



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

---

# **Genetics and genomics of cultivated eggplants and wild relatives**

---

*PhD dissertation by  
Pietro Gramazio*

*Advisors  
Jaime Prohens Tomás  
Santiago Vilanova Navarro*

*Valencia, February 2018*



*“Nothing in life is to be feared, it is only to be understood.  
Now is the time to understand more, so that we may fear less.”*

**Maria Salomea Skłodowska-Curie**



## **Agradecimientos**

Aunque parezcan mucho menos, ya siete son los años que llevo en el COMAV. Los años pasan volando cuando el ambiente de trabajo es estimulante, constructivo, apasionante, gratificante, motivador, intrigante, creativo, integrador, cooperativo, productivo, positivo, divertido..., familiar.

Todo empezó aquella tarde del 10 de noviembre del 2010, cuando **Lola Raigón**, a quien va mi eterna gratitud, me presentó la persona que iba a cambiar mi vida, Jaime Prohens.

**Jaime**, la admiración, el respeto y la gratitud que siento por ti no se pueden expresar en unas pocas frases. Eres la persona que más me ha enseñado, motivado y animado en estos años. Has conseguido sacar de mí unas cualidades que ni siquiera había imaginado tener. Tu inteligencia, sabiduría, humildad, paciencia, hacen de ti un verdadero líder. Gracias por estar pendiente en cada momento, espero colaborar contigo durante muchos años más. Eres un ejemplo a seguir y fuente de inspiración y guía para mi y muchas personas.

Mi formación doctoral no sería completa si no hubiera sido por tí, **Santi**. Inteligente, simpático, caótico, siempre de buen humor y "runner". Gracias por todo el conocimiento, las buenas ideas, las clases de Genómica, las risas y los videos "frikis" que has compartido conmigo. Gracias por alegrar nuestros días.

**Mariola**, aunque burocráticamente no has podido ser mi co-directora, te agradezco el apoyo, motivación y alegría de estos años. Aunque ahora pasamos todo el día en los despachos, echo de menos los primeros años de laboratorio, cuando pasábamos horas y horas hablando de lo que fuera. Junto a Jaime y Santi, eres un pilar de "Los Berenjen@s".

Gracias a mi "hermano mayor" **Javi**, por los muchos momentos de amistad pasados juntos, por tu inteligencia, por compartir tus experiencias, por tu buen gusto musical, tu escepticismo, por ser un gran amigo.

"Los Berenjen@s" se han ido multiplicando y muchas son las entradas que hemos recolectado en estos años alrededor del mundo. Me hubiera gustado escribir algo sobre cada uno de vosotros, pero necesitaría otro capítulo de tesis, así que os agradeceré de forma conjuntas.

Gracias **Abou, Adriana, Agustín, Alba E., Alba L., Alessio, Andrea, Arantxa, Arunika, Bianca, Brice, Camil, Carolina, Celia, Celine, Cristina, Dani, David, Dionís, Dora, Edgar, Elena C., Elena R., Elisa, Francesco, Giulio, Gloria, Ioana, Irene, Isa, Juan, Juan Pablo, Judith, Lahiru, Lara, Laura, Loles, Maria, Marieta, Michael, Muditha, Nacho, Nadia, Nirasha, Noelia, Orsi, Pelin, Prashant, Ramses, Ranil, Rory, Silia, Sorin, Trong**, por haber aportado vuestro granito de arena para la realización de esta tesis y de colaboraciones varias. Todos, los que todavía estáis por aquí y los que se han ido, habéis dejado una huella en el laboratorio y en mí (Pido disculpa si se me ha olvidado alguien, sois muchos los que se han pasado por el labo).

Gracias también a todos mis compañeros del COMAV, grandes profesionales que han contribuido a mi formación doctoral y personal y por haber compartido tantas horas, sobretodo en los almuerzos y comidas. Muchos de vosotros sois muy buenos amigos.

Especialmente quiero agradecer a los bioinformáticos (**Peio, Jose, Ximo y Pau**) por los consejos, explicaciones y paciencia cuando tenía dudas y me salían errores en la terminal y a **M<sup>a</sup>José Díez** por ser tan buena persona e investigadora y por dejarme colaborar en la asignatura de Mejora Genética Vegetal.

I also would like to thanks to all the colleagues and collaborators that strongly contributed to this thesis: **Sandra Knapp**, the Global Crop team (**Hannes Dempewolf** and **Benjamin Kilian**), the Italian team (**Sergio Lanteri, Ezio Portis, Lorenzo Barchi, Alberto Acquadro** and **Cinzia Comino**), **Rachel Meyer** and **Aureliano Bombarely**.

Un ringraziamento speciale a mia madre, per essere stata una donna così forte nei momenti difficili, per avermi cresciuto con dei valori sani, per avermi sostenuto e spronato in ogni mia scelta, per essere lí sempre quando c'è ne bisogno. Alle mie sorelle, i miei cognati e i miei nipoti, per l'affetto che mi dimostrano giornalmente nonostante la lontananza. A mio padre.

Ad **Antonella**,  
alla compagna di questo lungo viaggio, alla persona piú importante della mia vita, la quale mi sopporta tutti i giorni da ormai tanti anni, dedico questa tesi.

*So Long, and Thanks for All the Fish*





## Abstract

Feeding the future burgeoning population in a climate change scenario demands new breeding approaches and tools to develop new resource-efficient and resilient crop varieties. Among vegetable crops, eggplant (*Solanum melongena*) is recognized as an important food crop and as such is included in the Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture, which includes 34 crops considered as most relevant for mankind. Before the start of this thesis, few genetic and genomic tools and resources for eggplant breeding, which are reviewed in an introductory chapter, were available. We have recognized the importance of wild eggplant relatives, which have been barely used in eggplant breeding. In this respect, in order to make a more efficient use of wild relatives in plant breeding, we proposed an ambitious approach, called "introgressiomics", consisting of a systematic and massive development of materials carrying introgressions from crop wild relatives (CWRs), which usually are an unexplored and unexploited source of genetic variation for breeding traits. The works done in this thesis are related to the application of the introgressiomics approach to eggplant.

In this framework, the general objectives of the thesis are the development of genetic and genomic information and tools in eggplant genepool, using a multidisciplinary and multi-pronged approach to assist eggplant breeding in the development of new improved and resilient varieties using eggplant relatives as a source of variation. Specifically, in the first chapter of this thesis, we sequenced the transcriptome of two eggplant related species, the wild *Solanum incanum* and the cultivated *S. aethiopicum*, that have a great interest in eggplant breeding. The transcriptomes were assembled

in 83,905 and 87,084 unigenes for *S. incanum* and *S. aethiopicum* respectively, which were extensively structurally and functionally annotated. The variant call analysis identified tens of thousands intraspecific and interspecific polymorphisms, as well as around a thousand of SSRs in each species.

In the second chapter, a subset of those markers (11 SSRs and 35 SNPs) was tested for confirming their usefulness for genetic fingerprinting, diversity evaluation and the establishment of relationships in cultivated eggplant (common, scarlet and gboma) gene pools. We observed that SSRs and SNPs provided different results in the establishment of the relationships, suggesting that each marker type sampled different levels of genetic variation. However, although both markers provided a similar level of information, SNPs seem to provide a better resolution than SSRs for materials phylogenetically more distant.

In the third chapter, in order to broaden the estimation of the genetic diversity and genetic relationships among and within wild and cultivated species belonging to eggplant complexes, we performed a massive genotyping, by a genotype-by-sequencing approach, of 76 accessions belonging to 17 species from the primary, secondary and tertiary gene pool of common eggplant. Out of 75,399 polymorphic sites identified, 12,859 were associated to CDS regions and used to establish an exhaustive and detailed evaluation of the natural allelic diversity and genetic relationships in eggplant gene pool using three different approaches (hierarchical population structure, UPGMA-based dendrogram, and PCoA analysis).

In the fourth chapter, we developed an interspecific genetic map between *S. incanum* and *S. melongena*, linked to four previous eggplant maps and to one tomato map. A total of 243 molecular markers were successfully

mapped consisting of 42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs and one morphological polymorphic markers encompassed 1085 cM distributed in 12 linkage groups. Based on the synteny with tomato, the candidate genes involved in the core chlorogenic acid synthesis pathway in eggplant (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*), five polyphenol oxidase genes (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*), as well as two genes involved in fruit shape (*OVATE*, *SISUNI*), were mapped.

Finally, in the fifth chapter, we developed the first introgression line population (IL) in eggplant using *S. incanum* as a donor parent. A set of 73 selected advanced backcross materials were developed, which covered 99% of the *S. incanum* genome. Also, 25 fixed immortal ILs were obtained, each carrying a single introgressed fragment in homozygosity, altogether spanning 61.7% of the *S. incanum* genome. Sixty-eight candidate genes involved in drought tolerance were identified in the set of ILs.

In summary, in this thesis we have provided tools, knowledge, approaches and advanced breeding materials to develop enhanced and resilient eggplant varieties using the valuable genetic diversity of CWRs and the precision and efficiency of high-throughput techniques. We consider that the information presented in this work will be useful for the eggplant research community as well as for eggplant breeders in order to obtain a new generation of eggplant varieties able to cope with present and future breeding challenge.



## Resumen

Alimentar a la futura población en crecimiento en un contexto de cambio climático exige nuevos enfoques y herramientas de mejora genética para desarrollar nuevas variedades de cultivos que sean resilientes y eficientes en el uso de los recursos. Entre las hortalizas, la berenjena (*Solanum melongena*) es reconocida como un cultivo importante y como tal está incluida en el Anexo 1 del Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura, que incluye 34 cultivos considerados como los más relevantes para la humanidad. Antes del inicio de esta tesis, se disponía de pocas herramientas genéticas y genómicas para la mejora genética de la berenjena, las cuales se revisan en un capítulo introductorio. Asimismo, hemos reconocido la importancia de los parientes silvestres de la berenjena, que apenas se han utilizado en la mejora genética de la misma. A este respecto, para hacer un uso más eficiente de los parientes silvestres en la mejora genética, propusimos un enfoque ambicioso, llamado "introgressiomics", que consiste en un desarrollo sistemático y masivo de materiales que lleven introgresiones de parientes silvestres (CWR), los cuales generalmente son una fuente inexplorada e desaprovechada de variación genética. Los trabajos realizados en esta tesis están relacionados con la aplicación de este enfoque a la berenjena.

En este marco, los objetivos generales de la tesis son el desarrollo de la información y herramientas genéticas y genómicas en el acervo genético de la berenjena, utilizando un enfoque multidisciplinario y multifacético para la mejora genética en el desarrollo de nuevas variedades mejoradas y resistentes usando especies relacionadas como fuente de variación. Concretamente, en el primer capítulo de esta tesis, secuenciamos el

transcriptoma de dos especies relacionadas con la berenjena, la silvestre *Solanum incanum* y la especie cultivada africana *S. aethiopicum*, que tienen un gran interés en la mejora genética de la berenjena. Los transcriptomas se ensamblaron en 83,905 y 87,084 unigenes para *S. incanum* y *S. aethiopicum* respectivamente, los cuales fueron extensivamente anotados estructuralmente y funcionalmente. La búsqueda de variantes alélicas identificó decenas de miles de polimorfismos intraespecíficos e interespecíficos, así como alrededor de un millar de SSRs en cada especie.

En el segundo capítulo, un subconjunto de esos marcadores (11 SSRs y 35 SNPs) se analizaron para confirmar su utilidad para la huella genética, la evaluación de la diversidad y el establecimiento de relaciones entre las berenjenas cultivadas (común, escaleta y gboma). Observamos que los SSRs y los SNPs proporcionaron resultados diferentes en el establecimiento de las relaciones, lo que sugiere que cada tipo de marcador muestreó diferentes niveles de variación genética. Sin embargo, aunque ambos marcadores proporcionaron un nivel similar de información, los SNPs parecen proporcionar una mejor resolución que los SSRs para materiales filogenéticamente más distantes.

En el tercer capítulo, para ampliar la estimación de la diversidad genética y las relaciones genéticas entre y dentro de las especies silvestres y cultivadas pertenecientes al complejo berenjena, realizamos un genotipado masivo, mediante un enfoque de genotipado por secuenciación, de 76 accesiones pertenecientes a 17 especies del acervo genético primario, secundario y terciario de la berenjena común. De los 75,399 sitios polimórficos identificados, 12,859 se asociaron a regiones CDS y se usaron para establecer una evaluación exhaustiva y detallada de la diversidad alélica natural y las relaciones genéticas en el acervo genético de la berenjena

utilizando tres enfoques diferentes (estructura jerárquica de la población, dendrograma basado en UPGMA y análisis de PCoA).

En el cuarto capítulo, desarrollamos un mapa genético interespecífico entre *S. incanum* y *S. melongena*, vinculado a otros cuatro mapas anteriores de berenjena y a un mapa de tomate. Se mapearon con éxito un total de 243 marcadores moleculares que consistían en 42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs y un marcador morfológico, abarcando 1085 cM distribuidos en 12 grupos de ligamiento. Basándose en la sintenia con tomate, se localizaron los genes candidatos involucrados en la ruta metabólica de síntesis del ácido clorogénico en berenjena (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*), cinco genes de polifenol oxidasas (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*), así como dos genes implicados en la forma del fruto (*OVATE*, *SISUNI*).

Finalmente, en el quinto capítulo, desarrollamos la primera población de líneas de introgresión (ILs) en berenjena utilizando *S. incanum* como parental donante. Se desarrolló un conjunto de 73 materiales de retrocruzamiento avanzado, que cubrieron el 99% del genoma de *S. incanum*. Además, se obtuvieron 25 líneas de introgresión inmortales, cada una con un único fragmento introgresado en homocigosis, que en conjunto abarcaron el 61,7% del genoma de *S. incanum*. Además, sesenta y ocho genes candidatos involucrados en la tolerancia a la sequía se identificaron en el conjunto de IL.

En resumen, en esta tesis hemos proporcionado herramientas, conocimientos, enfoques y materiales avanzados de mejora genética para desarrollar variedades de berenjenas mejoradas y resistentes, utilizando la valiosa diversidad genética de las especies silvestres y la precisión y eficiencia de las técnicas de secuenciación de alto rendimiento. Consideramos que la información presentada en este trabajo será útil tanto para la comunidad

investigadora que trabaja en berenjena, como para los mejoradores, con el fin de obtener una nueva generación de variedades de berenjena capaces de hacer frente a los retos de la mejora genética presente y futura.



## Resum

Alimentar la futura població en creixement en un context de canvi climàtic exigeix nous enfocaments i eines de millora genètica per desenvolupar noves varietats de cultius que siguin resilients i eficients en l'ús dels recursos. Entre les hortalisses, l'albergínia (*Solanum melongena*) és reconeguda com un cultiu important i com a tal està inclosa en l'Annex 1 del Tractat Internacional sobre els Recursos Fitogenètics per a l'Alimentació i l'Agricultura, que inclou 34 cultius considerats com els més rellevants per a la humanitat. Abans de l'inici d'aquesta tesi, es disposava de poques eines genètiques i genòmiques per a la millora genètica de l'albergínia, les quals es revisen en un capítol introductori. Així mateix, hem reconegut la importància dels parents silvestres de l'albergínia, que tot just s'han utilitzat en la millora genètica de la mateixa. Referent a això, per fer un ús més eficient dels parents silvestres a la millora genètica, vam proposar un enfocament ambiciós, anomenat "introgressiomics", que consisteix en un desenvolupament sistemàtic i massiu de materials que portin introgressions de parents silvestres (CWR), els quals generalment són una font inexplorada i desaprofitada de variació genètica. Els treballs realitzats en aquesta tesi estan relacionats amb l'aplicació d'aquest enfocament a l'albergínia.

En aquest marc, els objectius generals de la tesi són el desenvolupament de la informació i eines genètiques i genòmiques en el patrimoni genètic de l'albergínia, utilitzant un enfocament multidisciplinari i multifacètic per a la millora genètica en el desenvolupament de noves varietats millorades i resistents emprant espècies relacionades com a font de variació. Concretament, en el primer capítol d'aquesta tesi, seqüenciem el transcriptoma de dues espècies relacionades amb l'albergínia, la silvestre

*Solanum incanum* i l'espècie cultivada africana *S. aethiopicum*, que tenen un gran interès en la millora genètica de l'albergínia. Els transcriptomes s'ensamblaren en 83,905 i 87,084 unigenes per *S. incanum* i *S. aethiopicum* respectivament, els quals van ser extensivament anotats estructuralment i funcionalment. La recerca de variants al·lèliques va identificar desenes de milers de polimorfismes intraespecífics i interespecífics, així com al voltant d'un miler de SSRs en cada espècie.

En el segon capítol, un subconjunt d'aquests marcadors (11 SSRs i 35 SNPs) es van analitzar per confirmar la seva utilitat per l'empremta genètica, l'avaluació de la diversitat i l'establiment de relacions entre les albergínies conreades (comú, escarlata i gboma). Observem que els SSRs i els SNPs van proporcionar resultats diferents en l'establiment de les relacions, el que suggereix que cada tipus de marcador va mostrejar diferents nivells de variació genètica. No obstant això, encara que tots dos marcadors van proporcionar un nivell similar d'informació, els SNPs semblen proporcionar una millor resolució que els SSRs per materials filogenèticament més distants.

En el tercer capítol, per a ampliar l'estimació de la diversitat genètica i les relacions genètiques entre i dins de les espècies silvestres i conreades pertanyents al complex albergínia, vam realitzar un genotipat massiu, mitjançant un enfocament de genotipat per seqüenciació, de 76 accessions pertanyents a 17 espècies del patrimoni genètic primari, secundari i terciari de l'albergínia comuna. Dels 75,399 llocs polimòrfics identificats, 12,859 es van associar a regions CDS i es van usar per a establir una avaluació exhaustiva i detallada de la diversitat al·lèlica natural i les relacions genètiques en el patrimoni genètic de l'albergínia utilitzant tres enfocaments

diferents (estructura jeràrquica de la població, dendrograma basat en UPGMA i anàlisi de PCoA).

En el quart capítol, desenvolupem un mapa genètic interespecífic entre *S. incanum* i *S. melongena*, vinculat a altres quatre mapes anteriors d'albergínia i a un mapa de tomàquet. Es mapejaren amb èxit un total de 243 marcadors moleculars que consistien en 42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs i un marcador morfològic, van abastar 1085 cM distribuïts en 12 grups de lligament. Basant-se en la sintènia amb tomàquet, es van localitzar els gens candidats involucrats en la ruta metabòlica de síntesi de l'àcid clorogènic en albergínia (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*), cinc gens de polifenol oxidases (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*), així com dos gens implicats en la forma del fruit (*OVATE*, *SISUNI*).

Finalment, en el cinquè capítol, desenvolupem la primera població de línies de introgressió (ILS) en albergínia utilitzant *S. incanum* com parental donant. Es va desenvolupar un conjunt de 73 materials de retrocruztament avançat, que van cobrir el 99% del genoma de *S. incanum*. A més, es van obtenir 25 línies de introgressió immortals, cadascuna amb un únic fragment introgressado en homozigosi, que en conjunt van abastar el 61,7% del genoma de *S. incanum*. A més, seixanta-vuit gens candidats involucrats en la tolerància a la sequera es van identificar en el conjunt d'IL. En resum, en aquesta tesi hem proporcionat eines, coneixements, enfocaments i materials avançats de millora genètica per desenvolupar varietats d'albergínies millorades i resistents, utilitzant la valuosa diversitat genètica de les espècies silvestres i la precisió i eficiència de les tècniques de seqüenciació d'alt rendiment. Considerem que la informació presentada en aquest treball serà útil tant per a la comunitat investigadora que treballa en albergínia, com per als milloradors, per tal d'obtenir una nova generació de

varietats d'albergínia capaces de fer front als reptes de la millora genètica present i futura.





# TABLE OF CONTENTS

<b>INTRODUCTION.....</b>	<b>31</b>
--------------------------	-----------

<b>Genomic Tools for the Enhancement of Vegetable Crops: A Case in Eggplant.....</b>	<b>1</b>
--	----------

Abstract.....	2
---------------	---

Introduction.....	5
-------------------	---

Genome assemblies.....	7
------------------------	---

Transcriptomes and RNA-based studies.....	10
---	----

Mapping studies, experimental populations, and genotyping methods.....	16
--	----

Future direction for genetics and genomics tools for eggplant breeding.....	31
---	----

<b>Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change.....</b>	<b>37</b>
---	-----------

Abstract.....	38
---------------	----

Introduction.....	41
-------------------	----

Crop wild relatives for broadening the genetic base of crops.....	44
---	----

The introgressiomics approach.....	49
------------------------------------	----

Identification of CWRs for introgressiomics.....	55
--	----

Interspecific hybridization and backcrossing.....	59
---	----

Development of introgressiomics populations.....	65
--	----

Genomic tools and new plant breeding techniques for introgressiomics.....	72
---	----

Moving the introgressed material into the breeding pipeline .....	75
Conclusions.....	77
<b>OBJECTIVES .....</b>	<b>81</b>
<b>RESULTS .....</b>	<b>87</b>

**Chapter 1: Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding ..... 89**

Abstract.....	90
Introduction.....	93
Materials and methods .....	98
Plant material .....	98
RNA extraction for Illumina sequencing.....	100
Sequence data analysis and De novo assembly .....	100
Structural and functional annotation.....	101
Mapping transcriptomes against eggplant genome.....	102
Molecular markers discovery .....	103
SNVs.....	103
SSRs.....	104
Results and Discussion .....	106
Illumina paired-end sequencing and EST assembly .....	106
Annotation of <i>S. incanum</i> and <i>S. aethiopicum</i> transcriptomes .....	113
Molecular markers discovery and validation.....	121



Single nucleotide variations (SNVs) .....	121
SSRs.....	127
Conclusions.....	132
Additional files .....	134

**Chapter 2: Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants..... 139**

Abstract.....	140
Introduction.....	142
Material and methods.....	146
Plant materials .....	146
DNA extraction.....	149
SSR genotyping .....	149
SNP genotyping .....	150
Data analyses .....	154
Results and discussion .....	155
Validation of the SSR and SNP markers for fingerprinting .	155
SSR and SNP diversity statistics .....	159
SSR and SNP-based genetic relationships.....	164
Cluster analyses .....	169
Conclusions.....	171

**Chapter 3: Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes ..... 175**

Abstract.....	176
Introduction.....	178
Material and methods.....	183
Plant materials .....	183
Library construction and sequencing.....	187
Sequence analysis .....	187
Genetic relationships analysis and population structure.....	188
Results and discussion .....	190
Sequencing and SNPs identification.....	190
Genetic relationships analysis and population structure.....	194
PCoA analyses .....	198
Within-groups PCoA analyses.....	204
Conclusions.....	208
Additional files .....	211

**Chapter 4: Location of the chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant..... 213**

Abstract.....	214
Introduction.....	216
Material and methods.....	221
Plant materials .....	221

Development of the SMIBC genetic map.....	223
DNA extraction.....	223
COSII markers .....	223
SSR markers .....	224
AFLP markers.....	225
Sequence search for candidate genes and in silico comparison.....	225
Additional markers .....	227
Linkage analysis and map construction .....	227
Macro-syteny between SMIBC and other genetic linkage maps.....	227
Results.....	229
Genetic map construction .....	229
COSII analysis .....	231
SSRs analysis.....	232
AFLPs analysis .....	232
Mapping of CGA pathway genes .....	233
PAL (phenylalanine ammonia lyase).....	233
C4H (cinnamate 4-hydroxylase).....	234
4CL (4-hydroxycinnamoyl-CoA ligase).....	234
HCT (hydroxycinnamoyl-coA shikimate/quinic hydroxycinnamoyl transferase) .....	235
C3'H (p-coumaroyl ester 3'-hydroxylase) .....	236
HQT (hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase).....	236

Mapping of PPO genes .....	237
Mapping of other genes and traits of agronomic importance	251
Synteny and orthologous candidate genes with other maps .	254
Conclusion .....	264

**Chapter 5: Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background..... 267**

Abstract .....	268
Introduction.....	270
Materials and Methods.....	275
Plant material .....	275
Breeding scheme.....	278
DNA extraction.....	281
Genotyping methods .....	281
COSII marker analysis.....	283
SSR marker analysis .....	283
SNP marker analysis.....	284
GBS analysis.....	285
Positioning candidate genes for drought tolerance .....	286
Results.....	287
Development of the ABs and ILs .....	287
Candidate genes for drought tolerance .....	296
Discussion.....	298

<b>GENERAL DISCUSSION .....</b>	<b>309</b>
Genomic information available in the eggplant genepool .....	312
Molecular markers identification .....	314
Diversity evaluation and the establishment of relationships in eggplant genepool .....	316
Development of genetic and genomic tools using CWRs .....	320
Future perspectives .....	325
<b>CONCLUSIONS .....</b>	<b>329</b>
<b>REFERENCES.....</b>	<b>335</b>



# **INTRODUCTION**









# Genomic Tools for the Enhancement of Vegetable Crops: A Case in Eggplant

Gramazio P<sup>1\*</sup>, Prohens J<sup>1</sup>, Plazas M<sup>2</sup>, Mangino G<sup>1</sup>, Herraiz J<sup>1</sup>, García-Fortea E<sup>1</sup>, Vilanova S<sup>1</sup>

<sup>1</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

<sup>2</sup> Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas - Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

\* Corresponding author: [piegra@upv.es](mailto:piegra@upv.es)

**Citation:** Gramazio P, Prohens J, Plazas M, Mangino G, Herraiz J, García-Fortea E, Vilanova S. (2017). Genomic tools for the enhancement of vegetable crops: a case in eggplant. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 46(1):1–13.

<https://doi.org/10.15835/nbha46110936>

## ***Abstract***

Dramatic advances in genomics during the last decades have led to a revolution in the field of vegetable crops breeding. Some vegetables, like tomato, have served as model crops in the application of genomic tools to plant breeding but other important crops, like eggplant (*Solanum melongena*), lagged behind. The advent of next generation sequencing (NGS) technologies and the continuous decrease of the sequencing costs have allowed to develop genomic tools with a greatly benefit for no-model plants such as eggplant. In this review we present the currently available genomic resources in eggplant and discuss their interest for breeding. The first draft of eggplant genome sequence and the new upcoming improved assembly, as well as the transcriptomes and RNA-based studies represent important genomic tools. The transcriptomes of cultivated eggplant and several wild relatives of eggplant are also available and have provided relevant information for the development of markers and understanding biological processes in eggplant. In addition, a historical overview of the eggplant genetic mapping studies, performed with different types of markers and experimental populations, provides a picture of the increase over time of the precision and resolution in the identification of candidate genes and QTLs for a wide range of stresses, and morphoagronomic and domestication traits. Finally, we discuss how the development of new genetic and genomic tools in eggplant can pave the way for increasing the efficiency of eggplant breeding for developing improved varieties able to cope with the old and new challenges in horticultural production.

**Keywords:** experimental populations, genetics, genomics, genotyping strategies, QTLs, sequencing, *Solanum melongena*

**Abbreviations:**

MAS: Marker-assisted selection

QTL: Quantitative trait locus

NCBI: National center for biotechnology information

RIL: Recombinant inbred line

UTR: Untranslated region

SNV: Single nucleotide variant

SSR: Simple sequence repeat

SNP: Single nucleotide polymorphism

INDEL: Insertion/deletion

AFLP: Amplified fragment length polymorphism

RAPD: Random amplification of polymorphic DNA

LG: Linkage group

COS: Conserved orthologous set

RFLP: Restriction fragment length polymorphism

PIC: Polymorphic information content

NGS: Next generation sequencing

CSS: Conserved syntenic segments

BC: Backcross

CAPS: Cleaved amplified polymorphic sequence

RAD: Restriction-site-associated

GBS: Genotyping-by-sequencing

SOL: Orthologous genes in *Solanum*

LD: Linkage disequilibrium

GWA: Genome-wide association

CGA: Chlorogenic acid

*PAL: Phenylalanine ammonia lyase*

*C4H: Cinnamate 4-hydroxylase*

*4CL: 4-hydroxycinnamoyl-CoA ligase*

*HCT: Hydroxycinnamoyl-coA shikimate/quinate hydroxycinnamoyl transferase*

*C3'H: P-coumaroyl ester 3'-hydroxylase*

*HQT: Hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase*

*PPO: Polyphenol oxidase enzyme*

NIL: Near isogenic line

IL: Introgression line

MAGIC: Multi-parent advanced generation inter-cross

SBG: Sequence based genotyping

SPET: Single primer enrichment technology.

## **Introduction**

Eggplant (*Solanum melongena* L.,  $2n = 2x = 24$ ), also known as common eggplant or brinjal eggplant, belongs to the Leptostemonum clade within the *Solanum* genus (the “spiny” solanums) and is the second most important solanaceous fruit crop in total production after tomato (*S. lycopersicum* L.) (Knapp et al., 2013). Eggplant has undergone a constant increase in yield (2.7-fold) and total production (8.7-fold) in the last fifty years, although the largest increases have been recorded in the last decade (FAOSTAT, 2017).

Common eggplant is widespread and consumed worldwide even though in some areas, mainly in the African continent, two other cultivated eggplants namely scarlet (*S. aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplants are locally important (Lester and Daunay, 2003; Plazas et al., 2014a).

The breeding and genomics revolution of the last fifteen years has resulted in the development of large amounts of information of interest for breeding, allowing the enhancement of many vegetable crops. Even though eggplant is the sixth most important vegetable in production in the world for many years it has lagged behind in the development and use of genomic tools compared to other important Solanaceae crops like tomato, potato or pepper (Barchi et al., 2011; Hurtado et al., 2013). Tomato, being the closest model crop species with many enhanced genomic tools, has been used as reference for eggplant in many studies (Doganlar et al., 2002a; Wu et al., 2009; Barchi et al., 2012). In this respect, the scientific genomic achievements in tomato have allowed fishing in its gene pool and taking advantage of the extraordinary genetic wealth of its wild relatives

(Viquez-Zamora et al., 2013; Aflitos et al., 2014). In this way, tomato breeders have been able to find genetic solutions to some biotic and abiotic threats and achieving improvements to cope with unfavorable agricultural environments in a climate change scenario (Bolger et al., 2014; Thapa et al., 2015).

