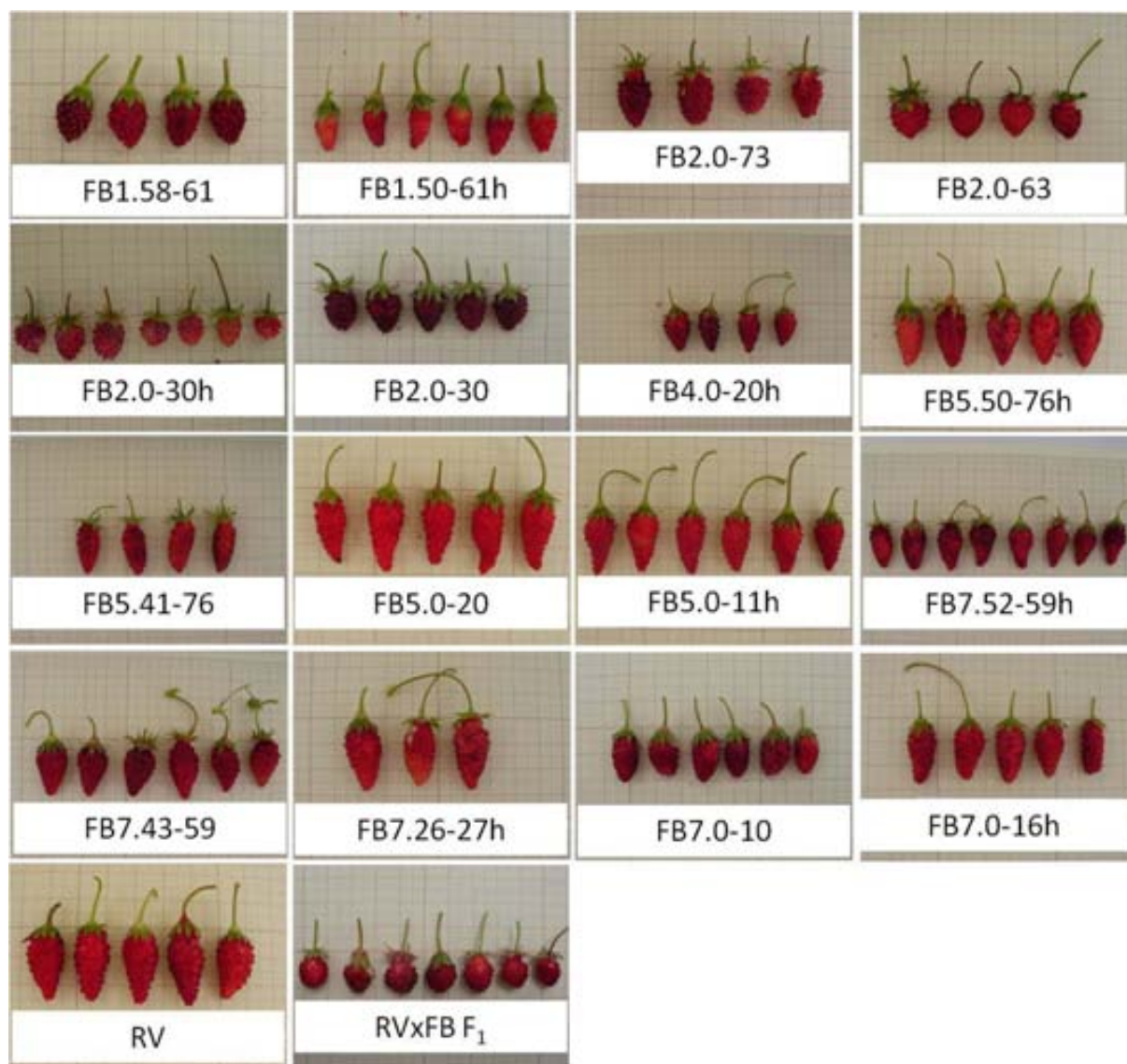


Figura 47. Imágenes de los frutos de las líneas evaluadas. Fotografías de algunos de los frutos empleados en la evaluación de los caracteres de fruto en esta investigación. En la parte de abajo de la imagen se muestran frutos del parental recurrente, RV, y del individuo F_1 empleado para generar la población de líneas de introgresión. Las líneas FB1.18-61, FB1.50-61h, FB2.0-73, FB2.0-63, FB2.0-30h, FB5.50-76h, FB5.41-76 y FB7.0-10 presentan frutos cuya dimensión principal es significativamente más pequeña que RV. Las líneas FB1.58-61, FB1,50-61h, FB5.0-11h y FB7.0-10 presentan frutos cuya anchura es significativamente menor a la de los frutos de RV. Las líneas FB2.0-73, FB2,0-63, FB2.0-30h y FB2.0-30 presentan frutos cuya forma es significativamente más redondeada que RV, mientras que FB5.0-11h presenta frutos significativamente más alargados que el control RV.

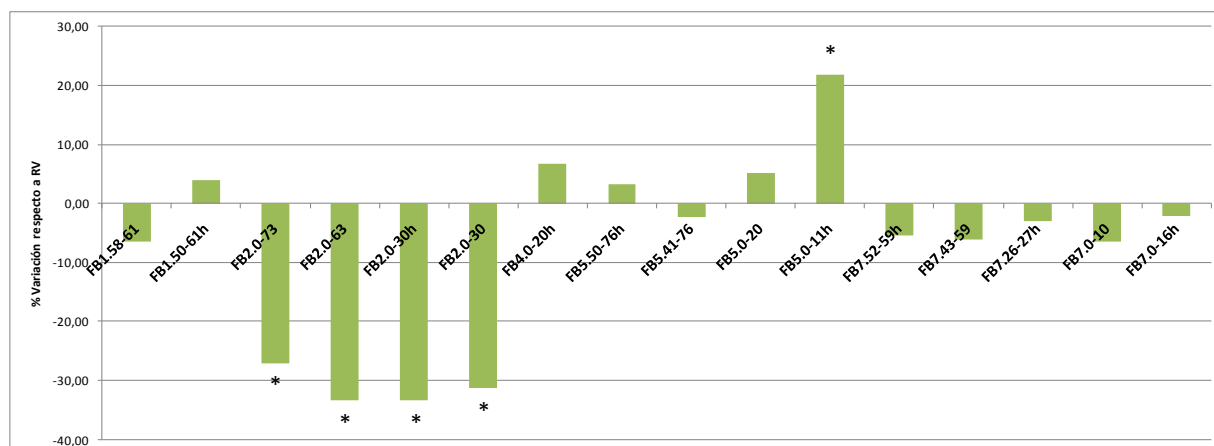


Figura 48. Datos obtenidos para el análisis de la forma de los frutos analizados. Porcentaje de variación observado respecto al valor control (RV) para el ratio L/A en 16 líneas. Las diferencias estadísticamente significativas se señalan con un asterisco (*).

3.9.6 EVALUACIÓN DE OTROS CARACTERES

La resistencia al virus del retorcimiento de la hoja (SCV) fueron examinados en la colección en diferentes momentos durante el verano de 2010. En función de la gravedad de los síntomas, a cada planta se le asignó una puntuación siguiendo una escala del 0 al 3, donde 0 representa la ausencia de síntomas y 3 representa el máximo grado de afección por el virus (fig. 49). Tal y como sucedió con las mediciones de síntomas de oidio, las líneas FB1.0-6, FB2.12-63, FB3.0-94, FB3.8-94, FB3.54-94, FB4.58-78 y FB4.0-58 no fueron evaluadas. El resultado de las mediciones figura en la tabla 26. Los resultados no muestran un patrón claro de resistencia al virus, si bien las líneas FB1.33-61, FB2.0-30, FB4.58-78, FB5.0-20, 5.0-11h, 7.52-59h y 7.0-10 presentaron síntomas leves.

Tabla 26. Caracterización de la colección de líneas de introgresión para los síntomas de infección por SCV. Grupo de ligamiento introgresado, nombre de la línea, y gravedad de la afección en diferentes mediciones efectuadas durante el verano de 2010. Las líneas señaladas en gris no fueron evaluadas

LÍNEA	28 jul	17 ago	30 ago	8 sep	23 sep	FENOTIPO	LÍNEA	28 jul	17 ago	30 ago	8 sep	23 sep	FENOTIPO
FB1.0-6h	-	-	-	-	0	-	FB4.0-20h	0	0	0	1	0	tolerante
FB1.33-61	0	1	1	1	0	tolerante	FB4.0-58	-	-	-	-	-	-
FB1.50-61h	0	1	1	2	1	tolerante	FB5.0-76	3	0	2	1	2	tolerante
FB1.58-61	0	2	2	2	0	tolerante	FB5.20-76h	1	0	3	3	3	tolerante
FB2.0-73	1	0	3	1	1	tolerante	FB5.50-76h	2	1	1	3	2	tolerante
FB2.0-63	1	0	1	2	0	tolerante	FB5.41-76	0	0	1	1	2	tolerante
FB2.12-63	-	-	-	-	-	-	FB5.0-20	0	0	1	1	1	tolerante
FB2.12-45	1	1	2	2	1	tolerante	FB5.0-11h	0	1	0	1	1	tolerante
FB2.0-30h	1	0	2	2	2	tolerante	FB6.0-39h	0	1	1	2	1	tolerante
FB2.0-30	1	0	0	1	1	tolerante	FB6.0-101h	0	1	1	2	0	tolerante
FB3.0-94	-	-	-	-	-	-	FB7.52-59h	0	0	1	1	1	tolerante
FB3.8-94	-	-	-	-	0	-	FB7.43-59	0	2	1	1	1	tolerante
FB3.54-94	-	-	-	-	0	-	FB7.26-27h	0	2	1	1	1	tolerante
FB3.0-8	0	3	0	0	0	tolerante	FB7.0-10	0	0	1	1	0	tolerante
FB4.58-78	-	1	-	1	1	-	FB7.0-16h	0	2	2	1	1	tolerante



Figura 49. Síntomas de la infección por SCV. Imágenes de la línea FB5.20h (480.029.072) presentando graves síntomas de la infección por el virus del retorcimiento de la hoja (SCV).

4. DISCUSIÓN

4.1 ESTUDIO GENÉTICO DE LAS POBLACIONES GENERADAS

La distribución de las introgresiones a lo largo del genoma de *Fragaria* diploide en las poblaciones *backcross* y en las familias BC2 ⊗ generadas durante el desarrollo de las líneas de introgresión no responde a un patrón homogéneo, tal como se esperaría para poblaciones de este tipo en ausencia de presión selectiva. Se han detectado fuertes niveles de distorsión para la segregación de marcadores en las poblaciones BC1 y en las familias BC2 ⊗ (tablas 20 y 22). La distorsión más intensa se detectó en los GLs 2, 5 6 y 7, con prevalencia de los genotipos BB y AB para los grupos 2, 5 y 6, y de los genotipos AA y AB para el GL7. Parece pues, que no existe una preferencia general para los alelos de uno de los parentales. Este fenómeno contradice a los observados en otros estudios de desarrollo de colecciones completas de NILs en tomate (Monforte y Tanksley 2000a; Chetelat y Meglic, 2000), donde se detecta un déficit de los alelos de la especie silvestre. Puesto que en esta investigación no se ha ejercido ninguna presión selectiva intencionadamente, y dado que en la generación de las líneas se han empleado diferentes familias BC2 en las que los patrones de distorsión coinciden, y que en su mayoría coinciden también con los patrones de distorsión encontrados en la población F₂ del mapa de referencia FV x FB, la frecuencia diferencial de ciertos alelos sobre otros podría deberse al efecto de la selección sobre caracteres de eficacia reproductiva como la viabilidad gamética o zigótica (Knox y Ellis 2002). Otros trabajos realizados sobre poblaciones similares y creadas con el mismo fin (generación de NILs), procedentes de cruzamientos interespecíficos de lechuga y tomate (Jeuken y Lindout 2004; Canady *et al.* 2005), constataron la prevalencia de alelos del parental recurrente sobre los del donante. Ambos equipos consiguieron revertir la distorsión en algunas de estas regiones al invertir el orden del cruzamiento, es decir, cuando emplearon como donante de polen al parental que había actuado como receptor en los cruzamientos con segregación no mendeliana. En nuestro caso, la población BC1 se obtuvo empleando polen del parental donante y, en la generación de las poblaciones BC2, se empleó como donante de polen al parental recurrente. Las poblaciones BC2 generadas en esta investigación fueron de tamaño poblacional pequeño (45 individuos en promedio), si bien en ellas se observa una fuerte reducción de los valores de distorsión respecto a la población BC1 y las familias BC2 ⊗, con valores muy próximos a los esperados para una segregación mendeliana, a excepción del extremo inferior del GL5, que sigue apareciendo muy distorsionado. El marcador EMFvi099 que mostró un ratio de segregación de 43:57 en la familia BC2 480, obtuvo un ratio de 18:82 en la población BC1 y de 1:51:48 en la familia 480.21 ⊗. El hecho de que la selección de ciertos genotipos sea unidireccional

en la mayoría de los casos apoya la hipótesis del *fitness* reproductivo como generador de presión selectiva (Jeuken y Lindhout. 2004).

La inviabilidad de ciertos genotipos, como los observados en la familia 460,014 ♂ (fig. 39) apuntan a factores polínicos de autoincompatibilidad que actúan en homocigosis recesiva para los alelos *F. bucharica*, que es una especie autoincompatible. Recientemente se ha publicado la existencia de dos locus de incompatibilidad, *S* y *T*, en los grupos 1 y 6 de *Fragaria* diploide. Ambos genes codifican para RNAsas específicas de estilo (Bošković *et al.* 2010). *F. nubicola*, presenta autoincompatibilidad cuando los loci *S* y *T* están en un fondo genético apropiado, incluso cuando sólo uno de estos factores es expresado. Este fondo genético incluye un factor adicional denominado *M*, que provoca autoincompatibilidad cuando se presenta en heterocigosis y se expresa en los estilos. Existe, además, un factor polínico, *SFB*, cuya posición no ha sido determinada y que podría estar jugando un papel clave en la viabilidad de los cigotos. Se da la circunstancia que, en la población FV x FB, lejos de encontrar un comportamiento similar al encontrado aquí, existe un exceso de alelos *F. bucharica* frente a los alelos *F. vesca* para los loci situados en el GL7. Este fenómeno se explicaría por una diferente composición alélica para este locus en los individuos F₁ que originaron ambas poblaciones. Serían necesarios análisis más profundos para establecer el origen de este fenómeno.

4.2 ESTRUCTURA GENÓMICA DE LA COLECCIÓN DE LÍNEAS DE INTROGRESIÓN.

La colección de líneas de introgresión desarrollada en esta Tesis Doctoral está formada por 30 individuos que cubren el 97,4% del genoma de *F. bucharica* en forma de introgresiones solapantes de un tamaño medio de 43 cM, sobre un fondo genético procedente del parental *F. vesca*. Las poblaciones de mapeo son un recurso básico para elucidar las bases genéticas de los caracteres cuantitativos. En esta investigación hemos desarrollado la primera población de líneas de introgresión de la familia *Rosaceae*, que permitirá la identificación y el estudio de las regiones del genoma que más influyen en los distintos parámetros de calidad del fruto y de la planta. Aunque deficiente en algunos aspectos –parental recurrente con loci segregantes y cobertura incompleta del genoma– esta colección es sin duda la mejor herramienta genómica generada hasta la fecha para el estudio del germoplasma de *Fragaria*. Ha quedado sin cubrir una pequeña región de aproximadamente 16 cM en el GL1, pero se dispone de líneas BC1 con introgresiones en la zona para la futura generación de NILs cubriendo la zona. La futura autofecundación y M.A.S. de recombinantes a partir de las líneas que componen la colección permitirá crear una segunda colección de líneas de introgresión, formada exclusivamente por NILs.

Las líneas seleccionadas por mantener introgresiones en el GL7 presentan un mayor porcentaje de heterocigosis que el resto; La muerte de uno de los individuos seleccionados y la búsqueda de introgresiones homocigóticas que rebasaran la posición 26 cM (marcador EMFvi008) provocó que se desestimaran genotipos homocigóticos que bien podrían haber formado parte de la colección final. Los 38 *bins* definidos por las introgresiones de esta población proporcionan una resolución media de 16,9 cM para el mapeo de nuevos marcadores. Aun siendo bajo, (el *bin set* de *Fragaria*, formado por sólo 6 plantas, genera *bins* con una resolución media de 6 cM) este nivel de resolución es suficiente para mapear QTLs con mayor precisión que con poblaciones clásicas (Tanksley 1993; Hyne *et al.* 1995; Mackay, 2001) para después diseccionarlos genéticamente mediante el desarrollo de sub-NILs por generación de nuevos recombinantes. Disponemos de líneas con introgresiones heterocigóticas para todos los *bins* que permitirán el estudio de estos QTLs en una sólo generación, sin necesidad de recurrir a poblaciones F₂, consiguiendo un considerable ahorro de tiempo. Además, mediante la caracterización fenotípica de un subconjunto de sólo 12 líneas podemos localizar rápidamente en qué grupo(s) de ligamiento se encuentran los QTLs más interesantes para el carácter que se desea mejorar. Este grupo de líneas contiene 12 introgresiones principales y 3 contaminantes y cubre la misma distancia genética que la colección completa. El fenotipado de estas pocas líneas puede ser muy útil cuando, por motivos diversos, no sea posible analizar la población completa.

Hemos detectado heterocigosis en el parental recurrente para un 14,7% de los loci analizados. Este fenómeno podría provocar que diferencias fenotípicas entre las distintas NILs debidas a la segregación en regiones del fondo genético sean confundidas con las diferencias causadas por las introgresiones. Las variedades comerciales son, por lo general, heterocigóticas para la mayoría de sus loci, ya que la heterosis suele ser, en gran medida, la responsable del alto valor comercial de la variedad en cuestión. El hecho de que la variedad 'Reine des Vallées' se haya reproducido recurrentemente mediante semillas desde su obtención, anterior a 1963 (Darrow, 1966), garantiza que los alelos segregantes no condicionan la calidad del fruto. La generación de sub-NILs segregantes para cualquier QTL resolvería los conflictos generados por la segregación del fondo genético.

4.3 DISEÑO EXPERIMENTAL

Si comparamos el modo de obtención de esta colección con los empleados por otros autores, encontramos varias diferencias significativas. En el desarrollo de una colección de NILs en melón Eduardo *et al.* (2005) emplearon como donantes de introgresiones un grupo de 8 LDHs. A pesar de partir de una colección de LDHs, los autores debieron llegar hasta la sexta generación para obtener

NILs. Pestsova *et al.* (2001), que partían de un set de líneas de sustitución de trigo, necesitaron entre 4 y 5 generaciones. Eshed y Zamir (1994) emplearon, como nosotros, una F_1 , que retrocruzaron hasta en 12 ocasiones. En la obtención de NILs de lechuga Jeuken y Lindhout (2004) partían de una F_1 y necesitaron llegar hasta la quinta generación de retrocruzamiento. Existen varios factores que determinan el número de generaciones necesario para aislar introgresiones entre las que destacan el tamaño del genoma (y más concretamente el número de cromosomas de la especie) y la resolución de los mapas genéticos disponibles.