On the contrary, up to now in common eggplant, as well as in scarlet and gboma eggplant, the use of wild relatives for the breeding purpose have been negligible and often purely academic (Rotino et al., 2014). The availability of a large array of molecular markers obtained by genomic techniques allows a very efficient marker-assisted selection (MAS), by selecting the genes or quantitative trait loci (QTLs) of interest and facilitating the removal of undesirable traits of wild relatives and saving time and resources compared to the conventional breeding approach (Morrell et al., 2012; Brozynska et al., 2016). Even though the gap between eggplant and the crops with more genomic resources is still wide, a clear turnaround in this trend has occurred in the last few years (Hirakawa et al., 2014; Portis et al., 2015; Kouassi et al., 2016; Plazas et al., 2016; Salgon et al., 2017).

In this paper, we review the available genomic resources in eggplant gene pool, including experimental populations that are under development, as well as the future perspectives and directions to exploit the full potential of these tools for basic and applied research in this crop.



## **Genome assemblies**

Sequencing of the genome of a crop is essential to dramatically accelerate crop improvement (Davey et al, 2012). A high quality reference genome opens the way to access to information of the complete set of genes, to the different layers of regulatory elements and to the basic genomic architecture, allowing precise structural and functional comparisons between species (Feuillet et al, 2011). As of June 2017 there is only one genome draft publicly available (SME\_r2.5.1) in the National Center for Biotechnology Information (NCBI) (SRA accession: DRR014074 and DRR014075) (Hirakawa et al., 2014).

The accession used for the whole-genome shotgun sequencing was ‘Nakate-Shinkuro’, an important traditional Asian-type cultivar that has been used in the development of some modern commercial cultivars. For sequencing (~144X), a combined genomic libraries approach was employed, consisting of paired-end (insert size of 200–300 bp) and 2 Kb Illumina mate-pair insert size (Table 1). The transcriptome of two other accessions ‘AE-P03’ and ‘LS1934’ was also sequenced to improve the de novo assembly (Li et al., 2010). The final assembly consisted of 33,873 scaffolds, covering about 74% (833.1 Mb) of the estimated length of the genome (~1.1Gb) (Arumuganathan et al., 1991) with an N50 parameter of 64.5 Kb. The total number of genes predicted was 42,035, of which 4,018 were described as being exclusive of eggplant, quite a large number if compared with more completed genomes like the last version of tomato Heinz 1706 (SL3.0 version, 34,879 genes in ITAG3.10, [https://solgenomics.net/organism/Solanum\\_lycopersicum/genome](https://solgenomics.net/organism/Solanum_lycopersicum/genome)) or the last version of *Arabidopsis thaliana* genome (Araport11 version, 27,655

Protein Coding Genes) (Cheng et al., 2016). A large number of different simple sequence repeats (SSRs) motifs and repeats (83,401) was identified by Hirakawa et al. (2014), as well as 4,536 single nucleotide polymorphisms (SNPs) from a microarray among the accessions used for the study.

This first genomic resource, which is quite far to be a comprehensive work, paved the way to understanding the genomic architecture of the eggplant and allowed to perform comparisons with other important Solanaceae crops, like potato, tomato and pepper (Potato Genome Sequencing Consortium, 2011; The Tomato Genome Consortium, 2012; Qin et al., 2014), whose genomes are more complete and annotated, and also with model species like *Arabidopsis thaliana* (Cheng et al., 2016).

The development of another eggplant genome by the Italian Eggplant Genome Sequencing Consortium has been presented in several scientific meetings (Barchi et al., 2016), but at the time of writing this work has not yet been published. According to the authors, a high-quality reference genome has been de novo assembled through Illumina sequencing (~155X), using different genomic libraries sizes (from 270 bp to 10 Kb), of the inbred eggplant line “67/3”, which was used as a male parent of a 157 F6 recombinant inbred lines (RILs) mapping population (Table 1). In addition, to improving the assembly, its transcriptome was also sequenced (Barchi et al., 2016). Moreover, the hybrid assembly was combined with the sequencing of the female parent “305E40” (35X) and the rest (1X) of the RILs mapping population and with a high-resolution restriction optical map. The final result consisted of 12 pseudomolecules, spanning ~1.2Gb with an L50 of >3Mb. The functional annotation

resulted in  $\sim 40\text{K}$  protein-coding genes, confirming the estimation of Hirakawa et al. (2014).

## ***Transcriptomes and RNA-based studies***

RNA sequencing is another essential genomic resource (Ozsolak and Milos, 2001). Probably due to the reduced economic and bioinformatics efforts compared to genome assembly and the variety of study approaches and aims, RNA-based studies are in general more abundant. In fact, in RNA-based studies the results depend on the genetic expression at a certain stage of development in a specific plant tissue (Waterhouse and Helliwell, 2003).

Regarding common eggplant, up to now, few studies using RNA sequencing have been carried out. One of them was the de novo assembly of the whole transcriptome from root, stem and young leaves samples (SRA accession: SRR1104129) (Yang et al., 2014). From paired-end 2x100 bp libraries, these authors retrieved 15 M of raw reads, which were assembled after a filtering process in 44,672 transcripts and 34,174 unigenes, a similar number to the protein-coding genes from the genome annotation (Hirakawa et al., 2014). In addition to performing structural and functional annotation, a comparison was performed using a set of 4,900 orthologs with other 11 plant species, including model plants and Solanaceae crops like tomato and potato, allowing to estimate the time of divergence between them. Furthermore, using the Plant Resistance Gene database ([http://prgdb.crg.eu/wiki/Main\\_Page](http://prgdb.crg.eu/wiki/Main_Page)) a set of 621 resistance genes were identified. In the same study (Yang et al., 2014), the transcriptome of the wild eggplant relative *S. torvum* Sw., a species belonging to the tertiary eggplant gene pool (Syfert et al., 2016), was also assembled (SRA accession: SRR1104128). *Solanum torvum*, also known as turkey berry, is of great interest for eggplant breeding since it is resistant

to a wide range of soil-borne diseases like root-knot nematodes, *Ralstonia solanacearum*, *Verticillium dahliae*, and *Fusarium oxysporum* f. sp. *melongenae* (Gousset et al., 2005; Yamaguchi et al., 2010). For *S. torvum*, the assembly statistics and annotation were slightly different to that of common eggplant, being the most relevant difference the number of unigenes (38,185 for *S. torvum* versus 34,174 for common eggplant) (Table 1).

Another de novo transcriptome assembly was released in 2016 in order to identify putative allergens in eggplant fruit (Ramesh et al., 2016; SRA accession: SRR1291243). In this study, total RNA was extracted from fruit peel and flesh, a different plant material compared to Yang et al. (2014) whole transcriptome, and a total of 48 putative allergens and 526 B-cell linear epitopes were identified from the 149,224 transcripts assembled (Table 1). Of these 40,752 showed significant similarity with predicted proteins in tomato and potato. In addition to common eggplant, the whole transcriptome of some eggplant relatives has been released. One of them was scarlet eggplant (*S. aethiopicum* accession BBS135) (Gramazio et al., 2016a; SRA accession: SRR2229192), the second most important cultivated eggplant, which is common in sub-Saharan Africa, as well as, in some areas of Brazil, Caribbean and south of Italy (Sunseri et al., 2010). In the same study, a de novo whole transcriptome of *Solanum incanum* L. accession MM577 (SRA accession: SRR2289250), a wild relative of common eggplant which is considered a powerful source of phenolics and tolerant to some abiotic stresses such as drought (Knapp et al. 2013), was also assembled. Due to the high amount of reads obtained (more than 100 M), the number of assembled unigenes in Gramazio et al. (2016a) was relatively high (87,084 for *S. aethiopicum* and 83,905 for *S.*

*incanum*), probably due to high representation of 3' or 5' untranslated regions (UTRs) and intron sequences from non-mature mRNAs as a consequence of a deep coverage (Table 1). On the other hand, the number of annotated unigenes in protein databases was similar to other *Solanum* crops (34,231 and 30,630 for *S. aethiopicum* and *S. incanum* respectively). In this study, molecular marker discovery was performed, identifying a total of 1,248 microsatellites for scarlet eggplant and 976 for *S. incanum*. In addition, intraspecific and interspecific single nucleotide variant (SNV), SNPs and insertions/deletions (INDELs), were identified not only in *S. aethiopicum* and *S. incanum* but also in *S. melongena* and *S. torvum*, using for comparison the raw reads generated in the Yang et al. (2014) study.

The most recent transcriptome released in the eggplant genepool was from *S. aculeatissimum* Jacq., a wild relative of eggplant resistant to verticillium wilt (Zhou et al., 2016). Two different libraries were constructed, for the control and for the infected roots with *Verticillium dahliae*, each of them giving 28 M of raw reads. A total of 64,413 and 71,291 unigenes were obtained for the control and infected roots, respectively, which were functionally annotated, resulting in 17,645 of them differentially expressed (11,696 upregulated and 5,949 downregulated).

Apart from the whole transcriptomes, in common eggplant, other RNA-based studies were performed. Yang et al. (2013) identified miRNA from eggplant involved in the process of infection by *V. dahliae* through deep-sequencing of two small RNA libraries, using control and infected seedlings. From 5,940 miRNA identified, 220 belonged to Solanaceae species and two new miRNA were eggplant specific. The authors identified a total of 33 differentially expressed miRNA between the two

libraries (28 downregulated and 5 up-regulated), which were strongly involved in the *V. dahliae* infectious process. The most recent RNA-based study (SRA: SRR3479276, SRA: SRR3479277), which is still unpublished and only available in NCBI database (<https://www.ncbi.nlm.nih.gov/sra/SRR3479277/>), identified miRNA involved in the difference in heterostyly in short-morph and long-morph eggplant pistils using high-throughput small RNA microarray and degradome sequencing. From the 686 miRNAs identified, 10 were differentially expressed and were determined as pistil development-related miRNAs.

**Table 1.** Metrics for genome and RNA-based studies assemblies performed in the eggplant genepool.

Species	Aim of the study	Genetic material	Plant material	Library	Sequencing platform	Raw reads (M)	Coverage	Final assembly	Marker discovery	NCBI accession	Source
<i>S. melongena</i>	Genome assembly	DNA	Leaves	PE (200–300 bp) MP (2Kb)	Illumina HiSeq 2000	1,323	144.3X	33,873 scaffolds	4,536 SNPs 83,401 SSRs	DRR014074 DRR014075	Hirakawa <i>et al.</i> , 2014
<i>S. melongena</i>	Genome assembly	RNA	Leaves, roots, fruits, flowers	454 GS FLX library	Roche 454 GS FLX	1.4	0.48X				
<i>S. melongena</i>	Genome assembly	DNA	Leaves	Several library sizes from PE (270 bp) to MP (10Kb)	Illumina	-	155X 35X 1X	12 pseudo molecules	-	-	Barchi <i>et al.</i> , 2016
<i>S. melongena</i>	Genome assembly	RNA	Leaves	-	Illumina	-	-				
<i>S. melongena</i>	Whole transcriptome assembly	RNA	Leaves, root, stem	PE (72 bp)	Illumina HiSeq 2000	30	-	38,185 unigenes	-	SRR1104129	Yang <i>et al.</i> , 2014
<i>S. torvum</i>	Whole transcriptome assembly	RNA	Leaves, root, stem	PE (72 bp)	Illumina HiSeq 2000	54	-	34,174 unigenes	-	SRR1104128	



<i>Melongena</i>	Identify putative allergens	RNA	Fruit	PE (100 bp)	Illumina HiSeq 2000	89	-	149,224 transcripts	-	SRR1291243	Ramesh <i>et al.</i> , 2016
<i>S. aethiopicum</i>	Whole transcriptome assembly	RNA	Leaf, floral bud, fruit	PE (300 bp)	Illumina HiSeq 2000	114	-	87,084 unigenes	164,127 SNVs 976 SSRs	SRR2229192	Gramazio <i>et al.</i> , 2016a
<i>S. incanum</i>	Whole transcriptome assembly	RNA	Leaf, floral bud, fruit	PE (300 bp)	Illumina HiSeq 2000	105	-	83,905 unigenes	12,396 SNVs 1,248 SSRs	SRR2289250	
<i>S. aculeatissimum</i>	Whole transcriptome assembly	RNA	Root	-	Illumina HiSeq 2000	28	-	69,824 unigenes	-	SRS1383901 SRS1383902	Zhou <i>et al.</i> , 2016
<i>S. melongena</i>	Identification of MiRNA	RNA	Seedlings	Small RNA libraries	Illumina MiSeq	30	-	5,940 miRNA	-	SRR833801 SRR833802	Yang <i>et al.</i> , 2013
<i>S. melongena</i>	Identification of MiRNA	RNA	Pistil	Small RNA libraries	-	-	-	686 miRNAs	-	SRR3479276 SRR3479277	Wang, 2017

## ***Mapping studies, experimental populations, and genotyping methods***

Gene mapping establishes a connection between a trait under study and one or more chromosomal regions (Sevon et al., 2005). Although genetic maps have been constructed since the first decade of the 20th century (Brown, 2006), the first genetic map in eggplant was released only in 2001 (Nunome et al., 2001). This intraspecific map was built using 181 dominant markers (93 amplified fragment length polymorphism (AFLPs) and 88 random amplification of polymorphic DNA (RAPDs)) using a population of 168 F2 progenies, resulting in 21 linkage groups (LGs) and spanning 779.2 cM (Table 2). The aim of this study was to identify genetic regions involved in fruit shape and color development, which were associated with regions in LG2 and LG7 respectively.

Low frequency of DNA polymorphism in eggplant (Doganlar et al., 2002a; Wu et al., 2009) and the tendency of some markers to cluster, like AFLPs and RAPDs (Alonso-Blanco et al., 1998; Nilsson et al., 2007), generated a high number of LG, which did not correspond to the basic chromosome number in eggplant.

One year later (2002), an interspecific genetic map was developed, using *S. melongena* and the wild relative *S. linneanum* Hepper & P.M.L. Jaeger as parents, in order to increase the general low DNA polymorphism observed in mapping populations derived from an intraspecific cross (Doganlar et al., 2002a). The aim of this study was to compare tomato and eggplant maps in order to evaluate synteny and identify rearrangements between the two *Solanum* species occurred during the domestication process. To achieve that objective, single-copy tomato cDNA, genomic tomato DNA, and tomato conserved orthologous set (COS) restriction fragment length

polymorphism (RFLP) markers were assessed, which were previously mapped in a tomato map and used to establish a synteny between the tomato and potato genome (Tanksley et al. 1992; Fulton et al. 2002). The map, which spanned 1,480 cM along 12 LG (Table 2), confirmed the high collinearity between the tomato and eggplant and identified 25 rearrangements, as well as 125 QTLs related to 22 domestication traits (fruit size, shape, and color and plant prickliness) (Doganlar et al., 2002b) and 18 morphological traits related to leaf, flower, and fruit size, shape, appearance, and development (Frary et al., 2003).

Another interspecific genetic map using the eggplant wild relative *S. sodomium* L. [= *S. linnaeanum*], was developed by Sunseri et al. (2003). This map, that was built to achieve markers linked to *Verticillium* tolerance using 48 F<sub>2</sub> progenies derived from an interspecific cross between a susceptible and tolerant parental, consisted of 273 markers (156 AFLPs and 117 RAPDs) distributed along 12 LG and 736 cM with a small average distance between them (2.7 cM, Table 2). An improved version of the Nunome et al. (2001) genetic map was released in 2003 (Nunome et al., 2003a), where seven SSRs markers were added through screening an eggplant genomic library of dinucleotide motifs, in order to merge and reduce LGs. In fact, the number of LGs decreased from 21 to 17, while the map length and the average marker density remained almost the same (Table 2).

At the same time, an eggplant genomic library was developed for screening trinucleotides motifs (Nunome et al., 2003b). SSRs, also known as microsatellites, are highly polymorphic, genomic (gSSRs) more than genic (EST-SSRs), within and across the species, generally displaying a higher polymorphic information content (PIC) compared to the other molecular markers (Kalia et al., 2011). In addition, microsatellites are codominant,

robust, highly reproducible, abundant and quite-well spread along the genome (Varshney et al., 2005), Nevertheless, before the advent of next generation sequencing (NGS) their identification through the development of genomic libraries required a quite high degree of expertise, which implied a significant investment of time and resources (Fernandez- Silva et al., 2013). Nowadays, the identification of thousands of gSSRs and EST-SSRs by DNA and RNA sequencing has become cost-effective and effortless by virtue of the overwhelming advancements in sequencing platforms (Xiao et al., 2013; Goodwin et al., 2016).

Other genomic libraries to discover microsatellites in eggplant were developed by Stàgel et al. (2008), Vilanova et al. (2012) and Nunome et al. (2009), the latter in order to improve the previous versions of their genetic map (Nunome et al., 2001, 2003) by adding 245 SSRs, thus reducing to 14 the number of LGs and increase the total length to 959 cM (Table 2).

An enhanced version of Doganlar et al. (2002) genetic map was obtained by adding 110 COSII and 5 tomato-derived markers and performing a detailed synteny with tomato through inferred the position of additional 522 COSII markers (Wu et al., 2009) (Table 2). The high number of common markers allowed to estimate the time of divergence in 12 million years and from genomic structural standpoint a minimum of 24 inversions and 5 chromosomal translocations occurred between the two species, and 37 conserved syntenic segments (CSSs) where the order of genes/markers have been well preserved were detected.

In order to map the resistance to *F. oxysporum* (gene *Rfo-sal*), Barchi et al. (2010) developed an intraspecific genetic map using a 141 F2 population from 305E40, a double haploid line obtained through another culture of a backcross material (BC7) using *S. aethiopicum* as a donor parent carrying the

*Rfo-sal* gene for resistance to *Fusarium*, and the line 67/3, derived from an intraspecific cross between an Italian and Chinese cultivars and susceptible to *F. oxysporum*. The map spanned 718.7 cM across 12 LGs with an average marker density of 3.0 cM and was composed of 212 AFLPs, 22 SSRs developed by Stägel et al. (2008), 1 RFLP and three *Rfo-sal* cleaved amplified polymorphic sequence (CAPS) markers which co-segregated in LG1 (Table 2).

The same F2 population was also used to identify QTLs related to anthocyanin content (Barchi et al., 2012), as well as, 105 QTLs associated with twenty yield, fruit, and morphological traits (Portis et al., 2014) and 29 QTLs with seventeen traits (fruit qualitative and health-related compounds) (Toppino et al., 2016), this time employing a large set of SNPs identified by restriction-site-associated (RAD) tags sequencing (Barchi et al., 2011), which led the total map length to 1390 cM, quite similar to that of Wu et al. (2009) and Doganlar et al. (2002) maps (Table 2).

In fact, when markers that tend to cluster, like AFLPs, are replaced with markers that are more dispersed, like SNPs, the total length of genetic maps usually increases (Rafalski, 2002). SNPs, like SSRs, are codominant, robust and easy to identify by NGS platform through sequencing (Davey et al., 2011; Scheben et al., 2016), and in addition have the advantage that are more abundant, ubiquitous and easy to automate than SSRs (Thomson et al., 2014; Kim et al., 2016), although are less informative (Filippi et al., 2015; Gonzaga, 2015).

SSRs are generally considered as dispersed markers, even though in some species tend to concentrate more frequently in heterochromatic regions and it is unlikely to cover all the genomic regions assessing just SSRs (Hong et al., 2007; Shirasawa et al., 2010). On the other hand, it has been reported

that the validation of SNPs by high-throughput SNP genotyping, like genotyping-by-sequencing (GBS), RAD tag sequencing or similar, can be 100-fold faster and 75% less expensive than SSRs detection through an agarose or polyacrylamide gels or capillary sequencing (Jones et al., 2007; Yan et al., 2010). For all these reasons SNPs markers have quickly replaced SSRs in the last few years.

Fukuoka et al. (2012) in an effort to represent genomic region overlooked in the previous maps, mapped a considerable set of SNPs identified from a set of 4,754 orthologous genes in *Solanum* (SOL) developed from 16K eggplant, 47K tomato, and 57K potato unigenes. An integrated intraspecific map was built from two F2 populations, LWF2 and ALF2, and 952 markers (639 SNPs and 313 SSRs) along 12 LGs, resulting in 1,285 cM and an average marker density of 1.4 cM, covering 1.5 times the genomic region represented in Nunome et al. (2009) (Table 2). The same set of SNPs developed by Fukuoka et al. (2012) and the SSRs developed by Nunome et al. (2009) were used to build two intraspecific maps from two F2 populations, ALF2 and NAF2, to identify QTLs involved in parthenocarpy (Miyatake et al., 2012). The two main QTLs detected (*Cop3.1* and *Cop8.1*) were identified in both maps, which presented different LGs (12 versus 15), length (1,414 versus 1,153), and markers mapped (132 SNPs and 118 SSRs versus 125 SSRs and 49 SNPs), although shared one parent (parthenocarpic line AE-P03) (Table 2).

The first intraspecific population of RILs in eggplant was developed from the resistant bacterial wilt *R. solanacearum* line AG91-25 (MM960), derived from the Turkish line MM127 and a *S. aethiopicum* Aculeatum Group accession, and the commercial type line MM738 (Lebeau et al., 2013), which were previously used by Doganlar et al. (2002a). That population was used to

dissect the genetic control of resistance to *R. solanacearum* to four strains of phylotype I and identify genes/QTLs through developing a genetic map, which spanned 884 cM along 18 LGs (Table 2). Although, the map had a low saturation due to extremely low polymorphism rate, with only 119 markers mapped and most of which AFLPs, a major gene was identified (*ERs1*) in LG2. The low level of polymorphism in the RILs population showed with AFLPs drastically changed when the individuals were sequenced by GBS, identifying 1,779 filtered SNPs and allowing to build a high-density genetic map for screening four new strains of *R. solanacearum*, apart from the previous ones of Lebeau et al. (2013), belonging to phylotypes I, IIA, IIB and III (Salgon et al., 2017). The overall map length and the average marker density had been increased to 1,085 and 4.4 cM, respectively, and the LGs number had been reduced from 18 to 14 (Table 2), which lead to identifying a major QTL at the bottom of LG 9 that controls three phylotype I strains, corresponded to the previously identified major gene *ERs1* of Lebeau et al. (2013), and two other minor QTLs on LG 2, associated with partial resistance to strains of phylotypes I, IIA, III, and on LG 5, controlling the strains of phylotypes IIA and III.

This was a clear example of the high potential of the recent advances in genomics which improves extremely the resolution and precision of the genetic studies.

An additional study about of *Fusarium* wilt resistance (*F. oxysporum* f. sp. *melongenae*) was performed for mapping a resistant locus using different F2 and F3 populations, LWF2, EWF2, ALF2, LWF3, and EWF3, a backcross inbred lines population, ALBIL, and three genetic maps, AL2010, LW2012, and EW2012, previously developed by Fukuoka et al. (2012) and Hirakawa et al. (2014) (Table 2). A major resistance locus (*FMI*) was

identified at the end of chromosome 2, at the exact same position of *Rfo-sal* from *S. aethiopicum* gr. Gilo (Toppino et al., 2008), suggesting they might be orthologous.

The first attempt of association mapping based on linkage disequilibrium (LD) was performed with 141 eggplant accession from different countries and 105 SSRs developed by Nunome et al. (2009) to investigate nine fruit traits (Ge et al., 2013, Table 2). The analysis performed revealed a total of 49 marker associations related to eight phenotypic traits and 24 SSRs, being the total variation explained ranged from 4.5 to 22.8%.

A larger set of 191 accessions, including breeding lines, old varieties, and landraces from the Mediterranean basin and Asia, were investigated by Cericola et al. (2014) based on genome-wide association (GWA) approach for anthocyanin pigmentation and fruit color at two locations over two years using 314 SNPs developed by Barchi et al. (2011). A total of 56 associations were found between SNPs and anthocyanin content and fruit color-related traits, which were clustered into 12 groups and scattered over nine chromosomes, being eight of the groups overlapping with known QTL and demonstrating in that way the effectiveness of GWA approach. In addition, synteny with tomato allowed the identification of the genomic regions associated with anthocyanin accumulation in LGs 2, 5, and 12. Using the same association panel of 191 accessions and set of 314 SNPs (Table 2), Portis et al. (2015) examined the phenotype/genotype associations related to 33 traits (fruit, plant and leaf morphology traits) identifying 194 association to 30 traits, which involved 79 SNP loci in 39 distinct regions distributed across the 12 LGs.

A further improvement of Doganlar et al. (2002a) and Wu et al. (2009) maps was performed by increasing the number of markers to 864, by adding



400 AFLPs and 117 RFLPs, and using a larger F2 population (108 individuals) (Doganlar et al., 2014). In that way, the overall map length remained almost the same but the marker average distance decreased from 6.1 cM to 1.8 cM (Table 2). On the other hand, the improved map precision led the authors to revise the number of rearrangements between eggplant and tomato from 29 (five translocations and 24 inversions) (Doganlar et al., 2002a; Wu et al., 2009) to 33 (19 translocations and 14 inversions). Simultaneously, Frary et al. (2014) revised the QTLs identification by Doganlar et al. (2002b) and Frary et al. (2003), taking the opportunity of the improved map of Doganlar et al. (2014) to interrogate 58 F2 individuals (*S. linnaeanum* line MM195 and *S. melongena* line MM738) for 42 traits identifying 71 QTLs, twenty-two of which were novel.

In order to exploit the great genetic diversity of the wild relatives, a new interspecific map (named SMIBC) was developed using *S. incanum* MM577, which, as has been mentioned previously, is considered a powerful source of phenolics and tolerant to some abiotic stresses such as drought (Knapp et al., 2013).

In fact, with the objective of increasing the content of chlorogenic acid (CGA) in eggplant, which is usually the main phenolic compound (Whitaker et al., 2003; Prohens et al., 2013), a first set of introgression lines in the eggplant gene pool has been developed using *S. incanum* as a donor parent (Gramazio et al., 2016b). Thus, to track and easily introgress the alleles involved in the content of CGA (*phenylalanine ammonia lyase* (PAL), *cinnamate 4-hydroxylase* (C4H), *4-hydroxycinnamoyl-CoA ligase* (4CL), *hydroxycinnamoyl-coA shikimate/quinate hydroxycinnamoyl transferase* (HCT), *pcoumaroyl ester 3'-hydroxylase* (C3'H), and *hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase* (HQT)) in the genetic background of

eggplant, they were successfully mapped in different LGs using 91 BC1 individuals (Gramazio et al., 2014). In addition, five polyphenol oxidase enzymes (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*), which may be involved in the browning of the fruit flesh by oxidation of CGA and other phenols, mapped in a cluster in LG 8, as well as, candidate genes important in domestication such as fruit shape (*OVATE*, *SISUN1*) and prickliness. The mapping was assisted by the use of synteny of the orthologous genes in tomato using Tomato-EXPEN 2000 map (Fulton et al., 2002).

Furthermore, apart from tomato, SMIBC established a macro-synteny with four other eggplant maps (Nunome et al., 2009; Wu et al., 2009; Barchi et al., 2012; Fukuoka et al., 2012) by using shared markers. SMIBC spanned 1,085 cM along 12 LGs with a total of 243 markers (42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs and one morphological marker) (Table 2).

A few months later, at the end of 2014, the first draft of eggplant genome was finally released online (Hirakawa et al., 2014), where, apart from the genomic sequence, an integrated linkage map was also constructed from two F2 population (EWF2 and LWF2), which were previously used for mapping by Nunome et al. (2001) and Fukuoka et al. (2012). The map presented a total of 795 markers (574 SNPs and 221 SSRs) for an overall map length of 1,280 cM along 12 LGs and achieving the highest average marker density so far (0.7 cM) (Table 2).

In order to exploit genomic resources and genetic data for key agronomic traits, Rinaldi et al. (2016) performed a syntenic relationship and QTL orthology among eggplant, tomato, and pepper using their respective genome assemblies, although the genome sequences of the three species are different in coverage, assembly quality, and percentage of anchorage. While the comparison between tomato and pepper was quite comprehensive due to

the high quality of their assembly, the comparison with eggplant and tomato and eggplant and pepper was less exhaustive due to the impossibility to localize the physical position of the most eggplant QTL. Nevertheless, most of the previous rearrangements previously detected were confirmed and new ones were identified, even though an enhanced version of eggplant genome could have improved the precision of the analysis.

**Table 2.** Statistics of mapping studies performed in the eggplant gene pool

Source	Type of map	Aim of the study	Population	Parentals	Markers	Map length (cM)	Linkage groups	Average distance between markers (cM)
Nunome et al. 2001	Intraspecific genetic map	Fruit shape, color development	168 F2	EPL-1 ( <i>S. melongena</i> , Japan) WCGR112-8 ( <i>S. melongena</i> , India)	93 AFLPs 88 RAPDs	779	21	4.9
Doganlar et al. 2002a	Interspecific genetic map	Syntheny with tomato	58 F2	MM195 ( <i>S. linnaeanum</i> ) MM738 ( <i>S. melongena</i> )	232 RFLPs	1,480	12	7.6
Sunseri et al. 2003	Interspecific genetic map	<i>Verticillium</i> wilt tolerance	48 F2	<i>S. sodomium</i> [= <i>S. linnaeanum</i> ] Buia ( <i>S. melongena</i> )	156 AFLPs 117 RAPDs	736	12	2.7
Nunome et al. 2003	Intraspecific genetic map	Fruit shape, color development	120 F2	EPL-1 ( <i>S. melongena</i> , Japan) WCGR112-8 ( <i>S. melongena</i> , India)	101 RAPDs 54 AFLPs 7 SSRs	716	17	4.9

Nunome et al. 2009	Intraspecific genetic map	SSR development	94 F2	EPL-1 ( <i>S. melongena</i> , Japan) WCGR112-8 ( <i>S. melongena</i> , India)	245 SSRs	959	14	4.3
Wu et al. 2009	Interspecific genetic map	Syntheny with tomato	58 F2	MM195 ( <i>S. linnaeanum</i> ) MM738 ( <i>S. melongena</i> )	232 RFLPs 110 COSII 5 tomato-derived markers	1,535	12	6.1
Barchi et al. 2010	Intraspecific genetic map	<i>F. oxysporum</i> resistance	141 F2	305E40 (DH, from <i>S.</i> <i>melongena</i> and <i>S.</i> <i>eathiopicum</i> ) 67/3 ( <i>S. melongena</i> )	212 AFLPs 22 SSRs 1 RFLP 3 CAPS	718	12	3.0
Barchi et al. 2012	Intraspecific genetic map	Anthocyanin content	156 F2	305E40 (DH, from <i>S.</i> <i>melongena</i> and <i>S.</i> <i>eathiopicum</i> ) 67/3 ( <i>S. melongena</i> )	339 SNPs 33 SSRs 27 COSII 11 RFLPs 3 CAPS 2 HRM	1,390	12	3.8

Fukuoka et al. 2012	Intraspecific genetic map (LW2010 x AL2010 = LWA2010)	Development of <i>Solanum</i> orthologous (SOL) gene sets	90 F2 (LWF2)  93 F2 (ALF2)	LS1934 ( <i>S. melongena</i> ) WCGR112-8 ( <i>S. melongena</i> , India)  LS1934 ( <i>S. melongena</i> ) AE-P03 ( <i>S. melongena</i> )	639 SNPs 313 SSRs	1,285	12	1.4
Miyatake et al. 2012	Intraspecific genetic map	QTL identification for parthenocarpy	135 F2 (ALF2)  93 F2 (NAF2)	LS1934 ( <i>S. melongena</i> ) AE-P03 ( <i>S. melongena</i> ) Nakate-Shinkuro ( <i>S. melongena</i> ) AE-P03 ( <i>S. melongena</i> )	132 SNPs 118 SSRs  125 SSRs 49 SNPs	1,414	12	-
Lebeau et al. 2013	Intraspecific genetic map	Genetic control of resistance to <i>R. solanacearum</i>	178 F6	MM738 ( <i>S. melongena</i> ) AG91-25 (MM960, <i>S. melongena</i> )	91 AFLPs 26 SSRs 2 SRAPs	884	18	8.8
Ge et al. 2013	Association mapping (GWA)	association analysis of fruit traits	141 eggplant accessions	128 from China and the rest from USA, India, Japan, Italy, Malaysia.	105 SSRs	-	-	-

Cericola et al. 2013	Association mapping (GWA)	Anthocyanin pigmentation and fruit color	191 eggplant accessions	Breeding lines, old varieties, and landraces from the Mediterranean basin and Asia	314 SNPs	-	-	-
Doganlar et al. 2014	Interspecific genetic map	Syntheny with tomato	108 F2	MM195 ( <i>S. linnaeanum</i> ) MM738 ( <i>S. melongena</i> )	400 AFLPs 348 RFLP 116 COSII	1,518	12	1.8
Gramazio et al. 2014	Interspecific genetic map (SMIBC)	Mapping chlorogenic acid biosynthesis pathway and polyphenol oxidase genes	91 BC1	MM577 ( <i>S. incanum</i> ) AN-S-26 ( <i>S. melongena</i> )	99 SSRs 88 AFLPs 42 COSII 9 CAPS 4 SNPs 1 morphological	1,085	12	4.4
Hirakawa et al. 2014	Intraspecific genetic map (AL2010 x EWF2 x LWF2 = LWAE2012)	Draft genome sequence of eggplant	120 F2 (EWF2)  90 F2 (LWF2)	EPL-1 ( <i>S. melongena</i> , Japan) WCGR112-8 ( <i>S. melongena</i> , India)  LS1934 ( <i>S. melongena</i> ) WCGR112-8 ( <i>S. melongena</i> , India)	574 SNPs 221 SSRs	1,280	12	0.7

Portis et al. 2015	Association mapping (GWA)	Fruit, plant and leaf morphological traits	191 eggplant accessions	Breeding lines, old varieties, and landraces from the Mediterranean basin and Asia	314 SNPs	-	-	-
Miyatake et al. 2016	QTL mapping (AL2010, LW2012, EW2012)	Mapping of a resistance locus against <i>F. oxysporum</i> f. sp. <i>melongenae</i>	90 F2 (LWF2) 120 F2 (EWF2) 93 F2 (ALF2) 87 F3 (LWF3) 120 F3 (EWF3) 186 BILs (ALBIL)	LS1934 × WCGR112-8 EPL-1 × WCGR112-8 LS1934 × AE-P03 Single-seed descent from LWF2 Single-seed descent from EWF2 (LS1934 × AE-P03) × LS1934	-	-	-	-
Salgon et al., 2017	Intraspecific genetic map	Genetic control of resistance to <i>R. solanacearum</i>	180 F6	MM738 ( <i>S. melongena</i> ) AG91-25 (MM960, <i>S. melongena</i> )	867 SNPs 139 AFLPs 28 SSRs 1 SRAP	1,518	14	1.4



***Future direction for genetics and genomics tools for eggplant breeding***

In the last ten years from the genetic and genomic investigation standpoint, eggplant has shortened the gap with other important major crops like tomato, potato, and pepper. Nevertheless, more genomic resources are needed for an efficient breeding in order to develop new improved varieties that have to deal with a changing climate scenario and new and threatening biotic stresses, and to understand the genetic diversity, evolutionary history, domestication, and ecology in eggplant (Gramazio et al., 2016a). In this light, a high-quality genome sequence from an inbred line is absolutely necessary, in which the scaffolds should be assembled in chromosomes or pseudomolecules from sequence-based genetic and physical maps.

The sequencing technologies using long-reads such as PacBio combined with a huge amount of small-reads of Illumina could achieve a satisfactory result (Mavromatis et al., 2012), although long-read sequencing technologies combined with physical mapping approaches, like optical mapping, Hi-C and the Dovetail Chicago methods, are offering new and most accurate solution to genome assembly (Yuan et al., 2017).

An accurate reference genome assembly would encourage many research groups to sequence the genomes and the transcriptomes of their genotypes of interest or wild eggplant related species with a modest amount of economic resources, obtaining a meaningful information for a wide range of studies.

For instance, more attempts to develop experimental populations using wild relatives could be addressed if more information would be

available for marker assisted selection to introgress efficiently genomic fragments of allied species.

Up to now, the attempts to introgress wild relatives traits has been performed only sporadically and even though 25 allied eggplant species have been employed in interspecific cross with *S. melongena* (Rotino et al., 2014), only one set of introgression lines has been developed using *S. incanum* as a donor parent (Gramazio et al., 2016b) and no other biparental and multiparental populations are available.

The delay in developing populations using wild relatives is staggering compared with other crops like tomato, in which more than 96 genes and QTLs has been introgressed from 14 different wild relatives, comprising several sets of near isogenic lines (NILs), introgression lines (ILs) and RILs, as well as a multi-parent advanced generation intercross (MAGIC) population (Pascual et al., 2015; Redden et al. 2015).

A reliable reference genome would indeed speed up the development and precision of these populations in eggplant. Barrantes et al. (2014) developed most of the lines of an ILs between tomato and *S. pimpinellifolium* L. from a BC3S1 generation, four generations versus eight required to develop the set of ILs with *S. incanum* (Gramazio et al., 2016b). Both tomato and *S. pimpinellifolium* parents had advanced assemblies that allowed designing arrays to perform high-throughput genotyping since the early generations of ILs population development.

In the last few years, sequence based genotyping (SBG) technology, which includes methods for the simultaneous polymorphism discovery and genotyping like GBS, RAD, ddRAD and related methods, has been a relative economical way to produce information in a crop and

facilitate genotyping before a reference genome was available (Elshire et al., 2011).

Barchi et al. (2011) RAD approach identified ~10,000 SNPs and 2,000 putative SSRs from the parents of a mapping population with a reference genome. Similarly, a GBS was performed in a 180 F6 RILs population identifying 1,779 filtered SNPs and improved drastically the quality of an intraspecific genetic map (Salgon et al., 2017). These genotyping through sequencing methods constitute a really powerful tools to increase precision and accelerate genetic and genomic studies, although the legal dispute for exploited the patent of these methods driven up the prices of these technologies turning them economical prohibitive for many research groups.

Currently, many efforts are being dedicated to developing an alternative to SBG technologies, although many of them required prior knowledge. The information obtained by sequencing is necessary to design primers to interrogate the genomic regions of interest for allelic discovery or genotyping. Each targeting platform differs in throughput, cost, probes, multiplexing, the number of target regions and much more customizable parameters. Among the most-used alternatives are Sequenom MassARRAY iPLEX platform (Gabriel et al., 2009), single primer enrichment technology (SPET) (NuGEN, San Carlos, USA), TruSeq Amplicon Sequencing and Nextera Target Enrichment (Illumina, San Diego, USA), KASP SNP genotyping (LGC Genomics, UK).

A new promising approach, rAmpSeq, was announced at the end of 2016 as a robust genotyping platform using repetitive sequences (Buckler et al., 2016). This method used conserved regions to design PCR primers for amplifying thousands of middle repetitive regions and

interrogate thousands of markers. The authors affirmed that the cost per sample can be less than \$2 per sample, which would allow to genotype thousands of samples for a very reasonable cost. Other advantages are the use of PCR without high requirements of DNA quality and quantity and less PCR competition among amplicons due to fairly similar length and composition of the repetitive. On the other hand, compared to the SBG technologies, rAmpSeq identifies fewer markers, required prior information and generally screens intergenic regions, as well as, more challenging bioinformatic analysis. This approach can revolutionize the breeding and conservation biology in the immediate future, even though at moment further improvement are required and optimization in more crops a part of maize.

The genomics revolution, that has led a perspective change in our comprehension of evolution, domestication, genetic architecture and much more aspects, is far from slowing down. The new achievements in sequencing technologies and their decreasing in cost are accelerating the development of high-quality genome reference assemblies, high-throughput genotyping and markers-assisted breeding selection that is reflecting in a greater overall understanding of species and new improved varieties adapted to new upcoming scenarios.

In the near future sequencing 100s or 1,000s of samples will become routinely and affordable including for non-model species and for resource and infrastructure-limited institutions in the developing world. This will undoubtedly speed up and revolutionize the breeding of eggplant for the development of a new generation of cultivars with dramatically improved yield, quality and resilience.

***Acknowledgements***

This work has been funded in part by the initiative “Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives” which is supported by the Government of Norway. This project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew and implemented in partnership with national and international gene banks and plant breeding institutes around the world. For further information see the project website: <http://www.cwrdiversity.org/>. Funding has also been received from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) and from Spanish Ministerio de Economía, Industria y Competitividad and Fondo Europeo de Desarrollo Regional (grant AGL2015-64755-R from MINECO / FEDER). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral (Programa FPI de la UPV-Subprograma 1/2013 call) contract. Mariola Plazas is grateful to Ministerio de Economía, Industria y Competitividad for a post-doctoral grant within the Juan de la Cierva programme (FCJI-2015-24835). Giulio Mangino is grateful to Conselleria d’Educació, Investigació, Cultura i Esport de la Generalitat Valenciana for a pre-doctoral grant within the Santiago Grisolia programme (GRISOLIAP / 2016/012).



## **Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change**

Prohens J<sup>1\*</sup>, Gramazio P<sup>1</sup>, Plazas M<sup>1</sup>, Dempewolf H<sup>2</sup>, Kilian B<sup>2</sup>, Díez MJ<sup>1</sup>, Fita A<sup>1</sup>, Herraiz FJ<sup>1</sup>, Rodríguez-Burruezo A<sup>1</sup>, Soler S<sup>1</sup>, Knapp S<sup>3</sup>, Vilanova S<sup>1</sup>

<sup>1</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

<sup>2</sup> Global Crop Diversity Trust, Platz der Vereinten Nationen 7, Bonn, Germany

<sup>3</sup> Department of Life Sciences, Natural History Museum, London SW7 5BD, United Kingdom

\* Corresponding author: [jprohens@btc.upv.es](mailto:jprohens@btc.upv.es)

**Citation:** Prohens J, Gramazio P, Plazas M, Dempewolf H, Kilian B, Díez MJ, Fita A, Herraiz FJ, Rodríguez-Burruezo A, Soler S, Knapp S, Vilanova S (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213(7):158.

<https://doi.org/10.1007/s10681-017-1938-9>

## *Abstract*

The need to boost agricultural production in the coming decades in a climate change scenario requires new approaches for the development of new crop varieties that are more resilient and more efficient in the use of resources. Crop wild relatives (CWRs) are a source of variation for many traits of interest in breeding, in particular tolerance to abiotic and biotic stresses. However, their potential in plant breeding has largely remained unexploited. CWRs can make an effective contribution to broadening the genetic base of crops and to introgressing traits of interest, but their direct use by breeders in breeding programs is usually not feasible due to the presence of undesirable traits in CWRs (linkage drag) and frequent breeding barriers with the crop. Here we call for a new approach, which we tentatively call ‘introgressiomics’, which consists of mass scale development of plant materials and populations with introgressions from CWRs into the genetic background of crops. Introgressiomics is a form of pre-emptive breeding and can be focused, when looking for specific phenotypes, or un-focused, when it is aimed at creating highly diverse introgressed populations. Exploring germplasm collections and identifying adequate species and accessions from different gene pools encompassing a high diversity, using different strategies like the creation of germplasm diversity sets, Focused Identification of Germplasm Strategy (FIGS) or gap analysis, is a first step in introgressiomics. Interspecific hybridization and backcrossing is often a major barrier for introgressiomics, but a number of techniques can be used to potentially overcome these and produce introgression populations. The generation of chromosome substitution lines (CSLs), introgression lines (ILs), or multi-



parent advanced inter-cross (MAGIC) populations by means of marker-assisted selection allows not only the genetic analysis of traits present in CWRs, but also developing genetically characterized elite materials that can be easily incorporated in breeding programs. Genomic tools, in particular high-throughput molecular markers, facilitate the characterization and development of introgressiomics populations, while new plant breeding techniques (NPBTs) can enhance the introgression and use of genes from CWRs in the genetic background of crops. An efficient use of introgressiomics populations requires moving the materials into breeding pipelines. In this respect public-private partnerships (PPPs) can contribute to an increased use of introgressed materials by breeders. We hope that the introgressiomics approach will contribute to the development of a new generation of cultivars with dramatically improved yield and performance that may allow coping with the environmental changes caused by climate change while at the same time contributing to a more efficient and sustainable agriculture.

**Keywords:** crop wild relatives, plant genetic resources, introgression breeding, hybridization, backcrossing, genomics

**Abbreviations:**

CWR: Crop wild relative

FIGS: Focused identification of germplasm strategy

CSL: Chromosome substitution line

IL: Introgression line

MAGIC: Multi-parent advanced inter-cross

NPBTs: New plant breeding techniques

PPPs: Public-private partnerships

ITPGRFA: International treaty on plant genetic resources for food and agriculture

DivSeek: Diversity Seek

ToMV: Tomato mosaic virus

TYLCV: Tomato yellow leaf curl virus

SCREAMing: Systematic Creation of Really Exotic Abnormal Material

RIL: recombinant inbred line

NGS: Next generation sequencing

PPPP: Public-private pre-breeding partnerships

***Introduction***

The expected increasing demand of plant products in the coming decades, with an estimated need of the doubling in relation to 2005 levels of agricultural production by 2050 (Tilman et al., 2011) in a climate change scenario, represents a formidable challenge for plant breeders. The current yearly rate of yield increase for major crops is clearly insufficient to meet this goal (Ray et al., 2013). Forecasts of agricultural productivity indicate that agriculture will be one of the sectors most affected by climate change (Rosenzweig et al., 2014). The negative effects of climate change on agricultural productivity will probably be greatest in tropical and subtropical areas (Knox et al., 2012; Rosenzweig et al., 2014), where a large part of the human population lives in marginal conditions in developing countries. Climate change will undoubtedly increase both permanent and occasional abiotic stresses (drought, high temperatures, salinity) in a large part of the global agricultural land and has the potential to drive abandonment of cultivation and desertification in some regions of the world. In addition, climate change is expected to result in increased biotic stress, as many pests and diseases will potentially migrate to areas where they were not present due to environmental limitations (Bebber et al., 2013).

Although many improvements have been made in increasing the efficiency of inputs used in agriculture (energy, water, agrochemicals) in the near future higher yields will have to be obtained with less input (Godfray et al., 2010; Ray et al., 2013). Also, many agricultural lands have become salinized or have lost fertility due to mismanagement (Fita et al., 2015). The availability of new arable land is not only limited but it is

undesirable to increase the agricultural land area on a global scale given that conservation of biodiversity is important for maintenance of functioning ecosystems and genetic input to agriculture in the longer term.

Meeting the challenges of increasing agricultural production in the face of climate change will require new strategies to develop new crop varieties with increased resilience to climate change-related stresses (Fita et al., 2015). Significant advances have been made in developing varieties tolerant to abiotic or biotic stresses but mostly for monogenic or oligogenic traits. The genetic variation available in cultivated species for adaptation to climate change is generally limited and it may be impossible to achieve major advances by resorting only to alleles present in the cultivated genepool (Hajjar and Hodgkin, 2007; Vincent et al., 2013; Dempewolf et al., 2014; Barshefsky et al., 2014).

Breeders have long recognized the value of crop wild relatives (CWRs) as sources of novel beneficial variation for resistance or tolerance to stresses (Bessey, 1906; Harlan and de Wet, 1971; Hajjar and Hodgkin, 2007). In this respect, most of the modern varieties of several important crops, like wheat or tomato, carry introgressions from wild relatives (Menda et al., 2014; Wulff and Moscou, 2014). CWRs are wild species that are able to exchange genes with the cultivated taxa through sexual or somatic hybridization (Harlan and de Wet, 1971, Maxted et al., 2006). CWRs are phylogenetically closely related to the crop and often encompass great genetic diversity. Some CWRs develop well under marginal conditions (low rainfall, high temperatures, high pressure of pests and diseases) that would be highly stressful for elite varieties of crops (Dwivedi et al., 2008; Dempewolf et al., 2014; Warschefsky et al., 2014). The enormous potential of crop wild relatives for breeding new varieties

adapted to climate change has largely remained unexploited, however, and the use of CWRs in breeding has largely focused on introgressing loci for disease resistance, while tolerance to abiotic stress has been little explored (Hajjar and Hodgkin, 2007; Maxted and Kell, 2009).

Recently an initiative has been launched aiming at adapting agriculture to climate change through the use of crop wild relatives (CWRs) for improving the cultivated gene pool of 29 crops included in the Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (Dempewolf et al., 2014). This project is aiming at collecting, preserving, evaluating, and preparing the use of CWRs with the ultimate aim to broaden the genetic base and thereby adapt our crops to climate change. Other initiatives, like Diversity Seek (DivSeek) are aimed at unlocking the potential of the diversity of crops and wild relatives present in genebanks (Meyer, 2015)

In this paper we present a novel strategy, tentatively named ‘introgressiomics’, for the systematic exploitation of the native variation present in CWRs for an efficient adaptation of our crops to climate change.

## ***Crop wild relatives for broadening the genetic base of crops***

Plant domestication is a unique and complex evolutionary process in which natural and artificial selection resulted in new plants adapted to human needs (Meyer and Purugganan, 2013). Domesticated plants generally present a syndrome of distinctive morphological, physiological and reproductive features compared to their wild ancestors (Hammer, 1984). For example in cereals, non-shattering phenotypes, with larger seeds, synchronized ripening, change in the plant architecture and loss of seed dormancy are characteristic traits of domesticates (Salamini et al., 2002). In other crops, such as those used for their fruits, tubers or leaves, domestication involved loss of bitterness and anti-nutritional or toxic compounds, removal or reduction of physical defence mechanisms like prickles, and gigantism of the organs used by humans (Meyer et al., 2012a). One extreme example is that of giant pumpkins (*Cucurbita maxima*) with record weights of over a ton for a single fruit (Savage et al., 2015). Many crops also experienced changes in the reproductive system that isolated them from the CWRs (Meyer and Purugganan, 2013). In other cases crossability has been maintained within the genepool (Jarvis and Hodgkin, 1999; Meyer and Purugganan, 2013). Autogamy, which allows fixation and maintenance of selected characteristics, has been favoured in a number of crops (Meyer et al., 2012a). Increased yield potential is also a general feature of domestication.

On many occasions domestication has involved genetic bottlenecks, resulting from a founder effect (domestication from a limited number of individuals), reproductive isolation and/or from strong selection pressures during domestication or crop improvement (Dempewolf et al.,

2012; Meyer and Purugganan, 2013). This resulted in a narrow genetic base for most crops as compared to their closest wild progenitors and CWRs (Dwivedi et al., 2008; Hajjar and Hodgkin, 2008; Meyer and Purugganan, 2013; Dempewolf et al., 2014). Modern plant breeding has generally led to a further reduction of genetic diversity in comparison to modern elite varieties and landraces (Tanksley and Nelson, 1996; de Wouw et al., 2010). Notable exceptions to this trend are crops where modern varieties carry introgressions from CWRs. For example, in tomato, modern varieties are genetically more diverse than local landraces due to the multiple introgressions, usually for resistance to diseases, from CWRs (Sim et al., 2012b). Nonetheless, the diversity of modern tomato is much lower than that present in its CWRs (Aflitos et al., 2014; Dodsworth et al., 2016).

A narrow toolbox of alleles in the elite varieties limits options for to better adapt crops to climate change. CWRs are often adapted to environments that are more stressful than artificial agricultural conditions. The large genetic diversity present in CWRs thus may constitute a major source of variation for improving crops with higher resilience (Warschefsky et al., 2014).

From a breeder's point of view, the utilization of CWRs present some significant challenges (Hajjar and Hodgkin, 2007; Dwivedi et al., 2008; Meyer and Purugganan, 2013). For example, crossing barriers and low hybrid fertility or sterility represent major barriers for the use of some species of CWRs in breeding programs. Although well-adapted to their natural environment, CWRs often contain a range of undesirable traits to agricultural conditions (low yield, undesirable physical and chemical defence systems, unpleasant flavour, lack of adaptation to cultivated

conditions, etc.) (Meyer and Purugganan, 2013). In many cases, these “wild” traits are dominant and polygenic, and thus challenging to select against in the breeding programs. Linkage drag due to reduced recombination is another issue that is frequently observed in introgression programs. Once introgressed into the cultivated genetic background, the CWR chromosomal fragments are challenging to break into smaller components. These fragments often contain genes that confer undesirable phenotypes – and often these are linked to the gene/s controlling the traits of interest (Tanksley and Nelson, 1996; Wendler et al., 2015). In addition, phenotypes of interest in a CWR may not be expressed in a cultivated genetic background. Nonetheless, use of CWRs in breeding has allowed significant improvements in a number of crops (Hajjar and Hodgkin, 2007). Maxted and Kell (2009) list 183 CWR taxa of a total of 29 crops that have been used for the transfer of traits of interest to the crop, while Dempewolf et al. (2017) lists 4,175 potential or confirmed uses of CWR taxa in crop improvement research, spread across 127 crops and 970 CWR taxa. In several cases the use of CWRs in breeding research has not resulted in their use in the development of cultivated varieties (Kilian et al., 2011). Maxted and Kell (2009) list only 39 CWR taxa that have been utilized for the development of advanced cultivars in nine major cereal and legume crops. Furthermore, most uses of CWRs for improvement of these nine crops are related to resistance to diseases and pests (61%), while their use for the improvement of tolerance to abiotic stresses (16%) or yield (7%) has been much lower. Other uses (16%) mostly involve quality traits. In tomato, most of the introgressions from wild taxa have been aimed at disease resistance, with 18 resistances having been introduced from seven different species (Díez and Nuez, 2008). In sum, the utilization of CWRs



has mostly been restricted to resistance or tolerance to pests and diseases, while other potential uses, like adaptation to abiotic stresses, have been largely neglected. With their adaptations to challenging environments, the utilization of CWRs represents a largely untapped opportunity for breeders to improve the adaptation of crops to abiotic stresses such as drought, salinity and high temperatures (Dwivedi et al., 2008; Dempewolf et al., 2014; Fita et al., 2015).

The ‘genepool’ concept was established by Harlan and de Wet (1971) to denote differences between CWR taxa with regards to the ease of exchanging genes with the crop. Typically, the primary genepool includes CWRs that can be easily crossed with the crop and the offspring is fertile. Secondary genepool CWRs are those that can be crossed with the cultivated species, although sometimes the degree of success is low, and/or the offspring may present low fertility. Finally, the tertiary genepool is composed of CWRs which have strong reproductive barriers with the crop and obtaining hybrids may require specific techniques such as embryo rescue, use of bridge species, stigma excision, or the use of pollen mixtures. Furthermore, hybrids between the crop and tertiary genepool species are often sterile and polyploidization may be needed to restore fertility in some cases. Thus the use of tertiary genepool CWRs for crop improvement is more challenging (Harlan and de Wet, 1971; Khush and Brar, 1992; Dwivedi et al., 2008). Given that crossabilities have not been sufficiently studied for numerous CWRs, phylogenetic relationships are sometimes used to assign CWRs to the respective genepool (Maxted et al., 2016).

CWRs are not only of interest for their use in breeding, but can also be used directly. For example, CWRs with high vigour, resistance to

diseases or with higher tolerance to abiotic stresses can be used as rootstocks in fruit and vegetable crops (Schwarz et al., 2010). The wild eggplant *Solanum torvum* is commonly used for eggplant grafting due to its resistance to multiple diseases and high vigour (Ranil et al., 2015). CWRs may also be of direct use for creating new cultivated species, such as for tritordeum, which is an amphiploid hybrid between durum wheat and the wild *Hordeum chilense* (Martín et al., 1999). Tritordeum has good tolerance to drought and high temperatures (Villegas et al., 2010).

Although many CWRs are of interest for adapting our crops to climate change, many CWRs are also threatened by climate change and human impact, as demonstrated for CWRs of cowpea, peanut and potato, among others (Jarvis et al., 2008). The collection and conservation of the CWR diversity continues to be an urgent priority (Maxted and Kell, 2009; Dempewolf et al., 2014).

### ***The introgressomics approach***

As CWRs cannot be directly incorporated into commercial breeding programs, pre-breeding activities have to be initiated as a first step for the utilization of CWRs (Longin and Reif, 2014). Most pre-breeding works use CWRs as donors of novel genetic diversity and is based on the following rationale:

(1) *ad hoc*, whenever an urgent breeding need appears (most commonly the occurrence of a new disease or pest), sources of useful variation are being explored. Screening usually starts within the primary genepool, mostly among the cultivars and landraces, sometimes including the closest CWRs. If unsuccessful, then screening is expanded to secondary and tertiary genepools.

(2) Once a suitable source of variation is detected, a crossing (and backcrossing) program is initiated to introgress the gene(s) conferring the desirable trait into the crop. Typically a single CWR donor species and most commonly just a single accession (Hajjar and Hodgkin, 2007) is considered.

Depending on the trait and introgressed fragment, this conventional approach can be time consuming and can require several cycles of backcrossing and selection, followed by selfing and then again further cycles of selection. Examples can be found in tomato, where most modern commercial varieties of tomato harbor resistances to nematodes or viral diseases caused by *Tomato mosaic virus* (ToMV) or *Tomato yellow leaf curl virus* (TYLCV) (Díez and Nuez, 2008). Resistance to nematodes conferred by the *Mi* locus was introduced in the 1940s from *Solanum peruvianum* accession PI128657 (Smith, 1944), while resistance to ToMV

conferred by *Tm2<sup>2</sup>* was introduced in the 1960s from *S. peruvianum* accession number PI128650 (Alexander, 1963), and later the resistance to TYLCV conferred by *Ty-1* was provided by *S. chilense* accession LA1969 in the 1990s (Zamir et al., 1994). Another successful example of alien gene introgression was the introduction of the *Lr19* leaf rust resistance allele, from *Thinopyrum ponticum* to wheat (Sharma and Knott, 1966). This gene not only confers rust resistance to wheat, but also increases yield, biomass and grain number (Bedő and Láng, 2015).

Alternative approaches have been proposed. McIntosh (1992), for example, suggested the development of ‘pre-emptive’ breeding populations to introgress resistance loci to wheat rusts with the aim of having promising materials ready for the rapid generation of resistant cultivars – in case the predominant rust races changed or a major resistance locus break down. In this way breeders could quickly develop new resistant varieties. However, this procedure bears significant risks and is expensive. Breeders need a clear strategy, a long time horizon and significant human and financial resources. A modification to this original pre-emptive breeding approach is the creation of ‘pre-breeding populations’ by crossing the crop with one or several CWRs. Valkoun et al. (2001) created pre-breeding populations of wheat containing introgressions from five different CWRs from genera *Triticum* and *Aegilops*. Some of these carry beneficial traits under certain environments, such as earliness, short plant stature, drought tolerance or resistance to several rusts. Introgression lines (ILs) contain one or a few introgressed genome fragments from a CWR (Zamir, 2001; Gur and Zamir, 2004). Sets of ILs with overlapping fragments of different size can be exploited for

different purposes, including for the analysis of the genetic basis for traits of (Zamir, 2001; Alseikh et al., 2013; Guerrero et al., 2016).

Here we suggest a novel approach for the development of pre-breeding materials, tentatively called ‘introgressiomics’ (Figure 1). ‘Introgressiomics’ consists of a *mass scale systematic development of plant materials and populations carrying introgressions of genome fragments obtained from (mostly wild) crop relatives into the genetic background of crops that may allow developing new generations of cultivars with improved properties*. ‘Introgressiomics’ is aiming at the massive generation of introgression materials for future (foreseen and unforeseen) needs and therefore may be considered as an advanced form of pre-emptive breeding. The ultimate aim of introgressiomics is to provide breeders with a significantly enlarged genetic pool from which new generations of cultivars adapted to future challenges in crop production can be rapidly obtained. Importantly, the materials developed through an introgressiomics approach are ready to be directly incorporated into breeding pipelines. Such materials will facilitate the work of breeders, especially with respect to the challenge of breeding for complex traits (Cattivelli et al., 2008; Dempewolf et al., 2014).