La colección de líneas de introgresión presentada aquí ha sido obtenida en sólo dos generaciones de retrocruzamiento y una autofecundación, gracias a una temprana selección asistida por marcadores en una amplia población BC1. El hecho de que el número básico de cromosomas de *Fragaria* diploide sea bajo (7) posibilitó realizar una selección inicial con sólo 14 marcadores moleculares localizados en los extremos de los cromosomas, descartando así más del 70% de los individuos BC1. La caracterización progresiva de los individuos seleccionados permitió ir descartando individuos hasta reducir la población de trabajo hasta 8 individuos con un máximo de 4 introgresiones por planta que cubren la totalidad del genoma. A pesar de que se obtuvieron familias BC2 de tamaño poblacional menor al esperado, el número de introgresiones en los individuos seleccionados disminuyó un 55,3% de media, pasando de 4 introgresiones en la línea 473 a 1 en la línea 473,061, por ejemplo. Esta importante reducción permitió obtener NILs con introgresiones en heterocigosis en sólo dos generaciones, y líneas de introgresión que se depuraron en una fase posterior de autofecundación. En especies con genomas pequeños como es el caso de *Fragaria* diploide, o con tiempos de generación largos, como es el caso de los árboles, es beneficioso hacer una selección temprana sobre poblaciones grandes (Eduardo *et al.* 2005; Picañol *et al.* 2007). En especies con tamaños de genoma grandes, como es el caso de la mayoría de los cereales o la lechuga y tiempos de generación cortos, parece más apropiado realizar varias generaciones de retrocruzamiento antes de empezar a hacer M.A.S. para favorecer la fijación de las regiones segregantes. Sin embargo, cuando se analiza el genoma de una población que ha pasado por varios ciclos de retrocruzamiento, es muy probable que el poder de detección de introgresiones disminuya, ya que éstas disminuyen su tamaño a medida que recombinan durante las sucesivas meiosis. Sería positivo realizar, en este tipo de especies, una caracterización genómica inicial en la que se identifique el número y posición de las introgresiones segregantes en las líneas iniciales para poder comprobar, al final del proceso de selección, que no hay rastro de introgresiones contaminantes.

4.4 CARACTERIZACIÓN PRELIMINAR DE LA COLECCIÓN DE LÍNEAS DE INTROGRESIÓN

Durante la construcción de la colección de líneas de introgresión se ha podido identificar una serie de caracteres cualitativos agronómicamente interesantes cuya posición genómica era conocida, como el desarrollo de estolones. La obtención de la línea FB2.12-45, una NIL que cubre una región de 33 cM en la que se encuentra el locus *r*, que permite a la planta desarrollar estolones, supondrá, en caso de que el resto de características agronómicas y organolépticas de la variedad se mantengan inalteradas, una mejora directa sobre la calidad de la planta, que podrá ser propagada vegetativamente, ahorrando tiempo y dinero en la reproducción de un cultivo que, pese a ser perenne, se renueva cada año para preservar el vigor de la variedad y la salud de la planta. En caso de que las características agronómicas de la línea difieran de las de la variedad comercial, el desarrollo de sub-NILs con introgresiones de menor tamaño debería solventar el problema.

La línea empleada como parental recurrente en este trabajo, 'Reine des Vallées' una variedad de floración remontante. En la colección de líneas de introgresión desarrollada aquí hemos incluido dos individuos FB6.0-39h Y FB6.0-101h, con introgresiones en el GL6, que manifiestan un hábito de floración distinto al de 'Reine des Vallées'. Dado que en la fecha actual, cuando han transcurrido más de 400 días desde su siembra, estas líneas no han florecido, incluso tras someterse a tratamientos de inducción floral, estamos en condiciones de afirmar que alguno de los factores genéticos que gobiernan el ciclo de floración en *Fragaria* diploide están localizados en una región situada entre las posiciones 0 y 54,3 cM del GL6. El requerimiento de frío para la floración tienen efectos dominantes sobre el genotipo *F. vesca*. Sargent *et al.* (2004a) localizaron el gen *s*, que proporciona el carácter *semperflorens* la accesión FDP815 de *F. vesca*, en la posición 29.6cM del GL6 en la primera versión del mapa de referencia, pero este marcador no ha vuelto a ser incluido en ninguna de las versiones posteriores de dicho mapa y tampoco ha aparecido ligado a ninguno de los loci analizados en esta Tesis Doctoral, indicando que la neutralidad respecto al fotoperiodo/termoperiodo son caracteres cuantitativos, como sucede en el fotoperiodo de las especies octoploides (Weebadde *et al.* 2008). La figura 50 muestra la posición de los QTLs localizados en esta caracterización preliminar de la colección de líneas de introgresión.

Se han detectado dos tipos de anomalías en la floración de las líneas con introgresiones en la región inferior del GL5. La línea FB5.20-76h nunca desarrolló flores. No hemos observado que las introgresiones heterocigóticas en el GL5 detectadas en las líneas BC1 y BC2 estuvieran asociadas a requerimientos de vernalización para el desarrollo de flores. Las líneas FB5.0-76, FB5.50-76 y FB5.41-

76, con introgresiones homocigóticas comunes entre las posiciones 37 y 76 cM del GL5, desarrollaron flores 18,5 días más tarde que la media en promedio y detuvieron el desarrollo de flores cuando los días comenzaron a acortarse. Por tanto, parece claro que esta región alberga otro QTL relacionado con el periodo de floración en *Fragaria* diploide. La floración tardía ha sido estudiada en otras especies de la familia Rosaceae como el almendro (Ballester *et al.* 2001; Sánchez Pérez *et al.* 2007) el melocotón (Dirlewanger *et al.* 1999), la manzana (Lawson *et al.* 1995). Ninguno de los 5 QTLs descritos para este carácter en el genoma de *Prunus* presenta sintenia con el GL5 de *Fragaria*. Sin embargo, otros estudios recientes llevados a cabo en *F. x ananassa* han identificado un QTL para el tiempo de floración situado en uno de los grupos de ligamiento homólogos al GL5 de *Fragaria* diploide (Zorrilla-Fontanesi *et al.* 2010b).

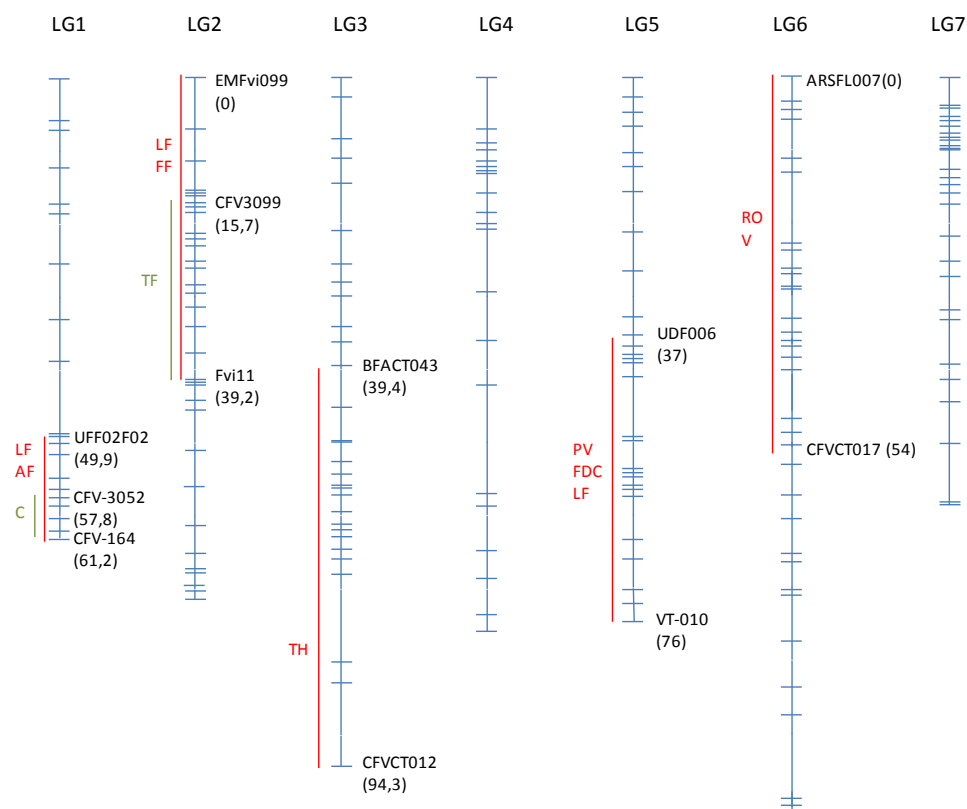


Figura 50. Posición genómica de los caracteres localizados en esta investigación. La posición de los QTLs y caracteres cualitativos localizados aparece señalada junto a los grupos de ligamiento del genoma de *Fragaria* diploide. Las posiciones están definidas por los marcadores moleculares flanqueantes y sus posiciones, en centiMorgans (cM). TF: Tallos florales engrosados, V: Vernalización, FDC: Floración en día corto, C: Clorosis de la hoja, PV: Periodo vegetativo, RO: Resistencia a oidio, TH: Tamaño de las hojas, LF: Longitud del fruto, AF: Anchura del fruto, FF: Forma del fruto.

Tanto en el GL5 como en el GL6, se constata la necesidad de aumentar el número de líneas con introgresiones en la zona para evaluar los efectos de ambos QTLs. La disponibilidad de un sólo individuo por cada genotipo imposibilita el análisis de tipo ANOVA, necesario para el análisis de caracteres cuantitativos en poblaciones de este tipo. Sin embargo, la obtención de nuevos individuos en la siguiente generación permitirá acotar más la posición de los factores genéticos que intervienen en la fenología de la floración. La constatación de este fenómeno sentaría las bases para el estudio de los los gene(s) que participan en este carácter y sus interacciones, que supondría un avance en el conocimiento de la formación del fruto de fresa diploide.

En este estudio hemos localizado una región genómica de 55 cM responsable de la resistencia al oidio en *Fragaria* diploide. El oidio es una patología causada por varios hongos pertenecientes al orden *Erysiphales*. La resistencia a esta enfermedad ha sido ampliamente estudiada en la familia *Solanaceae* especialmente en tomate, donde se han descrito las respuestas mono y poligénicas que intervienen a la respuesta a la infección por *Oidium neolycopersici*, causada por los genes *OI-1*, *OI-2* y *OI-4* (Bai *et al.* 2004; Li *et al.* 2006, 2007). El gen *OI-4*, localizado en el cromosoma 6 del genoma del tomate, proporciona una completa resistencia al hongo que se hereda de manera dominante. Los mecanismos moleculares responsables de la resistencia a oidio en fresa octoploide han sido estudiados *via* transformación genética con los genes *pcht28* y *ch5B*, que codifican para quitinasas de judía y tomate respectivamente. La expresión de estos genes en fresa aumenta la resistencia a los hongos *Verticillium dahliae* y *Botrytis cinerea* (Chalavi *et al.* 2003; Ricardo *et al.* 2006), si bien estos transgenes no han sido evaluados para la resistencia a *Oidium fragariae* y *Spaerotheca masularis*, que son los mayores agentes causantes de la enfermedad en *Fragaria*. Un estudio funcional de los genes candidatos para este carácter localizados en la región superior del GL6 podría esclarecer la base molecular y genética de la resistencia a esta enfermedad, con objeto de reducir el uso de productos fitosanitarios en los cultivos de fresa.

El tamaño de las hojas, define la capacidad fotosintética de una planta y tiene un impacto importante en muchos caracteres agronómicos interesantes como el rendimiento, la calidad del fruto, la resistencia a patologías y respuestas al estrés. Hemos encontrado un QTL para el tamaño de los foliolos en la región inferior del GL3. Las líneas con introgresiones en esta zona presentan foliolos cuya dimensión principal es, en promedio, un 57,9% menor a la de los foliolos de *F. vesca*. Esta reducción del tamaño de los foliolos va acompañada de una reducción importante en el tamaño de los peciolos, por lo que podemos concluir que este QTL determina el tamaño de las hojas. Las bases genéticas que determinan el tamaño de las hojas ha sido ampliamente estudiado en *A. thaliana* y

otras brasicáceas, así como en especies cultivadas de la familia *Fabaceae*, *Solanaceae*, y *Poaceae* (Pérez-Pérez *et al.* 2010). Sin embargo, no existen estudios detallados acerca de este fenómeno en ninguna especie de la familia *Rosaceae*. Durante el desarrollo de una colección de NILs en *Prunus* se han detectado dos QTLs para la forma de la hoja (Picañol *et al.* 2007). Las hojas de las líneas con introgresiones de 'Texas' en regiones de los grupos GL1, GL2 y GL5 del mapa de referencia de *Prunus* son significativamente más anchas que las del parental recurrente, 'Earlygold', proporcionando una mayor superficie fotosintética. Ninguno de estos grupos de ligamiento es sinténico con el GL3 de *Fragaria* (Vilanova *et al.* 2008).

En muchas especies vegetales, la domesticación ha resultado en una arquitectura de la planta fuertemente modificada, especialmente en lo referido a la morfología de los frutos. Forma y tamaño del fruto fueron los rasgos sometidos a una mayor selección durante la domesticación. El control genético de estos caracteres ha sido estudiado extensivamente en tomate, posibilitando el clonaje posicional de dos QTLs con efectos mayores para el peso del fruto (*fw2.2*) y la forma (*ovate*) (Frary *et al.* 2000; Liu *et al.* 2002) que han sido transferidos a otras especies de solanáceas, e incluso se han definido varios descriptores matemáticos para algunos atributos de la forma del fruto (van der Knapp *et al.* 2002, van der Knaap y Tanksley 2003). Los estudios acerca de las bases genéticas de este fenómeno en la familia *Rosaceae* son escasos. Zhang *et al.* (2006) demostraron que la evolución del tamaño del fruto en *Pyrus pirifolia* durante la domesticación procede de cambios en la capacidad de dividirse de las células después de la polinización. Se han descrito numerosos QTLs para el tamaño del fruto en el género *Prunus* (Dirlewanget *et al.* 1999; Etienne *et al.* 2002b; Quilot *et al.* 2004; Zhang *et al.* 2010) En este trabajo hemos localizado 3 QTLs que intervienen en la longitud de los frutos de fresa, 1 QTL con repercusión en la anchura del fruto y otro con efectos en la forma. El QTL localizado en el GL1 se corresponde con una región sinténica con el grupo de ligamiento 6 de *Prunus*, donde se localiza uno de los QTLs más consistentes para el tamaño del fruto en *Prunus* (Zhang *et al.* 2010). La confirmación de la existencia de estos QTLs en futuros estudios podría suponer un aliciente para el estudio de caracteres cuantitativos de morfología del fruto de fresa y de otras especies de la familia *Rosaceae*.

DISCUSIÓN GENERAL

Dado que los resultados obtenidos en esta investigación han sido sujetos a discusión en las distintas secciones de esta Tesis, la discusión general se centrará en señalar las principales aportaciones de este trabajo al desarrollo de la investigación genética y genómica en el género *Fragaria* y, por extensión, en la familia *Rosaceae*, desde una perspectiva histórica y con un pronóstico para el futuro más próximo.

1. GENÉTICA DE LAS ESPECIES DIPLOIDES DE *FRAGARIA*

La genética de *Fragaria* diploide fue estudiada por primera vez hace casi un siglo por Richardson (1914, 1918, 1920, 1923), que fue seguido por Jones (1955) y después por Brown y Wareing (1965). Estos últimos demostraron que tres caracteres monogénicos recesivos en *F. vesca*, (desarrollo de estolones (*r*), floración remontante (*s*) y color del fruto (*c*)) seguían una segregación mendeliana independiente. Treinta años después, gracias al desarrollo de los primeros estudios de genética molecular en *Fragaria* llevados a cabo en la Universidad de New Hampshire a mediados de los años 90 (Williamson *et al.* 1995; Yu y Davis, 1995), se demostró la existencia de ligamiento entre marcadores moleculares y los caracteres descritos por Brown y Wareing (1965). En poco tiempo apareció el primer mapa de ligamiento para el género, donde se definieron por primera vez los 7 grupos de ligamiento (Davis y Yu, 1997), que constituye la base del desarrollo de la genética molecular y genómica estructural de *Fragaria* que conocemos hoy. En su trabajo sobre del control genético y ambiental de la floración en fresa, Battey *et al.* (1998) describieron los patrones de floración en *Fragaria* a través de una aproximación basada en el ligamiento genético y propusieron *F. vesca* como organismo modelo para la investigación en plantas perennes. Basándose en el trabajo de Davis y Yu (1997), Deng y Davis (2001) comprobaron el potencial del mapa genético de *Fragaria* diploide para el aislamiento de genes, describiendo la cosegregación del locus *c* con el gen *F3H*, implicado en la ruta de biosíntesis de las antocianinas. Con la publicación del primer mapa de microsatélites en *Fragaria* diploide (Sargent *et al.* 2004a), los trabajos en fresa diploide se multiplicaron y el equipo dirigido por el Dr. Simpson en *East Malling Research Center* (EMR) pasó a liderar las investigaciones en genética molecular en *Fragaria* diploide. Las colaboraciones internacionales entre este grupo y otros grupos franceses (INRA), italianos (Universidad de Udina), estadounidenses (Universidad de Florida y USDA) y el Institut de Recerca i Tecnologia Agroalimentàries (IRTA) de Barcelona permitieron la ampliación del mapa genético, la incorporación de marcadores funcionales y la transferabilidad de marcadores a otras especies de rosáceas (Sargent *et al.* 2006, 2007).

El equipo dirigido por los doctores Pere Arús y Amparo Monfort publicó su primer trabajo sobre el desarrollo de un set de marcadores moleculares SSR en *Fragaria* diploide, desarrollado fundamentalmente en el IRTA (Monfort *et al.* 2006). Tras colaborar en varias ocasiones con el Dr. Sargent en la ampliación del mapa diploide y el desarrollo de estrategias de mapeo selectivo, el equipo del IRTA desarrolló un amplio trabajo en el que se describen las relaciones de sintenia y colinearidad entre los géneros *Prunus* y *Fragaria* (Vilanova *et al.* 2008).

En 2008 aparecen los primeros estudios comparativos entre genomas de diferente nivel de ploidía del género *Fragaria* empleando marcadores de los mapas diploide y octoploide de (Rousseau-Gueutin *et al.* 2008, Sargent *et al.* 2009), mostrando que ambos genomas son esencialmente sinténicos, abriendo las puertas a la transferabilidad de estudios entre ambos niveles. Este fenómeno proporcionó las bases para la creación del objetivo principal de este trabajo de Tesis Doctoral, que no es otro que la búsqueda de evidencias sobre la localización de los factores genéticos que determinan las características más interesantes del fruto de fresa de cara a la mejora genética mediante el desarrollo de una colección de NILs cubriendo la totalidad del genoma de *F. vesca*.