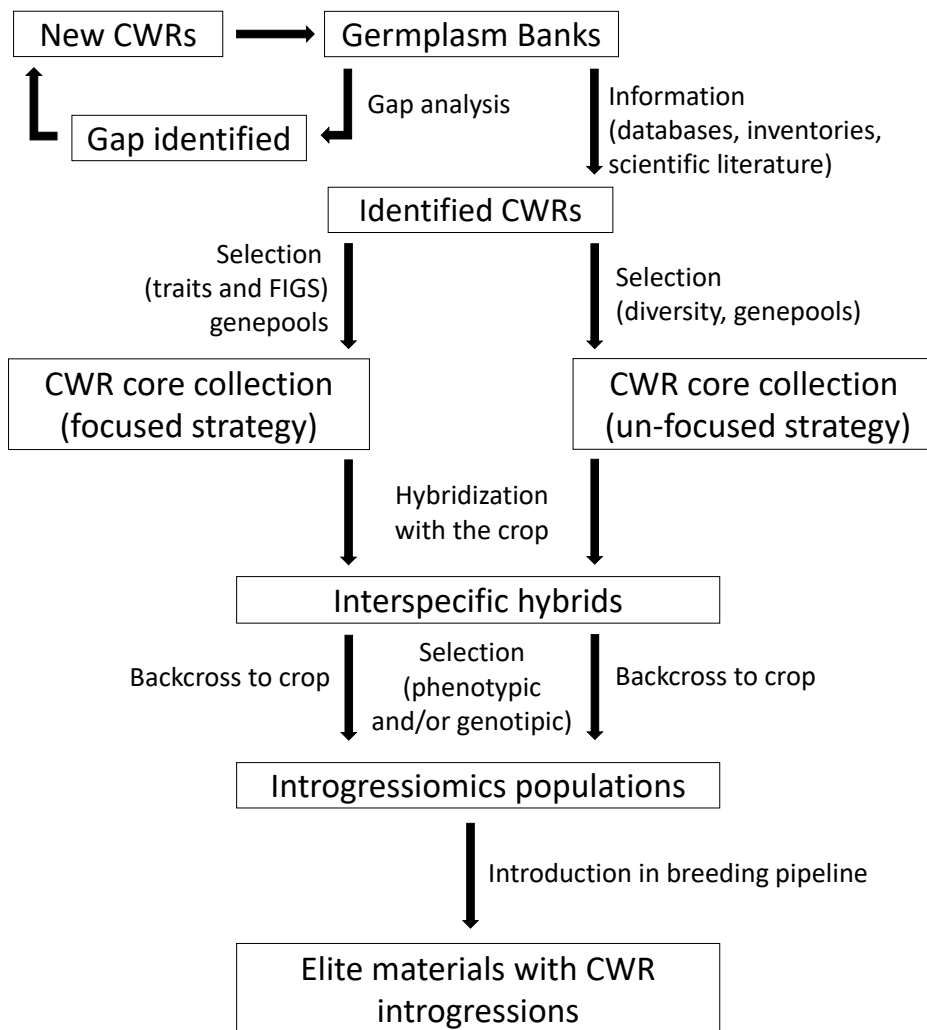
The first steps introgressiomics program are as follows: 1) identification of CWRs to be used in the program; 2) hybridization and backcrossing of the crop with a number (as large as possible) of CWRs from different genepools using special techniques when needed; 3) development of multiple special introgression populations containing introgressed fragments from one or several CWRs using genomic tools; 4) creating repositories of the introgressiomics populations and materials and

databases with phenotypic and genomic information; and lastly, 5) moving the materials into breeding pipelines (Figure 1).

An important aspect of introgressiomics is that populations may be created consisting of multiple genomic fragments obtained from different CWR sources. This approach of ‘pyramiding of genomic regions of interest’ would be beneficial for some breeding programs aiming at improving complex traits such as yield-related parameters.

Introgressiomics is similar, although more ambitious, than the approach proposed by Warschefsky et al. (2014), who pointed to the “*need for systematic efforts to introgress broad subsets of wild relative diversity to incorporate the range of useful adaptations for disease resistance, abiotic stress tolerance, and other agronomic challenges that are required in order to increase the resiliency and productivity of agriculture in the 21<sup>st</sup> century*”. Warschefsky et al. (2014) suggested a five-step approach similar to that of introgressiomics consisting of: 1) build comprehensive collections of CWRs, 2) sequence them, 3) create sets of purpose-driven hybrid populations and characterize them phenotypically, 4) develop a predictive network of genotype-phenotype associations, and 5) deploy the identified phenotypes into crop breeding pipelines. Their proposal presents substantial differences in its conception than the introgressiomics approach, as can be observed by comparing our Figure 1 with the Figure 1 of Warschefsky et al. (2014). The SCREAMing (Systematic Creation of Really Exotic Abnormal Material) approach has been proposed by Bert Vandenberg (pers. comm.) to develop pre-breeding populations using CWRs, but a comprehensive description of the approach has not been published yet.

Introgressiomics, as proposed here can be: i) focused and directed at specific phenotypes, which will determine what CWRs and types of populations are required, or ii) un-focused, in which the objective is the generation of introgression materials encompassing the maximum genetic diversity possible.



**Figure 1.** Outline of the introgressiomics pre-breeding approach workflow process, from the use of genetic resources to the development of elite materials with introgressions from crop wild relatives (CWRs).



## ***Identification of CWRs for introgressiomics***

The identification of CWRs to be utilized for introgression in the genetic background of the crop is the first step in introgressiomics (Figure 1). This largely depends on the strategy to be used (“focused” vs. “un-focused”) and also on the availability of CWRs (Vincent et al., 2013; Castañeda-Álvarez et al., 2016). In general, maximizing genetic diversity to the extent possible is a priority, as this increases the likelihood of capturing beneficial allelic diversity for breeding (Tanksley and McCouch, 1997; Porch et al., 2013). The number of CWR accessions to be used for introgressiomics depends on the resources available, the scale and time horizon, and objectives of the introgressiomics program. In order to select and identify CWRs, the ‘Harlan and de Wet’ Crop Wild Relatives Inventory (<http://www.cwrdiversity.org/checklist/>; Vincent et al., 2013) is a helpful resource. Currently, 173 crop gene pools are described. Previous and potential use cases of CWRs are also presented (Dempewolf et al., 2017).

One option for the identification of CWRs for ‘un-focused’ introgressiomics is the creation of germplasm diversity sets. This type of germplasm sets allows for a more rational use of accessions, while adjusting the number of accessions that can be realistically used (McKhann et al., 2004). Ideally, these germplasm subsets should include CWR representatives of all gene pools. However, germplasm diversity sets of CWRs merely based on diversity (genetic, phenotypic, origins, environmental, or a combination of them) might result in underrepresentation of CWRs from the primary gene pool, which generally are typically less variable than CWRs from the secondary or tertiary

genepools. Adjustments may be made in the germplasm diversity sets to ensure a good representation of primary genepool CWRs, which are the easiest to use for introgression breeding (Harlan and de Wet, 1971).

For ‘focused’ introgressomics, germplasm sets based only on diversity are not appropriate, as for this approach it is important that the germplasm sets are enriched for material which is harbouring alleles that may contribute to improving the target trait/s. For example, in creating introgressomics populations aimed at improving drought tolerance, emphasis should be given to include CWRs known to be tolerant to drought. In this case, strategies, like the Focused Identification of Germplasm Strategy (FIGS) could help selecting potentially beneficial material. FIGS is based on the assumption that accessions carry adaptive traits that reflect the selection pressures to which these were subjected in the environment during evolution (Street et al., 2016). Therefore, the retrieval of climatic and environmental data from databases such as WorldClim (<http://www.worldclim.org>; Hijmans et al., 2005), WorldGrids (<http://worldgrids.org>), or the CGIAR Consortium for Spatial Information (<http://www.cgiar-sci.org>), using the geographical coordinates may help in identifying promising CWR accessions. The FIGS strategy has been successfully applied for crops like wheat and beans (Bari et al., 2012; Khazaei et al., 2013). However, the potential of FIGS for selecting CWRs for introgression breeding is yet to be fully exploited (Street et al., 2016). A similar approach consisting of the use of eco-geographical data and expert assessment has allowed the identification of CWRs of sunflower to improve tolerance to various abiotic stresses (Kantar et al., 2015). Also, the utilization of historic characterization and evaluation data obtained *ex*

*situ* may be appropriate to identify CWR accessions of potential interest for breeding.

The creation of CWR sets requires the exploration of publicly available *ex situ* germplasm repositories as well as private working collections, when possible. Unlike accessions of the cultivated species, CWRs are often underrepresented in *ex situ* genebanks (Maxted and Kell, 2009; Castañeda-Álvarez et al., 2016), which means that relevant diversity may not be immediately available for introgressomics. This clearly reduces the potential for introgressing genes of interest from CWRs. The detection of potential collection gaps for CWRs is a first order step in order to fully exploit CWRs (Dempewolf et al., 2014). Gap analysis of germplasm collections is a strategy that was proposed by Maxted et al. (2008) and consists of identifying the taxa and geographical areas underrepresented in *ex situ* collections using taxonomic, genetic, eco-geographical, and threatened status data and other assessments. Additionally, comparison of georeferenced herbarium specimen data with the geographical coordinates of collection of the accessions conserved in *ex situ* collections allows identification of priority CWR taxa for collection as well as conservation areas as has been demonstrated in beans (Ramírez-Villegas et al., 2010) or eggplant (Syfert et al., 2016). In this respect, the Crop Wild Relative Occurrence Database (Centro Internacional de Agricultura Tropical, 2017) contains over five million records of germplasm accessions and herbarium databases on CWR and is of great interest for identification of gaps in the present germplasm collections.

Combining genomics and phenomics information might further assist the selection of CWRs for introgressomics, as shown by Vosman et al. (2016) for CWRs of cabbage resistant to the cabbage aphid. The

combination of phenotypic and genomics data from introgressiomics populations can be used for ‘introgressiomics by design’, which is an extension of the Breeding by Design<sup>TM</sup> approach (Peleman and van der Voort, 2003). The ‘introgressiomics by design’ approach would consist in mapping loci of agronomic interest from different donor CWR and using crossing and pyramiding schemes (Gur and Zamir, 2015) to introduce several favourable introgressions from different CWRs in a single line with the crop genetic background (Peleman and van der Voort, 2003).

### ***Interspecific hybridization and backcrossing***

A central step for introgressiomics is the hybridization between the crop and the CWR to create introgression populations. Theoretically, the crossability between the crop and all taxa within the primary genepool, which often comprise also the genome donors of the crop, should not present more difficulties than for intra-specific hybridizations (Harlan and de Wet, 1971, Jones, 2003, Maxted et al., 2006; Vincent et al., 2013). Although occasionally and depending on the crop, hybridization can be challenging for several reasons, such as non-synchronous flowering, cross-incompatibilities or fertility issues (Rieseberg and Carney, 1998; Dwivedi et al., 2008; Jones, 2003; Vincent et al., 2013).

Interspecific hybridization between the crop and CWR taxa from the secondary and tertiary genepools (wide or distant crosses) are often more challenging and hampered by pre-zygotic and post-zygotic reproductive barriers (Zenkteler, 1990; Khush and Brar, 1992) (Figure 2). In addition, although it may be possible to obtain hybrids, sterility issues may limit the development of introgressiomics populations. One important pre-zygotic reproductive barrier is pollen-style incompatibility (Dwivedi et al., 2008). As a consequence pollen does not germinate. When ‘unilateral’ incompatibility exists, hybrids can be obtained using the female parent in which the pollen is able to germinate (Figure 2). Unilateral incompatibility can frequently be observed when crossing self-compatible with self-incompatible species. The cross is successful when using the self-compatible species as a female parent (Dhaliwal, 1992). For example when crossing the cultivated tomato (*Solanum lycopersicum*) with its wild relative *Solanum peruvianum* the cross is unsuccessful when

using the self-incompatible *S. peruvianum* species as female parent, but possible when using it as male parent. In the former case the barrier is pre-zygotic and the pollen tube can not penetrate the style. The reciprocal cross is possible, although the embryo must be rescued and cultured *in vitro* to prevent the embryo abortion (Hogenboom, 1984). When ‘bilateral’ incompatibility exists, several techniques have proven useful to overcome it (Rieseberg and Carney, 1998; Dwivedi et al., 2008). Among others, pollen mixtures of compatible pollen, inactivated or not, stigma exertion, physical or chemical treatments to the stigmas, or *in vitro* fertilization can be applied (Zenkter, 1990; Khush and Brar, 1992; Dwivedi et al., 2008). Pollen mixture and treatment of stigmas with  $H_3BO_3$  and  $GA_3$  were used by Picó et al. (2002) to obtain offspring from crosses between the cultivated tomato and the distant wild relatives *S. chilense* and *S. peruvianum*.

If the male gamete is able to reach the ovary and to fertilize the female gamete in the ovule, post-zygotic barriers may prevent obtaining a viable hybrid seed. Most important post-zygotic barriers are caused by differences in number of chromosomes, ploidy levels, chromosomal alterations (rearrangements, duplications, inversions or translocations), embryo-endosperm incompatibility and hybrid lethality (Khush and Brar, 1992). These post-zygotic barriers (Figure 2) may result in embryo abortion, preventing the development of viable seed. However, different techniques can be applied to try to avoid post-zygotic barriers. For example, when crossing individuals of different ploidy levels, the use of technologies that modify the ploidy level can contribute to the development of viable hybrids. Most frequently used techniques to obtain parental lines with the same level of ploidy are anther culture to produce

haploid individuals (chromosome complement reduced to a half), or the duplication of the genome with colchicine (Khush and Brar, 1992; Rieseberg and Carney, 1998; Dwivedi et al., 2008), Also, embryo rescue, at an early stage of development (i.e. before it aborts due to either embryo-endosperm incompatibility or due to abnormal development resulting from genetic imbalance or other alterations), has been a successful tool in producing interspecific hybrids between crops and CWRs (Khush and Brar, 1992; Sharma et al., 1996).

When pre-zygotic and post-zygotic barriers cannot be overcome to obtain hybrid plants, the use of bridge species has to be considered (Shivanna and Bahadur, 2015). These species are compatible with one or both target taxa for interspecific hybridization. Once the interspecific hybrid between one of the taxa and the bridge species has been obtained, F1 plants can then be crossed with the other species - or with the other interspecific hybrid between the bridge species and the other parent. In this way, it has been possible to transfer genes from some CWRs to cultivated species. For example, the wild strawberry *Fragaria vesca* is a potential bridge species for introgression breeding of strawberry (*Fragaria x ananassa*) from the wild *F. nilgerriensis*, *F. nubicola*, *F. pentaphylla* and *F. viridis* (Bors and Sullivan, 2005). Somatic hybridization may be an alternative to obtain interspecific hybrids, as long as regeneration of plants is possible from hybrid somatic cells (Johnson and Veilleux, 2000). In this case the hybrid plant may be a polyploid having the full genome complements of both parental species. In some cases, like bread wheat, where crossing between elite varieties and CWR is frequently unsuccessful, inhibition of crossing is under the control of a few genes, so

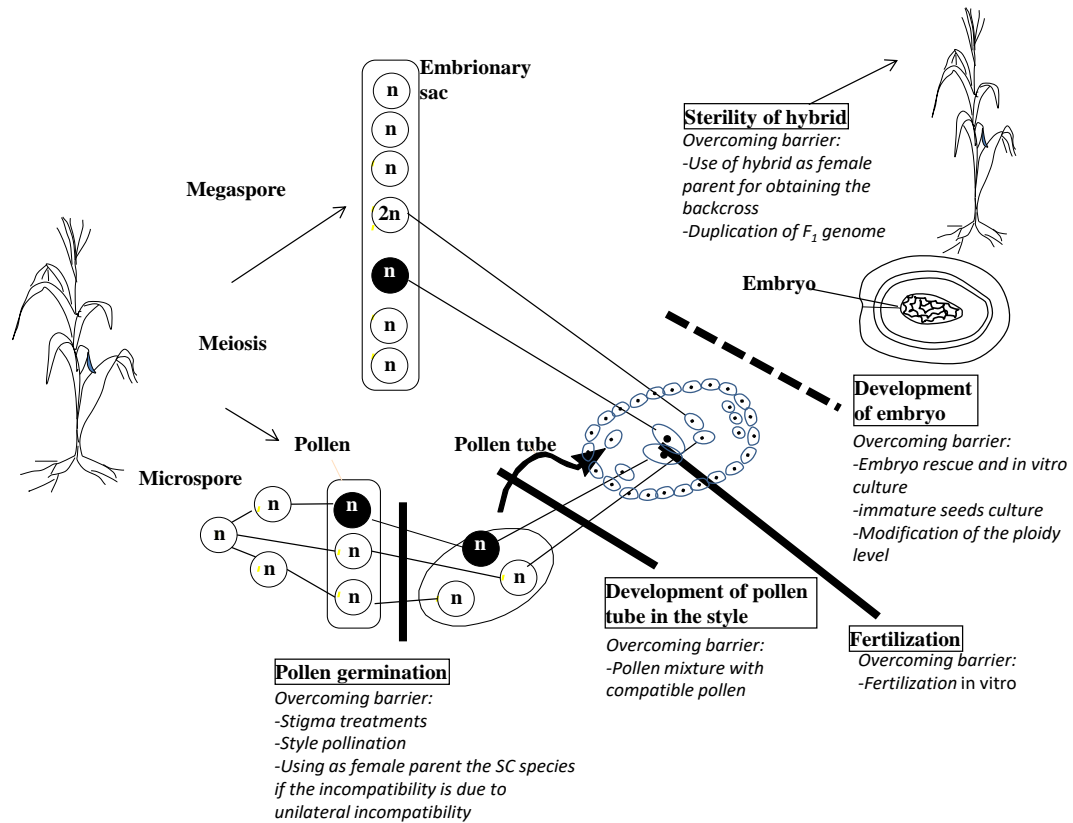
that wheat varieties or stocks homozygous for the crossability alleles *kr1* and *kr2* have increased crossability with CWR (Alfares et al., 2009).

Once interspecific hybrids have been obtained, backcross generations are produced (Zamir, 2001; Gur and Zamir, 2004). Although interspecific hybrids may be viable, they may be partially or completely sterile due to irregular chromosome pairing due to different ploidy levels or a low degree of synteny that results in non-viable gametes (De Storme and Mason, 2014). Some approaches can be used to increase the success rate of obtaining viable progeny from hybrids with reduced or low sterility. One strategy is to use the low-fertility hybrid as a female parent, since for the fertilization the pollen must have a high vigour to germinate and reach the ovule - a requirement not needed for the female gamete. This applies also to other plants from subsequent backcross generations. One way to recover fertility of the hybrid is duplicating its genome (Khush and Brar, 1992; Rieseberg and Carney, 1998; Shivanna and Bahadur, 2015). In this way normal chromosome pairing can be restored, although the hybrid plant will be polyploid, which may be an obstacle for obtaining subsequent generations due to different ploidy levels. To prevent this, a common technique is to duplicate also the genome of the recurrent cultivated parent to obtain backcross progeny and to restore the diploid status at a later stage. For example, Toppino et al. (2008) introgressed of resistance to *Fusarium* wilt from *S. aethiopicum* into the genetic background of eggplant (*S. melongena*) by backcrossing the tetraploid somatic hybrid to tetraploid eggplant to obtain the first backcross generation, which was subsequently returned to the diploid stage by anther culture. Generally, the loss of fertility in interspecific hybrids can be recovered in backcross generations, with increasing levels of fertility as the genome of the recurrent cultivated



parent is being recovered (Wall, 1970). However, occasionally the phenomenon of selective chromosome elimination of the donor parent may occur, complicating the introgression of fragments of CWR (Dwivedi et al., 2008). In other cases, recombination in the hybrids between chromosomes of the cultivated species and the CWR is suppressed or reduced, which makes introgression more difficult (Bedő and Láng, 2015). In this way, in wheat the gene *Ph1* suppresses pairing and recombination of wheat and alien chromosomes; however, in plants that are nullisomic for the *Ph1* gene, or in *ph1b* mutant stocks, homoeologous wheat and alien chromosomes can pair and recombine, which facilitates introgression from CWR in the cultivated wheat (Friebe et al., 2012).

Introgressiomics generally is conceived as the introgression of nuclear genes from wild species into a cultivated genetic background. However, introgression of cytoplasmic genes is also possible to produce alloplasmic hybrids with the nuclear genome of the cultivated species and the cytoplasm of the wild species (Khan et al., 2015). In this case, the cultivated species has to be used as male parent to ensure that the wild cytoplasm is maintained during the process. Nevertheless, it has to be taken into account that the interaction between the nuclear genome of the recurrent parent (cultivated species) and the cytoplasm of the donor (wild species) may result in cytoplasmic sterility (Prakash et al., 2001; Dwivedi et al., 2008; Khan et al., 2015), which may not be desirable if fertile plants of the cultivated species are needed. In that case, the cultivated species should be used as female parent to recover their cytoplasm, at some point during the backcross program.



**Figure 2.** Pre- and post-zygotic barriers difficulting interspecific hybridization and introgression breeding with crop wild relatives (CWRs) and strategies to overcome them.

### ***Development of introgressiomics populations***

Once interspecific hybrids have been obtained, introgressiomics populations (Figure 3), have to be developed. Some of the most commonly used populations, which contain genome fragments from CWR, are chromosome substitution lines (CSLs) and introgression lines (ILs). These are considered ‘immortal’ populations as they can be maintained by selfing, based on backcrossing the hybrid to the recurrent cultivated parent for several generations (Zamir, 2001; Lippman et al., 2007). CSLs consist of lines harboring the full genome complement of the crop except for one chromosome pair that corresponds to the wild donor parent (Cavanagh et al., 2008). CSLs have been extensively used in wheat breeding (Kilian et al. 2011; Khlestkina, 2014) and allow ascribing genes and traits to specific chromosomes. Although deleterious genes may be present in the introgressed pair of chromosomes, in contrast to CSLs ILs harbour the full genome of the crop, except for a small chromosomal segment of a donor parent, typically a CWR (Zamir, 2001). The development of collections of ILs, can be of great utility for breeders, because given that the introgression represents only a part of a chromosome, it may contain less deleterious alleles than CSLs (Gur and Zamir, 2004; Lippman et al., 2007). Both CSLs and ILs can be obtained through repeated backcrossing of the hybrid to the recurrent parent. Molecular markers and/or complementary cytogenetic techniques such as genome *in situ* hybridization (GISH) help tracking the introgressed fragments and thus support the selection of beneficial materials for subsequent backcross cycles (Gupta et al., 2016). A final step in obtaining ILs is selfing or obtaining doubled haploids to fix the introgressed fragment in a homozygous state (Herzog et al., 2014). Also,

ILs can be obtained from CSLs by crossing with the recurrent parent and subsequent selection of individuals in which recombination has taken place (Cavanagh et al., 2008). Similarly sub-ILs (Figure 3) can be obtained from ILs to shorten the introgressed fragment to reduce linkage drag (Alkeesh et al., 2013). A further advantage of ILs is the ability to intercross favourable traits that are present in different ILs for pyramiding favourable alleles (Gur and Zamir 2015). Advanced backcross populations, in which no marker assisted selection has been carried out during population development, may also be of interest as introgressiomics populations (Tanksley and Nelson, 1996; Cowling et al., 2009). In these populations the percentage of the donor genome will be reduced on average by half in each backcross cycle. Typically, these advanced backcross populations contain several fragments of the donor in several parts of the genome.

Other types of populations can be utilized to obtain ‘introgressiomics populations’. For example, recombinant inbred lines (RILs) obtained after crossing one cultivated species and a CWR followed by several generations of selfing have been very useful to dissect traits of interest present in CWR (Peleg et al., 2009; Salinas et al., 2013). However, RILs carry on average 50% of the wild parent, which in most cases are unsuitable for commercial breeding programs without a pre-breeding pipeline in place. An alternative to the backcross method is to develop multi-parental populations, such as multi-parent advanced generation inter-cross (MAGIC), which consist of multi-parent RIL populations (Cavanagh et al., 2008; Pascual et al., 2016).

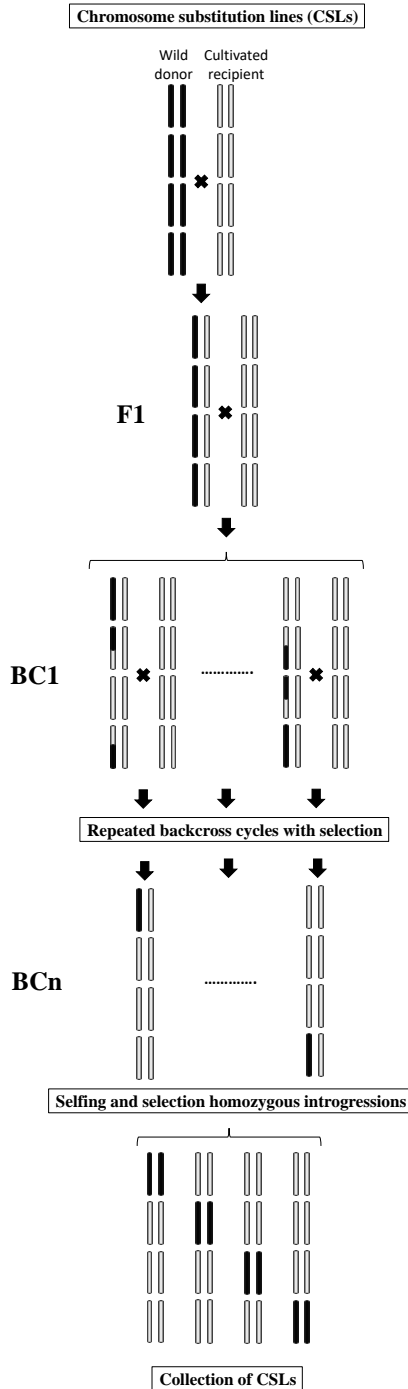
A more simplistic approach for the development of introgressiomics populations involves the development of several populations, each of which contains introgressions from a single donor

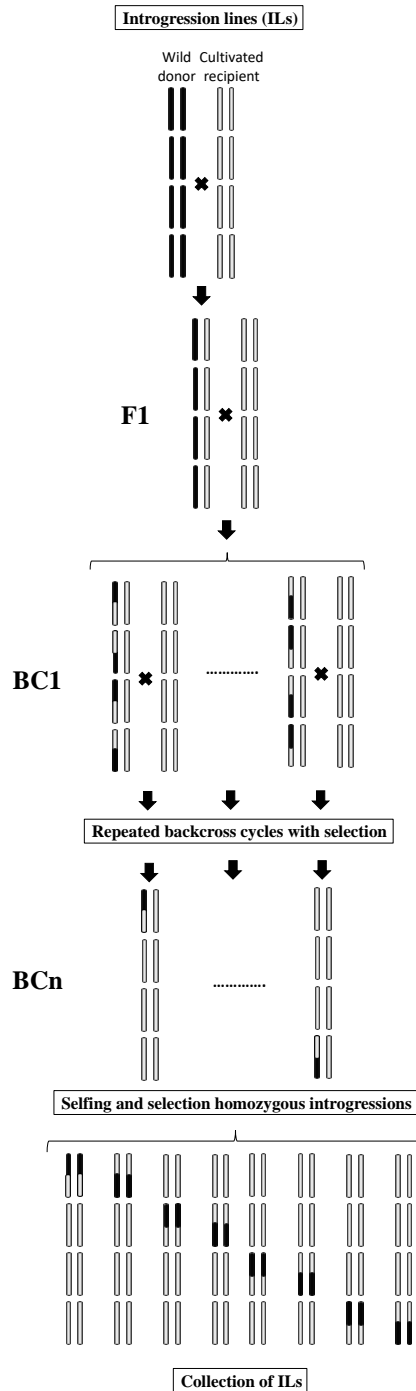
CWR parent. Introgressiomics populations may contain introgressions from several CWRs. For example, interspecific hybrids between two CWRs or double hybrids between four CWRs can be used as starting point to develop advanced backcross generations that may contain genome fragments from several CWRs. Also, more than one CWR parent can be included in MAGIC populations to generate RILs that have genomic fragments of several wild species.

One of the main challenges to the development and use of introgressiomics populations is linkage drag, based on reduced recombination at introgressed fragments (Tanksley and Nelson, 1996; Wendler et al., 2015). The use of molecular markers allows selecting individuals carrying recombined introgressed fragments (Alkeesh et al., 2013). Special mating designs, such as sib-mating, can be considered in the backcross scheme to enhance recombination (Wall, 1970; Liu et al., 1996; Rieseberg et al., 1996). Also, the use of congruency backcrossing, where backcrosses towards the recurrent parent are alternating to backcrosses to the donor parent to obtain fertile materials can be used to increase recombination (Haghighi and Ascher, 1988; Muñoz et al., 2004). Recently, Wendler et al. (2015) proposed crossing ILs with overlapping fragments originating from different donors to obtain progenies exhibiting higher levels of recombination within the introgressed fragment. However, in some occasions, due to the lack of synteny, recombination is prevented (Tanksley and Nelson, 1996). For example, in tomato, repeated attempts to reduce the introgression size associated to *Tomato yellow leaf curl virus* (TYLCV) resistance gene *Ty-1* introgressed from *S. chilense* failed due to lack of recombination caused by an chromosomal translocation (inversion) (Verlaan et al., 2011). In these cases reducing the linkage drag is not

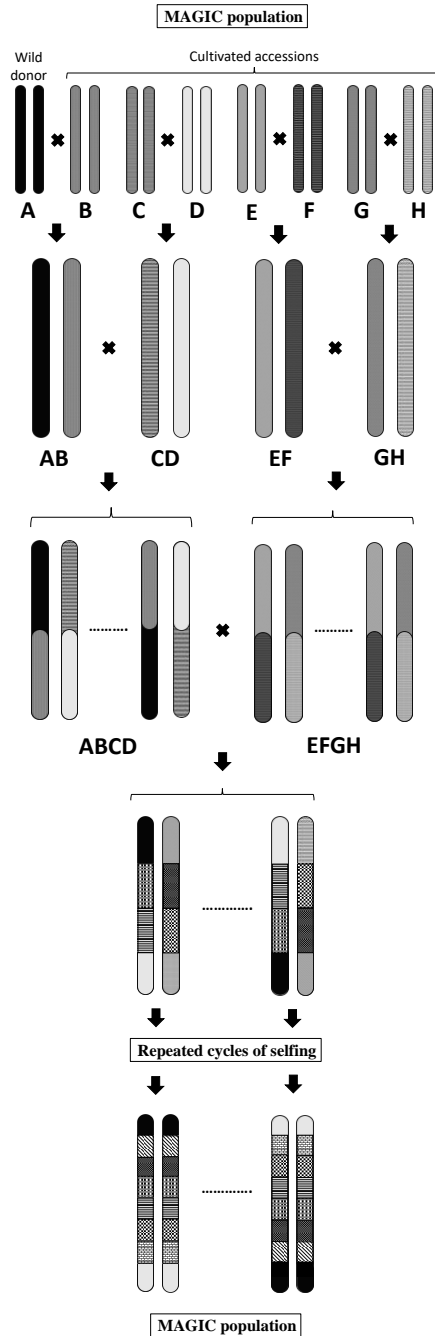
possible by using standard crossing schemes. One alternative to no or low levels of recombination due to the lack of crossing-over is the development of addition lines, containing an extra chromosome from a donor. In this way it may be possible to recover some plants in which a fragment of the donor is introgressed in the genome of the recipient parent (Friebe et al., 1996; Jacobsen and Schouten, 2007).