La conversión de marcadores moleculares poco resolutivos, en su mayoría RFLPs, en nuevos marcadores de tipo PCR, de uso sencillo y de bajo coste, junto con la localización de nuevos marcadores obtenidos mediante la colaboración con otros laboratorios, ha permitido la creación de una nueva versión del mapa de ligamiento FV x FB, de alta resolución, formada exclusivamente por marcadores de tipo PCR, cuya densidad es un 25% mayor a la de la versión publicada por Sargent *et al.* (2008). Este mapa ha sido el marco de referencia para el análisis genómico desarrollado en la segunda parte de esta Tesis Doctoral, en la caracterización genómica de las poblaciones generadas durante el desarrollo de una colección de líneas de introgresión en *Fragaria* diploide. La conversión de marcadores RFLP en marcadores de tipo PCR (15 SSRs y 1 SNP) y la inclusión de estos marcadores en un nuevo mapa de ligamiento permite analizar de manera rápida la composición genómica de poblaciones segregantes de gran tamaño poblacional, evitando técnicas de hibridación, que son costosas y requieren grandes cantidades de ADN de alta calidad, además del empleo de materiales radioactivos. Además, la naturaleza de estos marcadores ha permitido mejorar la resolución de su mapeo, obteniendo un mapa con unas distancias genéticas entre marcadores más homogéneas que las presentadas por Sargent *et al.* (2006) y Vilanova *et al.* (2008). Recientemente ha sido publicada una nueva versión del mapa de referencia FV x FB (Ruiz-Rojas *et al.* 2010) que posee una resolución mayor a la del mapa presentado aquí. Este mapa cubre una distancia genética tan sólo un 2,4% mayor que el nuestro, si bien se observan algunas diferencias en el orden relativo de algunos

marcadores moleculares empleados en el desarrollo de las líneas de introgresión, localizados en el extremo inferior del GL5 (VT-010, EMFvi018, EMFv024 y EMFn010). Se da la circunstancia que ninguna de las líneas de introgresión seleccionadas para contener introgresiones en el GL5 presentan puntos de recombinación en esta región. Sin embargo, Ruiz-Rojas *et al.* (2010) prolongan la región inferior del GL5 en 25,9 cM con la adición de cuatro nuevos marcadores moleculares que no han sido evaluados en esta investigación. Sería pues necesario comprobar la isogenicidad de las líneas para estos marcadores en próximas investigaciones.

La colección de líneas de introgresión elaborada en esta Tesis será de gran utilidad en futuros estudios sobre la herencia de caracteres agronómicos y de calidad del fruto de fresa. El poder de esta herramienta es tan sólo comparable al de la colección de mutantes por inserción de T-DNA desarrollada por Ruiz-Rojas *et al.* (2010). La elección de un genoma autoincompatible como donador de introgresiones, el de la especie silvestre *F. bucharica*, que únicamente se reproduce vegetativamente, impedía conocer su comportamiento fenológico en etapas tempranas del desarrollo. Este fenómeno provocó un retraso considerable en los cruzamientos, dado el elevado requerimiento de frío para inducir la floración en esta especie. Sin embargo, gracias a esta elección hemos podido establecer un nuevo locus de auto-incompatibilidad en el género que no ha sido descrito con anterioridad. En una caracterización preliminar, hemos localizado cuatro caracteres cualitativos y ocho QTLs para importantes caracteres agronómicos. Algunos de ellos pueden ser de gran importancia en mejora genética, como los relacionados con la floración (vernalización y floración en día corto), resistencia a oidio y forma del fruto. La aproximación empleada no es más que un primer acercamiento al estudio de estos factores, que deben ser estudiados en detalle mediante la evaluación de todas las líneas de la colección, la generación de nuevas líneas más homocigóticas, con introgresiones más pequeñas, la obtención de varios individuos por genotipo y el análisis de las líneas bajo condiciones ambientales diferentes. Creemos que este tipo de estudios proporcionarán conocimientos básicos sobre la transmisión de los caracteres biológicos, agronómicos y de calidad nutricional que más interesan al sector fresero, como son las propiedades nutricionales, la resistencia a patógenos, la productividad o la forma y tamaño del fruto, que han sido desconocidos hasta ahora, por lo que confiamos en que tenga aplicaciones directas sobre la mejora genética del cultivo, tanto de la especie diploide como del híbrido octoploide, *Fragaria x ananassa*.

2. HERRAMIENTAS GENÓMICAS EN LAS ESPECIES DIPLOIDES DEL GÉNERO *FRAGARIA*

El estudio de la genómica en especies modelo, particularmente en *Arabidopsis thaliana*, ha proporcionado una gran cantidad de conocimientos, facilitando el desarrollo de nuevos recursos de análisis genético y molecular en otras especies. El desarrollo de la genómica en *Fragaria* ha estado tradicionalmente lastrado por el éxito de otra especie perenne de la familia de las rosáceas, *Prunus persica*, con un mayor arraigo en la comunidad científica debido a su condición de árbol perteneciente al género *Prunus*, de gran importancia económica, y por ello ha sido elegida como especie modelo en la genómica de rosáceas, de manera que el conocimiento genómico de este género es mucho mayor al que se tiene del género *Fragaria*. El melocotonero, como especie modelo, posee ciertas características negativas de las que *Fragaria* diploide carece, como la mala regeneración después de la transformación (Shulaev *et al.* 2008), que impide desarrollar un gran número de técnicas de genómica funcional; El género *Prunus* no es un buen candidato para realizar estudios de complementación, básicos en la demostración de la función génica, o para el desarrollo de plataformas de experimentación en genética reversa como los mutantes de inserción de T-DNA. *Fragaria*, en cambio, posee numerosas ventajas respecto a otras especies de la familia *Rosaceae*. Se trata de una planta pequeña, con un ciclo reproductivo muy rápido y un tiempo de generación que va de las 12 a las 20 semanas. Además, su genoma es pequeño (206 Mpb) (Folta y Davis, 2006), comparable en tamaño al de *A. thaliana* (157 Mpb), y responde bien a la transformación genética y a la regeneración tisular (El Mansouri *et al.* 1996; Oosumi *et al.* 2006).

A través de iniciativas como ISAFRUIT, desarrollada por la Unión Europea para la mejora de las variedades de fruta actuales, especialmente en lo referente a frutales de hueso, se están dando grandes avances en el estudio de la genómica de varias especies de la familia *Rosaceae*. Estos resultados deben ser transferidos al resto de especies de la familia en los próximos años, y para ello, las regiones sinténicas entre los diferentes géneros deben ser claramente establecidas. En los últimos años, la genómica comparada ha revelado que, en ocasiones, especies próximas comparten la posición de algunos QTLs y genes mayores, con importantes repercusiones en la mejora genética de especies cultivadas, como es el caso del tiempo de floración y la respuesta a la vernalización en dos especies de la familia *Brassicaceae* (*A. thaliana* y *Brassica rapa*), gobernados por el gen *BrFLC2* (Zhao *et al.* 2010), o la resistencia a insectos en sorgo y maíz, controlada por el mismo locus en ambas especies de gramíneas (Satish *et al.* 2009). La colocalización de genes candidatos para la calidad de fruto de melocotón en el genoma de *Fragaria* diploide, desarrollada en esta investigación podrá ser empleada para el análisis funcional de estos genes. La localización de estos genes candidatos

mediante *bin mapping* proporciona nuevos marcadores moleculares al mapa de referencia de *Fragaria* que pueden ser de gran utilidad, ya que proceden de regiones codificantes del genoma de *P. persica*. La futura colocación de algunos de estos marcadores con QTLs para la calidad del fruto de fresa mediante la evaluación de una colección de NILs puede ser crucial para la identificación y caracterización de los factores genéticos que intervienen en la fisiología y fenología de la fresa, como ya ha sucedido en otras especies con recursos genómicos similares como la lechuga (Argyris *et al.* 2010), el arroz (Shan *et al.* 2009) o el melón (Moreno *et al.* 2008).

La generación y caracterización de la única genoteca de BACs en *Fragaria vesca* de uso público, así como la creación de un sistema de pools y superpools de ADN de dicha genoteca, supone un gran avance alcanzado tras varios años de investigaciones llevadas a cabo en distintos laboratorios europeos y americanos. La publicación de este trabajo pone a disposición de la comunidad científica una plataforma para el aislamiento y caracterización de secuencias de ADN de alto peso molecular. En pleno desarrollo de las nuevas tecnologías de secuenciación de ADN, la creación de mapas físicos basados en la superposición de *contigs* obtenidos por caracterización de BACs parece una tecnología ya superada. Los genomas de *Fragaria vesca* y *Prunus persica* están siendo ensamblados y anotados actualmente, de manera que la estructura de los genes relevantes para la calidad del fruto en ambas especies dejará de ser pronto un problema para la comunidad científica. La obtención de la secuencia completa del primer genoma de una especie de *Fragaria* es, sin duda, un hito histórico para la comunidad científica. Sin embargo, la tecnología actual no permite establecer con claridad una secuencia única, sino un conjunto de 272 *scaffolds* o fragmentos no solapantes que abarcan un 87,4% de la secuencia genómica completa (Shulaev *et al.* enviado). El desarrollo de un set de *pools* de BACs que permite localizar clones que contengan cualquier gen de interés del genoma de *Fragaria* diploide, sin necesidad de recurrir técnicas de hibridación, supone un avance respecto a los recursos actuales existentes, tanto en el género *Prunus* como en *Fragaria*, ya que permite la rápida identificación de los clones de interés para establecer el mapa físico de cualquier región genómica interesante con una baja inversión de tiempo y dinero. Además, la secuenciación y anclaje de BACs que contengan secuencias no ensambladas en el borrador del genoma podría colaborar a eliminar los *gaps* que actualmente existen.

3. POTENCIAL DE LA COLECCIÓN DE NILs EN *FRAGARIA* DIPLOIDE PARA EL ESTUDIO DE CARACTERES FENOTÍPICOS

A partir del análisis preliminar de la genoteca de NILs hemos detectado alelos de *F. bucharica* que pueden ser muy útiles en la disección del control genético de algunos caracteres de calidad de fruto, tal y como se ha hecho en otras especies como el tomate (Eshed y Zamir, 1995), melón (Eduardo *et al.* 2007) o maíz (Pea *et al.* 2009). Nuestra colección, formada por 30 líneas que contienen principalmente introgresiones homocigóticas procedentes de la especie silvestre *Fragaria bucharica*, localizadas en un fondo genético homogéneo perteneciente a la variedad 'Reine des Vallées' de *Fragaria vesca* es la primera población disponible para la localización de QTLs en fresa diploide. El potencial de esta herramienta va mucho más allá, dado que sus características genéticas facilitan el estudio de las interacciones entre QTLs y entre QTLs con el medio ambiente o los efectos de la heterosis en el género (Fernandez-Silva *et al.* 2009) y abren la puerta al estudio de los distintos perfiles metabólicos que pudieran encontrarse en las diferentes líneas (Overy *et al.* 2005). Los cruzamientos entre NILs darán respuesta a la cuantificación de los efectos de las interacciones entre QTLs. Para estimar los efectos QTL x ambiente, deberían cruzarse las distintas NILs con otras accesiones de *Fragaria* diploide que proporcionen distintos entornos genéticos (Monforte *et al.* 2001, Fernández-Silva *et al.* 2010). Otra ventaja de esta colección es que se han desarrollado en una variedad comercial, por lo que algunas líneas podrían ser directamente empleadas en programas de mejora genética de fresa diploide; la obtención de líneas que se reproducen vegetativamente mediante el desarrollo de estolones es una mejora directa sobre la variedad, si bien este carácter aparece ligado a otros no deseables en todas las líneas obtenidas, como es la producción de frutos más pequeños y de forma redondeada. Probablemente, los factores genéticos que determinan ambos caracteres podrán ser separados en líneas recombinantes con introgresiones no solapantes que permitirán, además, el mapeo fino de estos caracteres la caracterización del/de los gene(s) implicados. El estudio de las líneas que presentan un periodo vegetativo más largo también podrá ser tenido en cuenta para la generación de variedades de floración más temprana, mediante la incorporación de regiones genómicas localizadas en el GL5 procedentes de variedades precoces en el genoma de variedades comerciales de fresa diploide, o mediante la incorporación de regiones homólogas al GL5 en el genoma octoploide (grupos Va (Va-f*/Va-m), Vb (Vb-f/Vb-m), Vc (Fc-f/Vc-m) y Vd(Vd-f*/m) del mapa de Rousseau-Gueutin *et al.* 2008 y grupos RG5-A, RG5-B, RG5-C, RG5-D, HA5-1, HA5-B, HA5-C y HA5-D del mapa de Sargent *et al.* 2009).

Hemos demostrado la idoneidad de la colección de líneas de introgresión en *Fragaria* para el estudio de caracteres complejos como la forma y el tamaño del fruto o el periodo de floración. Estudios realizados con genotecas de NILs en tomate han permitido clonar uno de los genes mayores implicados en la variabilidad del contenido de azúcares del fruto (Fridman *et al.* 2000), y han permitido mapear QTLs para el aroma (Tadmor *et al.* 2002) y el color (Liu *et al.* 2003) del fruto de tomate. El desarrollo de NILs en melón está siendo crucial en el esclarecimiento del carácter climatérico de algunas accesiones (J. Vegas-Renedo, en preparación). Por tanto, es probable que a través de la colección de líneas de introgresión desarrollada aquí también puedan estudiarse caracteres complejos como el aroma o la vida media de los frutos, cruciales en la mejora genética de fresa, y caracteres de calidad nutricional como el contenido en azúcares y ácidos orgánicos, vitaminas y polifenoles de acción antioxidante. La colección tiene una resolución de 43 cM aproximadamente. Aunque es un valor alto, se dispone de líneas con introgresiones heterocigóticas más pequeñas que podrán ser incluidas en la colección tras una generación de autofecundación y M.A.S. para acotar aún más la posición de los QTLs e identificar las causas genéticas de la variación presente en la colección.

La eficacia del mapeo fino de estos caracteres dependerá de la saturación del mapa que utilicemos. En estos momentos, nuestro laboratorio, en colaboración con EMR (UK) se están empleando métodos de secuenciación masiva de alto rendimiento (sistema 454) para el desarrollo de un nuevo mapa de marcadores SNPs en *Fragaria* diploide (Monfort *et al.* en preparación). La ubiquidad de este tipo de marcadores en el genoma garantiza la disponibilidad de marcadores moleculares en las regiones de interés agronómico, por lo que la caracterización de NILs segregantes podría conducir al mapeo fino de algunos caracteres interesantes y, en un futuro, a su posible clonación.

4. PERSPECTIVAS DE FUTURO EN LA GENÓMICA DEL GÉNERO *FRAGARIA*

Las actuales restricciones en el uso de pesticidas en los cultivos, la creciente preocupación de los consumidores por un consumo responsable y respetuoso con el medio ambiente, el aumento de la demanda de alimentos con actividad antioxidante en la dieta y las futuras consecuencias de la amenaza del cambio climático constituyen un reto para la mejora de los cultivos actuales. El avance en la mejora de variedades pasa inevitablemente por la explotación del germoplasma natural como reserva del acervo genético de las especies adaptadas a diferentes medios. La diversidad natural presente en las especies diploides del género *Fragaria* ha sido caracterizada en diferentes trabajos a lo largo de los últimos cincuenta años (Staudt, 1962; Naruhashi y Iwata, 1988; Staudt y Dickoré, 2001; Sargent *et al.* 2004b). El género *Fragaria* presenta un alto grado de fertilidad interespecífica, incluyendo especies de diferente nivel de ploidía, que posibilita la introgresión de regiones genómicas procedentes del germoplasma silvestre en cultivares élite mediante M.A.S. (Olbricht *et al.* 2009). Además, las especies silvestres de *Fragaria* se caracterizan por unas concentraciones de compuestos volátiles mucho más altas que las especies cultivadas (Ulrich *et al.* 2007). Por todo ello, el género *Fragaria* constituye un taxón excepcional para la aplicación de las técnicas de análisis empleadas en genética molecular en la mejora genética clásica de variedades. La culminación de la anotación de los genes del genoma de *Fragaria vesca* constituirá la herramienta clave que permitirá la identificación de los genes implicados en los caracteres más interesantes de cara a la mejora genética.

Este objetivo no podrá alcanzarse sin antes identificar las regiones genómicas implicadas a través del uso de herramientas como las desarrolladas aquí. El uso de estos recursos como fuente de variabilidad para algunos caracteres complejos requiere el uso de potentes plataformas de fenotipado que permitan realizar análisis metabólicos profundos, así como estudios de la expresión génica que deben desarrollarse en paralelo (Kopka *et al.* 2004). El auge que la metabolómica está experimentando en esta década será una de las claves para la comprensión de la variación genotípica de las especies cultivadas, y debe ser integrada a los estudios genéticos para la comprensión de las complejas interacciones entre alelos, genes y genomas que se dan en los ecosistemas y que han sido seleccionados durante la evolución natural de las especies. Este tipo de investigaciones está aún en sus inicios y la utilidad de *Fragaria* diploide como parte de un modelo integrado para la familia *Rosaceae* debe ser confirmado. Sin embargo, el potencial de *Fragaria* diploide para el estudio de la Bioquímica vegetal, la Fisiología del desarrollo del fruto y la Fitopatología es enorme, y debe ser explotado a través de herramientas como las desarrolladas en esta Tesis en el futuro más próximo.

CONCLUSIONES

1. La base de datos genómicos obtenidos a partir de la tecnología de secuenciación masiva (454-GS-FLX), *Strawberry Genome Assembly version 8*, se ha revelado como una fuente muy útil para el desarrollo de nuevos marcadores moleculares próximos a secuencias conocidas, con una tasa de éxito del 76%.
2. Los nuevos marcadores desarrollados han contribuido a la mejora del mapa de referencia de *Fragaria* diploide, aumentando su poder de resolución en un 25%, principalmente en los extremos de los grupos de ligamiento y reduciendo el tamaño medio de los *gaps* o regiones sin cobertura analizados en un 68%.
3. El nuevo mapa de ligamiento está formado exclusivamente por marcadores de tipo PCR, por lo que puede ser usado en toda su extensión para el análisis genético de poblaciones de gran tamaño, con una resolución media de un marcador cada 1,87 cM.
4. Tras la adición de 16 marcadores desarrollados para sustituir otros ya existentes y 108 marcadores moleculares cedidos por otros laboratorios, el mapa de *bins* de *Fragaria* diploide queda formado por 49 bins, cuyo tamaño medio es de 8,9 cM.
5. La caracterización de una genoteca de 18.432 BACs con insertos procedentes del genoma de *Fragaria vesca* de un tamaño medio de 85 kb y una cobertura de 7,6x ha permitido el anclaje de 102 BACs al mapa genético de *Fragaria*, sentando las bases para la generación del primer mapa físico del género.
6. La elaboración de un sistema de *pools* y *superpools* de ADN de la genoteca CUFvAB es un método eficaz, rápido, seguro y económico para el aislamiento de secuencias de alto peso molecular del genoma de *Fragaria* mediante amplificación por PCR.
7. La identificación y localización de una serie de genes candidatos que pueden estar jugando un papel importante en la calidad del fruto de fresa nos ha permitido aumentar en un 36,8% el número de genes de función conocida en el mapa de referencia de *Fragaria* diploide.
8. La transferabilidad de algunos marcadores moleculares derivados de los genes candidatos seleccionados nos ha permitido confirmar las relaciones de sintenia ya descritas entre los géneros *Prunus* y *Fragaria* y definir nuevas regiones sinténicas.
9. La colección de líneas de introgresión desarrollada a partir de la variedad comercial *F. vesca* 'Reine des Vallées' con introgresiones de la especie silvestre *F. bucharica* reúne las cualidades necesarias para la identificación y el estudio de la herencia de los caracteres que más influyen en los distintos parámetros de calidad del fruto y de la planta de fresa.

10. La estrategia empleada en la elaboración de esta colección, consistente en elaborar pocas poblaciones segregantes de gran tamaño poblacional y M.A.S temprana se ha revelado como una aproximación adecuada para el desarrollo de NILs en pocas generaciones.
11. La colección de líneas de introgresión en *Fragaria* está compuesta por 30 líneas que cubren el 97% del genoma de *F. bucharica* en forma de introgresiones solapantes de un tamaño medio de 43 cM y permite el mapeo de caracteres con una resolución media de 17,7 cM.
12. Los altos niveles de distorsión encontrados en el cruzamiento *F. vesca* x *F. bucharica* son atribuibles a la selección gamética o zigótica y pueden ser revertidos mediante la inversión del cruzamiento.
13. El grupo de ligamiento 7 de *Fragaria* diploide contiene una región directamente relacionada con la autoincompatibilidad en *F. bucharica* que no había sido descrita hasta la fecha.
14. Se han desarrollado líneas de introgresión para caracteres agrónomicamente interesantes, como el desarrollo de estolones, que puede suponer una mejora directa sobre la variedad 'Reine des Vallées'.
15. El análisis preliminar de la colección ha permitido localizar en el genoma de *Fragaria* diploide doce loci relacionados con caracteres cuantitativos y cualitativos de gran importancia económica, implicados en distintos parámetros de calidad del fruto y de la planta de fresa como la vernalización, la duración del periodo vegetativo, la floración en día corto, el tamaño de las hojas, la forma y tamaño del fruto o la resistencia al oidio.

BIBLIOGRAFÍA GENERAL

- Abbot AG, Lecouls AC, Want Y, Georgi L, Scorza R, Reighard G (2002)** Peach: the model genome for Rosaceae genomics. *Acta Horticulturae* **592**:199-203
- Agarwal M, Shrivastava N, Padh H (2008)** Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**:617-631
- Aharoni A, O'Connell AP (2002)** Gene expression analysis of strawberry aquene and receptacle maturation using DNA microarrays. *Journal of Experimental Botany* **53**:2073-2087
- Ahmadi H, Bringham RS, Voth V (1990)** Modes of inheritance of photoperiodism in *Fragaria*. *Journal of the American Society for Horticultural Science* **115**:146-152
- Akiyama Y, Yamamoto Y, Ohmido N, Ohshima M, Fukui K (2001)** Estimation of the nuclear DNA content of strawberries (*Fragaria* spp.) compared with *Arabidopsis thaliana* by using dual-step flow cytometry. *Cytologia* (Tokyo) **66**:431-436
- Albani MC, Battey NH, Wilkinson MJ (2004)** The development of ISSR-derived SCAR markers around the SEASONAL FLOWERING LOCUS (SFL) in *Fragaria vesca*. *Theoretical and Applied Genetics* **109**: 571-579
- Aleza P, Juárez J, Ollitrault P, Navarro L (2009)** Production of tetraploid plants of non apomictic citrus genotypes. *Plant Cell Reproduction* **28**(12):1837-1846
- Allard RW (1953)**. Inheritance of some seed coat colours and patterns in lima beans. *Hilgardia* **22**:167-177
- Almenar E (2005)** Envasado activo de fresas silvestres. Tesis doctoral. Departamento de Bromatología, Salud Pública y Medicina Legal. Universitat de Valencia (España).
- Alonso-Blanco C, Koornneef M (2000)** Naturally occurring variation in *Arabidopsis*: An underexploited resource for plant genetics. *Trends in Plant Science* **5**:22-29
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997)** Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**(17):3389-3402
- Andres-Lacueva C, Shukitt-Hale B, Galli RL, Jauregui O, Lamuela-Raventos RM, Joseph JA (2005)** Anthocyanins in aged blueberry-fed rats are found centrally and may enhance memory. *Nutritional Neuroscience* **8**:111-120
- Areshchenkova T, Ganai MW (1999)** Long tomato microsatellites are predominantly associated with centromeric regions. *Genome* **42**:536-544
- Argyris J, Truco MJ, Ochoa O, McHale L, Dahal P, Van Deynze A, Michelmore RW, Bradford KJ (2010)** A gene encoding an abscisic acid biosynthetic enzyme (LsNCED4) collocalizes with the high temperature germination locus Htg6.1 in lettuce (*Lactuca sp.*) *Theoretical and Applied Genetics* DOI: 10.1007/s00122-010-1425-3.
- Arulsekhar S, Bringham RS (1981)** Genetic model for the enzyme marker PGI in diploid California *Fragaria vesca* Its variability and use in elucidating the mating system. *Journal of Heredity* **72**: 117-120
- Arulsekhar S, Bringham RS, Voth V (1981)** Inheritance of PGI and LAP isozymes in octoploid cultivated strawberries. *Journal of the American Society for Horticultural Science* **106**:679-683
- Ashley MV, Wilk JA, Styan SMN, Craft KJ, Jones KL, Feldheim KA, Lewers KS, Ashman TL (2003)** High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana* Mill. (Rosaceae). *Theoretical and Applied Genetics* **107**: 1201-1207
- Atkinson CJ, Dodds PA, Ford YY, Le Mière J, Taylor JM, Blake PS, Paul N. (2006)** Effects of cultivar, fruit number and reflected photosynthetically active radiation on *Fragaria x ananassa* productivity and fruit ellagic acid and ascorbic acid concentrations. *Annals of Botany (London)* **97**(3):429-41

- Bai Y, van der Hulst R, Huang CC, Wei L, Stam P, Lindhout P** (2004) Mapping *Ol-4*, a gene conferring resistance to *Oidium neolycopersici* and originating from *Lycopersicon peruvianum* LA2172, requires multi-allelic, single-locus markers. *Theoretical and Applied Genetics* **109**:1215-1223
- Bailey DW** (1971) Recombinant inbred strains. *Transplantation* **11**:325-327
- Baird WV, Estager AS, Wells JK** (1994) Estimating nuclear-DNA content in peach and related diploid species using laser flowcytometry and DNA hybridization. *Journal of the American Society of Horticultural Sciences* **119**:1312-1316
- Ballester J, Socias i Company R, Arús P, de Vicente MC** (2001) Genetic mapping of a major gene delaying blooming time in almond. *Plant Breeding* **177**:69-72
- Bassil NV, Gunn M, Folta K, Lewers K** (2006) Microsatellite markers for *Fragaria* from 'Strawberry Festival' expressed sequence tags. *Molecular Ecology Notes* **6**: 473-476.
- Batley NH, Le Miere P, Tehranifar A, Cekic C, Taylor S, Shriver KJ, Hadley P, Greenland AJ, Darby J, Wilkinson MJ.** (1998). Genetic and environmental control of flowering in strawberry. In: Cockshull KE, Gray D, Seymour GB, Thomas B, eds. *Genetic and environmental manipulation of horticultural crops*. Wallingford: CAB International, 111-131.
- Bellido-Cabello, ML** (2005) Estudio y caracterización molecular de genes relacionados con la maduración del fruto de fresa. Tesis Doctoral. Universidad de Córdoba.
- Bentley S, Bassam BJ** (1996) A robust DNA amplification fingerprinting system applied to analysis of genetic variation within *Fusarium oxysporum fsp cubense*. *Journal of Phytopathology*. **144**:207-213
- Bernanncchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, López J, Petiard V, Sayama H, Uhlig J, Zamir D, Tanksley S** (1998) *Theoretical and Applied Genetics* **97**:170-180
- Bickford PC, Gould T, Briederick L, Chadman K, Pollock A, Young D, Shukitt-Halle B, Joseph J** (2000) Antioxidant-rich diets improve cerebellar physiology and motor learning in aged rats. *Brain Research* **866**:211-217
- Bink M, Voorrips R, Van de Weg E, Jansen H,** (2007) Statistical tools for QTL mapping in multiple, pedigreed populations. Eucarpia XOO Fruit Section Symposium, September 16-20, 2007, Zaragoza, Spain.
- Blair MW, Pedraza F, Buendia HF, Gaitán Solís E, Beebe SE, Gepts P, Tohme J** (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L). *Theoretical and Applied Genetics* **107**:1362-1374
- Blanco A, Simeone R, Gadaleta A** (2006) Detection of QTLs for grain protein content in durum wheat. *Theoretical and Applied Genetics* **112**(7):1195-1204
- Bliss FA, Arulsekar S, Foolad MR, Becerra V, Gillen AM, Warburton ML, Dandekar AM, Kocsisne GM, Mydin KK** (2002) An expanded genetic linkage map of *Prunus* based on an interspecific cross between almond and peach. *Genome* **45**:520-529
- Bolon YT, Bindu J, Cannon SB, Grahan MA, Diers BW, Farmer AD, May GD, Muehlbauer GJ, Specht JE, Tu ZJ, Weeks N, Xu WW, Schoemaker RC, Vance CP** (2010) Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biology* **10**:41
- Bombarely A** (2007) Estudio mediante el uso de herramientas genómicas y bioinformáticas del desarrollo y maduración del fruto de fresa. Tesis doctoral. Universidad de Málaga.

- Bombarely A, Merchante C, Csukasi F, Cruz-Rius E, Caballero JL, Medina-Escobar N, Blanco.Portales R, Botella MA, Muñoz-Blanco J, Sánchez-Sevilla JF, Valpuesta V** (2010) Generation and analysis of ESTs from strawberry (*Fragaria x ananassa*) fruits and evaluation of their utility in genetic and molecular studies. *BMC Genomics* **17**;11(1):503
- Bonet J, Lopez-Girona E, Sargent DJ, Muñoz-Torres MC, Monfort A, Abbott AG, Arús P, Simpson DW, Davik J** (2009) The development and characterisation of a bacterial artificial chromosome library for *Fragaria vesca*. *BMC Research Notes* **2**:188
- Bošković RI, Sargent DJ, Tobutt KR** (2010) Genetic evidence that two independent S-loci control RNase-based self-incompatibility in diploid strawberry. *Journal of Experimental Botany*, **61**(3):755–763.
- Botstein D, White RL, Skolnick M, Davis RW** (1980) Construction of a genetic map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32**:314-331
- Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF** (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *BioTechniques* **31**(1):24-6,28
- Bringhurst RS, Arulsekar S, Hancock JF, Voth V** (1981) Electrophoretic characterization of strawberry cultivars. *Journal of the American Society for Horticultural Science* **106**:684–687
- Bringhurst, RS** 1990. Cytogenetics and evolution in American *Fragaria*. *Hortscience.*, **25**, 879-881.
- Brouwer DJ, St. Clair DA** (2004) Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *Theoretical and Applied Genetics* **108**:628-638
- Brown T, Wareing PF** (1965) The genetical control of the everbearing habit and three other characters in varieties of *Fragaria vesca*. *Euphytica* **14**:97-112
- Burke J, Wang H, Hide W, Davison DB** (1998) Alternative gene form discovery and candidate gene selection from gene indexing projects. *Genome Research* **8**:276-290
- Burke J, Davison D, Hide W** (1999) d2_cluster: a validated method for clustering EST and full-length cDNA sequences. *Genome Research* **9**:1135-1142
- Burns MJ, Barnes SR, Bowman JG, Clarke MHE, Werner CP, Kearsey MJ** (2003) QTL analysis of an intervarietal set of substitution lines in *Brassica napus*: (i) Seed oil content and fatty acid composition. *Heredity* **90**:39-48
- Burr B, Burr Fa, Thompson KH, Albertson MC, Stuber CW** (1988) Gene mapping with Recombinant Inbreds in Maize. *Genetics* **118**:519-526
- Bushman BS, Larson SR, Mott IW, Cliften PF, Wang RRC, Chatterton JC, Hernández AG, Ali S, Kim RW, Thimmapuram JT, Gong G, Liu L, Mikel MA** (2008) Development and annotation of perennial *Triticaceae* ESTs and SSR markers. *Genome* **51**:779-788
- Canady MA, Meglic V, Chetelat RT** (2005) A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. *Genome* **48**:685-697
- Capomaccio S, Barone P, Reale L, Veronesi F, Rosellini D** (2009) Isolation of genes from female sterile flowers in *Medicago sativa*. *Sex Plant Reproduction* **22**:97-107.
- Carrillo JM, Rousset M, Qualset CO, Kasarda DD** (1990) Use of recombinant inbred lines of wheat for study of associations of high-molecular-weight glutenin subunit alleles to quantitative traits. *Theoretical and Applied Genetics* **79**:321-330

- Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C** (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. *Journal of Experimental Botany* **55**:1671-1685
- Causse M, Chaïb J, Lecomte L, Buret M, Hospital F.** (2007) Both additivity and epistasis control the genetic variations for fruit quality traits in tomato. *Theoretical and Applied Genetics* **115**:429-442
- Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EH, Gardiner SE, Korban SS** (2008) Development of a set of SNP markers present in expressed genes of the apple. *Genomics*. Nov;**92**(5):353-358.
- Chaim AB, Borovsky Y, Rao GU, Tanyolac B, Paran I** (2002) *fs3.1*: a major fruit shape QTL conserved in Capsicum. *Genome* **46**: 1-9
- Chalavi V, Tabaeizadeh Z, Thibodeau P** (2003) Enhanced resistance to *Verticillium dahliae* in transgenic strawberry plants expressing a *Lycopersicon chilense* chitinase gene. *Journal of the American Society of Horticultural Sciences* **128**:747-753
- Chaparro JX, Werner DJ, Omalley D, Sederoff RR** (1994) Targeted mapping and linkage analysis of morphological isozyme, and RAPD markers in peach. *Theoretical and Applied Genetics* **87**:805-815
- Chetelat RT, Meglic V** (2000) Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*).
- Chung MJ, Lee SH, Sung NJ** (2002) Inhibitory effects of whole strawberries, garlic juice or kale juice on endogenous formation of N-nitrosodimethylamine in humans. *Cancer letters* **182**:1-10
- Chung CL, Longfellow JM, Walsh EK, Kerdieh Z, Van Esbroeck GV, Balint-Kurti P, Nelson RJ** (2010) Resistance loci affecting distinct stages of fungal pathogenesis: use of introgression lines for QTL mapping and characterization in the maize - *Setosphaeria turcica* pathosystem. *BMC Plant Biology* 2010, **10**:103 doi:10.1186/1471-2229-10-103
- Cipriani G, Testolin R** (2004) Isolation and characterization of microsatellite loci in *Fragaria*. *Molecular Ecology Notes* **4**: 366-368
- Cipriani G, Pinosa F, Bonoli M, Faedi W** (2006) A new set of microsatellite markers for *Fragaria* species and their application in linkage analysis. *Journal of Horticultural Science and Biotechnology* **81**: 668-675
- Civello PM, Powell ALT, Sabehat A Bennett AB** (1999) An expansin gene expressed in ripening strawberry fruit. *Plant Physiology* **121**(4):1273-1279
- Clifford MN, Scalbert A** (2000) Ellagitannins—nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture* **80**:1118-1125
- Congiu L, Chicca M, Cella R, Rossi R, Bernacchia G** (2000) The use of random amplified polymorphic DNA (RAPD) markers to identify strawberry varieties: A forensic application. *Molecular Ecology* **9**: 229-232
- Cuartero J, Rodriguez A** (1998) El punto de vista del mejorador público: I. Plantas reproducibles por semillas. Pp105-126. En: Los derechos de propiedad de las obtenciones vegetales. Eds: Nuez F, Llácer G, Cuartero J. Monografías INIA 103. Madrid.
- Damianaki A, Bakogeorgou E, Kampa M, Notas G, Hatzoglou A, Panagiotou S, Gemetzi C, Koroumalis E, Martin PM, Castanas E** (2000) Potent inhibitory action of red wine polyphenols on human breast cancer cells. *Journal of Cell Biochemistry* **78**:429-441
- Darrow GM** (1966) Strawberry Breeding and Industry on the European Continent. In: The strawberry: History, Breeding and Physiology. Holt, Rinehart and Winston, USA

- Davis TM, Yu H** (1997) A linkage map of the diploid strawberry, *Fragaria vesca*. *Journal of Heredity* **88**: 215–221
- Davis TM, DiMeglio LM, Yang RH, Styan SMN, Lewers KS** (2006) Assessment of SSR transfer from the cultivated strawberry to diploid strawberry species: Functionality, linkage group assignment, and use in diversity analysis. *Journal of the American Society for Horticultural Science* **131**: 506–512
- Davis TM, Bennetzen JL, Pontaroli AC, San Miguel P, Folta KM, Tombolato DCM, Shields M, Zhang Q**: (2006) Sequence samples from the diploid strawberry model species, *Fragaria vesca* . *Plant and Animal Genomes XIV conference, San Diego, CA, USA, - Genome Sequencing & ESTs*.
- Davuluri GR, van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Brummel DA, King SR, Palys J, Uhlig J, Bramley PM, Pennings HM, Bowler C** (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nature Biotechnology* **23**(7):890-895
- Dawe RK** (1998). Meiotic chromosome organization and segregation in plants. *Annual review of plant physiology and plant molecular biology* **49**:371–395
- Deleu W, González V, Monfort A, Bendahmane A, Puigdomènech P, Arús P, García-Mas J** (2007) Structure of two melon regions reveals high microsynteny with sequenced plant species. *Molecular Genetics and Genomics* **278**:611–622
- Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, Carde JP, Merillon JM, Hamdi S** (2006) Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant Physiology* **140**:499-511
- Deng C, Davis TM** (2001) Molecular identification of the yellow fruit color (c) locus in diploid strawberry: A candidate gene approach. *Theoretical and Applied Genetics* **103**: 316–322
- Dettoni MT, Quarta R, Verde I** (2001) A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. *Genome* **44**:783-790
- Díaz MN, Frei B, Vita JA, Keaney JF** (1997) Mechanisms of disease: antioxidants and arteriosclerotic heart disease. *New England Journal of Medicine* **337**:408-416.
- Dirlewanger E, Moing A, Rothan C, Svanella L, Pronier V, Guye A, Plomion C, Monet R** (1999) Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch). *Theoretical and Applied Genetics* **98**:18-31
- Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arús P** (2004a) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proceedings of the National Academy of Sciences of the USA* **101**:9891-9896
- Dirlewanger E, Cosson P, Renaud C, Monet R, Poëssel JL, Moing A** (2005) New detection of QTLs controlling major fruit quality components in peach. *Acta Horticulturae* **713**:65-72
- Dirlewanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A** (2006) Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. *Tree Genetics and Genomes* **3**:1-13
- Doganlar S, Frary A, Daunay MC, Lester R, Tanksley SD** (2002) Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* **161**:1713-1726
- Doyle JJ, Doyle JL**. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**:13-15
- Dunnett CW** (1955) A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**:1096-1121