In polyploid crops, like bread wheat, an alternative to obtain introgressiomics populations using the diversity of CWR is the artificial synthesis of the allopolyploid using genetically diverse materials of the ancestor species as parental lines. For example, hexaploid wheats have been resynthesized through the hybridization of tetraploid wheats and *Aegilops* species (Bedó and Lang, 2015). Some of these materials were backcrossed to breeding lines and were found to be a source of variation for adaptation to several abiotic and biotic stresses (Trethowan and Mujeeb-Kazi, 2008). These synthetic introgression materials revealed a significant increase in genetic diversity (Warburton et al., 2006).









**Figure 3.** Scheme for the development of several types of introgressiomics populations: chromosome substitution lines (CSLs, above); introgression lines (ILs, below); and, multi-parent advanced inter-cross (MAGIC) lines (right).

## *Genomic tools and new plant breeding techniques for introgressomics*

Tremendous advances in next generation sequencing (NGS) technologies have provided breeders with a wide array of genomic tools that facilitate the introgression of CWR fragments into crops (Baute et al., 2015; Kole et al., 2015; Abberton et al., 2016; Brozynska et al., 2016). Some of the most important tools from genomics are the development of molecular markers distributed throughout the genome, the availability of dense genetic maps, the availability of reference genome sequences, transcriptome sequences, the discovery of regulatory elements, as well as gene annotations (Pérez-de-Castro et al., 2012). The use of these tools facilitates the identification of sources of variation and the development and characterization of introgressomics populations.

The availability of molecular markers that can be adapted to high-throughput genotyping platforms is of great utility for identifying those CWRs that may contribute beneficial diversity to the introgressomics approach. For example, the development of introgressomics populations in eggplant using wild relatives (Plazas et al., 2016) benefited from molecular marker studies that evaluated the genetic diversity and relationships between eggplant and a wide number of CWRs (Furini and Wunder, 2004; Vorontsova et al., 2013). This facilitated the selection of 15 eggplant CWR accessions from all gene pools for the initiation of an introgression breeding program (Plazas et al., 2016).

Molecular markers greatly facilitate the creation and characterization of CSL and IL populations. Marker-assisted foreground and background selection strategies - for those individuals that contain the target chromosomal segments, while at the same time selecting for

beneficial alleles in the recipient parental genome - are being applied in several crops (Zamir, 2001, Gur and Zamir, 2004; Pérez-de-Castro et al., 2012). In this way, collections of CSLs and ILs containing introgressed genome fragments of CWRs have been obtained in many crops (Gur and Zamir, 2004; Dwivedi et al., 2008; Khlestkina, 2014). The increased availability of high-density and mapped markers allows the fine mapping of the introgressed fragments, their extent as well as the breaking of ILs into sub-ILs by detecting recombinants (Lippman et al., 2007; Wendler et al., 2015; King et al., 2016).

Apart from providing molecular markers and identifying allelic variants, CWR transcriptome and genome sequences as well as resequencing studies can contribute other relevant information for introgressiomics. For example, synteny studies derived from the comparison of sequences of the crop and its CWRs may help identifying those genomic regions for which major chromosomal rearrangements have taken place during crop evolution and for which recombination will certainly be suppressed (Verlaan et al., 2011). Furthermore, sequence information facilitates allele mining in CWR collections in order to detect allelic variants of candidate genes controlling traits of interest (Ramkumar et al. 2016).

The use of the so-called ‘new plant breeding techniques’ can be useful for introgressiomics (Lusser et al., 2011; Hartung and Schiemann, 2014) as well. In this respect, cis-genesis (Jacobsen and Schouten, 2007) and genome editing, in particular based on CRISPR/Cas9 (Belhaj et al., 2013), are promising. Cis-genesis consists of the genetic transformation of a recipient parent, in the case of introgressiomics the crop, with isolated genes and their promoters from a crossable donor (CWR), without the

introduction of reporter or selectable markers from other organisms (Jacobsen and Schouten, 2007). In this way, by using cis-genesis, genes isolated from CWRs can be transferred in the genetic background of the crop without linkage drag (Tardi, 2016). Theoretically, by introducing genes isolated from CWRs from different gene pools it would be possible to produce collections of isogenic lines with genes from different donors, as well as to pyramid or to ‘stack’ genes from different species in a single genotype (Jo et al., 2014). This is particularly interesting in the case of secondary and tertiary gene pool species, with strong hybridization barriers (Khush and Brar, 1992; Dwivedi et al., 2008). For example, Jo et al. (2014) introduced two genes for resistance to late blight from the potato CWRs *S. stoloniferum* and *S. venturii* in several potato varieties. Several techniques exist for genome editing in plants (Cardi, 2016), but among them currently the most promising is the CRISPR/Cas9 system (Kole et al., 2015). In this way, it will potentially be possible to introgress traits whose expression depends on one or a few functional SNPs.

Cis-genic materials are genetically indistinguishable from those obtained by induced translocation breeding (Jacobsen and Schouten, 2007) and CRISPR/Cas edited genes cannot be distinguished from a natural mutation at the same locus (Belhaj et al., 2013). However, legal uncertainty and the risk of non-acceptance of these technologies by consumers restrict their current usefulness (Lusser et al., 2011; Hartung and Schiemann, 2014).

### ***Moving the introgressed material into the breeding pipeline***

The development of introgressomics materials may require several years depending on the crop and trait, among others. In most occasions, the development of these types of materials cannot be achieved within the time span of a single research project (which is on average 3-5 years). Therefore, on many occasions the public introgressomics programs may remain unfinished and the materials remain unused by the breeding sector. An important reason could be the lack of characterization and evaluation data, which is essential for breeders. In some cases, materials obtained in the public sector are available. For example, over 300 ILs and subIL of *S. pennellii* introgressed into cultivated tomato are available (Alkeesh et al., 2013). Also, the Tomato Genetic Resources Center (<http://tgrc.ucdavis.edu/>) maintains a large stock of pre-breeding materials readily usable by breeders.

We argue that long-term public-private pre-breeding partnerships (PPPPs) could bridge the gap between the development of introgressomics materials (pre-breeding) and their utilization in subsequent breeding programs (Lusser, 2014; Dempewolf et al., 2017). These PPPPs, which would involve scientists and breeders, could focus on characterising and evaluating the pre-bred materials in more detail, before breeders are willing to include them in their programs (Lusser, 2014; Warschefsky et al., 2014). Currently, phenotyping is the limiting factor in using introgressomics populations (Gur and Zamir 2004) and PPPPs could make a great contribution to an increased utilization by conducting more phenotyping experiments in diverse environments. Several successful examples of PPPPs are presented in Lusser (2014) or Moore (2015). For

an efficient utilization in breeding, plant materials and related information resulting as an outcome of public-private pre-breeding partnership programs should be publically accessible as a community resource, so that they can contribute to the development of new cultivars (Campi and Nuvolari, 2015). In this respect, global gateways such as Genesys (<https://www.genesys-pgr.org>) or international initiatives, like DivSeek, can make an important contribution to the sharing of phenotypic and genotypic characterization data of materials stored in genebanks (Meyer, 2015),

We propose that introgressiomics materials that clearly belong in the ‘pre-competitive domain’ should be made available under the terms of the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA). In this respect, one possibility could be to store introgressiomics materials in germplasm banks for long term conservation (Khoury et al., 2010). Clearly, the development of conservation strategies for pre-breeding materials including mapping populations, cytogenetic stocks and mutant collections would enhance the utilization of CWR and other sources of native diversity.

## ***Conclusions***

Introgression breeding from CWRs can make a major contribution to adapting our crops to climate change as well as to the development of new generations of crops with new and improved properties. The new approach we propose, ‘introgressiomics’, calls for an increased and most efficient use of CWRs in breeding by using the available genetic resources, introgression and population development schemes, phenotyping and genotyping tools and integration of data for obtaining elite materials that can be readily incorporated by breeders in their breeding programs. We hope that by using this holistic approach the practical use of CWRs in developing new commercial varieties with better characteristics will be enhanced. This may result in having more diverse, resilient and resource efficient crops that can contribute to a more sustainable and productive agriculture under the environmental changes resulting from climate change.

## *Acknowledgements*

This work was undertaken as part of the initiative “Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives”, which is supported by the Government of Norway. The Project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew and implemented in partnership with national and international gene banks and plant breeding institutes around the world. For further information see the project website: <http://www.cwrdiversity.org/>. This work has also been funded in part by European Union’s Horizon 2020 research and innovation programme under Grant agreement No 677379 (G2P-SOL) and from Spanish Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional (Grant AGL2015-64755-R from MINECO/FEDER, EU). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral (Programa FPI de la UPV-Subprograma 1/2013 call) contract.







# **OBJECTIVES**



The main objectives of this Ph.D. thesis are the development of genetic and genomic information and tools in the eggplant genepool and their use in establishing genetic relationships among and within cultivated eggplants and their wild relatives.

The knowledge produced in this thesis will be of great relevance for eggplant breeding and for paving the way for further studies in eggplant and may serve as a model for other species.

After a review of the state-of-the-art about the tools and genetic and genomic information available in eggplant together with our proposal of a new approach (introgressiomics) for using crop wild relatives in breeding for adaptation to climate change, we planned the following specific objectives:

1. Sequencing the transcriptome and performing a de novo assembly of two close relatives of the common eggplant, the wild *Solanum incanum* and the African cultivated scarlet eggplant *S. aethiopicum*.
  - 1.1. Structural and functional annotation of the assembled unigenes.
  - 1.2. Molecular marker discovery (SNPs, INDELs and SSRs).
  - 1.3. Comparison with two other available transcriptomes of *S. melongena* and *S. torvum*.
2. Validation of markers identified in the *S. incanum* and *S. aethiopicum* transcriptomes (SNPs and SSRs) in a collection of 48 accessions of the three cultivated eggplant species (*S. melongena*, *S. aethiopicum*

and *S. macrocarpon*) and their respective wild relatives (*S. incanum*, *S. anguivi* and *S. dasyphyllum*).

- 2.1. Comparison of SNPs and SSRs in order to estimate the best marker type in terms of reliability, time- and cost-effectiveness for genetic fingerprinting, diversity evaluation and the establishment of relationships in eggplant genepool.
3. High-throughput genotyping of 76 eggplant accessions from 17 species belonging to the primary, secondary and tertiary eggplant genepool using a genotyping-by-sequencing approach.
  - 3.1. Identification of a large set of SNPs amongst the species.
  - 3.2. Establishment of genetic relationships and phylogeny among phylogenetically very distant species from the eggplant genepool.
  - 3.3. Analysis of the effectiveness of high-throughput genotyping for establishing genetic relationships.
4. Development of an interspecific map between *S. melongena* and *S. incanum* (SMIBC).
  - 4.1. Development of a BC1 mapping population.
  - 4.2. Anchoring of the interspecific map SMIBC to a tomato genetic map and previous eggplant maps.

- 4.3. Mapping the candidate genes involved in the chlorogenic acid synthesis pathway, the genes involved in the fruit flesh browning and other genes of agronomic importance.
  - 4.4. Establishment of a macro-syteny between SMIBC interspecific genetic map and Tomato-EXPEN 2000 genetic map.
5. Development of an introgression line (IL) population between *S. melongena* and *S. incanum* from the BC1 population used for SMIBC map development.
- 5.1. Comparison between different genotyping methods and strategies for IL development.
  - 5.2. Positioning candidate genes for drought tolerance in fixed lines.





# **RESULTS**



**Chapter 1: Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding**

Gramazio P<sup>1\*</sup>, Blanca J<sup>1</sup>, Ziarsolo P<sup>1</sup>, Herraiz FJ<sup>1</sup>, Plazas M<sup>1</sup>, Prohens J<sup>1</sup>, Vilanova S<sup>1</sup>

<sup>1</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

\* Corresponding author: [piegra@upv.es](mailto:piegra@upv.es)

**Citation:** Gramazio P, Blanca J, Ziarsolo P, Herraiz FJ, Plazas M, Prohens J, Vilanova S (2016). Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding. *BMC Genomics* 1:300.

<https://doi.org/10.1186/s12864-016-2631-4>

## ***Abstract***

**Background:** *Solanum incanum* is a close wild relative of *S. melongena* with high contents of bioactive phenolics and drought tolerance. *S. aethiopicum* is a cultivated African eggplant cross-compatible with *S. melongena*. Despite their great interest in *S. melongena* breeding programs, the genomic resources for these species are scarce.

**Results:** RNA-Seq was performed with NGS from pooled RNA of young leaf, floral bud and young fruit tissues, generating more than one hundred millions raw reads per species. The transcriptomes were assembled in 83,905 unigenes for *S. incanum* and in 87,084 unigenes for *S. aethiopicum* with an average length of 696 and 722 bp, respectively. The unigenes were structurally and functionally annotated based on comparison with public databases by using bioinformatic tools. The single nucleotide variant calling analysis (SNPs and INDELs) was performed by mapping our *S. incanum* and *S. aethiopicum* reads, as well as reads from *S. melongena* and *S. torvum* available on NCBI database, against the eggplant genome. Both intraspecific and interspecific polymorphisms were identified and subsets of molecular markers were created for all species combinations. 36 SNVs were selected for validation in the *S. incanum* and *S. aethiopicum* accessions and 96% were correctly amplified confirming the polymorphisms. In addition, 976 and 1,278 SSRs were identified in *S. incanum* and *S. aethiopicum* transcriptomes respectively, and a set of them were validated.

**Conclusions:** This work provides a broad insight into gene sequences and allelic variation in *S. incanum* and *S. aethiopicum*. This work is a first step toward better understanding of target genes involved in metabolic pathways relevant for eggplant breeding. The molecular markers detected in this study could be used across all the eggplant genepool, which is of interest for breeding programs as well as to perform marker-trait association and QTL analysis studies.

**Keywords:** *Solanum incanum*, *S. aethiopicum*, eggplant genepool, de novo transcriptome, gene annotation, molecular marker discovery.

**Abbreviations:**

EC number: nzyme commission number

ESTs: expressed sequence tags

FISH: fluorescence in situ hybridization

GATK: genome analysis tool kit

GO term: gene ontology term

HRM: high-resolution melting

INDEL: insertion/deletions

KEGG: Kyoto encyclopedia of genes and genomes

NCBI: national center for biotechnology information

ncRNAs: noncoding RNAs

NGS: next generation sequencing

NR database: non-redundant protein database

ORF: open reading frame

SNP: Single nucleotide polymorphism

SNV: single nucleotide variation

SSR: simple sequence repeat

UTR: untranslated region

VCF: variant call format.

## **Introduction**

The global production of eggplants has considerably increased, especially in Asia and Africa, rising from 29 to 49 millions of tons in the last decade (Faostat 2015). Despite its economic importance, compared to other major vegetable crops, few efforts have been made to use related species for the genetic enhancement of common eggplant (*Solanum melongena* L.). In this respect, resistance and tolerance to biotic and abiotic stresses, as well as high levels of bioactive compounds have been found in *S. melongena* relatives (Collonnier et al., 2001b; Plazas et al., 2014b; Rotino et al., 2014), but they have not been widely used in breeding programs.

Although *S. melongena*, which was domesticated in Southeast Asia (Meyer et al., 2012a), is by far the economically most important cultivated eggplant, there are two other cultivated eggplant species of African origin, *S. aethiopicum* L. and *S. macrocarpon* L., which are major vegetable crops (Schippers et al., 2000; Maundu et al., 2009). The three species, together with a large number of wild species from the eggplant clade and the *anguivi* clade of *Solanum* subgenus *Leptostemonum* (Vorontsova et al., 2013), form part of the same genepool. Most of these species can be successfully hybridized with *S. melongena* (Rotino et al., 2014). Among them the wild *S. incanum* L. and the cultivated *S. aethiopicum* have been regarded as valuable sources of variation for *S. melongena* breeding (Mennella et al., 2010; Prohens et al., 2012; Prohens et al., 2013; Plazas et al., 2014a).

*Solanum incanum*, which has been regarded for a time as the wild ancestor of *S. melongena* (Lester et al., 1991), is naturally distributed in

desert and dryland areas in a broad area between northern Africa and the Middle East of Pakistan (Knapp et al., 2013). *S. incanum* is considered a powerful source of bioactive phenolics, mainly chlorogenic acid (5-O-caffeoyl-quinic acid) and to lesser extent N-(E)-caffeoylputrescine, 3-O-malonyl-5-O-(E)-caffeoylquinic acid and 5-O-malonyl-4-O-(E)-caffeoylquinic acid (Prohens et al., 2013), showing a content several times higher than that of cultivated eggplants (Prohens et al., 2013; Plazas et al., 2014). Moreover, *S. incanum* presents resistance at some fungal diseases, like *Fusarium oxysporum* and *Phomopsis vexans* and tolerance to abiotic stresses such as drought (Yamakawa et al., 1979; Lester et al., 1991; Collonnier et al., 2001b). Fully fertile interspecific hybrids have been obtained between *S. incanum* and *S. melongena* with a regular meiosis (Lester et al., 1991; Anis et al., 1994; Behera et al., 2002; Prohens et al., 2013), as well as a backcross population to *S. melongena* that has allowed the development of an interspecific genetic linkage map (Gramazio et al., 2014). *S. aethiopicum* is the second most important cultivated eggplant, and its cultivation is widespread in Africa, mainly in the west and central part, as well as in some parts of Caribbean, Brazil and south Italy (Schippers et al., 2000; Maundu et al., 2009; Sunseri et al., 2010). Generally, this species is divided in four cultivar groups, namely Aculeatum, Gilo, Kumba, and Shum (Lester et al., 1986), with the Gilo group, used for its edible oval to rounded fruits, the most important group in the *S. aethiopicum* complex (Lester et al., 1986; Sunseri et al., 2010; Plazas et al., 2014a). *S. aethiopicum* is of interest for *S. melongena* breeding as resistance to fungi (*Fusarium oxysporum*, *F. solani*, *Pythium vexans*, *Phytophthora parasitica*), bacteria (*Ralstonia solanacearum*), insect (*Leucinodes orbonalis*) as well as root-knot nematodes



(*Meloidogyne incognita*) has been found in different materials of this species (Cappelli et al., 1995; Rizza et al., 2002; Gisbert et al., 2011). Although different degrees of fertility have been found in interspecific hybrids between *S. melongena* and *S. aethiopicum*, backcrosses to *S. melongena* and introgression materials have been obtained (Rizza et al., 2002; Mennella et al., 2010; Prohens et al., 2012). As *S. aethiopicum* is a cultivated species, it does not present undesirable traits characteristic of wild species (e.g., prickliness, small fruit size, high content in saponins and glycoalkaloids, seed dormancy, etc.) that have to be removed in breeding programs. Also interspecific hybrids between *S. melongena* and *S. aethiopicum* are highly vigorous and of interest for being used as rootstocks of *S. melongena* (Gisbert et al., 2011).

Despite the importance of eggplants for security food for millions of people, genomics studies in this group have been limited. Only *S. melongena* has received some attention, with several intraspecific and interspecific genetic maps (Nunome et al., 2001; Doganlar et al., 2002b; Barchi et al., 2011; Fukuoka et al., 2012; Gramazio et al., 2014), collections of molecular markers (Nunome et al., 2009, Barchi et al., 2011; Vilanova et al., 2012) a set of 16000 unigenes (Nunome et al., 2001), a de novo transcriptome assembly (Yang et al., 2014), and a draft genome sequence (Hirakawa et al., 2014) being available. Genomic resources in *S. melongena* relatives are generally scarce. In this respect, while for *S. melongena* there are 126,715 DNA and RNA sequences deposited in NCBI (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) nucleotide database (on September 2015), of which 100,389 correspond to ESTs (expressed sequence tag) sequences, only 68 sequences are available from *S. aethiopicum*, none of which is an

ESTs. In *S. incanum*, a transcriptome assembly was released recently (GAYS 00000000.1), but it is still unpublished and no analyses have been released up to now.

The only exception concerns *S. torvum* Swartz, with 133,602 DNA and RNA sequences, of which 28,745 are ESTs. *Solanum torvum*, also known as turkey berry, is used as a vegetable and medicinal plant (Mohan et al., 2010; Gandhi et al., 2011; Jaiswal et al., 2012). Also *S. torvum* shows resistance to many plant and soil-borne disease, such as *F. oxysporum*, *Verticillium dahliae*, *P. parasitica*, *R. solanacearum*, and *Meloidogyne* spp. (Collonnier et al., 2001b; Gousset et al., 2005; Bagnaresi et al., 2013). Many efforts have been done to transfer these resistances through different biotechnological approaches, due to the high sterility of hybrids obtained via conventional crosses (Guri et al., 1988; Sihachakr et al., 1989; Collonnier et al., 2003; Kumchai et al., 2013; YiKui et al., 2014). However, at present, the main use of *S. torvum* is as a rootstock for eggplant (Takeda et al., 2007; Sabatino et al., 2013; Miceli et al., 2014).

Yang et al. (Yang et al., 2014) sequenced simultaneously the transcriptomes of *S. torvum* and *S. melongena*, providing valuable sets of unigenes and detailed information about the two species. However, this study did not include the discovery of molecular markers, which could have been of great assistance in the breeding programs within and between each species.

The aims of the present study are building two transcriptomes from *S. incanum* and *S. aethiopicum* through the generation of expressed sequence tags (ESTs) using RNA-Seq, providing genomic tools in these relatives of *S. melongena*. This will be the starting point for gene discovery, splicing patterns and other post-transcriptional modifications,

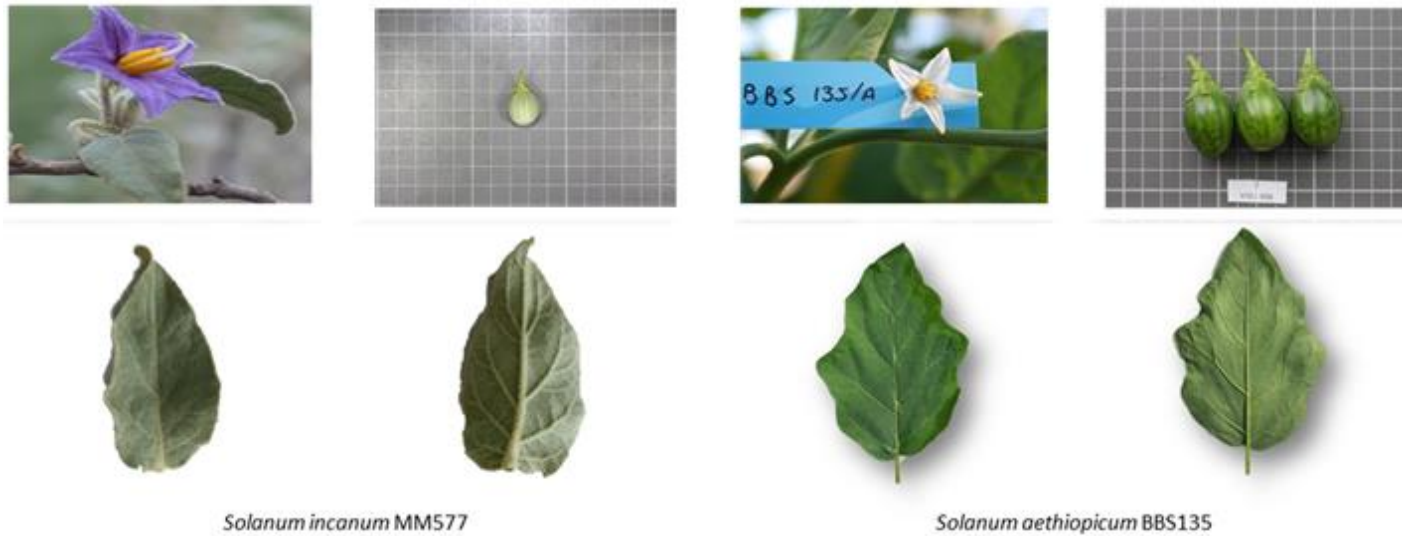
as well as expression levels of transcripts during development and under different conditions. Furthermore, the trimmed transcripts of *S. aethiopicum* and *S. incanum* and the transcripts of *S. melongena* and *S. torvum*, downloaded from NCBI database, were mapped against the eggplant genome to discover the molecular variations within and between species in order to create large subsets of markers directly applicable in breeding programs along to the eggplant genepool. All this information will contribute to the utilization of these species for *S. melongena* breeding, as well as to the enhancement of the cultivated, but neglected, *S. aethiopicum*.

## *Materials and methods*

### **Plant material**

The materials used were *S. incanum* accession MM577, which was collected in the wild in Israel and *S. aethiopicum* accession BBS135, which belongs to the Gilo cultivar group and was originally collected in Ivory Coast. *Solanum incanum* accession MM577 is a spiny plant with small green rounded fruit and purple corolla while the *S. aethiopicum* accession BBS135 is thornless and presents green obovoid fruits and white corolla (Figure 1). *Solanum incanum* accession MM577 has been used by our group as a parental to develop an interspecific genetic linkage map with *Solanum melongena* in which the candidate genes involved in the core chlorogenic acid synthesis pathway were mapped (Gramazio et al., 2014).

Plants of both accessions were grown in a greenhouse at the Universitat Politècnica de València (Valencia, Spain). Tissue samples were taken all at once at the stages of young leaf, floral bud and young fruit from a single plant and immediately frozen in liquid nitrogen and stored at -80 °C until used for RNA extraction.



**Figure 1.** Flower, fruit and leaf of *S. incanum* (left) and *S. aethiopicum* (right). The cells in the fruit picture grid have a size of 1 cm x 1 cm.

## **RNA extraction for Illumina sequencing**

About 100 mg of tissue were powdered in liquid nitrogen with a mortar and pestle. TRI Reagent® Protocol (Sigma-Aldrich, St. Louis, USA) was used for the total RNA extraction. In order to avoid DNA contamination DNase I Recombinant, RNase-free (Roche, Basel, Switzerland) was used. RNA integrity was confirmed by agarose electrophoresis and RNA quantification was performed using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA). Equal amounts of total RNA from each tissue were pooled for each accession and send to Macrogen Korea (Seoul, South Korea). After the construction of paired-end library (insert size of 300 bp), RNA-Seq was performed in HiSeq 2000 sequencer (Illumina, San Diego, USA). The raw sequences obtained are available in the Sequence Read Archive at the National Center for Biotechnology Information (NCBI) at the accession number (SRS1054263) for *S. incanum* and at the accession number (SRS1052489) for *S. aethiopicum*.

## **Sequence data analysis and De novo assembly**

The quality of the reads generated by Illumina was checked using the FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). In order to obtain high-quality data, the raw reads were pre-processed and trimmed using in-house developed software, NGS\_CRUMBS [[http://bioinf.comav.upv.es/ngs\\_crumbs](http://bioinf.comav.upv.es/ngs_crumbs)]. Through the different utilities the adapters used during the sequencing process were removed, as well as, low quality sequences, with a Phred quality score  $Q < 20$  and ambiguous

100

sequences with N. The trimmed reads were finally assembled into transcripts using Trinity (Grabherr et al., 2011), using default setting, which was specifically developed for de novo transcriptome assembly and for short-read RNA-Seq like Illumina HiSeq 2000.

In order to reduce the redundancy, the assembled transcriptomes were screened with CAP3 program (Huang and Madan, 2006). CAP3 uses base quality values, merging transcripts which overlap at least 200 bp with an identity of 99%. After that, to remove low complex sequences, the transcripts which have shown a DUST score less than 7 were masked (Morgulis et al., 2006). The estimation of transcript expression levels were calculated using RSEM software (RNA-Seq by Expectation-Maximization) (Li et al., 2011) and subsequently the most expressed transcripts of each Trinity transcript cluster were selected to create a set of unigenes for each species. Then both sets of unigenes were blasted (cut-off value of  $1e-20$ ) against the eggplant genome in order to obtain the physical position. The distribution of unigenes over eggplant genome was graphically depicted with Circos software (Krzywinski et al., 2009).

## **Structural and functional annotation**

The set of assembled transcripts was compared using BlastX (cut-off value of  $1e-20$ ) against four public protein databases in the following order: Swiss-Prot ([http://web.expasy.org/docs/swiss-prot\\_guideline.html](http://web.expasy.org/docs/swiss-prot_guideline.html)), ITAG2.4 ([ftp://ftp.sgn.cornell.edu/tomato\\_genome/annotation/](ftp://ftp.sgn.cornell.edu/tomato_genome/annotation/)), Arabidopsis (<http://www.arabidopsis.org/>) and UniRef90 (<http://www.ebi.ac.uk/uniprot/database/download.html>). If a transcript

gave a blast hit in the first database, no further searches were done, otherwise a second, third or fourth blast was performed.