- Edmé SJ, Glynn NG, Comstock JC** (2006) Genetic segregation of microsatellite markers in *Saccharum officinarum* and *S. Spontaneum*. *Heredity* **97**:366-375
- Eduardo I** (2005) Desarrollo de una genoteca de líneas casi-isogénicas (NILs) en melón y estudio de caracteres de calidad de fruto. Tesis doctoral. Universitat Autònoma de Barcelona
- Eduardo I, Arús P, Monforte AJ** (2005) Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI 161375. *Theoretical and Applied Genetics* **112**:139-148.
- Eduardo I, Arús P, Monforte AJ** (2007) Estimating the genetic architecture of fruit quality traits in melon using a genomic library of Near Isogenic Lines. *Journal of the American Society of Horticultural Sciences*. **132**(1):80-89
- El Mansouri I, Mercado JA, Valpuesta V, López-Aranda JM, Pliego-Alfaro F, Quesada MA** (1996) Shoot regeneration and *Agrobacterium*-mediated transformation of *Fragaria vesca* L. *Plant Cell Reports* **15**:642-646
- Elsik CG, Williams CG** (2001) Families of clustered microsatellites in a conifer genome. *Molecular Genetics and Genomics* **265**:535-542
- Eshed Y, Zamir D** (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* **79**:175-179
- Eshed Y, Zamir D** (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**:1147-1162.
- Essafi A, Díaz-Pendón JA, Moriones E, Monforte AJ, García-Mas J, Martín-Hernández AM** (2009) Dissection of the oligogenic resistance to Cucumber mosaic virus in the melon accession PI 161375. *Theoretical and Applied Genetics* **118**(2):275-284
- Etienne C, Moing A, Dirlewanger E, Raymond P, Monet R, Rothan C** (2002a) Isolation and characterization of six peach cDNAs encoding key proteins in organic acid metabolism and solute accumulation: involvement in regulating peach fruit acidity. *Physiologia Plantarum* **114**:259-270
- Etienne C, Rothan C, Moing A, Plomion C, Bodenes C, Svanella-Dumas L, Cosson P, Pronier V, Monet R, Dirlewanger E** (2002b) Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. *Theoretical and Applied Genetics* **105**:145-159
- Ewing B, Green P** (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**:186-194
- Ewing B, Hillier L, Wendl MC, Green P** (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**:175-185
- Fazio G, Staub JE, Stevens MR** (2003) Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. *Theoretical and Applied Genetics* **107**:864-874
- Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH** (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved noncoding sequences in orphan crops. *Plant Physiology* **140**:1183-1191
- Fernandez-Silva I, Moreno E, Eduardo I, Arús P, Álvarez JM, Monforte AJ** (2009) On the genetic control of heterosis for fruit shape in melón (*Cucumis melo* L.) *Journal of Heredity* **100**(2):229-235
- Fernandez-Silva I, Moreno E, Essafi A, Fergany M, García-Mas J, Martín-Hernández AM, Álvarez JM, Monforte, AJ** (2010) Shaping melons: agronomic and genetic characterization of QTLs that modify melon fruit morphology. *Theoretical and Applied Genetics* **121**:931-940

- Folta KM, Staton M, Stewart PJ, Jung S, Bies DH, Jesdurai C, Main D** (2005) Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (*Fragaria × ananassa*). *BCM Plant Biology*, **5**, 12–23
- Folta KM, Davis TM** (2006) Strawberry genes and genomics. *Critical Reviews in Plant Sciences* **25**:399-415.
- Foolad MR, Arulsekar S, Becerra V, Bliss FA** (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. *Theoretical and Applied Genetics* **91**:262-269
- Francia E, Tacconi G, Crosatti C, Barabaschi D, Bulgarelli D, Dall'aglio E, Vale G** (2005) Marker assisted selection in crop plants. *Plant cell tissue and organ culture* **82**:317-342
- Frary A, Nesbitt TC, Frary A, Grandillo S, Knaap E van der, Cong B, Liu J, Meller J, Elber R, Alpert K, Tanksley S** (2000) Cloning and transgenic expression of *fw2.2* : a quantitative trait locus key to the evolution of tomato fruit. *Science*, **289**:85-87
- Frary A, Doganlar S, Frampton A, Fulton T, Uhlig J, Yates H, Tanksley S** (2003) Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon chmielewskii* chromosome 1. *Genome* **46**(2):235-243
- Fraser CM, Rider LW, Chapple C** (2005) An expression and bioinformatics analyses of the *Arabidopsis* serine carboxypeptidase-like gene family. *Plant Physiology* **138**:1136-1148
- Fridman E, Pleban T, Zamir D** (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. *Proceeding of the National Academy of Sciences of the U.S.A.* **25**;97(9)4718-4723
- Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D** (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305**:1786–1789
- Fukino N, Ohara T, Monforte AJ, Sugiyama M, Sakata Y, Kuniyama M, Matsumoto S** (2008) Identification of QTLs for resistance to powdery mildew and SSR markers diagnostic for powdery mildew resistance genes in melón (*Cucumis melo* L.) *Theoretical and Applied Genetics*. DOI 10.1007/s00122-008-0885-1
- Gaxiola RA, Palmgren MG, Schumacher K** (2007) Plant proton pumps. *FEBS Letters* **581**:2204-2214
- Gelderman H** (1975) Investigation on inheritance of quantitative characters in animals by gene markers. 1. Methods. *Theoretical and Applied Genetics* **46**:319-330
- Georgi LL, Wang Y, Yvergnaux D, Ormsbee T, Iñigo M, Reighard G, Abbott AG** (2002) Construction of a BAC library and its application to the identification of simple sequence repeats in peach [*Prunus persica* (L.) Batsch]. *Theoretical and Applied Genetics*, **105**:1151-1158
- Georgi LL, Wang Y, Reighard GL, Mao L, Wing RA, Abbott AG** (2003) Comparison of peach and *Arabidopsis* genomic sequences: fragmentary conservation of gene neighborhoods. *Genome* **46**:268-276
- Gil MI, Tomas-Barberan FA, Hess-Pierce B, Kader AA** (2002) Antioxidant capacities, phenolic compounds, carotenoids, and vitamin C contents of nectarine, peach, and plum cultivars from California. *Journal of Agricultural and food chemistry* **50**:4976-4982
- Gil-Ariza DJ, Amaya I, Botella MA, Muñoz Blanco J, Caballero JL, López-Aranda JM, Valpuesta V, Sánchez-Sevilla F** (2006) EST-derived polymorphic microsatellites from cultivated strawberry (*Fragaria x ananassa*) are useful for diversity studies and varietal identification among *Fragaria* species. *Molecular Ecology Notes* **6**:1195-1197
- Gillaspy G, Ben-David H, Grisse W** (1993) Fruits: A developmental perspective. *Plant Cell* **5**:1439-1451

- Giovanonni JJ** (2001) Molecular biology of fruit maturation and ripening. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**:725-749
- Goldman IL, Paran I, Zamir D** (1995) Quantitative trait locus analysis of a recombinant inbred line population derived from the *Lycopersicon esculentum* x *Lycopersicon cheesmanii* cross. *Theoretical and Applied Genetics* **90**:925-932
- Gordon D, Abajian C, Green P** (1998) Consed: a graphical tool for sequence finishing. *Genome Research* **8**:195-202
- Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC** (1996) RFLP mapping in cultivated sugarcane (*Saccharum spp.*): Genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**:987-1000
- Gupta, PK, Varshney RK** (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163–185
- Hackauf B, Rudd S, van der Voort JR, Miedaner T, Wehling P** (2008) Comparative mapping of DNA sequences in rye (*Secale cereale* L.) in relation to the rice genome. *Theoretical and applied genetics* **118**:371-384.
- Hadonou AM, Sargent DJ, Wilson F, James CM, Simpson DW** (2004) Development of microsatellite markers in *Fragaria*, their use in genetic diversity analysis and their potential for genetic linkage mapping. *Genome* **47**: 429–438
- Haldane JBS** (1919) The combination of linkage values and the calculation of distances between loci of linked factors. *Journal of Genetics* **8**:229-309
- Hancock JF, Bringham RS** (1978) Inter-populational differentiation and adaptation in perennial, diploid species *Fragaria-vesca* l. *American Journal of Botany* **65**: 795–803
- Hancock JF, Bringham RS** (1979) Ecological differentiation in perennial, octoploid species of *Fragaria*. *American Journal of Botany* **66**:367–375
- Hancock JF** (1999) *Strawberries*. CABI Publishing, Oxon, UK, Pp25-46
- Hannum SA** (2004) Potential impact of strawberries on human health: A review of the science. *Critical Reviews in Food Science and Nutrition* **44**:1-17
- Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Mullet J** (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. *Journal of Experimental Botany* **58**(2):327-338
- Haymes KM, Henken B, Davis TM, van de Weg WE** (1997) Identification of RAPD markers linked to a *Phytophthora fragariae* resistance gene (Rpf1) in the cultivated strawberry. *Theoretical and Applied Genetics* **94**:1097–1101
- Haymes KM, van de Weg WE, Arens P, Maas JL, Vosman B, Den Nijs APM** (2000) Development of SCAR markers linked to a *Phytophthora fragariae* resistance gene and their assessment in European and North American strawberry genotypes. *Journal of the American Society for Horticultural Science* **125**:330–339
- Hayward MD, Bosemark NO, Romagosa I** (1993). Preface. En *Plant Breeding. Principles and prospects*. Chapman & Hall. London.
- Herrmann MG, Durtschi JD, Voelkerding KV, Wittwer CT** (2006) Instrument comparison for DNA genotyping by amplicon melting. *Journal of the Association for Laboratory Automation* **11**:273–277

- Hesse CO** (1975) Peaches. In: Janick J, Moore JN (eds) *Advances in fruit breeding*. Purdue Univ Press, West Lafayette, pp 285–335
- Hidalgo MA, Sánchez D, Hureau R, Moreau J, Monfort A, Arús P** (2009) The Inotalis breeding program (Planasa and Darbonne) - New trends and challenges in strawberry plant breeding. In: Proceedings of the VIth International Strawberry Symposium. (Ed. López-Medina, J) *Acta Horticulturae* **842**:431-434
- Hirvi T, Honkanen E** (1982) The volatiles of two new strawberry cultivars, 'Annelie' and Alaska Pioneer obtained by backcrossing of cultivated strawberries with wild strawberries, *Fragaria vesca*, Rügen and *Fragaria virginiana*. *Zeitschrift für Lebensmittel-Untersuchung und-Forschung* **175**:113-116
- Hiwasa K, Nakano R, Inaba A, Kubo Y** (2003) Expression analysis of genes encoding xyloglucan endotransglycosylase during ripening in pear fruit. *Acta Horticulturae* **628**:549-553
- Horn R, Lecouls AC, Callahan A, Dandekar A, Garay L, McCord P, Howad W, Chan H, Verde I, Main D, Jung S, Georgi L, Forrest S, Mook J, Zhebentyayeva T, Yu YS, Kim HR, Jesudurai C, Sosinski B, Arus P, Baird V, Parfitt D, Reighard G, Scorza R, Tomkins J, Wing R, Abbott AG** (2005) Candidate gene database and transcript map for peach, a model species for fruit trees. *Theoretical and Applied Genetics*, **110**:1419-1428
- Howad W, Yamamoto T, Dirlwanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P** (2005) Mapping with a few plants: using selective mapping for microsatellite saturation of the Prunus reference map. *Genetics* **171**:1305–1309
- Hu CG, Hao YJ, Honda C, Kita M, Moriguchi T** (2003) Putative PIP1 genes isolated from apple: expression analyses during fruit development and under osmotic stress. *Journal of Experimental Botany* **54**:2193–2194
- Hummer KE, Nathewet P, Yanagi T** (2009) Decaploidy in *Fragaria iturupensis* (Rosaceae). *American Journal of Botany* **96**(3):713-716
- Hunter RL, Market CL** (1975). Histochemical demonstration of enzymes separated by zone electrophoresis in starch genes. *Science* **125**:1294-1295.
- Hureau R, Moreau J, Hidalgo MA, Sánchez D** (2008) **The Inotalis breeding program (Planasa & Darbonne) – New trends and challenges in strawberry plant breeding.** In: VI International Strawberry Symposium 2008. Huelva, Spain. Book of Abstracts. (Ed. Junta de Andalucía). P 46.
- Hyne V, Kearsey MJ, Pike DJ, Snape JW** (1995) QTL analysis: Unreliability and bias in estimation procedures. *Molecular Breeding* **1**:273-282
- Immer FR, Henderson MT** (1943) Linkage studies in Barley. *Genetics* **28**:419-440
- James CM, Wilson F, Hadonou AM, Tobutt KR** (2003) Isolation and characterization of polymorphic microsatellites in diploid strawberry (*F. vesca* L.) for mapping, diversity studies and clone identification. *Molecular Ecology Notes* **3**:171–173.
- Jeuken MJW, Lindhout P** (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theoretical and Applied Genetics* **109**:394-401
- Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis THN, Flavell AJ** (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). *Genetics* **177**:2263–2275
- Jones JK**, (1955) Cytogenetic studies in the genera *Fragaria* and *Potentilla*. PhD thesis, Manchester University, UK.

- Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Weil C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschnieder R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmioli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A, (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* **3**: 381-390
- Joobeur T, Viruel MA, de Vicente MC, Jáuregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, I Batlle, Quarta R, Dirlewanger E, Arús P (1998) Construction of a saturated linkage map for *Prunus* using an almond x peach F₂ progeny. *Theoretical and Applied Genetics* **97**:1034-1041
- Jung S, Main D, Staton M, Cho IH, Zhebentyayeva T, Arús P, Abbott A (2006) Synteny conservation between the *Prunus* genome and both the present and ancestral *Arabidopsis* genomes. *BMC Genomics*, **7**:81
- Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D (2008) GDR (Genome Database for Rosaceae): integrated webdatabase for Rosaceae genomics and genetics data. *Nucleic Acids Research* **36**:D1034-D1040
- Kamei A, Tsuru M, Kubo N (2010) QTL mapping of clubroot resistance in radish (*Raphanus sativus* L.) *Theoretical and Applied Genetics* **120**:1021-1027
- Kaufmann H, Mattiesch L, Lorz H, Debener T (2003) Construction of a BAC library of *Rosa rugosa* Thunb. and assembly of a contig spanning *Rdr1*, a gene that confers resistance to blackspot. *Molecular Genetics and Genomics*, **268**:666-674
- Keuls M, (1956) The use of the 'studentized range' in connection with an analysis of variance. *Euphytica* **1**:112-122
- Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a Near-Isogenic Line population of *Arabidopsis thaliana* and comparison of mapping power with a Recombinant Inbred Line population. *Genetics* **175**:891-905
- Kim IJ, Lee BH, Jo J, Chung WI (2001) Sequence variability of nine cytosolic ascorbate peroxidases in polyploid strawberry. *DNA Seq* **11**:475-484
- Knox MR, Ellis THN (2002) Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. *Genetics* **162**:861-873
- Konieczny A, Ausubel FM (1993) Procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant Journal* **4**:403-410
- Kopisch-Obuch FJ, Diers BW (2006) Segregation at the SCN resistance locus *rhg1* in soybean is distorted by an association between the resistance allele and reduced field emergence. *Theoretical and Applied Genetics* **112**:199-207
- Kopka J, Fernie A, Weckwerth W, Gibon Y, Stitt M (2004) Metabolite profiling in plant biology: Platforms and destinations. *Genome Biology* **5**:109.3-109-9
- Kops J, van Hall HC (1844) *Flora batava of of Afbeelding en Beschrijving van Nederlandsche Gewassen*. Volúmen 8. Fuente: www.biolib.de
- Korbin M, Kuras A, Zurawicz E (2002) Fruit plant germplasm characterisation using molecular markers generated in RAPD and ISSR-PCR. *Cellular & Molecular Biology Letters*.**7**(2B):785-794
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann. Eugen.* **12**:172-175
- Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsey MJ (2002) STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *The Plant Journal* **31**(3), 355-364.

- Kubota N, Mimura H, Shimamura K** (2000) Differences in phenolic levels among mature peach and nectarine cultivars and their relation to astringency. *Journal of the Japan Society of Horticultural Science* **69**:35-39
- Kubota N, Yakushiji H, Nishiyama N, Mimura H, Shimamura K** (2001) Phenolic contents and L-phenylalanine ammonia-lyase activity in peach fruit as affected by rootstocks. *Journal of the Japan Society of Horticultural Science* **70**:151-156
- Kunihisa M, Fukino N, Matsumoto S** (2005) CAPS markers improved by cluster-specific amplification for identification of octoploid strawberry (*Fragaria x ananassa* Duch.) cultivars, and their disomic inheritance. *Theoretical and Applied Genetics* **110**: 1410-1418
- Kunihisa M, Ueda H, Fukino N, Matsumoto S** (2009) Genotyping of Strawberry (*Fragaria x ananassa* Duch.) cultivars by DNA markers: Interlaboratory study. *Journal of AOAC International*, **92**(3)896-906
- Kuras A, Korbin M, Zurawicz E** (2004) Comparison of suitability of RAPD and ISSR techniques for determination of strawberry (*Fragaria x ananassa* Duch.) relationship. *Plant Cell Tissue and Organ Culture* **79**: 189-193
- Kwon SW, Cho YC, Kim YG, Suh JP, Jeung JU, Roh JH, Lee SK, Jeon JS, Yang SJ, Lee YT** (2008) Development of near-isogenic Japonica rice lines with enhanced resistance to *Magnaporthe grisea*. *Molecules and Cells* **25**:407-416
- Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH** (2005) Identification and mapping of SNPs from ESTs in sunflower. *Theoretical and Applied Genetics* **111**:1532-1544
- Lander ES, Botstein** (1986) Mapping complex genetic traits in humans: New methods using a complete RFLP linkage map. *Quantitative Biology* **51**:49-62
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L** (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**(2):174-181
- Lander ES, Botstein** (1989) Mapping Mendelian factors underlying quantitative traits using RFLP markers. *Genetics* **121**:185-199
- Lawson DM, Hemmat M, Weeden N** (1995) The use of molecular markers to analyze the inheritance of morphological and developmental traits in apple. *Journal of the American Society of Horticultural Sciences* **120**:532-537
- Lea MA, Ibeh C, desBordes C, Vizzotto M, Cisneros-Zevallos L, Byrne DH, Okie WR, Moyer MP** (2008) Inhibition of growth and induction of differentiation of colon cancer cells by peach and plum phenolic compounds. *Anticancer Research* **28**:2067-2076
- Lebreton CH, Visscher PM, Haley CS, Semikhodskii A, Quarrie SA** (1998) A nonparametric bootstrap method for testing close linkage vs. pleiotropy of coincident quantitative trait loci. *Genetics* **150**:931-943
- Le Dantec L, Chagne D, Pot D, Cantin O, Garnier-Gere P, Bedon F, Frigerio JM, Chaumeil P, Leger P, Garcia V, Laigret F, De Daruvar A, Plomion C** (2004) Automated SNP detection in expressed sequence tags: statistical considerations and application to maritime pine sequences. *Plant Molecular Biology* **54**:461-470
- Lerceteau-Köhler E, Guerin G, Laigret F, Denoyes-Rothan B** (2003) Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (*Fragaria x ananassa*) using AFLP mapping. *Theoretical and Applied Genetics* **107**: 619-628
- Lerceteau-Köhler E, Guerin G, Denoyes-Rothan B** (2005) Identification of SCAR markers linked to *Rca2* anthracnose resistance gene and their assessment in strawberry germplasm. *Theoretical and Applied Genetics*. **111**: 862-870

- Lerceteau-Kohler E, Moing A, Guerin G, Renaud C, Maucourt M, Rolin D, Roudeillac P, Denoyes-Rothan B** (2006) QTL analysis for sugars and organic acids in strawberry fruits. *Acta Horticulturae* **708**:573-577
- Lewers KS, Saski CA, Cuthbertson BJ, Henry DC, Staton ME, Main DS, Dhanaraj AL, Rowland LJ, Tomkins JP** (2008) A blackberry (*Rubus* L.) expressed sequence tag library for the development of simple sequence repeat markers. *BMC Plant Biology* **8**:69
- Lewers KS, Styan SMN, Hokanson SC, Bassil NV** (2005) Strawberry GenBank-derived and genomic simple sequence repeat (SSR) markers and their utility with strawberry, blackberry, and red and black raspberry. *Journal of the American Society for Horticultural Science* **130**:102-115
- Lewis RS, Milla SR, Kernodle SP** (2007) Analysis of an introgressed *Nicotiana tomentosa* genomic region affecting leaf number and correlated traits in *Nicotiana tabacum*. *Theoretical and Applied Genetics* **114**:841-854
- Li JZ, Sjakste TG, Röder MS, Ganai MW** (2003) Development and genetic mapping of 127 new microsatellite markers in barley. *Theoretical and Applied Genetics* **107**:1021-1027
- Li ZK, Fu BY, Gao YM, Xu JL, Ali J, Lafitte HR, Jiang YZ, Rey JD, Vijayakumar CHM, Maghirang R, Zheng TQ, Zhu LH** (2004) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). *Plant Molecular Biology* **59**(1):33-52
- Li C, Bai Y, Jacobsen E, Visser R, Lindhout P, Bonnema G** (2006) Tomato defense to the powdery mildew fungus: differences in expression of genes in susceptible, monogenic- and polygenic resistance responses are mainly in timing. *Plant Molecular Biology* **62**(1-2):127-140.
- Li C, Bonnema G, Che D, Dong L, Lindhout P, Visser R, Bai Y** (2007) Biochemical and molecular mechanisms involved in monogenic resistance responses to tomato powdery mildew. *Molecular Plant and Microbe interactions* **20** (9):1161-1172
- Liebhart R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C** (2002) Development and characterization of 140 new microsatellites in apple (*Malus × domestica* Borkh.). *Molecular Breeding* **10**:217-241
- Lin J, Davis TM** (2000) S1 analysis of long PCR heteroduplexes: detection of chloroplast InDel polymorphisms in *Fragaria*. *Theoretical and Applied Genetics* **101**: 415-420
- Lin SY, Sasaki T, Yano M** (1998) Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. *Theoretical and Applied Genetics* **96**:997-1003
- Liu J, van Eck J, Cong B, Tanksley SD** (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the National Academy of Science of the USA* **99**:13302-13306
- Liu J, Cong B, Tanksley SD** (2003) Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the clones quantitative trait locus *fw2.2* controls fruit size. *Plant Physiology* **132**:292-299
- Llácer G** (2005) Problemática actual de la mejora genética de frutales en España. *Información Técnica Económica Agraria (ITEA)*, **101**(4):364-372
- Llop.Tous I, Dominguez-Puigjaner E, Palomer X, Vendrell M** (1999) Characterization of two divergent endo-beta-1,4-glucanase cDNA clones highly expressed in the nonclimateric strawberry fruit. *Plant Physiology* **119**(4):1415-1421

- López-Aranda JM, Miranda L, Romero F, De Los Santos B, Soria C, Medina JJ, Montes F, Vega JM, Páez JJ, Bascón J., et al.** (2004) Main results of trial son methyl bromide alternatives for strawberry fruit and runners produced in Spain. Proceedings of the Fifth International Conference on Alternatives to Methyl Bromide. September 27-30, Lisbon (Portugal). Pp 35-40
- López-Aranda JM** (2008) El cultivo de la fresa en Huelva. En: La Fresa de Huelva. Ed: Junta de Andalucía. Consejería de Agricultura y Pesca.
- Luo M, Wing RA** (2003) An Improved Method for Plant BAC Library Construction. *Plant Functional Genomics. Ed Erich Grotewoldl. Series: Methods in Molecular Biology*, **236**(3):19. 15
- Mackay MFC** (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**:303-339
- Marcel TC, Aghnoum R, Durand J, Varshney RK, Niks RE** (2007) Dissection of the barley 2L1.0 region carrying the 'Laevigatum' quantitative resistance gene to leaf rust using Near Isogenic Lines (NIL) and sybNIL. *Molecular Plant-Microbe Interactions* **20**(12):1604-1615
- Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitzel NO, Hillier L, Kwok PY, Gish WR** (1999) A general approach to single-nucleotide polymorphism discovery. *Nature Genetics* **23**:452-456
- Martins WS, Lucas DCS, Neves KFS, Bertoli DJ** (2009) WebSat - A Web Software for MicroSatellite Marker Development. *Bioinformatics* 2009, **3**(6):282-283
- Masia A, Zanchin A, Rascio N, Ramina A** (1992) Some biochemical and ultrastructural aspects of peach fruit-development. *Journal of the American Society of Horticultural Science* **117**:808-815
- Matus I, Corey A, Filichkin T, Hayes PM, Vales MI, Kling J, Riera-Lizarazu O, Sato K, Powell W, Waugh R** (2003) Development and characterization of recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. *Genome* **46**:1010-1023
- McClellan PE, Mamidi S, McConnell M, Chikara S, Lee R** (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics*, **11**:184
- Medina JJ** (2008) Origen del cultivo: Un pionero. En: *La fresa de Huelva*. Junta de Andalucía (Ed). Pp.17-22
- Melchinger AE, Utz HF, Schön CC** (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* **149**:383-403
- Melchinger AE, Piepho HP, Utz HF, Muminović J, Wegenast T, Törjék O, Altmann T, Kusterer B** (2007) Genetic basis of heterosis for growth-related traits in *Arabidopsis* investigated by testcross progenies of Near-Isogenic Lines reveals a significant role of epistasis. *Genetics* **177**:1827-1837
- Mezzetti B** (2009) GMO strawberry: Methods, Risk and Benefits. In: *Genetics and Genomics of Rosaceae*, Plant Genetics and Genomics: Crops and Models 6. Eds: Folta KM, Gardiner SE. Springer, NY, USA.
- Michaels SD, Amasino RMA** (1998) A robust method for detecting single nucleotide changes as polymorphic markers by PCR. *Plant Journal* **14**:381-385
- Miller ER, Apple LJ, Risby TH** (1998) Effect of dietary patterns on measures of lipid peroxidation. Results from a randomized clinical trial. *Circulation* **98**:699-710
- Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR, Hide WA** (1999) A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. *Genome Research* **9**:1143- 1155

- Moing A** (2000) Sugar alcohols as carbohydrate reserves in some higher plants. In: Gupta AK, Kaur N (eds) *Carbohydrate reserves in plants—synthesis and regulation*, vol 26. Elsevier Ludhiana, India, p 380
- Moing A, Svanella L, Rolin D, Gaudillere M, Gaudillere JP, Monet R** (1998) Compositional changes during the fruit development of two peach cultivars differing in juice acidity. *Journal of the American Society of Horticultural Science* **123**:770–775
- Monet R, Guye A, Roy M, Dachary N** (1996) Peach Mendelian genetics: a short review and new results. *Agronomie* **16**:321–329
- Monfort A, Vilanova S, Davis TM, Arus P** (2006) A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid *Fragaria* species and to *Fragaria* × *ananassa*. *Molecular Ecology Notes* **6**:197–200
- Monfort A, Aranzana MJ, Hidalgo MA, Sánchez D, Arús P** (2008) Phenotyping fruit nutritional quality parameters in the INOTALIS collection of strawberry genotypes. VI International strawberry symposium ISHS. March 2008. Huelva, Spain.
- Monforte AJ, Tanksley SD** (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome* **43**:803–813
- Monforte AJ, Tanksley SD** (2000b) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. *Theoretical and Applied Genetics* **100**:471–479
- Monforte AJ, Friedman E, Zamir D, Tanksley SD** (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization. *Theoretical and Applied Genetics* **102**:572–590
- Moreno E, Obando JM, Dos-Santos Noelia, Fernández-Trujillo JP, Monforte AJ, García-Mas J** (2008) Candidate genes and QTLs for fruit ripening and softening in melón. *Theoretical and Applied Genetics*. DOI 10.1007/s00122-007-0694-y
- Morgan TH**, (1911) Random segregation versus coupling in Mendelian inheritance. *Science* **36**: 718–719.
- Moyle LC, Graham EB** (2006) Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. *Molecular Biology and Evolution* **23**(5):973–980
- Naruhashi, N, Iwata, T, (1988)**. Taxonomic re-evaluation of *Fragaria nipponica* Makino and allied species. *Journal of Phytogeography and Taxonomy* **36**: 59–64
- Navarro C, Muñoz-Garmendia** (2005) Flora Ibérica. Plantas vasculares de la Península Ibérica e Islas Baleares, Vol 6. Real Jardín Botánico. Consejo Superior de Investigaciones Científicas, CISC. Madrid, Spain. Pp. 88–93.
- Neff MM, Neff JD, Chory J, Pepper AE** (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant Journal* **14**:387–392
- Nickerson DA, Tobe VO, Taylor SL** (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research* **25**:2745–2751
- Nier S, Simpson DW, Tobutt KR, Sargent DJ** (2006) Construction of a genetic linkage map of an interspecific diploid *Fragaria* BC1 mapping population (*F. vesca* 815 × [*F. vesca* 815 × *F. viridis* 903]) and its comparison to the *Fragaria* reference map (FVxFN). *Journal of Horticultural Science & Biotechnology* **81**:645–650
- Nitsch JP** (1959) Growth and morphogenesis of the strawberry as related to auxin. *Plant Physiology* **37**:211–215

- Nuez F** (2000) Introducción al análisis del ligamiento y cartografía de marcadores simples. En: Los marcadores genéticos en la mejora vegetal. Nuez F, Carrillo JM (eds). Universidad Politécnica de Valencia, Valencia, Pp. 182
- Ogundiwin EA, Marti C, Forment J, Pons C, Granell A, Gradziel TM, Peace CP, Crisosto CH** (2008a) Development of ChillPeach genomic tools and identification of cold-responsive genes in peach fruit. *Plant Molecular Biology* **68**:379-397
- Ogundiwin EA, Peace CP, Nicolet C, Rashbrook C, Gradziel T, Bliss F, Parfitt D Crisosto C** (2008b) Leucoanthocyanidin dioxygenase gene (PpLDOX): A potential functional marker for cold storage browning in peach. *Tree Genetics and Genomes* **4**(3):543-554
- Oikeh SO, Menkir A, Maziva-Dixon B, Welch RM, Galph RP, Gauch G** (2004) Environmental stability of iron and zinc concentrations in grain of elite early-maturing tropical maize genotypes grown under field conditions. *The Journal of Agricultural Science* **142**:543-551
- Olbricht K, Grafe C, Dunemann F, Würzburg F, Drewes-Alvarez R, Hoberg E, Ulrich, D.** (2009). Breeding potential of *Fragaria vesca* F. *Alba* in respect to volatile compounds. *Acta Horticulturae* 814:561-566
- OMS (Organización Mundial de la salud)** (2010) Estadísticas sanitarias mundiales 2010. Biblioteca de la OMS (Francia).
- Oosumi T, Gruszewski HA, Blischak LA, Baxter AJ, Wadl PA, Shuman JL, Veilleux RE, Shulaev V** (2006) High-efficiency transformation of the diploid strawberry (*Fragaria vesca*) for functional genomics. *Planta* **223**:1219-1230
- Overy SA, Walker HJ, Malone S, Howard TP, Baxter CJ, Sweetlove LJ, Hill SA, Quick WP,** (2005) Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. *Journal of Experimental Botany* **56**:287-296
- Palmieri L, Saviane A, Sordo M, Virzi A, Grando MS, Giongo L** (2008) Characterization of *Fragaria vesca* L. genotypes and through the use of SSR, SCAR and SNP markers. In: *Book of Abstracts VI International Strawberry Symposium ISHS*. Huelva, Spain, 3-7 March 2008. CAP Junta de Andalucía. 2008. Pp 173
- Paran I, Michelmore RW,** (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics* **85**: 985-993
- Paran I, Zamir D** (2003) Quantitative traits in plants: beyond the QTL. *Trends in Genetics* **19**(6):303-306
- Paterson AH, DeVerna JW, Lanini B, Tanksley SD** (1990) Fine Mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. *Genetics* **124**:735-742
- Pea G, Paulinesandra P, Canè MA, Savo-Sardaro ML, Landi P, Morgante M, Porceddu E, Pè ME, Frascaroli E** (2009) Recombinant near-isogenic lines: a resource for the mendelization of heterotic QTL in maize. *Molecular Genetics and Genomics* **281**:447-457
- Peace CP, Crisosto CH, Gradziel TM** (2005) Endopolygalacturonase: a candidate gene for Freestone and Melting flesh in peach. *Molecular Breeding* **16**:21-31
- Pérez-Pérez JM, Esteve-Bruna D, Micol JL** (2010) QTL analysis of leaf architecture. *Journal of Plant Research* **123**:15-23
- Perkins-Veazie P** (1995) Growth and ripening of strawberry fruit. *Horticultural Reviews* **17**:267-297
- Pestsova EG, Börner A, Röder MS** (2001) Development of a set of *Triticum aestivum* – *Aegilops tauschii* introgression lines. *Hereditas* **135**(2-3):139-143

- Picañol R, Howad W, Alegre S, Batlle I, Arús P** (2007) Introgressing genes from new sources into peach using whole-genome selection: Analysis of a large BC1 population (almond x peach)x peach. In: Twelfth Eucarpia Symposium on Fruit Breeding and Genetics. Book of Abstracts.
- Polans NO, Weeden NF, Thompson WF** (1985) Inheritance, organization and mapping of *rbcS* and *cab* multigene families in pea. *Proceedings of the National Academy of Sciences of the USA*. **82**:5083-5087
- Pontaroli AC, Rogers RL, Zhang Q, Shields ME, Davis TM, Folta KM, SanMiguel P, Bennetzen JL** (2009) Gene content and distribution in the nuclear genome of *Fragaria vesca*. *The Plant Genome*. **2**(1):93-101
- Poormohammad Kiani S, Grieu P, Maury P, Hewezi T, Gentzbittel L, Sarrafi A** (2007) Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, **114**(2):193-207
- Potter D, Luby, JJ, Harrison RE** (2000) Phylogenetic relationships among species of *Fragaria* (Rosaceae) inferred from non-coding nuclear and chloroplast DNA sequences. *Systematic Botany* **25**: 337–348
- Qin Y, Teixeira da Silva JA, Zhang L, Zhang S** (2008) Transgenic strawberry: State of the art for improved traits. *Biotechnology advances* **26**:219-232
- Quilot B, Wu BH, Kervella J, Genard M, Foulongne M, Moreau K** (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. *Theoretical and Applied Genetics* **109**:884-897
- Raman H, Zhang K, Clakir M, Appels R, Garvin DF, Maron LG, Kochian LV, Moroni JS, Raman R, Imtiaz M, Drake-Brockman F, Waters I, Martin P, Sasaki T, Yamamoto Y, Matsumoto H, Hebb DM, Delhaize E, Ryan PR** (2005) Molecular characterization and mapping of AL;T1, the aluminium-tolerance gene of bread wheat (*Triticum aestivum* L.) *Genome* **48**(5):781-791
- Ramsay L, Macaulay M, degli Ivanisovich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvevsson S, Morgante M, Massari M, Maestri E, Marmiroli N, Sjakste T, Ganai M, Powell W, Waugh W** (2000) A simple sequence repeat-based linkage map of barley. *Genetics* **156**:1997-2005
- Ramsay LD, Jennings DE, Kearsey MJ, Marshall DF, Bohuon EJ, Arthur AE, Lydiate DJ** (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. *Genome* **39**(3):558-567
- Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R, Abbott AG** (1995) Genetic-linkage mapping in peach using morphological, RFLP and RAPD markers. *Theoretical and Applied Genetics* **90**:503-510
- Reif JC, Kusterer B, Piepho HP, Meyer RC, Altmann T, Schön CC, Melchinger AE** (2009) Unraveling epistasis with triple testcross progenies of Near-Isogenic Lines. *Genetics* **181**:247-257
- Ricardo VG, Ricci JCD, Hernández L, Castagnaro AP** (2006) Enhanced resistance to *Botrytis cinerea* mediated by the transgenic expression of the chitinase gene *ch5B* in strawberry. *Transgenic Research* **15**:57-68
- Richardson, CW** (1914). A preliminary note on the genetics of *Fragaria*. *Journal of Genetics* **3**: 171-177
- Richardson, CW** (1918). A further note on the genetics of *Fragaria*. *Journal of Genetics* **7**: 167-170
- Richardson, CW** (1920). Some notes on *Fragaria*. *Journal of Genetics* **10**: 39-46
- Richardson, CW** (1923). Notes on *Fragaria*. *Journal of Genetics* **13**: 147-152