Subsequently, a functional annotation was realized using Blast2GO software (Conesa and Gotz, 2008) to assign at the transcripts the corresponding GO terms (Gene Ontology) (Trapnell et al., 2009) and EC number (Enzyme Commission number) (Webb, 1992). For this purpose a BlastX (cut-off value of  $1e-20$ ) was performed in the NR database (Non Redundant) and the resulting hits were mapped into gene ontology database to assign the correspondent annotation. Blast2GO was used also to obtain the KEGGs pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (version 73.0, January 1, 2015) (Kanehisa and Goto, 1999).

Additionally, best reciprocal hits with BlastN (cut-off value of  $1e-20$ ) were performed with tomato genome (version SL2.50) (Tomato Genome Consortium, 2012) to detect orthologs. On the other hand the tomato genome was employed to predict gene model and intron frames, using est2genome software (Kumar et al., 2012). Open reading frames (ORFs) were predicted with ESTScan program (Mott, 1997).

### **Mapping transcriptomes against eggplant genome**

The high-quality clean reads from our RNA-Seq experiment (*S. incanum* and *S. aethiopicum*) were aligned against the eggplant genome using the Top Hat program (Trapnell et al., 2009). The TopHat pipeline is very fast and specifically designed for detecting junctions even in genes transcribed at very low levels. Because only 20% of *S. torvum* reads mapped in eggplant genome with Top Hat, we decide to use a BWA (Li,



2013), a most suited mapper in the case of greater genetic distance. Subsequently, the reads were realigned using the GATK (Genome Analysis Tool Kit) software in order to split the reads (McKenna et al., 2010).

Raw paired-end reads from *Solanum melongena* and *S. torvum* were downloaded through the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). The *S. melongena* reads were deposited under accession number [SRA: SRR1104129] and *S. torvum* reads under accession [SRA: SRR1104128]. The raw sequences were processed and trimmed as described above. The FASTA sequence of the draft eggplant genome was downloaded from the Eggplant Genome Database (<http://eggplant.kazusa.or.jp/>).

## **Molecular markers discovery**

### *SNVs*

SNVs (Single nucleotide variations, SNPs and INDELs) were detected using the FreeBayes program (Wittwer et al., 2003), a bayesian haplotype-based SNP caller, using the Top Hat alignment. To verify the quality of SNP calling, three SNVs per eggplant chromosome were validated in the sequenced genotypes. Every SNV locus was checked by IGV software (<http://www.broadinstitute.org/igv/>), to select the most polymorphic loci with the higher coverage. Primers pairs were designed in flanking regions using Primers3 (<http://bioinfo.ut.ee/primer3-0.4.0/>).

HRM-based PCR was used to validate the SNPs in a LightCycler 480 Real-Time PCR (Roche, Basel, Switzerland). The reactions were performed in a 10  $\mu$ L: 5  $\mu$ L Master Mix 2X, 0.8  $\mu$ L MgCl<sub>2</sub> 25 mM, 0.25

$\mu$ L each primer, 1.7  $\mu$ L water and 2  $\mu$ L DNA 30 ng/ $\mu$ L with the following touchdown PCR program: denaturation at 95°C for 10 min, followed by 55 cycles of 10 s at 95°C, 15 s at 65°C (decreasing 1°C each cycle until 55°C) and of 15 s at 72°C, finally the melting at 1 min at 95 °C, 1min at 40°C, 1 sec at 60°C and rising the temperature at 0.02°C/s until 95°C.

Although all SNVs matched the quality criteria, not all of them seemed equally reliable. Different filters, developed by ours ([https://bioinf.comav.upv.es/ngs\\_crumbs/](https://bioinf.comav.upv.es/ngs_crumbs/)), have been applied to VCF file (Variant Call Format) in order to maximize the polymorphism validation. The settings of the filters are provided in the Additional file 13. The VKS filter was applied to differentiate INDELs from SNPs, the filter CS60 to detect if the SNV was closer than 60 nucleotides to another SNV, the filter CL60 to identify SNVs closer than 60 nucleotides to the transcript edge and the filter HV0.05 to determine if the region had more than 5 SNVs per 100 bases. All these filters allow selecting SVNs, with small amplicons size (e.g. 80-100 bp) and are suitable for manually validation such as with High Resolution Melting (HRM) as well as for high-throughput genotyping platform (Abajian, 1994; Daunay et al., 2008; Gupta et al., 2008). If the SNV is going to be genotyped by CAPS, the filter CEF will help to select common low-priced digestion enzymes.

### *SSRs*

The annotation of SSRs was carried out with Sputnik software (Ding et al., 2015), selecting the sequences containing  $\geq 9$  di-,  $\geq 6$  tri-, or  $\geq 4$  tetranucleotide motifs. The sequences of unigenes which contain EST-SSR were blasted against the eggplant genome database in order to know

their physical position while their region in the transcripts (ORFs, 3'-UTR and 5'-UTR) were detected using the Bedtools utilities (Quinlan and Hall, 2010). The representation of the EST-SSR distribution along the eggplant genome was performed with Circos software. One EST-SSR per eggplant chromosome was selected in order to validate them in the *S. incanum* and *S. aethiopicum* sequenced genotypes.

All selected SSRs for validation were checked via IGV viewer and primers pairs were designed with Primers3. The amplification of SSRs were performed by touchdown PCR in a final volume of 12  $\mu\text{L}$ : 7.21  $\mu\text{L}$  water, 1.2  $\mu\text{L}$  1X PCR buffer, 0.6  $\mu\text{L}$   $\text{MgCl}_2$  50 mM, 0.24  $\mu\text{L}$  dNTPs 10 mM, 0.3  $\mu\text{L}$  reverse primer 10  $\mu\text{M}$ , 0.06  $\mu\text{L}$  forward primer with M13 tail 10  $\mu\text{M}$ , 0.24  $\mu\text{L}$  fluorochrome (FAM, VIC, NED and PET) 10  $\mu\text{M}$ , 0.15  $\mu\text{L}$  Taq DNA Polymerase (5U/ $\mu\text{L}$ ), 2  $\mu\text{L}$  DNA template 20 ng/ $\mu\text{L}$  under the following cycling conditions: denaturation at 95°C for 3 min, followed by 10 cycles of 30 s at 95°C, 30 s at 65°C (with each cycle the annealing temperature decreasing 1°C), and of 30 s at 72°C. Products were subsequently amplified for 20 cycles at 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, with a final extension at 72°C for 5 min.

PCR products were diluted in formamide and analyzed on an automated DNA sequencer ABI PRISM 3100-Avant with a GeneScan 600LIZ (Applied Biosystems, California, USA) size standard. The data were analyzed using the GeneScan software (Applied Biosystems) to obtain the electropherograms and polymorphisms were analyzed with Genotyper DNA Fragment Analysis software (Applied Biosystems, California, USA).

## ***Results and Discussion***

### **Illumina paired-end sequencing and EST assembly**

The new generation of high-throughput sequencing platforms and the improved algorithms for de novo transcriptome assembly has allowed the availability of transcriptomes even in non-model organisms without a reference genome (Davey and Blaxter, 2010). In our case, in order to build a transcriptome of *S. incanum* and *S. aethiopicum*, mixed RNA from young leaf, floral bud and young fruit, was used in order to increase the heterogeneity and diversity of the transcripts. Two different libraries were constructed, one per species, and subsequently sequenced in a HiSeq 2000 sequencer (Illumina).

A total of 105,625,594 and 114,162,500 raw reads were obtained from *S. incanum* and *S. aethiopicum* respectively (Table 1), which have been deposited in the NCBI Sequence Read Archive (Bioproject SRP063088). After the filtering and trimming process, removing adapters and low Phred quality sequences, 91,579,142 and 99,012,712 high-quality sequences were obtained for *S. incanum* and *S. aethiopicum*. The trimmed reads were assembled into transcriptomes using Trinity software (Grabherr et al., 2011), generating 108,322 transcripts for *S. incanum* and 106,660 for *S. aethiopicum* (Additional file 1). Subsequently, in order to test the overall assembly quality, the clean reads were mapped against the transcriptomes using Bwa (Li et al., 2013), which is an ultrafast and memory-efficient mapper particularly good at aligning reads between 50 to 100 bp. The large number of reads properly mapped, specifically 94.3% for *S. incanum* and 95.9% for *S. aethiopicum*, confirmed the high quality of Trinity assembly. The total length of assembled transcripts for *S.*

106

*incanum* and *S. aethiopicum* was about 102 and 92 Mbp with an average length of 946 and 868 bp, respectively. In recent years the assemblies have been improved progressively as a result of the advances in sequencing platforms, especially Illumina (Lee et al., 2014; van Dijk et al., 2014; Faure et al., 2015).

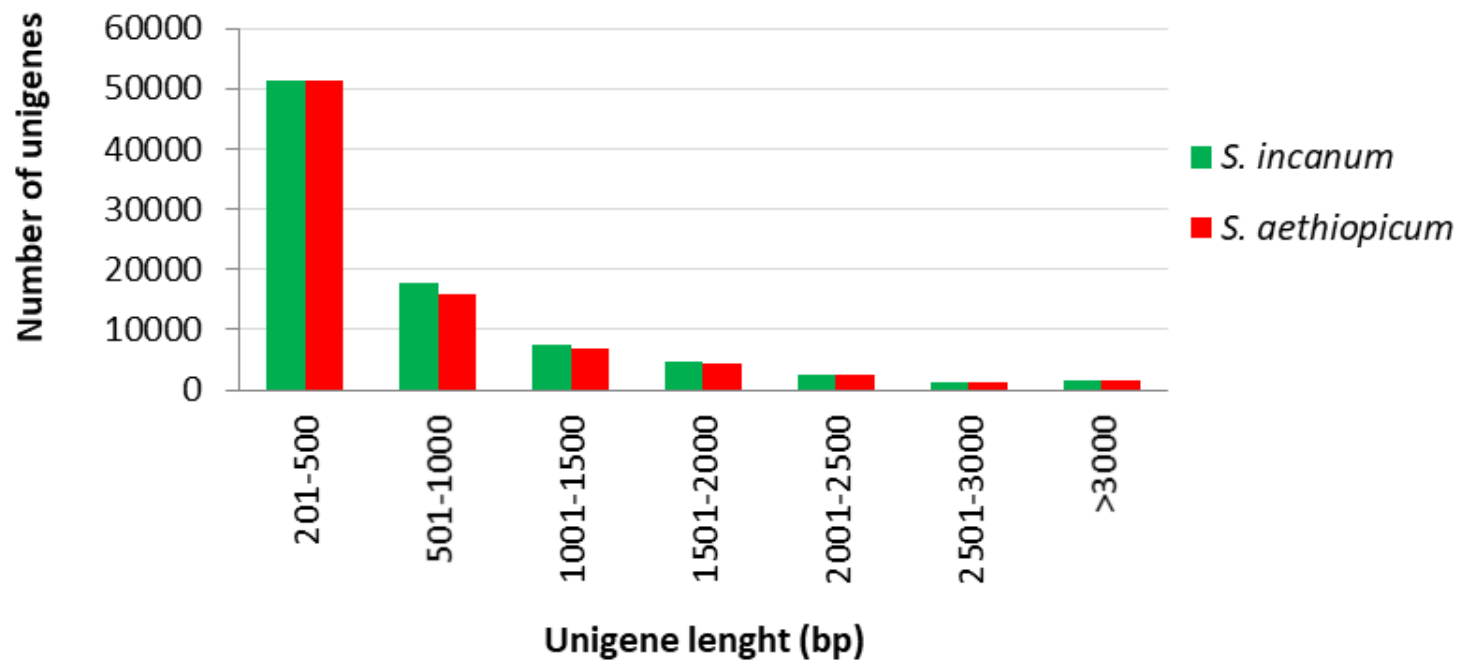
Trinity software determines splice variants (isoforms) and distinguishes transcripts from recently duplicated and identified allelic variants (Grabherr et al., 2011). To obtain a set of single-copy gene locus (unigene), only the most expressed transcript from the isoforms of each locus was chosen, using the RSEM software (Li and Dewey, 2011). A total of 83,905 unigenes were identified in *S. incanum* and 87,084 in *S. aethiopicum*, showing that 22.5% and 18.3% of transcripts were splice variants respectively (see Additional file 2).

Even though the 60% of unigenes had between 201 and 500 bp, more than 20% of them were longer than 1 kbp. The length distribution of unigenes is shown in Figure 2. In order to obtain the physical position of the unigenes, a BlastN against the *S. melongena* genome was performed and the results are illustrated in the Circos plot in Figure 3 and Additional file 3 (Krzywinski et al., 2009). The distribution of unigenes was uneven along the eggplant genome. Most unigenes clustered in areas which could correspond to the short and long arms of chromosomes, while a lower unigene density was observed in regions which could correspond to the centromere and pericentromeric regions, based on the observation of Doganlar et al. (2002). This gene distribution is similar to the observed in other species (Erayman et al., 2004; Mizuno et al., 2011). Our sets of unigenes are much higher than those obtained for *S. melongena* and *S. torvum*. *S. melongena* unigenes set consisted of 16,245 (Fukuoka et al.,

2010) and 34,174 unigenes (Yang et al., 2014), while in *Solanum torvum* 38,185 unigenes were obtained (Yang et al., 2014). A deeper sequencing and better coverage were probably the reasons of the greatest number of our *S. incanum* and *S. aethiopicum* datasets unigenes. In fact the number of our total residues was 58,447,674 for *S. incanum* and 62,899,378 for *S. aethiopicum*, higher than those obtained by Fukuoka et al. (2010) and Yang et al. (2014) in *S. melongena* (50,438,137 and 27,771,410, respectively), and Yang et al. (2014) in *S. torvum* (30,868,727). These higher numbers of unigenes has been observed in other plant transcriptomes recently published (Groves et al., 2015; Zhang J. et al., 2015; Zhang S. et al., 2015) in which the number of raw reads obtained was quite large.

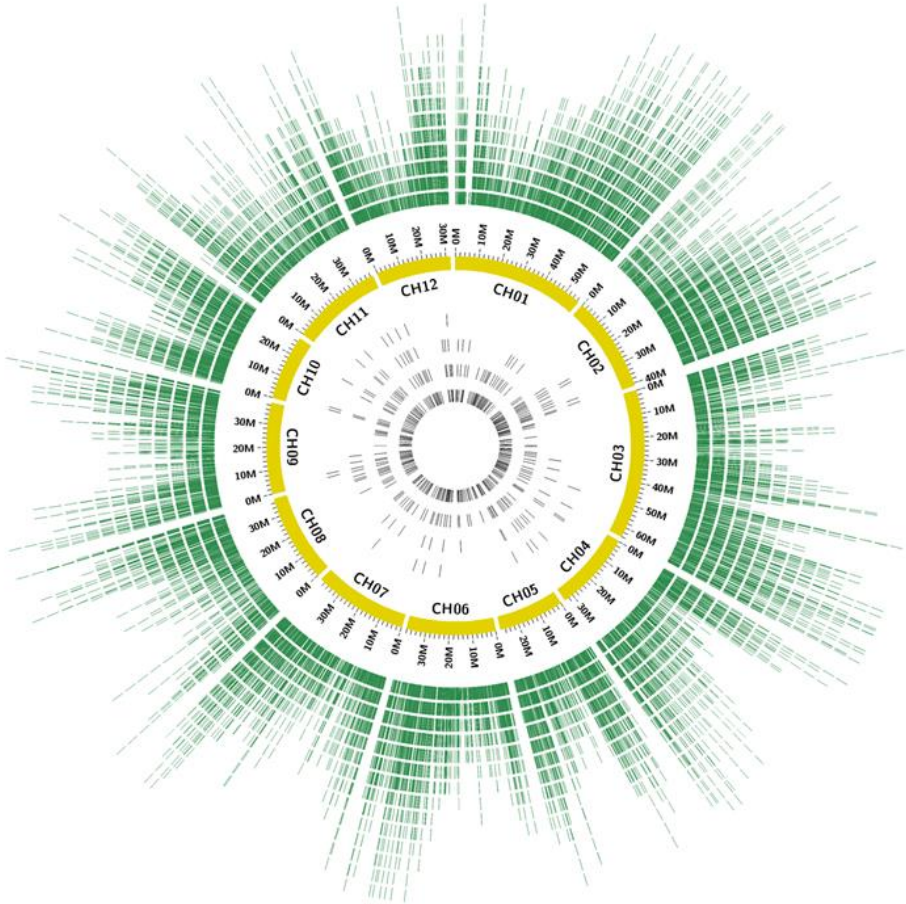
**Table 1.** Statistics of *S. incanum* and *S. aethiopicum* assembled transcripts and unigenes, using Trinity software. Data correspond to the results of RNA-Seq projects by HiSeq 2000 sequencer (Illumina) of total RNA from three tissues of *S. incanum* and *S. aethiopicum* and subsequent processing.

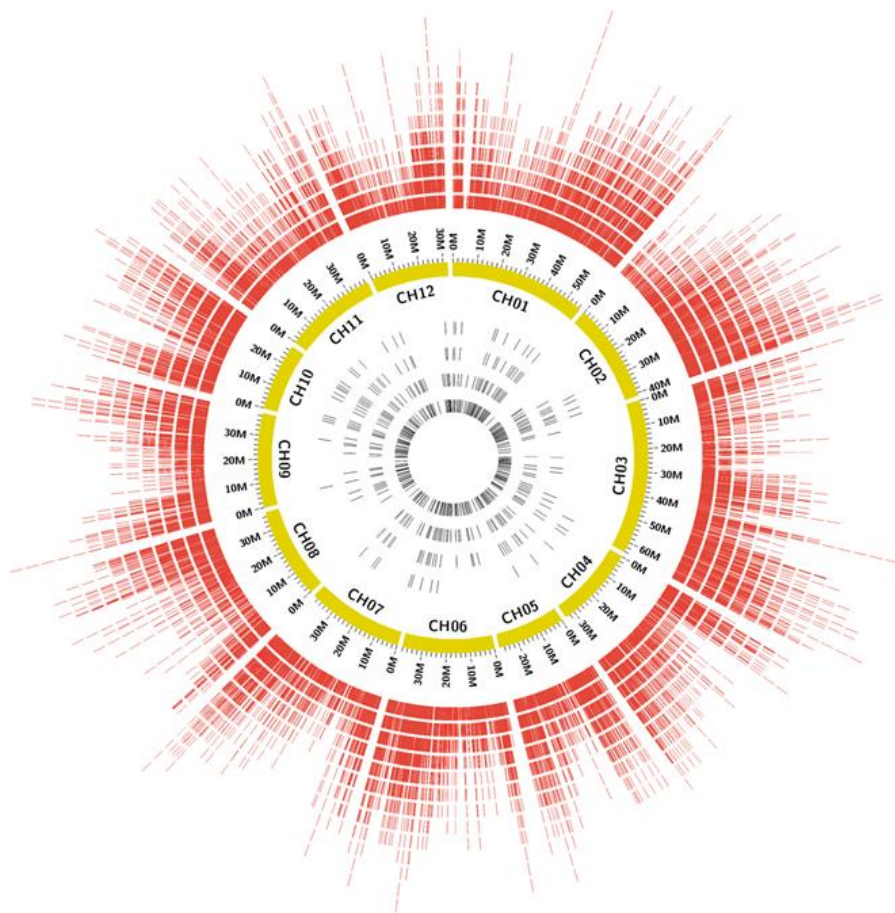
	<i>S. incanum</i>	<i>S. aethiopicum</i>
<b>Raw reads</b>	105,625,594	114,162,500
Sequence length	101	101
Mean sequence quality (Phred Score)	36.03	36.02
%CG	43	43
<b>High-quality reads</b>	91,579,142	99,012,712
Sequence length	70-101	70-101
Mean sequence quality (Phred Score)	36.88	36.87
%CG	42	42
<b>Transcript</b>	108,322	106,660
Max length	12,202	12,179
Average length	946	868
N50	1,693	1,455
Total residues	102,496,435	92,629,886
<b>Unigenes</b>	83,905	87,084
Max length	12,181	12,159
Average length	696	722
N50	1,153	1,139
Total residues	58,447,674	62,899,378



**Figure 2.** Length distribution (bp) of *S. incanum* and *S. aethiopicum* unigenes.







**Figure 3.** Distribution of *S. incanum* (previous page) and *S. aethiopicum* assembled unigenes and SSRs on *S. melongena* genome. In the outer histogram, depicted in green for *S. incanum* and in red for *S. aethiopicum*, are represented the assembled unigenes along the eggplant genome (yellow ring). Only unigenes which have given an E-value 0.0 via BlastN search were shown. In the inner histograms, in black, the distribution of the SSRs detected in each species is represented.

## Annotation of *S. incanum* and *S. aethiopicum* transcriptomes

Transcriptome annotation provides insight into the structural, functions and biological processes of assembled unigenes (Mutz et al., 2013). The functional annotation was performed using the assembled unigenes as query in BlastX searches against the three major protein databases and tomato protein database in this sequential order: Swiss-Prot ([http://web.expasy.org/docs/swiss-prot\\_guideline.html](http://web.expasy.org/docs/swiss-prot_guideline.html)), ITAG2.4 ([ftp://ftp.sgn.cornell.edu/tomato\\_genome/annotation/](ftp://ftp.sgn.cornell.edu/tomato_genome/annotation/)), Arabidopsis proteins (<http://www.arabidopsis.org/>) and Uniref90 (<http://www.ebi.ac.uk/uniprot/database/download.html>). Moreover a BlastX analysis was performed against the non-redundant (NR) protein database of NCBI and the GO terms and EC number have been assigned through the Blast2GO software (Conesa et al., 2008).

A total of 30,630 (36.5%) *S. incanum* and 34,231 (39.3%) *S. aethiopicum* have shown at least one hit in the protein databases. Most of the unigenes (57.5% for *S. incanum* and 57.3% for *S. aethiopicum*) were annotated using the manually reviewed Swiss-Prot database and ITAG 2.4 (34.9% and 33.2%) and less using Uniref90 (7.3% and 9.3%) and Arabidopsis protein database (0.3% and 0.2). The unigenes annotated under different protein database are reported in Additional file 4.

Even though the percentage of unigenes annotated in the protein databases seems to be quite low, the total number, 30,630 for *S. incanum* and 34,231 for *S. aethiopicum*, is consistent with the number of protein-coding genes described in tomato (34,727) (Tomato Genome Consortium, 2012) and in previous works in other plant species. For instance in pepper transcriptome (*Capsicum annum* L.) (Ashrafi et al., 2012) 24,003 out of

31,196 unigenes were annotated in protein databases as well as 32,410 out of 68,132 unigenes in *Oryza officinalis* Wall. ex Watt (Bao et al., 2015) and 34,368 out of 82,036 unigenes in litchi (*Litchi chinesis* Sonn.) (Lu et al., 2014). Similarly, in *S. melongena* and *S. torvum* 28,016 and 29,845 unigenes were annotated, respectively (Yang et al., 2014).

A large portion of hitless unigenes were short sequences between 200 and 500 bp. This huge set of non-annotated short sequences has been observed in recent published transcriptomes, in which large amount of raw reads have been obtained (Ranjan et al., 2014; Zhang J et al., 2015). Besides 3' or 5' untranslated regions and intron sequences from non-mature mRNAs, several authors have described that some of these sequences could be noncoding RNAs (ncRNAs) (Laurent et al., 2015). Unfortunately at the present the ncRNA is still in their early stages and just some mammalian entries were uploaded to IncRNA database (Laurent et al., 2015). Up to now, the only plant that has received some attention is *Arabidopsis thaliana* where 13,000 RNAs were found transcribed from intergenic regions (Liu et al., 2012, 2015a)

Gene ontology provides a systematic language in three key biological domains shared by all organisms: molecular function, biological process and cellular component to unify the representation of gene features across all species (Gousset et al., 2005). GO terms are structured as a graph and can be distributed in levels. Level 1 represents the most general categories and provides the most coverage, whereas higher levels provide more specific information and less coverage (Gousset et al., 2005). Briefly, level 1 is a general description of a process whereas higher levels provide a more specific description.

A total of 136,904 and 109,044 GO terms were assigned to 25,650 (30.5%) and 25,169 (28.9%) unigenes in *S. incanum* and *S. aethiopicum* respectively. The GO annotation results are presented in Additional file 5. The GO terms per unigenes ranged from 1 to 92 for *S. incanum* and from 1 to 55 for *S. aethiopicum*, although most of the unigenes have 1 to 10 GO terms (Figure 4). The unigenes were also annotated with EC number (Webb et al., 1992), which identifies the reactions they catalyze. EC numbers were assigned to 8,343 (9.9%) and 14,524 (16.6%) unigenes, varying from 1 to 9 per unigene, although almost 80% presented only one EC number (Figure 4).

The majority of GO terms (44.6% for *S. incanum* and 47.4% for *S. aethiopicum*) were related to biological processes (Figure 5). Most of them had a GO annotation level in the range of 4 to 10. Biological processes such as oxidation-reduction and metabolic processes, protein phosphorylation and regulation of transcription are usually specific of tissues in a developmental stage (Niederhuth et al., 2013). Molecular functions have been assigned to 30.7% and 35.4% of ontologies, most of them showing a GO annotation level of 3 to 9 and being the binding activities the most represented. The remaining 24.7% and 17.3% of annotated unigenes have shown a cellular component GO term, mostly related to nucleus, plasma membrane, cytosol, as well as chloroplast and mitochondria. The distribution of GO level for this category is quite uniform, with the exceptions of levels 5 and 8.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is an integrated database resource, which links genomic data with functional information to standardize gene annotation (Kanehisa and Goto, 1999). Using Blast2GO software the annotated unigenes were blasted against the

KEGG pathway database in order to dissect the molecular interaction among them. A total of 11,151 (13.2%) *S. incanum* unigenes were assigned to 146 KEGG biological pathways, involving 378 enzyme types and 879 EC numbers. Regarding *S. aethiopicum* 13,101 unigenes (15.0%) were ascribed to 147 KEGG pathways, including 356 different enzymes and 821 EC numbers. In *S. incanum* the three most enriched pathway were the biosynthesis of antibiotics, which included 713 unigenes (map01130), made mainly by phosphohexokinase (21 unigenes), dehydrogenase (18) and dehydrogenase (NAD<sup>+</sup>) (18) enzymes, Purine metabolism pathway (567 unigenes, map00230) was mostly represented by phosphatase (194 unigenes), adenylypyrophosphatase (130) and RNA polymerase (79) enzymes, and starch and sucrose metabolism pathway (516 unigenes, map map00500) was composed especially by pectin demethoxylase (85 unigenes), pectin depolymerase (64) and UDP synthase (54) enzymes. In *S. aethiopicum* the most enriched pathway was purine metabolism (map00230), which encompasses 1404 unigenes, mostly constituted by phosphatase (1160 unigenes), adenylypyrophosphatase (630) and RNA polymerase enzymes (92), followed by thiamine metabolism pathway (1172, map00730), made basically for phosphatase enzymes (1160 unigenes), and biosynthesis of antibiotics pathway (639 unigenes, map01130) including 168 enzyme types of which the most represented were dehydrogenase (20 unigenes), phosphohexokinase (18) and transaminase (16) enzymes. The biological pathways maps are reported in the Additional file 6 and the KEGG annotation results are compiled in the Additional file 7.

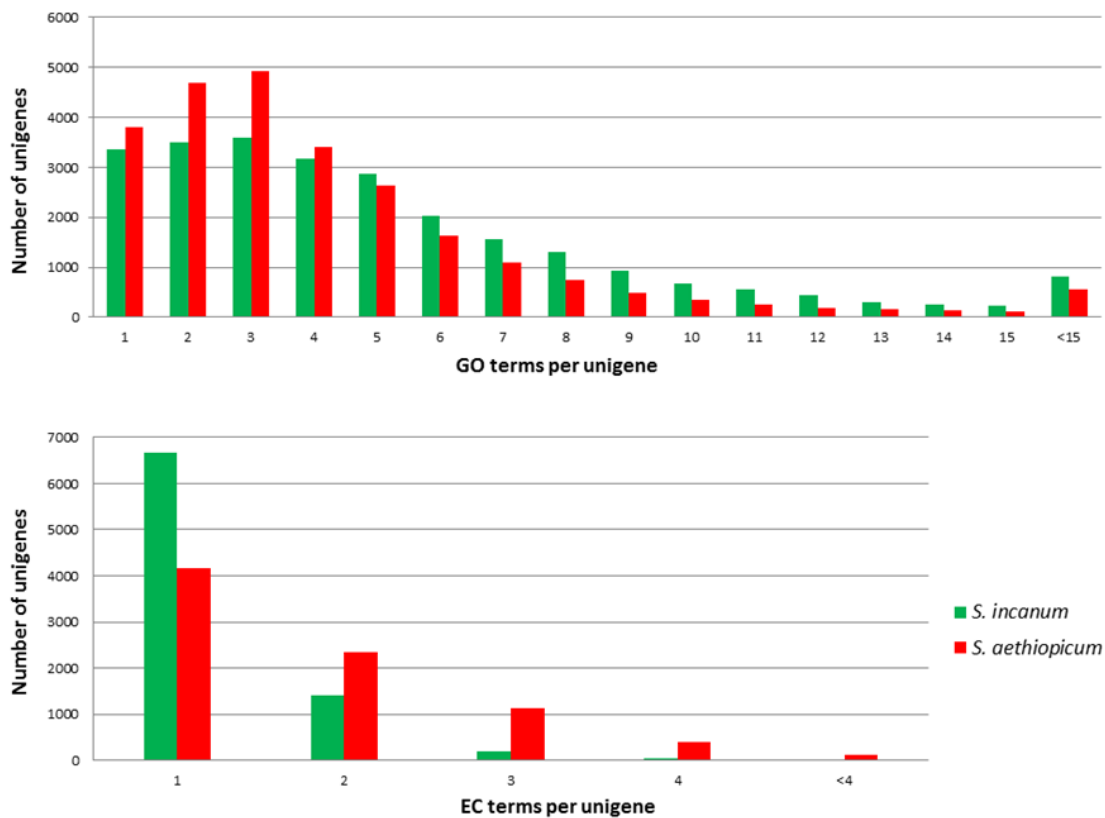
In order to establish a set of orthologs and gene model prediction, a best reciprocal Blast hits was performed with the tomato reference

genome (version SL2.50) (Tomato Genome Consortium; 2012). Up to now, the closest phylogenetically published genome of *S. incanum* and *S. aethiopicum* is the one of *S. melongena* (Yang et al., 2014). Even though this first version has provided valuable information for eggplant breeding programs, it is just a draft, requiring a deeper and most complete sequencing. On the other hand, the *S. lycopersicum* genome is the most complete and accurate in genus *Solanum*. In the last version, SL2.50, the tomato genome scaffolds were re-ordered and re-oriented, and the gap sizes between scaffolds were set using FISH and optical mapping. For that reason, apart from *S. melongena*, a BlastN was performed against *S. lycopersicum*. In the Additional file 8 the Blast hits results between the assembled unigenes and *S. lycopersicum* genome version SL2.50 are reported. A total of 16,388 (19.5%) and 17,630 (20.2%) unigenes have presented orthologs with tomato reference genome in *S. incanum* and *S. aethiopicum* respectively (Additional file 9), while between the two transcriptomes 46,498 orthologs were found.

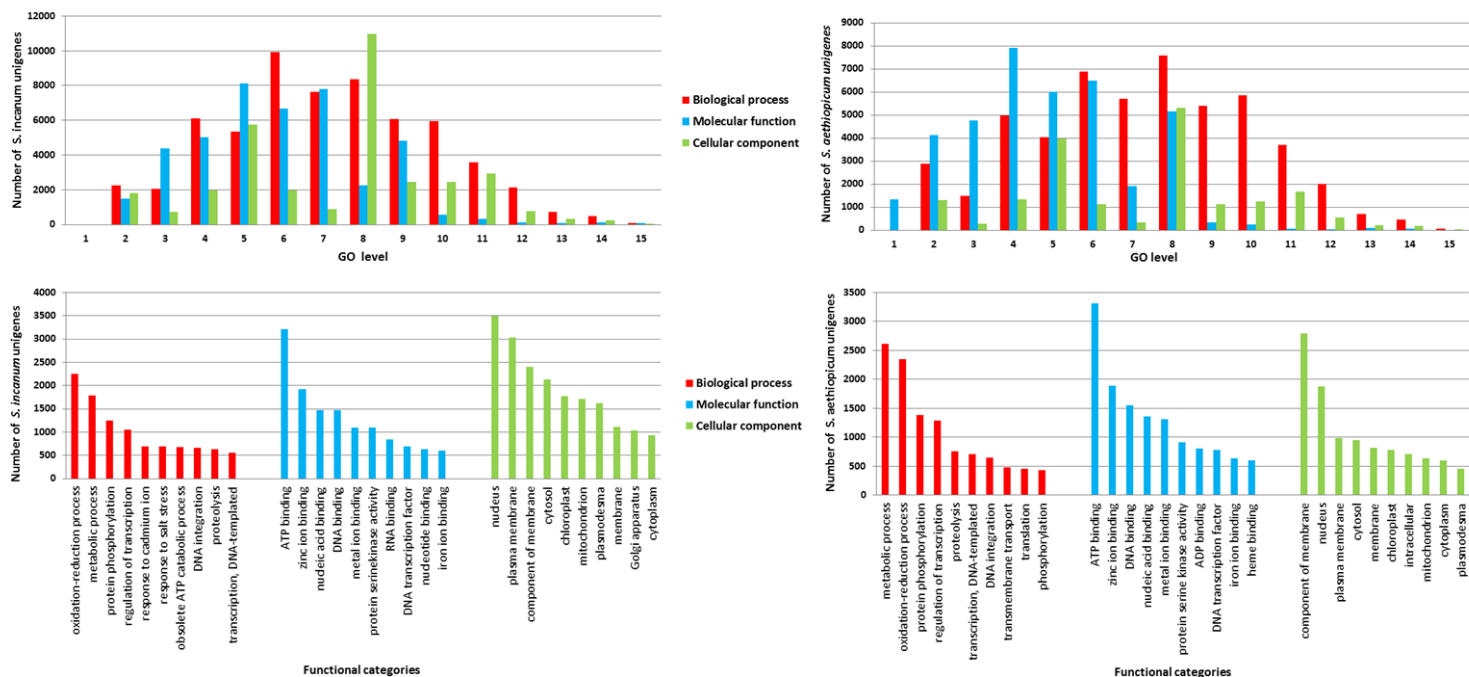
Regarding structural annotation, the ORF detection was performed using ESTScan software (Iseli et al., 1999), which predicted 35,943 ORFs (42.8% of the total unigenes residues) in *S. incanum*. The total number of nucleotides in *S. incanum* ORFs was 39,611,611 (67.7% of total *S. incanum* unigenes assembled residues). In *S. aethiopicum* 40,353 ORFs were predicted (46.3% of the total unigenes), which are constituted by 43,653,585 nucleotides (69.4% of total *S. aethiopicum* unigenes assembled residues). Furthermore using est2genome (Mott et al., 1997) the intron regions were detected, providing valuable information about gene structure as well as in the task of primers design, avoiding regions in the intron edges proximity. In 12,368 *S. incanum* unigenes (14.7% of the total

unigenes) 59,501 introns were predicted, while 65,996 introns were detected in 13,661 *S. aethiopicum* unigenes (15.6%). Both species have exhibited an average of 4.8 introns per unigene with a maximum of 49 for *S. incanum* and 56 for *S. aethiopicum*. The low percentage of introns detected in the unigenes is probably due to that the 60% of unigenes had between 201 and 500 bp and for their small size did not contain any intron.





**Figure 4.** Number of GO terms per unigene (top) and number of EC terms per unigene (bottom) distribution in *S. incanum* and *S. aethiopicum* transcriptomes.



**Figure 5.** GO level distribution (top) and functional categories annotation (bottom) of the *S. incanum* (left) and *S. aethiopicum* (right) unigenes. On the top are shown the distributions of GO level in the following functional categories: biological process, molecular function and cellular component. On the bottom are represented the descriptions of the most abundant gene products for each functional category of GO ontologies: biological process, molecular function and cellular component

## Molecular markers discovery and validation

### *Single nucleotide variations (SNVs)*

During the last decade Next-Generation Sequencing (NGS) techniques have allowed the development of large molecular marker collections with modest investments even in non-model species (Kumar et al., 2012; Zalapa et al., 2012). These collections enable the location of thousands of single polymorphisms along the genome as well as the development of high-density genetic maps, arrays and genotyping assays (Ramos et al., 2009; Robbins et al., 2011; Sim et al., 2012b).

Although genomic resources have been developed in eggplant (e.g., Barchi et al., [2011], Hirakawa et al., [2014], Yang et al., [2014]), few genomic information is available for closely related species. In the present study, large subsets of SNPs and INDELS have been identified to assist efficiently plant breeding projects and diversity studies. The SNP calling was performed for the species sequenced in the present study (*S. incanum* and *S. aethiopicum*) and for the two other species of the eggplant genepool (*S. melongena* and *S. torvum*), whose transcriptomes have already been sequenced (Yang et al., 2014). The reads of the four transcriptomes have been mapped against the eggplant genome and SNPs detected using Freebayes SNP caller (Garrison and Marth, 2012), as detailed in Materials and Methods section. The complete information of SNP calling is provided in Additional file 11. A set of 36 SNVs (Single nucleotide variations), three per eggplant chromosome, were selected for validation in the *S. incanum* and *S. aethiopicum* accesions using the HRM technique (Wittwer et al., 2003). Of these, a total of 96% were correctly

amplified and in all of them polymorphisms detected have been confirmed (see Additional file 12).

For each of the four species, the intraspecific and interspecific polymorphisms were identified by filtering the VCF file through the species ID number. In addition the INDELs were separated from the SNPs applying the VKS filter (it is not a SNP). The results of SNP calling are reported in Table 2. *S. aethiopicum* presented the highest value of intraspecific SNVs, with 159,571 SNPs and 4,556 INDELs. Many less intraspecific polymorphisms (12,396, of which 11,861 were SNPs and 535 INDELs) were identified in *S. incanum*. This suggests that the *S. aethiopicum* accession used presented a larger degree of heterozygosity than the *S. incanum* accession. Finally, in *S. melongena* and *S. torvum* (Yang et al., 2014) 2780 SNVs (2660 SNPs and 120 INDELs) and 25,147 SNVs (18,829 SNPs and 6,318 INDELs) intraspecific polymorphisms were discovered respectively. The high level of intraspecific variation in *S. aethiopicum*, *S. torvum* and *S. incanum* in comparison with *S. melongena* are probably due to the higher degree of autogamy and breeding selection for uniformity of the latter (Daunay et al., 2008).

The interspecific SNVs were detected in the comparisons between two species or three species at the same time. An interspecific SNV were selected when the species were homozygous for a specific allele, but different allele from one species to another (Table 2). The Circos plot in Figure 6 shows the six combinations of interspecific SNVs comparison and the location of the variations in the eggplant genome. As in the case of unigenes distribution along the *S. melongena* genome, the SNVs are less represented in areas which could correspond to centromere and more represented in regions which could correspond to the chromosome arms.

*Solanum torvum* presented the greater number of interspecific SNVs. A total of 1,648,949 polymorphisms (1,376,721 SNPs and 272,228 INDELs) have been detected in *S. torvum*, 588,764 of which are with *S. incanum*, 556,394 with *S. melongena* and 503,791 with *S. aethiopicum*. Out of the four species, *S. torvum* is the most phylogenetically distant (Vorontsova et al., 2013); this explains the large number of interspecific polymorphisms. *Solanum melongena* presented the second most abundant set of interspecific polymorphisms, 737,622 SNVs (639,234 SNPs and 98,388 INDELs). The comparison with *S. aethiopicum* detected 75,451 SNVs and 105,777 with *S. incanum*.

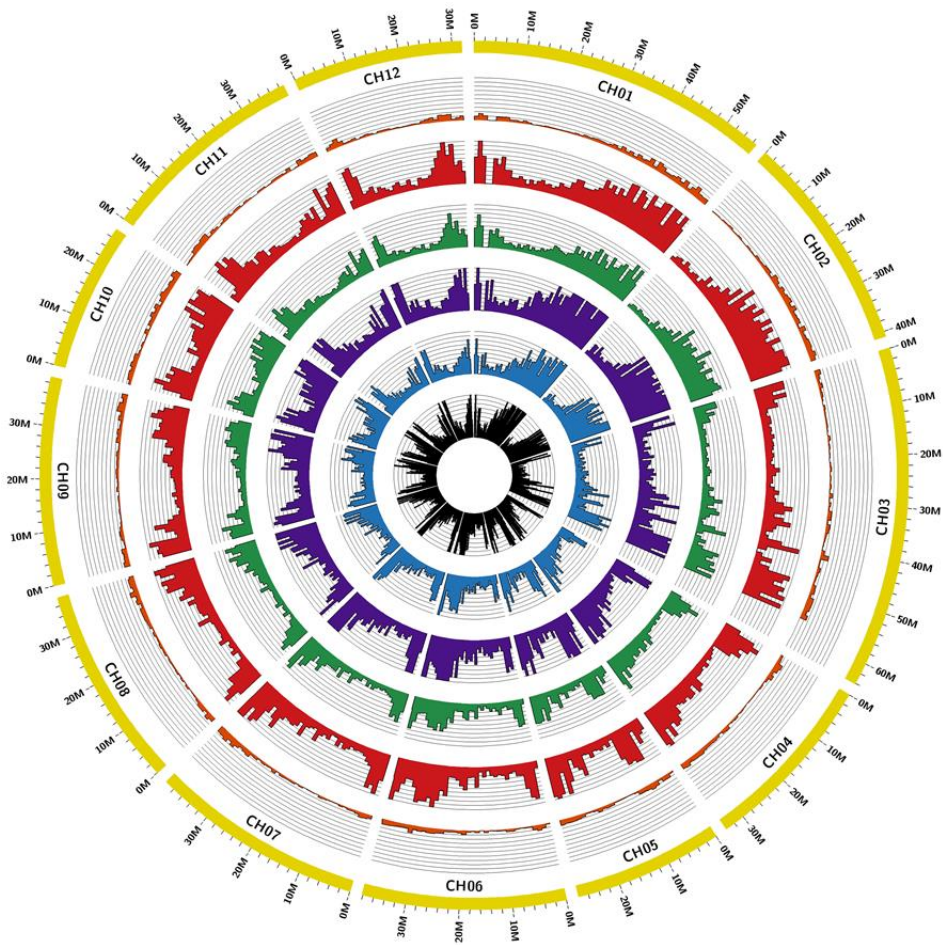
*S. melongena* and *S. incanum*, being members of the eggplant clade (Vorontsova et al., 2013) we would expect less polymorphisms between these two species than between *S. melongena* and *S. aethiopicum*, which belongs to the *anguivi* clade. This discrepancy could be explained by the high amount of *S. aethiopicum* intraspecific polymorphisms, which reduces the differences in the number of SNVs between *S. melongena* and *S. aethiopicum*. Finally the lowest variation was found between *S. incanum* and *S. aethiopicum* (15,162 SNVs). The interspecific SNVs were substantially less abundant when three species were compared. The majority of polymorphisms were INDELs rather than SNPs, due to the preferential bi-allelic nature of the latter (Amar et al., 2011; Pariasca-Tanaka et al., 2015). The collections of variations have been larger in the combinations which included *S. melongena* and *S. torvum*, and smaller with *S. incanum* and *S. aethiopicum*. In addition 50 INDELs were detected when the four species were compared all together.

Subsequently, all intraspecific and interspecific SNVs detected in the four species were filtered in order to create subsets of the most suitable

and effective variations for genotyping assays, both manually and with high throughput platforms, such as High Resolution Melting and GoldenGate Assay (Wittwer et al., 2003; Fan et al., 2006; Gupta et al., 2008). These variations have well-spaced positions, distanced more than 60 bp from another polymorphism and from the edges of the assembled transcripts. Furthermore, they could be detected by digestion with commonly used and cheap restriction enzymes.

**Table 2.** Single nucleotide variations statistics for the *S. incanum* and *S. aethiopicum* transcriptomes. SNVs (SNPs and INDELS) have been identified by the Freebayes SNP caller. The interspecific and intraspecific variations, as well as all species combination, have been determined by filtering ID (identification) species number. Every species and species comparison present an unambiguous identification number. The INDELS have been separated from SNPs applying the VKF filter to the total SNVs. The filtered SNPs and INDELS have been obtained by adding the CS60, HV0.05, CL60 and CEF filters.

<b>SNVs intraspecific variations</b>	<b>INDELS</b>	<b>SNPs</b>	<b>Total SNVs</b>
<i>S. incanum</i>	535	11,861	12,396
filtered	28	385	
<i>S. aethiopicum</i>	4,556	159,571	164,127
filtered	312	5,804	
<i>S. melongena</i>	120	2,660	2,780
filtered	6	57	
<i>S. torvum</i>	6,318	18,829	25,147
filtered	26	90	
<b>SNVs interspecific variations</b>	<b>INDELS</b>	<b>SNPs</b>	<b>Total SNVs</b>
<i>S. incanum</i> and <i>S. aethiopicum</i>	586	14,576	15,162
filtered	29	649	
<i>S. incanum</i> and <i>S. melongena</i>	3,673	102,104	105,777
filtered	253	4,184	
<i>S. incanum</i> and <i>S. torvum</i>	96,799	491,965	588,764
filtered	760	3,995	
<i>S. aethiopicum</i> and <i>S. melongena</i>	2,392	73,059	75,451
filtered	165	3,277	
<i>S. aethiopicum</i> and <i>S. torvum</i>	83,106	420,685	503,791
filtered	604	3,229	
<i>S. melongena</i> and <i>S. torvum</i>	92,323	464,071	556,394
filtered	722	3,682	
<i>S. incanum</i> , <i>S. aethiopicum</i> and <i>S. melongena</i>	108	41	149
filtered	8	2	
<i>S. incanum</i> , <i>S. aethiopicum</i> and <i>S. torvum</i>	908	198	1,106
filtered	10	3	
<i>S. incanum</i> , <i>S. melongena</i> and <i>S. torvum</i>	8,022	1,265	9,287
filtered	118	15	
<i>S. aethiopicum</i> , <i>S. melongena</i> and <i>S. torvum</i>	6,986	886	7,872
filtered	102	12	
<i>S. incanum</i> , <i>S. aethiopicum</i> , <i>S. melongena</i> and <i>S. torvum</i>	50	0	50
filtered	2	0	



**Figure 6.** Distribution of interspecific SNVs graphically depicted on eggplant genome. The histogram distributions, represented in different colors, illustrate the interspecific variations, two species at a time, along the eggplant genome (outer circle, yellow ring). From outer to inner are shown: *S. incanum* vs *S. aethiopicum* comparison (orange histogram), *S. incanum* vs *S. melongena* comparison (red histogram), *S. aethiopicum* vs *S. melongena* comparison (green histogram), *S. incanum* vs *S. torvum* comparison.



## SSRs

The transcriptomes were examined to discover SSRs (Expressed Sequence Tag-Simple Sequence Repeat) made of di-, tri-, tetra-nucleotide motifs by using the Sputnik software (Abajian et al., 1994). In *S. incanum*, a set of 976 SSRs were identified in 954 unigenes; i.e., 1.1% of the unigenes contained at least one microsatellite, while in *S. aethiopicum* 1,708 SSRs were detected in 1628 unigenes (1.8%). The total number of SSRs yielded was lower than obtained in other studies (Chen et al., 2015; Ding et al., 2015), probably due to the stringent criteria used to obtain high confident markers. The microsatellites identified are summarized in Table 3, while their location on eggplant genome is shown in Figure 3.

The range of SSRs length varied between 16 and 72 nucleotides in *S. incanum* and between 16 and 74 in *S. aethiopicum* with an average value of 21 and 24 nucleotides respectively. The most represented SSRs motifs in both species corresponded to AG, AAG and AAAG in agreement to the observation in other crops (Zeng et al., 2010; Gao et al., 2014; Pinosio et al., 2014). Trinucleotide repeat motifs were the most abundant (63.6% and 62.7%), followed by dinucleotide (26.6% and 28.4%) and tetranucleotide (9.8% and 8.9%) repeats. The prevalence of trinucleotide motifs is well documented in literature in eggplant (Stàgel et al., 2008; Barchi et al., 2011) as well as in other crops (Blanca et al., 2011; Schlautman et al., 2015). Metzgar et al. (2000) hypothesized that non-triplet SSRs show higher risks of frameshift mutation in coding regions and the selection against these mutations would reduce their chances of fixation. Otherwise the selection against frameshift events does not occur in SSRs with a repeat length divisible by three (tri- and hexanucleotide repeats). Depending on

their position in the gene, SSRs can be involved in different processes. The genic SSRs in 5'-UTR are implicated in gene transcription and gene translation while in 3'-UTR are implied in gene silencing and transcription slippage. In introns, SSRs can activate and inactivate genes (Li et al., 2014). The analysis of localization revealed that most of SSRs were located in ORFs, 33.5% for *S. incanum* and 32.7% for *S. aethiopicum*, and less in the UTRs (Table 4). In ORFs the trinucleotides repeats were the most abundant (88.7% for *S. incanum* and 87.7 for *S. aethiopicum*), ensuring the conservation of coding capacity and better protection against big changes in frameshift which might cause dramatic effects. On the other hand dinucleotides and tetranucleotides were more abundant in the UTRs, showing no great differences between the 5' and 3'. These results are consistent with previous study in other species (Gonzalez-Ibeas et al., 2007; Blanca et al., 2011).

One polymorphic EST-SSR was selected per chromosome to be validated in *S. incanum* and *S. aethiopicum* accessions used for the transcriptome sequencing. Eleven out of the 12 markers were correctly amplified and resulting polymorphic between the two species (Additional file 12). This indicates that the SSRs discovered in this study will provide a valuable set of molecular markers to disclose the intraspecific and interspecific variability across the eggplant genepool. The high rate of SNPs and SSRs correctly amplified is an indirect evaluation of good transcriptome assembly.

SSRs present some advantages in comparison with genomic DNA-based markers. For instance, their development from RNA-Seq projects is low-cost (Durand et al., 2010). SSRs are in many cases functional markers, when changes in allelic repeats affect functions and the phenotype.

Functional markers permit a direct allele selection, if there is a clear association for a target trait. Furthermore, the flanking sequences of SSRs are more evolutionary conserved than genomic SSRs (Zhou et al., 2014; Kumar et al., 2015), allowing considerable cross-species transferability. On the basis of its properties, the uses of EST-SSR are multiple, allowing genome and comparative mapping, genetic diversity analysis, QTL studies, gene tagging, association mapping and functional genomics (Varshney et al., 2005; Poczai et al., 2013).

**Table 3.** SSRs statistics corresponding to the *S. incanum* and *S. aethiopicum* transcriptomes. Di-, tri- and tetranucleotide repeats and motifs identified in the *S. incanum* and *S. aethiopicum* assembled unigenes are indicated.

SSRs repeat motifs	<i>S. incanum</i>			<i>S. aethiopicum</i>		
	SSRs	%	Unigenes	SSRs	%	Unigenes
<b>Dinucleotide</b>	260	26.6	258	362	29.0	342
AG/CT	160	61.5		187	51.7	
AT/TA	63	24.2		98	27.1	
AC/GT	36	13.8		76	20.9	
CG/GC	1	0.3		1	0.3	
<b>Trinucleotide</b>	621	63.6	609	776	62.2	755
AAG/CTT	169	27.2		263	33.8	
AAC/GTT	145	23.3		173	22.2	
AAT/ATT	82	13.2		106	13.7	
ATC/GAT	73	11.7		63	8.1	
AGG/CCT	53	8.5		54	7.0	
ACC/GGT	39	6.2		41	5.3	
AGC/GCT	34	5.4		40	5.2	
CCG/CGG	12	1.9		17	2.2	
ACT/AGT	10	1.6		10	1.3	
ACG/CGT	4	0.6		9	1.2	
<b>Tetranucleotide</b>	95	9.8	93	110	8.8	110
AAAG/CTTT	29	30.5		36	32.7	
AAAT/TTTA	23	24.2		28	25.5	
AAAC/GTTT	10	10.5		5	4.5	
ACAT/ATGT	5	5.2		9	8.2	
AAGG/CCTT	4	4.2		3	2.8	
ATCC/GGAT	4	4.2		1	0.9	
AACC/GGTT	3	3.1		4	3.6	
AATG/CATT	3	3.1		2	1.8	
AATT/AATT	3	3.1		6	5.5	
AGGG/CCCT	3	3.1		1	0.9	
Others motifs	8	8.4		15	13.6	
<b>Total</b>	976		954	1,248		1,270

**Table 4.** SSRs localization in the *S. incanum* and *S. aethiopicum* transcriptomes. The SSRs detected in the transcriptomes were checked for their position in the unigenes (ORF, 5'UTR and 3'UTR) according to the nucleotide repeats. When no precise position was established the SSRs were defined as "Other".

<i>S. incanum</i>	di-SSRs		tri-SSRs		tetra-SSRs		total	
	N°	%	N°	%	N°	%	N°	%
5'-UTR	63	42.2	66	44.3	20	13.5	149	15.2
ORF	25	7.6	291	88.7	11	3.7	327	33.5
3'-UTR	40	28.8	83	59.7	16	11.5	139	14.2
Other	132	36.5	181	50.1	48	13.4	361	37.1
Total	260	26.6	621	63.6	95	9.8	976	100

<i>S. aethiopicum</i>	di-SSRs		tri-SSRs		tetra-SSRs		total	
	N°	%	N°	%	N°	%	N°	%
5'-UTR	74	43.0	75	43.6	23	13.4	172	13.8
ORF	29	7.1	358	87.8	21	5.1	408	32.7
3'-UTR	73	38.0	106	52.3	13	6.7	192	15.3
Other	186	39.0	237	49.8	53	11.2	476	38.2
Total	362	29.0	776	62.2	110	8.8	1,248	100

## ***Conclusions***

In this study we present a de novo assembly and analysis of *S. incanum* and *S. aethiopicum* transcriptomes obtained by RNA-Seq. As a result of the annotation of these two common eggplant relatives, a broad overview of expressed genes was obtained. The annotation of the two transcriptomes has provided valuable information on function and structure of the assembled unigenes, which will allow the detection of candidate genes for important breeding traits in eggplant gene pool. The large amount of intraspecific and interspecific molecular markers, genic SSRs and SNVs (SNPs and INDELS), identified in our transcriptomes and in the *S. melongena* and *S. torvum* transcriptomes (Yang et al., 2014), will be extremely helpful for the breeding programs, although a deeper comparison between the four transcriptome it would be of great interest. In particular filtered SNVs will allow accurate genotyping assays through high throughput platforms or arrays increasing the efficiency and rapidity of the programs. Overall, the information produced in this study provides a valuable genomic resource in two non-model species, opening the door to further studies as gene tagging, comparative mapping, association mapping for enhancing eggplant genomics and breeding.

## **Acknowledgements**

The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the PAB (Andalusian Bioinformatics Platform) center located at the SCBI of the University of Malaga ([http:// www.scbi.uma.es/site/](http://www.scbi.uma.es/site/)). Pietro Gramazio is grateful to the Universitat Politècnica de València for a pre-doctoral (Programa FPI de la UPVSubprograma 1/2013 call) contract. This work has been partially funded by Spanish Ministerio de Economía y Competitividad and FEDER (grant AGL2015-64755-450 R).

### *Additional files*

**Additional file 1.** *Solanum incanum* and *S. aethiopicum* assembled transcripts (compressed folder). The file provides the fasta sequences of the 108,322 *S. incanum* and 106,660 *S. aethiopicum* transcripts.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM1\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM1_ESM.7z)

**Additional file 2.** List of *Solanum incanum* and *S. aethiopicum* unigenes (compressed folder). The file provides the list of the 83,905 most expressed single-copy *S. incanum* and 87,084 *S. aethiopicum* transcripts in fasta format.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM2\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM2_ESM.7z)

**Additional file 3.** *Solanum incanum* and *S. aethiopicum* unigene physical positions (compressed folder). Blast search results in order to obtain the *S. incanum* and *S. aethiopicum* unigen positions using *S. melongena* draft genome (cut-off e-value 1e-20).

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM3\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM3_ESM.7z)

**Additional file 4.** *S. incanum* and *S. aethiopicum* unigenes potentially encoding proteins (compressed folder). The file contains the unigene protein annotations performed using the three major public and the tomato protein databases (Swiss-Prot, ITAG 2.4, Arabidopsis, Uniref90).



[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM4\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM4_ESM.7z)

**Additional file 5.** GO terms and EC number unigene annotation (compressed folder). The file provides a list of GO terms and EC numbers assigned to *S. incanum* and *S. aethiopicum* unigenes.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM5\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM5_ESM.7z)

**Additional file 6.** Biological pathway maps (compressed folder). The file provides the biological pathway maps, obtained by KEGG searches.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM6\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM6_ESM.7z)

**Additional file 7.** Details of KEGG annotation (compressed folder). The files provide the names of biological pathways, number of unigenes involved, enzymes types and EC numbers of *S. incanum* and *S. aethiopicum* unigenes KEGG annotation.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM7\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM7_ESM.7z)

**Additional file 8.** Position in tomato reference genome (compressed folder). The file provides the physical position of *S. incanum* and *S. aethiopicum* unigenes in tomato through Blast against *S. lycopersicum* genome, version SL2.5.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM8\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM8_ESM.7z)

**Additional file 9.** *Solanum incanum* and *S. aethiopicum* orthologs (compressed folder). The file provides the lists of *S. incanum* and *S. aethiopicum* unigenes which presented orthologs in *S. lycopersicum* obtained through best reciprocal hits using BlastN.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM9\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM9_ESM.7z)

**Additional file 10.** ORFs and introns annotation (compressed folder). The GFF3 format file provides the ORFs and introns detected in *S. incanum* and *S. aethiopicum* unigenes.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM10\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM10_ESM.7z)

**Additional file 11.** SNP calling results. The file in VCF format provides the list of SNVs (SNPs and INDELs) identified in *S. incanum*, *S. aethiopicum*, *S. melongena* and *S. torvum* transcriptomes by mapping against *S. melongena* genome. For each SNV are indicated their position in eggplant genome scaffold, the allele of reference (eggplant genome allele) and the alternative allele (the transcriptomes allele), the quality of the SNV, the filters applied and detailed information about SNP calling process for each SNV and the different kind of transcriptome ID combinations in order to filter the intraspecific and interspecific polymorphisms. The different filter applied for in silico selection were: CS60 (the SNV is closer than 60 nucleotides to another SNV), HV0.05 (the region has more than 5.0 SVNs per 100 bases), CL60 (the SNV is

closer than 60 nucleotides to the reference edge), CEF (SNV is not a CAP detectable by the enzyme: cheap\_ones), VKF (it is not an SNP).

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM11\\_ESM.7z](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM11_ESM.7z)

**Additional file 12.** Validated markers. The file provides the tables of SNPs and SSRs experimentally validated. For each marker are provided the unigene which came from, the position in the correspondent eggplant scaffold, the detected alleles and the primers used for validation.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM12\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM12_ESM.xlsx)

**Additional file 13.** Filters settings. The file provide the settings of the filters that have been applied to VCF file (Variant Call Format) in order to filter the SNVs.

[https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864\\_2016\\_2631\\_MOESM13\\_ESM.txt](https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-016-2631-4/MediaObjects/12864_2016_2631_MOESM13_ESM.txt)



## **Chapter 2: Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants**

Gramazio P<sup>1</sup>, Prohens J<sup>1\*</sup>, Borràs D<sup>1</sup>, Plazas M<sup>2</sup>, Herraiz FJ<sup>1</sup>, Vilanova S<sup>2</sup>

<sup>1</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

<sup>2</sup> Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas - Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

\* Corresponding author: [jprohens@btc.upv.es](mailto:jprohens@btc.upv.es)

### **Citation:**

Gramazio P, Prohens J, Borràs D, Plazas M, Herraiz FJ, Vilanova S (2017). Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants. *Euphytica* 213: 264.