- Roberts MR** (2003) 14-3-3 proteins find new partners in plant cell signalling. *Trends in Plant Sciences* **8**:218-223
- Rokas A, Williams B, King N, Carroll S** (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**:798-804
- Rong J, Bowers JE, Schulze SR, Waghmare VN, Rogers CJ, Pierce GJ, Zhang H, Estill JC, Paterson AH** (2005) Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. *Genome Res*, **15**:1198-1210
- Rothan C, Etienne C, Moing A, Dirlwanger E, Raymond P, Monet R** (1999) Plant gene register PGR 99-126. Isolation of a cDNA encoding a metallothionein-like protein (accession no. AJ243532) expressed during peach fruit development. *Plant Physiology* **121**:311 (Electronic Plant Gene Register)
- Rousseau-Gueutin M, Gaston A, Ainouche A, Ainouche ML, Olbricht K, Staudt d G, Richard L, Denoyes-Rothan B** (2009) Tracking the evolutionary history of polyploidy in *Fragaria L.* (strawberry): 3 New insights from phylogenetic analyses of low-copy nuclear genes. *Molecular Phylogenetics and Evolution* **51**:515–530
- Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson DW, Arús P, Guérin G, Denoyes-Rothan B** (2008) Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. *Genetics* **179**:2045–2060
- Rozen S, Skaletsky HJ** (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386
- Ruiz-Rojas JJ, Sargent DJ, Shulaev V, Dickerman AW, Pattison J, Holt SH, Ciordia A, Veilleux RE** (2010) SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca L.* *Theoretical and Applied Genetics*. DOI 10.1007/s00122-010-1322-9
- Saha S, Jenkins JN, Wu J, McCarty JC, Gutiérrez OA, Percy RG, Cantrell RG, Stelly DM** (2006) Effects of chromosome-specific introgression in upland cotton on fiber and agronomic traits. *Genetics* **172**:1927-1938
- Saladié M, Matas JA, Isaacson T, Jenks MA, Googwin M, Niklas KJ, Xiaolin R, Labavitch JM, Shackel KA, Fernie AR, Lytovchenko A, O'Neill MA, Watkins CB, Rose JCK** (2007) A reevaluation of the key factors that influence tomato fruit softening and integrity. *Plant Physiology* **144**:1012-1028
- Salentijn EMJ, Aharoni A, Schaart JG, Boone MJ, Krens FA** (2003) Differential gene expression analysis of strawberry cultivars that differ in fruit-firmness. *Physiologia Plantarum* **118**: 571–578
- Sambrook J, Fritsch E, Maniatis T** (1989) *Molecular cloning: A laboratory manual* 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sánchez-Pérez R, Howad W, Dicenta F, Arús P, Martínez-Gómez P** (2007) Mapping major genes and quantitative trait loci controlling agronomic traits in almond. *Plant breeding* **126**:310-318
- Sanger F, Nicklen S, Coulson R** (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**(12):5463-5467
- Sargent DJ, Hadonou AM, Simpson DW** (2003) Development and characterisation of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry. *Molecular Ecology Notes* **3**: 550–552
- Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH, Simpson DW** (2004a) A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid *Fragaria*. *Theoretical and Applied Genetics* **109**:1385-1391.

- Sargent DJ, Geibel M, Hawkins JA, Wilkinson MJ, Battey NH, Simpson DW** (2004b) Quantitative and qualitative differences in morphological traits revealed between diploid *Fragaria* species. *Annals of Botany* **94**: 787–796
- Sargent DJ** (2005) A genetic investigation in diploid *Fragaria*. PhD Thesis. School of Plant Sciences, the University of Reading.
- Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arús P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH** (2006) An enhanced microsatellite map of diploid *Fragaria*. *Theoretical and Applied Genetics* **112**:1347-1359
- Sargent DJ, Rys A, Nier S, Simpson DW, Tobutt KR** (2007) The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. *Theoretical and Applied Genetics* **114**: 373–384
- Sargent DJ, Cipriani G, Vilanova S, Gil-Ariza D, Arús P, Simpson DW, Tobutt KR, Monfort A** (2008) The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map. *Genome* **51**:120-127
- Sargent DJ, Davis TM, Simpson DW** (2009a) Strawberry (*Fragaria spp.*) Structural Genomics. In: Folta KM, Gardiner SE (eds). *Genetics and Genomics of Rosaceae*. Springer, New York, NY. Pp.443.
- Sargent DJ, Fernández-Fernández F, Ruiz-Rojas JJ, Southerland BG, Passey A, Whitehouse AB, Simpson DW**, (2009b) A genetic linkage map of the cultivated strawberry (*Fragaria x ananassa*) and its comparison to the diploid *Fragaria* reference map. *Molecular Breeding* **24**:293-303
- Sargent DJ, Marchese A, Simpson DW, Howad W, Fernández- Fernández F, Monfort A, Arús P, Evans KM, Tobutt KR** (2009c) Development of “universal” gene-specific markers from *Malus* spp. cDNA sequences, their mapping and use in synteny studies within Rosaceae. *Tree Genetics and Genomes* **5**:133-145
- Satish K, Srinivas G, Madhusudhana R, Padmaja PG, Nagaraja-Reddy R, Murali Mohan S, Seetharama N** (2009) Identification of quantitative trait loci for resistance to shoot fly in sorghum [*Shorghum bicolor* (L.) Moench] *Theoretical and Applied Genetics* **119**(8):1425-1439
- Sax K** (1923) The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**:552-560
- Sears ER, Loegering WQ, Rodenhiser HA** (1957) Identification of chromosomes carrying genes for stem-rust resistance in four varieties of wheat. *Agronomy Journal* **49**:208-212
- Semel YJ, Nissenbaum N, Menda M, Zinder U, Krieger Issman N, Pleban T, Lippman Z, Gur A, Zamir D** (2006) Overdominant quantitative trait loci for yield and fitness in tomato. *Proceedings of the National Academy of Sciences of the USA*, **103**(35):12981-12986
- Senter SD, Callahan A** (1990) Variability in the quantities of condensed tannins and other major phenols in peach fruit during maturation. *Journal of Food Science* **55**:1585-1587
- Shan JX, Zhu MZ, Shi M, Gao JP, Hin HX** (2009) Fine mapping and candidate gene analysis of *spd6*, responsible for small panicle and dwarfness in wild rice (*Oryza rufipogon* Griff.) *Theoretical and Applied Genetics* **119**(5):827-836
- Shimomura K, Hirashima K** (2006) Development and characterization of simple sequence repeats (SSR) as markers to identify strawberry cultivars (*Fragaria x ananassa* Duch.) *Journal of the Japanese Society for Horticultural Science* **75**: 399–402
- Shiratake K, Martinoia E** (2007) Transporters in fruit vacuoles. *Plant Biotechnology* **24**:127-133

- Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arus P, Dandekar AM, Lewers K, Brown SK, Davis TM, Gardiner SE, Potter D, Veilleux RE** (2008) Multiple models for Rosaceae genomics. *Plant Physiology* **147**:985-1003
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler T, Veilleux RE, Folkerts O, L. Delcher AL, Jaiswal P, Liston A, Mane SP, Burns P, Mockaitis K, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JC, Rees DJG, Williams K, Holt SH, Dickerman A, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Filichkin AAS, Velasco R, Troggio M, Viola R, Borodovsky M, Ashman TL, Aharoni A, Bennetzen J, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant Jr. DW, Fox SE, Givan SA, Naithani S, Christoffels A, Salama DY, Carter J, Lopez Girona E, Zdepki A, Wang W, Kerstetter RA, Salzberg SL, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Folta KM** (2010) Short and sweet. The genome of a diploid strawberry. Submitted to *Science*.
- Simchen G, Hugerat Y** (1993). What determines whether chromosomes segregate reductionally or equationally in meiosis? *BioEssays* **15**: 1–8
- Smith HO** (1970) Nucleotide sequence specificity of restriction endonucleases. *Science* **205**:455-462
- Soria C, Medina JJ, Sánchez-Sevilla JF, Bartual R, Refoyo A, Gálvez J, Miranda L, Villalba R, Ariza MT, López-Aranda JM** (2009) Current situation of the spanish public strawberry breeding program. In: Proceedings of the VIth International Strawberry Symposium. (Ed. López-Medina JA) *Acta Horticulturae* **842**:487-490
- Spigler RB, Lewers KS, Main DS, Ashman TL** (2008) Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity* **101**:507-517
- Staub JE, Serquen FC, Gupta M** (1996) Genetic markers, map construction, and their application in plant breeding. *HortScience* **31**:729–741
- Staudt, G** (1962). Taxonomic studies on the genus *Fragaria*, typification of *Fragaria* species known at the time of Linnaeus. *Canadian Journal of Botany* **40**: 212-218
- Staudt G** (1989) The species of *Fragaria*, their taxonomy and geographical distribution. *Acta Horticulturae* **256**:23-33
- Staudt G** (1999) *Systematics and geographical distribution of the American strawberry species: Taxonomic studies in the genus Fragaria (Rosaceae: Potentilleae)*. University of California publications in botany, USA.
- Staudt, G, Dickoré, WB, (2001)**. Notes on Asiatic *Fragaria* species: *Fragaria pentaphylla* Losinsk. and *Fragaria tibetica* spec. nov. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* **123**: 341-354
- Staudt G** (2006) Himalayan species of *Fragaria* (Rosaceae), *Botanische Jahrbücher für Systematik* **126** pp. 483–508
- Staudt G** (2008) Strawberry Biogeography, Genetics and Systematics. *Acta Horticulturae* **842**(1):71-83
- Sturtevant AH** (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology* **14**: 43-59.
- Sugimoto T, Tamaki K, Matsumoto J, Yamamoto Y, Shiwaku K, Watanabe K** (2005) Detection of RAPD markers linked to the everbearing gene in Japanese cultivated strawberry. *Plant Breeding* **124**: 498–501
- Szalma SJ, Hostert BM, LeDeaux JR, Stuber CW and Holland JB** (2007) QTL mapping with near-isogenic lines in maize. *Theoretical and Applied Genetics* (2007) **114**:1211-1228
- Takahashi M, Matsuda F, Margetic N, Lathrop M** (2003) Automated identification of single nucleotide polymorphisms from sequencing data. *Journal of Bioinformatics and Computational Biology* **1**:253-265

- Tamdor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E** (2002) Identification of malodorus, a wild species allele affecting tomato aroma that was selected against during domestication. *Journal of Agriculture and Food chemistry* **50**:2005-2009
- Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ** (2002) Simple sequence repeat map of the sunflower genome. *Theoretical and Applied Genetics* **105**:1124-1136
- Tanhuanpää P, Kalendar R, Schulman AH, Kiviharju E** (2008) The first doubled haploid linkage map for cultivated oat. *Genome* **51**:560-569
- Tanksley SD, Young ND, Paterson AH, Bonierbale MW** (1989). RFLP mapping in plant breeding: new tools for and old science. *Biotechnology* **7**: 257-264.
- Tanksley SD, Ganai MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND** (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**:1141-1160
- Tanksley SD** (1993). Mapping polygenes. *Annual Review of Genetics* **27**: 205-233
- Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed T, Petiard V, Lopez J, Beck-Bunn T** (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics* **92**:213-224
- Tanksley SD, McCouch SR** (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**:1063-1066
- Tautz D, Renz M** (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**:4127
- Taylor BA** (1978) Recombinant inbred strains: In: Morse HC, (Ed.) *Origins of Inbred Mice*. Academic Press, New York. Pp 423-438
- Tian F, Stevens NM, Buckler ES IV** (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proceedings of the National Academy of Sciences of the USA* **16**:9979-9986
- Tomas-Barberan FA, Gil MI, Cremin P, Waterhouse AL, Hess-Pierce B, Kader AA** (2001) HPLC-DAD-ESIMS analysis of phenolic compounds in nectarines, peaches, and plums. *Journal of Agricultural and Food Chemistry* **49**:4748-4760
- Toojinda T, Baird E, Booth A, Broers L, Hayes P, Powell W, Thomas W, Vivar H, Young G** (1998) Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. *Theoretical and Applied Genetics*, **96**(1):123-131
- Trainotti L, Spolaroe S, Ravanello A, Baldan B, Casadoro G** (1999) A novel E-type endo-beta-1,4-glucanase with a putative cellulose-binding domain is highly expressed in ripening strawberry fruits. *Plant Molecular Biology* **40**(2):323-332
- Tsuda T, Yamaguchi M, Honda C, Moriguchi T** (2004) Expression of anthocyanin biosynthesis genes in the skin of peach and nectarine fruit. *Journal of the American Society of Horticultural Sciences* **129**:857-862
- Tucker GA** (1993) Introduction. In: Seymour GB, Taylor JE, Tucker GA (eds) *Biochemistry of fruit ripening*. Chapman & Hall, London, pp 3-43
- Ulrich D, Komes D, Olbricht K, Hoberg E** (2007) Diversity of aroma patterns in wild and cultivated *Fragaria* accessions. *Genetic Resources and Crop Evolution* **54**:1185-1196

- Urbanczyk-Wochniak E, Usadel B, Thimm O, Nunes-Nesi A, Carrari F, Davy M, Blasing O, Kowalczyk M, Weicht D, Polinceusz A, Meyer S, Stitt M, Fernie AR** (2006) Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. *Plant Molecular Biology* **60**:773-792
- Van der Knaap E, Lippman ZB, Tanksley SD** (2002) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. *Theoretical and Applied Genetics* **104**:241-247
- Van der Knaap E, Tanksley SD** (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in yellow stuffer tomato. *Theoretical and Applied Genetics* **107**:139-147
- Van Ooijen JW** (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theoretical and Applied Genetics* **84**:803-811
- Van Ooijen JW**, (2006) Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.
- Varshney RK, Graner A, Sorrells ME** (2005) Genic microsatellite markers in plants: features and applications. *TRENDS in Biotechnology* **23**, 48-55
- Vilanova S, Arús P, Sargent DJ, Monfort A** (2008) Synteny conservation between two distantly-related Rosaceae genomes: Prunus (the stone fruits) and Fragaria (the strawberry). *BMC Plant Biology* **8**:67
- Viruel MA, Sánchez D, Arús P** (2002) An SSR and RFLP linkage map for the octoploid strawberry (*Fragaria × ananassa*). In: Plant, Animal and Microbe Genomes. 10th Conference, San Diego, California, USA. http://www.intl-pag.org/10/abstracts/PAGX_P660.html
- Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD** (2000) Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics* **155**:407-420
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M** (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**(21):4407-4414
- Wadman M** (2008) James Watson's genome sequenced at high speed. *Nature* **452**: 788
- Waldron J, Peace CP, Searle IR, Furtado A, Wade N, Graham MW, Carroll BJ** (2002). Randomly amplified DNA fingerprinting (RAF): a culmination of DNA marker protocols based on arbitrarily primed PCR. *Journal of Biomedicine and Biotechnology* **2**: 141-150
- Wan XY, Wan JM, Su CC, Want CM, Shen WB, Li JM, Wang HL, Jiang L, Liu SJ, Chen LM, Yasui H, Yoshimura A** (2004) QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. *Theoretical and Applied Genetics* **110**:71-79
- Warburton ML, Becerra-Velasquez VL, Goffreda JC, Bliss FA** (1996) Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach. *Theoretical and Applied Genetics* **93**:920-925
- Warrington CV, Zhus S, Parrot WA, All JN, Boerma HR** (2008) Seed yield of near-isogenic soybean lines with introgressed quantitative trait loci conditioning resistance to corn earworm (Lepidoptera: Noctuidae) and soybean looper (Lepidoptera: Noctuidae) from PI 229358. *Journal of Economic Entomology* **101**(4):1471-1477
- Weber JL** (1990) Informativeness of human (dC-dA)n. (dG-dT)n polymorphisms. *Genomics* **7**, 524-530
- Weebadde CK, Wang D, Finn CE, Lewers KS, Luby JJ, Bushakra J, Sjulín TM, Hancock JF** (2008) Using a linkage mapping approach to identify QTL for day-neutrality in the octoploid strawberry. *Plant Breeding* **127**: 94-101
- Wehrhahn C, Allard RW** (1965) The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. *Genetics* **51**:109-119

- Williamson SC, Yu H, Davis TM** (1995) Shikamate deshydrogenase allozymes: Inheritance and close linkage to fruit color in the diploid strawberry. *Journal of Heredity* **86**:74-76
- Wu F, Tanksley SD** (2010) Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* **11**:182
- Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD** (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* **83**:294-300
- Wu SB, Tavassolian I, Rabiei G, Hunt P, Wirthensohn M, Gibson JP, Ford CM, Sedgley M** (2009) Mapping SNP-anchored genes using high-resolution melting analysis in almond. *Molecular Genetics and Genomics* **282**:273-281
- Wu SB, Wirthensohn MG, Hunt P, Gibson JP, Sedgley M** (2008) High resolution melting analysis of almond SNPs derived from ESTs. *Theoretical and Applied Genetics* **118**:1-14
- Wyman AR, White R** (1980) A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences of the USA*, **77**:6754-6758.
- Xu M, Song J, Cheng Z, Jiang J, Korban SS** (2001) A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 and contig construction for positional cloning of the apple scab resistance gene *Vf*. *Genome*, **44**:1104-1113
- Xu K, Xu X, Fulkao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ** (2006) *Sub1A* is an ethylene-response-factor like gene that confers submergence tolerance to rice. *Nature* **442**:705-708
- Yang L, Jin G, Zhao X, Zheng Y, Xu Z, Wu W** (2007) PIP: a database of potential intron polymorphism markers. *Bioinformatics* **23**:2174-2177
- Yao YX, Li M, Liu Z, Hao YJ, Zha H** (2007) A novel gene, screened by cDNA-AFLP approach, contributes to lowering acidity on fruit in apple. *Plant Physiology and Biochemistry* **173**:44-54
- Yesudas CR, Sharma H, Lightfoot DA** (2010) Identification of QTL in soybean underlying resistance to herbivory by Japanese beetles (*Popilla japonica*, Newman). *Theoretical and Applied Genetics* **121**(2):353-362
- Yu H, Davis TM** (1995) Genetic linkage between runnering and phosphoglucoisomerase allozymes, and systematic distortion of monogenic segregation ratios in diploid strawberry. *Journal of American Horticultural Sciences*. **120**:687-690
- Zafra-Stone S, Yasmin T, Bagchi M, Chatterjee A, Vinson JA, Bagchi D** (2007) Berry anthocyanins as novel antioxidants in human health and disease prevention. *Molecular nutrition & food research* **51**:675-683.
- Zamir D** (2001) Improving plant breeding with exotic genetic libraries. *Nature Review. Genetics* **2**:983-989
- Zhang HB, Wu C** (2001) BAC as tools for genome sequencing. *Plant Physiology and Biochemistry* **39**:195-209
- Zhang C, Tanabe K, Wang S, Tamura F, Yoshioka A, Matsumoto K** (2006) The impact of cell division and cell enlargement on the evolution of fruit size in *Pyrus pyrifolia*. *Annals of Botany* (London) **98**:537-543
- Zhang Y, Luo L, Liu T, Xu C, Xing Y** (2009) Four rice QTL controlling number of spikelets per panicle expressed the characteristics of single Mendelian gene in near isogenic backgrounds. *Theoretical and Applied Genetics* **118**:1035-1044
- Zhang G, Sebolt AM, Sooriyapathirana SS, Wang D, Bink MCAM, Olmstead JW, Iezzoni AF** (2010) Fruit size QTL analysis of an F₁ population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. *Tree Genetics & Genomes* **6**:25-36

- Zhao T, Palotta M, Langridge P, Prasad M, Graner A, Schulze-Lefert P, Koprek T** (2006a) Mapped Ds/T-DNA launch pads for functional genomics in barley. *Plant Journal* **47**:811–826
- Zhao B, Deng Q-M, Zhang Q-J, Li J-Q, Ye S-P, Liang Y-S, Peng Y, Li P** (2006a) Analysis of segregation distortion of molecular markers in F2 population of rice. *Acta Genetica Sinica* **33**:449–457
- Zhao J, Kulkarni V, Liu N, Pino del Carpio D, Bucher J, Bonnema G** (2010) *BrFLC2* (Flowering locus C) as a candidate gene for a vernalization response QTL in *Brassica rapa*. *Journal of Experimental Botany* **61**(6):1817-1825
- Zhebentyayeva TN, Swire-Clark G, Georgi LL, Garay L, Jung S, Forrest S, Blenda AV, Blackmon B, Mook J, Horn R, Howad W, Arus P, Main D, Tomkins JP, Sosinski B, Baird WV, Reighard GL, Abbott AG** (2008) A framework physical map for peach, a model Rosaceae species. *Tree Genetics and Genomes* **4**:745–756
- Zhou R, Zhu Z, Kong X, Huo N, Tian Q, Li P, Jin C, Dong Y, Jia J** (2005) Development of wheat near-isogenic lines for powdery mildew resistance. *Theoretical and Applied Genetics* **110**:640–648
- Zorrilla-Fontanesi Y, Cabeza A, Torres A, Botella MA, Valpuesta V, Monfort A, Sánchez-Sevilla JF, Amaya I** (2010a) Development and bin mapping of strawberry genic-SSRs in diploid *Fragaria* and their transferability across the Rosoideae subfamily. *Molecular Breeding* DOI 10.1007/s11032-010-9417-1
- Zorrilla-Fontanesi Y, Cabeza A, Domínguez P, Medina JJ, Valpuesta V, Denoyes-Rothan B, Sánchez-Sevilla FJ, Amaya I** (2010b) Identification of QTLs for agronomical and fruit quality traits in strawberry (*Fragaria x ananassa*) 28th international horticultural congress (IHC 2010). Book of Abstracts. Lisboa (Portugal).

ANEXOS

ANEXO II. ABREVIACIONES

AFLP: *Amplified Fragment Length Polymorphism*. Polimorfismo de longitud de fragmentos amplificados.

BLAST: *Basic Local Alignment Search Tool*. Herramienta de búsqueda de alineamientos de tipo local.

BC: *Backcross*. Retrocruzamiento

BAC: *Bacterial Artificial Chromosome*. Cromosoma artificial bacteriano.

BIL: *Backcross Inbred Line*. Línea de introgresión por retrocruzamiento.

BSA: *Bulked-Segregant Analysis*. Análisis de mezclas de ADN segregantes.

CAPS: *Cleaved Amplified Polymorphic Sequence*. Secuencia polimórfica amplificada y fragmentada.

CG: *Candidate Gene*. Gen candidato

CHEF gel: *Clamped homogenous electric fields*. Gel de campo eléctrico homogéneo fijo.

CRAG: *Centre de Recerca en Agrigenòmica*. Centro de investigación en Agrigenómica.

CSSL: *Chromosome Segment Substitution Line*. Línea de sustitución de segmentos cromosómico.

cM: centiMorgan.

DAB: *Days After Blooming*. Días después de la floración.

DAF: *DNA Amplification Fingerprinting*. Identificación genética mediante amplificación de DNA.

dCAPS: *Derived Cleaved Amplified Polymorphic Sequence*. Secuencia derivada de un CAPS.

DHL: *Double Haploid Lines*. Líneas doble-haploides.

EMBL: *European Molecular Biology Laboratory*. Laboratorio europeo de Biología Molecular.

eQTL: *Expression QTL*. QTL de expresión.

EST: *Expressed Sequence Tag*. Marcador de secuencia expresada.

FAO: *Food and Agricultural Organization of the United Nations*. Organización para los alimentos y la agricultura de las Naciones Unidas.

FVxFB: Población de referencia *Fragaria vesca* x *Fragaria bucharica*.

FG: *Fragaria linkage group*. Grupo de ligamiento de *Fragaria*.

GL: Grupo de ligamiento

hiTAIL-PCR: *High-Efficiency Thermal Asymmetric Interlaced PCR*. PCR termal de entrelazado asimétrico de alta eficiencia.

HRM: *High Resolution Melting*. Hibridación de alta resolución.

Indel: *Insertion/Deletion polymorphism*. Polimorfismo de inserción/delección.

IRTA: *Institut de Recerca i Tecnologia Agroalimentàries*. Instituto de Investigación y Tecnología Agroalimentarias.

ISSR: *Inter-Simple Sequence Repeat*. Inter-microsatélite.

JxF: Población de referencia 'Ferjalou-Jalousia[®]' x 'Fantasia'

LD: *Linkage disequilibrium*. Desequilibrio de ligamiento.

LG: *Linkage group*. Grupo de ligamiento

LOD: *Log Of Odds*. Logaritmo de probabilidades.

- MAS:** *Marker Assisted Selection*. Selección asistida por marcadores.
- NCBI:** *National Center for Biotechnology Information*. Centro Estadounidense de información biotecnológica.
- NIL:** *Near Isogenic Line*. Línea casi isogénica.
- NOAA:** *National Oceanic and Atmospheric Administration*. Administración oceánica y atmosférica nacional.
- PCR:** *Polymerase Chain Reaction*. Reacción en cadena de la polimerasa.
- PFGE:** *Pulse Field Gel Electrophoresis*. Electroforesis en gel sobre campo pulsante.
- PG:** *Prunus linkage group*. Grupo de ligamiento de *Prunus*.
- QTL:** *Quantitative Trait locus*. Locus de carácter cuantitativo.
- RAF:** *Randomly amplified DNA Fingerprinting*. Identificación genética de DNA amplificado aleatoriamente.
- RAPD:** *Random Amplified Polymorphic DNA*. DNA polimórfico amplificado aleatoriamente.
- RCSL:** *Recombinant chromosome substitution line*. Línea de sustitución de cromosomas recombinantes.
- RE:** *Restriction enzyme*. Enzima de restricción.
- RIL:** *Recombinant Inbred Line*. Línea recombinante consanguínea.
- RNase:** *Ribonuclease*. Ribonucleasa.
- RFLP:** *Restriction Fragment Length Polymorphism*. Polimorfismo de longitud de fragmentos de restricción.
- SCAR:** *Sequence Characterized Amplified Region*. Región amplificada de secuencia caracterizada.
- SDRF:** *Single Dose Restriction Fragment*. Fragmento de restricción de herencia dominante
- SL:** *Substitution Line*. Línea de sustitución.
- SNP:** *Single Nucleotide Polymorphism*. Polimorfismos de sustitución nucleotídica.
- SSD:** *Single Seed Descendent*. Método de semilla única.
- SSR:** *Simple Sequence Repeat*. Secuencia repetida en tandem, (microsatélite).
- STS:** *Sequence-specific tag*. Marcador de secuencia específica.
- TRD:** *Transmission Ratio Distortion*. Tasa de distorsión en la transmisión de marcadores.
- VBI:** *Virginia Bioinformatics Institute*. Instituto de Bioinformática de Virginia.
- YAC:** *Yeast Artificial Chromosome*. Cromosoma artificial de levadura.

ANEXO III. LISTADO DE FIGURAS Y TABLAS

Figura 1. Distribución de <i>Fragaria vesca</i>	4
Figura 2. Distribución de las especies asiáticas descritas en el género <i>Fragaria</i>	5
Figura 3. Morfología general de una planta de fresa	6
Figura 4. Producción mundial de fresa en 2007	10
Figura 5. The diploid <i>Fragaria</i> reference map FV x FB	34
Figura 6. Mapa de ligamiento FV x FB generado por Villanova <i>et al.</i> (2008)	45
Figura 7. Amplificación de ADN genómico con marcaje fluorescente	48
Figura 8. Mapa de ligamiento de fresa diploide (FV x FB) tras la localización de los marcadores desarrollados en esta investigación	58
Figura 9. Valores de distorsión de los marcadores localizados en el mapa FVxFB en esta investigación	59
Figure 10. Mapa de bins en FV x FB	62
Figure 11 Schematic outline of the plasmid pIndigo-356	76
Figure 12. Analysis of BAC clones with PFGE	76
Figure 13. Three-fold BAC DNA pooling strategy	77
Figure 14. BAC screening of an SSR locus, ARSFL007, in two different DNA bulked superpools	78
Figure 15. Genetically-anchored BAC clones	79
Figure 16. Comparison of EST and PCR product sequences for CGPPB3316 in peach	102
Figure 17. Comparison of EST and PCR product sequences for CGPPB5769 in peach	102
Figure 18. CG mapped in the peach JxF F ₂ progeny	106
Figure 19. CG Mapped in <i>Fragaria</i> FV x FB bin set	106
Figure 20. Agarose gels of the seven CAPS markers derived from CGs in peach	107
Figure 21. Agarose gels of two INDEL markers derived from CGs in strawberry	107
Figura 22. Representación esquemática del genoma de diferentes tipos de líneas de introgresión	120
Figura 23. Representación gráfica del genoma de la colección de NILs desarrollada en melón	122
Figura 24. Parentales empleados en la creación de una colección de NILs en <i>Fragaria</i>	128
Figura 25. Distribución de los SSRs analizados en las poblaciones BC1 y BC2	132
Figura 26. Detección de introgresiones en el genoma de <i>Fragaria vesca</i> con marcadores CAPS	134
Figura 27. Detección de introgresiones en el genoma de <i>Fragaria vesca</i> con marcadores SSR	135
Figura 28. Diseño experimental para la obtención de una colección de NILs en <i>Fragaria</i>	138
Figura 29. Características morfológicas de la generación F ₁	139
Figura 30. Análisis posicional de las introgresiones detectadas en la población backcross 1	141
Figura 31. Representación gráfica del genoma de los individuos seleccionados en la población BC1	142
Figura 32. Representación gráfica de los genotipos seleccionados en la población BC2	146
Figura 33. Esquema de los GLs 1 y 5 de los genotipos seleccionados en la familia 306.004	147
Figura 34. Esquema del GL1 de los individuos con genotipo seleccionado en la familia 505.011	148
Figura 35. Análisis de los individuos pertenecientes a la familia 480.021	148
Figura 36. Esquema del GL3 de los individuos con genotipo seleccionado en la familia 473.061	149

Figura 37. Esquema de los GLs 2 y 4 de los individuos con genotipo seleccionado en la familia 90.009	149
Figura 38. Esquema del GL5 de los individuos con genotipo seleccionado en la familia 480.029	150
Figura 39. Análisis de los individuos pertenecientes a la familia 460.014	151
Figura 40. Representación gráfica de los bins del GL5	155
Figura 41. Genotipos y bins generados por las 30 líneas seleccionadas para formar parte de la colección de NILs	156
Figura 42. Imágenes de las 30 líneas de introgresión seleccionadas en esta investigación	157
Figura 43. Fenotipos mostrados por algunas líneas de introgresión	158
Figura 44. Variación detectada para la longitud de los folíolos de las líneas de introgresión analizadas	162
Figura 45. Imágenes de las hojas de algunas líneas de introgresión	162
Figura 46. Imágenes de los frutos de las líneas evaluadas	164
Figura 47. Datos obtenidos para el análisis de las dimensiones longitudinal y transversal de los frutos analizados	165
Figura 48. Datos obtenidos para el análisis de la forma de los frutos analizados	166
Figura 49. Síntomas de la infección por SCV	167
Figura 50. Posición genómica de los caracteres localizados en esta investigación	173
Tabla 1. Especies del género <i>Fragaria</i> , ploidía y distribución geográfica	4
Tabla 2. Microsatellite markers developed in <i>Fragaria</i> species	30
Tabla 3. Locus name, gene annotation and map position of mapped genes in the <i>Fragaria</i> genome	32
Tabla 4. Desarrollo de cebadores específicos para la optimización de marcadores RFLP y SSR	50
Tabla 5. Datos obtenidos a partir del genotipado de los parentales y <i>Bin set</i> de la población FV x FB	53
Tabla 6. Segregación de los marcadores desarrollados en este trabajo en la población FV x FB y comparación con la segregación obtenida para los marcadores RFLP y SSR originales	55
Tabla 7. Análisis de la segregación en FV x FB de los marcadores obtenidos	57
Tabla 8. Características de los 49 <i>bins</i> de FV x FB descritos en esta investigación	61
Tabla 9. <i>Fragaria</i> molecular markers and associated BAC clones	81
Tabla 10. Peach candidate genes potentially involved in fruit quality	94
Tabla 11. Origin of the peach candidate genes selected	98
Tabla 12. Primer pairs designed from peach CG sequences	99
Tabla 13. Characterization of PCR products obtained with the CG primer pairs in peach and strawberry	104
Tabla 14. <i>Fragaria</i> -specific primers for a subset of CGs	105
Tabla 15. PCR products obtained with <i>Fragaria</i> -specific primers for a subset of CGs	105
Tabla 16. CG polymorphism and mapping in peach and diploid <i>Fragaria</i>	105
Tabla 17. Marcadores moleculares analizados en esta investigación	133
Tabla 18. Caracterización genómica de la población <i>Backcross 1</i>	140
Tabla 19. Características de las familias BC2	143
Tabla 20. Segregación de marcadores obtenida en las familias BC2	144
Tabla 21. Características de las familias generadas mediante autofecundación de las líneas BC2 seleccionadas	147
Tabla 22. Segregación de marcadores obtenida en las familias BC2	152

Tabla 23. Segregación de loci segregantes identificados en el parental <i>F. vesca</i> en los individuos F1-30, BC1, BC2 y BC2 ⊗ seleccionados para el análisis genómico final	153
Tabla 24. Periodo vegetativo de los individuos que componen la colección de líneas de introgresión	160
Tabla 25. Caracterización de la colección de líneas de introgresión para los síntomas de infección por oidio	161
Tabla 26. Caracterización de las líneas de introgresión para los síntomas de infección por SCV	167

Sólo me queda agradecer a todos los que, quizá sin saberlo, han participado en esta Tesis Doctoral.

En primer lugar quiero agradecer especialmente a Joana Ribes su colaboración en tantos momentos durante estos cuatro años y medio.

Me quedo corto si digo que esta Tesis es tan tuya como mía. Te deseo todo lo mejor.

Gracias a todos los que me habéis aconsejado y enseñado, los que habéis compartido conmigo vuestra experiencia en esto tan complicado que es hacer Ciencia en España.

Os debo mucho. Gracias a Montse Saladié, Santi Vilanova, Claudio Meneses, Sole Martos, Edu Moreno, Wim Deleu, Mourad Mnejja, Juan Vegas, Mireia González y Marta Pujol.

Gracias al Equipo Técnico del Laboratorio de Marcadores. Que nunca se os olvide que sois lo mejor del IRTA. Que tampoco lo olviden los demás.

Gracias también a las chicas de *in vitro*, que tanto me cuidan y regalan los oídos.

Gracias a todos los que han trabajado conmigo en el invernadero.

Gracias a Jesús, Victoria Framis, Gerard, Cristina Alonso, Gemma Gabarra, Josep María San Millán, Victor y Marisa.

Gracias a ‘mis chicos de prácticas’, Edgard, Silvia y David.

Gracias a todos los ‘Becarios Precarios’ que pelearon por dignificar la figura del investigador en formación,

Gracias, Ibo por regalarnos tu mes de octubre, absolutamente *priceless*. (y por tus NILs de melón, claro).

Gracias a todos los que me acogisteis tan cálidamente en Virginia Tech durante mi estancia.

Gracias a Mike Edwards y a su familia, a Richard Veilleux, Diego Cortés, Juan J. Ruiz Rojas, Norma C Manrique, Joel Shuman, Jenny Armenta. Kerri Mills, Bob, Ana Martins, Mihaela Babiceanu y por supuesto, a Judit Pampalona.

El nombre de Dan Sargent aparece en más de 70 ocasiones en este trabajo.

Dan, GRACIAS. Por tu generosidad, por tus consejos, por tu ayuda y por tu amistad.

Gracias a todos los que habéis confiado en mí. A los que habéis reído, llorado y sudado la camiseta conmigo durante estos años.

Gracias a Ali Essafi, Celia Guiu, Vane Alfaro, Puri, Dani Alonso, Jony Vega, Roger Picañol, Álvaro Moreno,

Fuensi García, Ana Sese, Celine Humbert, Andrea Blas, Michael Bourgeois, Gisela Mir, Angel Montejo, Montse Martín, Wener Howad, Elena López, Mohamed Fergany, Meihong Xu Raúl Tonda y Carmen Vilatoro.

Nunca olvidaré tantos buenos momentos compartidos con vosotros.

Gracias al equipo administrativo del IRTA y del CRAG por su ayuda, especialmente a Noemí Vila.

Gracias también a los ‘jefes’, Pere Arús, Pere Puigdomènech y Jordi García-Mas. Ha sido un placer formar parte de este equipo.

Gracias a M. Carmen Martínez por aceptar la tutoría de esta Tesis.

Gracias a los miembros del Tribunal por haber tenido a bien participar en él.

Buena suerte a todos los que empezáis nueva etapa en Bellaterra. Me voy sabiendo que esa, también es mi casa.

Y suerte también a todos los que, como yo, abandonáis el barco, Alí, Eudald, Juan, Álvaro, Claudio y Andrej

Gracias, Amparo Monfort, por tu confianza, tu optimismo, tu compromiso y tu empatía. Ha sido un lujo trabajar contigo.

He aprendido muchísimo contigo y espero haber estado a la altura.

Ojalá algún día pueda parecerme un poquito a ti.

Gracias a todos los amigos que me habéis levantado el ánimo cuando lo he necesitado.

Gracias, Leo por el tiempo robado, por estar siempre ahí y por todo lo demás.

Gracias a mi familia, por apoyarme siempre.