<https://doi.org/10.1007/s10681-017-2057-3>

## **Abstract**

Simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers are amongst the most common markers of choice for studies of diversity and relationships in horticultural species. We have used 11 SSR and 35 SNP markers derived from transcriptome sequencing projects to fingerprint 48 accessions of a collection of brinjal (*Solanum melongena*), gboma (*S. macrocarpon*) and scarlet (*S. aethiopicum*) eggplant complexes, which also include their respective wild relatives *S. incanum*, *S. dasyphyllum* and *S. anguivi*. All SSR and SNP markers were polymorphic and 34 and 36 different genetic fingerprints were obtained with SSRs and SNPs, respectively. When combining both markers all accessions but two had different genetic profiles. Although on average SSRs were more informative than SNPs, with a higher number of alleles, genotypes and polymorphic information content (PIC), and expected heterozygosity ( $H_e$ ) values, SNPs have proved highly informative in our materials. Low observed heterozygosity ( $H_o$ ) and high fixation index ( $f$ ) values confirm the high degree of homozygosity of eggplants. Genetic identities within groups of each complex were higher than with groups of other complexes, although differences in the ranks of genetic identity values among groups were observed between SSR and SNP markers. For low and intermediate values of pair-wise SNP genetic distances, a moderate correlation between SSR and SNP genetic distances was observed ( $r^2=0.592$ ), but for high SNP genetic distances the correlation was low ( $r^2=0.080$ ). The differences among markers resulted in different phenogram topologies, with a different eggplant complex being basal (gboma eggplant for SSRs and brinjal eggplant for SNPs) to the two others.

Overall the results reveal that both types of markers are complementary for eggplant fingerprinting and that interpretation of relationships among groups may be greatly affected by the type of marker used.

**Keywords:** eggplants, simple sequence repeats, single nucleotide polymorphisms, fingerprinting, genetic diversity, cluster analysis

**Abbreviations:**

SSRs: simple sequence repeats

NGS: new generation sequencing

SNPs: single nucleotide polymorphism

HRM: high resolution melting

*He*: expected heterozygosity

*Ho*: observed heterozygosity

f: fixation index

UPGMA: unweighted pair-group method using arithmetic averages

## ***Introduction***

Molecular markers based on DNA polymorphisms are of great utility for different applications in biological and agricultural sciences (Awise 2012; Grover and Sharma 2014). There is a wide array of available molecular markers (e.g., isozymes, RFLPs, RAPDs, ISSRs, AFLPs, SSRs, SNPs, etc.), which have different characteristics depending on their nature and the techniques employed to identify them. Among them, microsatellites or simple sequence repeats (SSRs) have been amongst the most used ones for germplasm management, selection and breeding (Kalia et al. 2011). SSRs are codominant, abundant, robust and highly polymorphic (Varshney et al. 2005). Although a few years ago their identification was relatively expensive and time-consuming through the development of genomic libraries, with the debut of NGS (New Generation Sequencing) platforms their isolation has become straightforward and cost-effective (Fernandez-Silva et al. 2013; De Barba et al. 2016; Zhan et al. 2016). In fact, thanks to the tremendous advances in sequencing of transcriptomes and genomes, hundreds or thousands of SSRs can be identified (Xiao et al. 2013; Goodwin et al. 2016). However, these large-scale sequencing projects also allow the identification of thousands to millions of molecular markers of single nucleotide polymorphism (SNPs) markers with a reasonable amount of resources (Van Tassel et al. 2008; Davey et al. 2011; Scheben et al. 2017). As occurs with the SSRs, SNPs are codominant, very reliable, ubiquitous and universal molecular markers, and although are generally less informative than SSRs (Yang et al. 2011; Filippi et al. 2015; Gonzaga, 2015), they are



much more abundant and easy to automate (Thomson et al. 2014; Kim et al. 2016).

SNPs are more uniformly spread across the genomes, while the SSRs tend to distribute more frequently in heterochromatic regions, although this strongly depends on the species (Li et al. 2002; Hong et al. 2007). In addition, the validation of SSRs is generally more time-consuming and expensive due to the need of detection through an agarose or polyacrylamide gels or capillary sequencing (Jones et al. 2007). In this respect, Yan et al. (2010) estimated that the resources and time needed to build a genetic map using a high-throughput SNP genotyping may be 75% cheaper and 100-fold faster than SSR gel-based methods.

Because of their different characteristics, SSRs and SNPs sample different levels of genetic diversity. In this respect, SSRs have a much higher mutation rate than SNPs (Hamblin et al. 2007; Coates et al. 2009; Fischer et al. 2017), because the mutations can be derived from a variation of the number of repeats, motif length or motif sequence (Ellegren, 2004; Kashi and King, 2006). In fact, the estimation of mutational rates of SSRs ( $1 \times 10^{-5}$ ) is several orders of magnitude higher than that of SNPs ( $1 \times 10^{-9}$ ) (Li et al. 1981; Kruglyak et al. 1998; Martínez-Arias et al. 2001). Although at the transcriptomic level the mutation rate of both SSRs and SNPs are lower than at the genomic level, the mutation rate of SSRs is also much higher than that of SNPs in the expressed sequences (Li et al. 2002).

Cultivated eggplants and their wild relatives are very variable (Vorontsova et al. 2013; Kaushik et al. 2016; Acquadro et al. 2017), with three different eggplant complexes (Daunay and Hazra, 2012; Plazas et al. 2014; Syfert et al. 2016), and may be an appropriate material for comparing SSR and SNP markers for fingerprinting, evaluation of genetic

identities and distances and for studying their relationships. Although the brinjal (or common) eggplant (*Solanum melongena* L.) is the most economically important, two other cultivated eggplant species exist, namely the gboma (*S. macrocarpon* L.) and the scarlet eggplants (*S. aethiopicum* L.), which are mostly grown in sub-Saharan Africa (Lester et al. 1990; Sunseri et al. 2010). *Solanum melongena* together with its ancestor (*S. insanum* L.) and other close relatives, like *S. incanum* L., are part of the so-called brinjal eggplant complex (Knapp et al. 2013). Similarly, the gboma eggplant together with its ancestor *S. dasyphyllum* Schumach. & Thonn. on one side, and the scarlet eggplant together with its ancestor *S. anguivi* Lam. on the other constitute, respectively, the gboma and scarlet eggplant complexes (Lester and Niakan, 1986; Bukenya and Carasco, 1994). Among the latter, the scarlet eggplant complex is more variable than the gboma eggplant and four cultivar groups (Aculeatum, Gilo, Kumba, and Shum) are considered within the cultivated *S. aethiopicum* (Lester and Daunay, 2003). In addition, intermediate forms between the wild *S. anguivi* and cultivated *S. aethiopicum* are also common (Plazas et al. 2014).

Although independently of the markers used the three eggplant complexes are often genetically differentiated clearly (Sakata and Lester, 1997; Furini and Wunder, 2004; Isshiki et al. 2008; Tumbilen et al. 2011; Acquadro et al., 2017), the genetic proximity and phylogenetic relationships among them frequently depend on the markers and plant materials used. In this respect, there is wide discrepancy in the studies over which of the three eggplant complexes is phylogenetically basal or genetically more distant to the two others (Sakata et al. 1991; Sakata and Lester, 1997; Furini and Wunder, 2004; Levin et al. 2006; Isshiki et al.

2008; Weese and Bohs, 2010; Tumbilen et al. 2011; Meyer et al. 2012; Vorontsova et al. 2013; Särkinen et al. 2013, Acquadro et al. 2017). Regarding the relationships within the hypervariable scarlet eggplant complex there have been few molecular studies evaluating them (Sunseri et al. 2010; Adeniji et al. 2013; Acquadro et al., 2017), but in general, they reveal that they present a low genetic differentiation.

Given the lack of studies comparing SSR and SNP markers in *Solanum* crops for genetic resources, breeding and phylogenetics, in this work we have genotyped a collection of accessions from the brinjal, gboma and scarlet eggplants with SSR and SNP markers obtained from the transcriptomes of *S. aethiopicum* and *S. incanum* (Gramazio et al. 2016). Although genomic markers tend to be more informative and accurate because are less prone to selection (Gadaleta et al., 2011), several studies with markers derived from transcriptome have shown that they are useful and reliable for phylogenetic and establishment of relationship among and within populations (Vogel et al., 2006; Castillo et al., 2008; Choudhary et al., 2009). Our objective is to confirm the potential utility of these markers as well as to compare them for fingerprinting, evaluation of genetic identities and distances, and for the establishment of relationships among these three groups.

## ***Material and methods***

### **Plant materials**

The collection of accessions used in this study encompasses accessions from different origins (African, Asian and European), as indicated in Table 1. All materials are maintained at COMAV germplasm bank (Universitat Politècnica de València, Spain). Thirty-nine accessions belonged to the scarlet eggplant complex and according to Lester and Niakan (1986) classification key and Plazas et al. (2014), morphological characterization, belong to *S. anguivi* (2), to *S. aethiopicum* groups Gilo (16), Kumba (7), Aculeatum (5), Shum (3), or to an intermediate group between *S. anguivi* and *S. aethiopicum* (8). Seven other accessions used in this study corresponded to the gboma complex, of which six are of *S. macrocarpon* and one of *S. dasyphyllum*. Finally, one accession of *S. incanum* and one *S. melongena* were used to represent the brinjal complex. The seeds were germinated following a protocol which is especially recommended for *Solanum* species that may present dormancy (Ranil et al. 2015).

**Table 1.** Plant materials used for molecular characterization with SNP and SSR markers of a collection of accessions of brinjal, gboma and scarlet eggplant complex, including the species and cultivar group and the country of origin.

Accession	Species/group	Origin
<i>Brinjal eggplant complex</i>		
MM577	<i>S. incanum</i>	Israel
AN-S-26	<i>S. melongena</i>	Spain
<i>Gboma eggplant complex</i>		
BBS117	<i>S. macrocarpon</i>	Ivory Coast
BBS168	<i>S. macrocarpon</i>	Ivory Coast
BBS171	<i>S. macrocarpon</i>	Ivory Coast
BBS178	<i>S. macrocarpon</i>	Ivory Coast
MM1153	<i>S. dasyphyllum</i>	Uganda
MM1558	<i>S. macrocarpon</i>	Malaysia
RNL0367	<i>S. macrocarpon</i>	Ghana
<i>Scarlet eggplant complex</i>		
AN05	<i>S. aethiopicum</i> group Gilo	Angola
AN39	<i>S. aethiopicum</i> group Gilo	Angola
AN67	<i>S. aethiopicum</i> group Gilo	Angola
BBS107	<i>S. aethiopicum</i> group Kumba	Ivory Coast
BBS110	<i>S. aethiopicum</i> group Kumba	Ivory Coast
BBS111	<i>S. aethiopicum</i> group Kumba	Ivory Coast
BBS114	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
BBS116	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
BBS119	<i>S. anguivi</i>	Ivory Coast
BBS125	<i>S. anguivi</i>	Ivory Coast
BBS131	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
BBS135	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS140	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS142	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS147	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS148	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
BBS151	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS159	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS170	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
BBS180	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast

BBS181	<i>S. aethiopicum</i> group Gilo	Ivory Coast
BBS184	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
BBS192	Intermediate <i>S. anguivi-S. aethiopicum</i>	Ivory Coast
INRA4	<i>S. aethiopicum</i> group Kumba	Senegal
IVIA026	<i>S. aethiopicum</i> group Gilo	Spain
MM457	<i>S. aethiopicum</i> group Aculeatum	Japan
MM585	<i>S. aethiopicum</i> group Kumba	Senegal
MM1207	<i>S. aethiopicum</i> group Kumba	Mali
MM1483	<i>S. aethiopicum</i> group Aculeatum	Unknown
PI413783	<i>S. aethiopicum</i> group Kumba	Burkina Faso
RAREGILO	<i>S. aethiopicum</i> group Gilo	Unknown
RNL0022	<i>S. aethiopicum</i> group Shum	Benin
RNL0187	<i>S. aethiopicum</i> group Aculeatum	Unknown
RNL0252	<i>S. aethiopicum</i> group Gilo	Ghana
RNL0288	<i>S. aethiopicum</i> group Gilo	Ghana
RNL0340	<i>S. aethiopicum</i> group Shum	Zambia
RNL0395	<i>S. aethiopicum</i> group Gilo	Liberia
UPV29014	<i>S. aethiopicum</i> group Gilo	Algeria
UPV29803	<i>S. aethiopicum</i> group Aculeatum	Argelia

---

## **DNA extraction**

Total genomic DNA was isolated from leaves of 3-4 true leaves stage plantlets, according to the CTAB protocol (Doyle and Doyle 1987) with slight modifications. The extracted DNA was dissolved in Milli-Q water and general quality was confirmed in agarose gel at 0.8%. After a concentration measurement using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA), the DNA was diluted at 30 ng/uL for PCR and High Resolution Melting (HRM) amplification.

## **SSR genotyping**

The identification of SSRs from the transcriptomes of one *S. aethiopicum* and one *S. incanum* accessions was reported in Gramazio et al. (2016) (Table 2). The selection of 11 highly reliable polymorphic SSRs was performed through filtering them for quality parameters and checking their coverage and length in the IGV visor (Thorvaldsdóttir et al. 2013). Primers pairs were designed using Primer3 tool (v. 0.4.0, <http://bioinfo.ut.ee/primer3-0.4.0/primer3/>).

The SSRs amplification was performed by PCR in a volume of 12  $\mu$ L including 7.21  $\mu$ L water, 1.2  $\mu$ L 1 $\times$  PCR buffer, 0.6  $\mu$ L MgCl<sub>2</sub> 50 mM, 0.24  $\mu$ L dNTPs 10 mM, 0.3  $\mu$ L reverse primer 10  $\mu$ M, 0.06  $\mu$ L forward primer with M13 tail 10  $\mu$ M, 0.24  $\mu$ L fluorochrome (FAM, VIC, NED and PET) 10  $\mu$ M, 0.15  $\mu$ L Taq DNA Polymerase (5U/ $\mu$ L), 2  $\mu$ L DNA template 20 ng/ $\mu$ L. The PCR program used was the following: 95 °C for 3 min for a denaturation, 30 cycles of 30 s at 95 °C followed by 30 s at 65 °C and of 30 s at 72 °C and finally 72 °C for 5 min for the last step of extension. The PCR products were subsequently diluted in formamide and sequenced by

capillary electrophoresis through an ABI PRISM 3100-Avant sequencer (Thermo Fisher Scientific, Waltham, USA) using a 600 LIZ GeneScan size standard (Thermo Fisher Scientific, Waltham, USA). The fragments were analyzed using the GeneScan software (Thermo Fisher Scientific, Waltham, USA) to obtain the electropherograms and polymorphisms were analyzed with Genotyper DNA Fragment Analysis software (Thermo Fisher Scientific, Waltham, USA).

### **SNP genotyping**

The identification of 35 SNPs and primer pairs design was similar to the one for SSRs as indicated above by using *S. aethiopicum* and *S. incanum* transcriptomes (Gramazio et al. 2016) (Table 2). Validation of SNPs was performed through Real-Time PCR in a LightCycler 480 (Roche, Basel, Switzerland). The reactions were performed in a 10  $\mu$ L volume comprising 5  $\mu$ L Master Mix 2X, 0.8  $\mu$ L MgCl<sub>2</sub> 25 mM, 0.25  $\mu$ L each primer, 1.7  $\mu$ L water and 2  $\mu$ L DNA 30 ng/ $\mu$ L with the following touchdown PCR program: denaturation at 95 °C for 10 min, followed by 55 cycles of 10 s at 95 °C, 15 s at 65 °C (decreasing 1 °C each cycle until 55 °C) and of 15 s at 72 °C, finally the melting step at 1 min at 95 °C, 1 min at 40 °C, 1 sec at 60 °C and rising the temperature at 0.02 °C/s until 95 °C.



**Table 2.** Characteristics of the SRR and SNP markers used, including the chromosome in which they are situated, the *S. incanum* and *S. aethiopicum* unigenes to which they correspond (Gramazio et al. 2016), the corresponding scaffold in the eggplant genome ((Hirakawa et al. 2014)), and the forward and reverse primers.

Marker	Chr.	<i>S. incanum</i> unigene	<i>S. aethiopicum</i> unigene	Scaffold eggplant genome	Forward primer	Reverse primer
<b>SSR markers</b>						
SSR_38227	1	SIUC38227_TC01	SAUC18592_TC02	Sme2.5_00192.1	TGCATAACCAACATTCAAACC	CCCTCCAAGCTCAAAGAAAG
SSR_37353	2	SIUC37353_TC01	SAUC65284_TC01	Sme2.5_01898.1	AGGGGTTTCGCTGAAACAAG	GTGTCACTCATCCCCTGTTG
SSR_37602	4	SIUC37602_TC01	SAUC23061_TC01	Sme2.5_01233.1	AAAAATGGAGGAATTAGAGAAGAAG	CAGCACCAGAAAACAACATGAG
SSR_31851	5	SIUC31851_TC01	SAUC57317_TC01	Sme2.5_00211.1	CATGGGAACCTTAGCCATGC	TTTTTGGTCTTTGGTTTTAGCTG
SSR_29668	6	SIUC29668_TC01	SAUC22524_TC01	Sme2.5_00163.1	GTGTTGAGCCAGTTGCAGAG	CCCCAAAACCAAGATCCAAAC
SSR_37966	7	SIUC37966_TC01	SAUC48208_TC01	Sme2.5_00502.1	GCAACTTCCTTCTCCATCTCC	TGAAACGGTGGCTTTAGCTC
SSR_21086	8	SIUC21086_TC01	SAUC64285_TC01	Sme2.5_00001.1	CGAGCCTGAAGAAGTTGTTG	AGACCCATCATCCAAATTCC
SSR_18542	9	SIUC18542_TC01	SAUC66830_TC02	Sme2.5_02687.1	CAAAATGGATGGGAGAGGAG	TTTCATGTGTGGGAGGGAAC
SSR_03112	10	SIUC03112_TC04	SAUC25971_TC01	Sme2.5_00173.1	GGTGGCATTGATCCAAGAAC	TCCCATTTTAGCAGCCTCAC
SSR_18317	11	SIUC18317_TC02	SAUC52997_TC01	Sme2.5_03669.1	CCTTTGGCAGTCACCATTTAG	TCATATGAGAAGCAACACTTTGG
SSR_37681	12	SIUC37681_TC01	SIUC37681_TC01	Sme2.5_00226.1	ACCGATGAAATCGACTCTGG	TCCATTCCTTCTGGACCATC
<b>SNP markers</b>						
SNP_38959	1	SIUC38959_TC01	SAUC82582_TC01	Sme2.5_03132.1	TGTGTGACTAGGACTTCATCCTC	GCCCTAGAAGGAGCTTTCATC
SNP_38905	1	SIUC38905_TC01	SAUC83481_TC01	Sme2.5_00529.1	TGAAGGAGAAGGACCAGCAG	TCAGCCCATATCAGATCTTGC
SNP_11564	2	SIUC11564_TC01	SAUC38669_TC01	Sme2.5_02714.1	AGGAGAATTGCAGAGTGATGC	TCGCAGCTCATAGCCATATTC
SNP_19191	2	SIUC19191_TC03	SAUC71621_TC01	Sme2.5_05438.1	AACCTCCCTAAAACCCCAAC	TGGCTCTGACAACCTGGAAATC

SNP_38971	2	SIUC38971_TC01	SAUC54450_TC02	Sme2.5_01610.1	GATGGTGGTTCTGCGGTATC	TAGGTTACCAGGCTCCATC
SNP_34715	3	SIUC34715_TC01	SAUC15119_TC01	Sme2.5_02159.1	CTAAGGGGCAGAGCTTCTTG	GACGCCAATAGTTAATAGAAGCTGC
SNP_00907	3	SIUC00907_TC04	SAUC70117_TC02	Sme2.5_01604.1	GGGAAAGAAAAGGAGGAATGG	AAATTTTGGATTTCATCATCTTC
SNP_27060	3	SIUC27060_TC05	SAUC62129_TC02	Sme2.5_04268.1	TGTTCTCACTCAATGTGTGCG	AGGTGCACCGATTCTTTCC
SNP_13379	3	SIUC13379_TC02	SAUC75690_TC01	Sme2.5_04555.1	CTTCTCTCCCACCAGGCTAC	TGGCAGCATACCAAATAGGC
SNP_23081	4	SIUC23081_TC01	SAUC07358_TC01	Sme2.5_06002.1	AAGTACCTCTGCAGCAACAGC	TCATCACCAAATCTCCATCG
SNP_39035	4	SIUC39035_TC01	SAUC20071_TC01	Sme2.5_03086.1	CCCGTTACTCAAGGGGATG	CACTGCCTTTCCAAATGAGG
SNP_23613	4	SIUC23613_TC02	SAUC38638_TC01	Sme2.5_00505.1	AAATCCAATTCACAGACATTGC	TGTTGATATCACCGACAACG
SNP_15567	4	SIUC15567_TC01	SAUC21364_TC01	Sme2.5_09958.1	TCAAATGAATGTGAGGAACAGG	TGGAAGAGGAAGAGGCTGAG
SNP_00676	5	SIUC00676_TC02	SAUC66850_TC01	Sme2.5_00697.1	CTCGGGGTCCAGAACTAGAA	CCTACTCCAGGGCTTCCTTC
SNP_19562	5	SIUC19562_TC02	SAUC02478_TC02	Sme2.5_09181.1	GCATCTCAATGTAAAAGCTTCC	GCCTTTGAGTCCGAGTTCAG
SNP_10686	6	SIUC10686_TC01	SAUC40320_TC01	Sme2.5_04479.1	GCACAATTAGCTGGTGTGG	AAGAGATTGTTGAAGAAAGACGTG
SNP_32044	6	SIUC32044_TC02	SAUC45004_TC02	Sme2.5_02660.1	GACCTAGGCAAGAACGAAGG	TGTAGGACGCTATCCCATTG
SNP_37940	6	SIUC37940_TC01	SAUC79062_TC01	Sme2.5_00001.1	CGGCTATGTACTTCATAACAGC	CCCAGAAATGATTTGCGAAG
SNP_30643	6	SIUC30643_TC05	SAUC05686_TC01	Sme2.5_00016.1	CACGGTCACTGCTTTCTCTG	AGATGGTGAGCCTTCCTACG
SNP_31222	7	SIUC31222_TC01	SAUC23026_TC01	Sme2.5_00673.1	CCTCCACCTACCCTCAACTC	AAGAAGCGCGAGTTGTTTCAG
SNP_01600	7	SIUC01600_TC01	SAUC78921_TC01	Sme2.5_02807.1	GGGAGGGTGGTAAAGGAGTG	GGTTTTCACTCAGCCGCTAC
SNP_23399	7	SIUC23399_TC02	SAUC11413_TC01	Sme2.5_00423.1	TAGAGATGGCCTCGGGAAG	GGAAGATAGATCAAAAACGAGCTG
SNP_17586	8	SIUC17586_TC01	SAUC46883_TC01	Sme2.5_00013.1	CTCCGGAATAAATGCAAACC	CCTGTCAATGGAGATGTTTCG
SNP_02438	8	SIUC02438_TC01	SAUC15038_TC01	Sme2.5_00858.1	GAGACAGGGGATGATGAAGG	GATGGCACATTGCACCTAAC
SNP_38436	8	SIUC38436_TC01	SAUC26635_TC02	Sme2.5_00391.1	TTGATGCAATAAAGGAAAGTGG	AGGCAGATGGGACACTCTTC
SNP_32294	9	SIUC32294_TC04	SAUC36933_TC01	Sme2.5_00104.1	TAGCAAAGCTTACGGCTGGTC	CAACTGAAGTGGCATGATGG
SNP_14499	9	SIUC14499_TC01	SAUC71552_TC02	Sme2.5_00504.1	CGGAACAAAAAGCTTTCAACC	ATGCTTCTTTGGGGCTAGAG

SNP_13910	9	SIUC13910_TC01	SAUC15875_TC01	Sme2.5_00488.1	CCCATAAGGTCGCGTAATTC	TCACCCGCAAACCTACTCTC
SNP_14015	10	SIUC14015_TC01	SAUC34299_TC01	Sme2.5_00190.1	TGCCCCATTTCTTCAACTTC	CAGCCATCTTCTCCTGGTAG
SNP_39817	10	SIUC39817_TC01	SAUC37197_TC01	Sme2.5_07454.1	TGTTGTGGACACGGCTACTC	CAAATGTTCTAGGCCCATTC
SNP_14306	11	SIUC14306_TC01	SAUC48392_TC01	Sme2.5_00442.1	TGGCATCAGCAGTCGTTG	CATGGGGAATTGAATTTTGG
SNP_39475	11	SIUC39475_TC01	SAUC08994_TC01	Sme2.5_09299.1	CACATTGGTGAAAGCCATTG	CTGGCTGCCTCTTGTGAG
SNP_29844	11	SIUC29844_TC04	SAUC64313_TC07	Sme2.5_00432.1	GCTCGCTTAGGATGAATTTCC	GCATATGGTGGAGGTGGTTC
SNP_14718	12	SIUC14718_TC01	SAUC85575_TC01	Sme2.5_00381.1	TCATGGGTTGCATTGTGAAC	TGCCGACGTAAAGGTCAATC
SNP_30456	12	SIUC30456_TC03	SAUC63329_TC01	Sme2.5_07259.1	ACTGGCCAAGCTTTTGCTAC	GTGTGGGCTCTAAGGGAATG

---

## Data analyses

The molecular marker analysis for SNPs and SSRs was performed using the software packages PowerMarker (Liu and Muse, 2005) and GenAlEx 6.5 (Peakall and Smouse, 2012). The following parameters were calculated using the PowerMarker package: number of alleles per locus, major allele frequency, number of genotypes, polymorphic information content (PIC) values calculated as  $PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j^2$ , where  $n$  is the total number of alleles detected,  $p_i$  the frequency of the  $i$ th allele, and  $p_j$  the frequency of the  $j$ th allele) (Botstein et al. 1980), expected heterozygosity ( $H_e$ ), calculated as  $H_e = 1 - \sum_{i=1}^n p_i^2$  (where  $p_i$  is the frequency of the  $i$ th allele) (Nei, 1972), observed heterozygosity ( $H_o$ ), calculated as the number of heterozygous alleles/number of alleles and fixation index ( $f$ ), calculated as  $f = 1 - (H_o/H_e)$  (Wright, 1965).

The consensus tree was calculated from the genetic similarity to illustrate the level of relatedness between the accessions using the UPGMA method (unweighted pair-group method using arithmetic averages) (Sneath and Sokal, 1973) and it was reconstructed using the software TreeView (Page, 2001). Branch support on the phenogram was tested by bootstrap analysis with 1000 replications using the PHYLIP version 3.67 software (Felsenstein, 2007). Bootstrap values of 50% or higher were used to indicate support for the phenogram topology at a node (Highton, 1993). The genetic distance matrix (Nei 1972) among the different accessions was calculated with the GenAlEx 6.5 software package for both for SSRs and SNPs. The correlation between pair-wise

genetic distances calculated for both markers was investigated by the Mantel test (Mantel, 1967) of matrix correspondence.

## ***Results and discussion***

### **Validation of the SSR and SNP markers for fingerprinting**

All SSR and SNP markers tested amplified and were polymorphic in the collection evaluated, although in a few cases there were missing data. In other studies in which eggplant SSR markers were developed de novo, the levels of polymorphism were not as high. For example, Vilanova et al. (2012) found an 85.5% of polymorphism in SSRs derived from a genomic library of *S. melongena*. A lower level of polymorphism (56.7%) was detected in the genetic SSRs derived from a library developed by Nunome et al. (2009), where 598 out of 1,054 markers were polymorphic. When comparing the SSR and SNP profile of the two accessions used for obtaining the transcriptomes of *S. aethiopicum* (accession BBS135) and *S. incanum* (accession MM577) (Gramazio et al. 2016), only one SNP marker (SNP\_14499) was found to be monomorphic among the two accessions. This confirms the high quality of the transcriptome sequences and *in silico* analysis performed by Gramazio et al. (2016) and reveals that both SSR and SNP markers obtained from *in silico* analyses of transcriptomes in materials of cultivated and wild eggplants are reliable and transferable to related species for being used in genotyping and fingerprinting. Other studies confirmed the reliability of the molecular markers discovered in silico from transcriptomes. For example, a set of SNPs identified in silico from a pepper (*Capsicum annuum* L.) transcriptome were validated in 43 pepper lines and accessions resulting

in a rate of 89.9% polymorphic markers (Ashrafi et al. 2012). Also 86.7% of a subset of SSRs identified in silico in a zucchini (*Cucurbita pepo* L.) transcriptome resulted polymorphic in a set of ten accessions of genus *Cucurbita*, nine of which were representative of the diversity within *C. pepo* and one accession *C. moschata* accession (Blanca et al. 2011). Up to now, few transcriptomes have been sequenced in genus *Solanum* and in just a few of them molecular markers have been identified and primers pairs designed in silico for a subsequent validation in a wide range of related materials. An example is the *S. dulcamara* L. transcriptome where the SSR identified in silico were validated in seven plants and all of them resulted polymorphic (D'Agostino et al. 2013).

When considering the 11 SSRs, a total of 34 different genetic profiles were found among the 48 accessions evaluated, while for the 35 SNPs, the number of different genetic profiles was 36 (Table 3). This confirms the highest discrimination potential of SSRs compared to SNPs for genetic fingerprinting (Hamblin et al. 2007; Varshney et al. 2007; Yang et al. 2011), as the number of different profiles obtained is almost the same with 11 SSRs or 35 SNPs. In this respect, several authors considered that in order to obtain a similar genetic power to discriminate individuals from different populations the number of SNPs required might be 8 to 15 times the number of SSRs (Hess et al. 2011; Yu et al. 2009).

When considering a combination of both SNP and SSR markers 47 unique genetic fingerprints were obtained for the 48 accessions, and only two of them (both from *S. aethiopicum* Gilo group) shared the same genetic profile for the markers that amplified in both of them. Although in general (Hu et al. 2011; Nandha and Singh, 2014; Thiel et al. 2003) and in the particular case of eggplant (Muñoz-Falcón et al. 2011) genomic SSRs

and SNPs are frequently more polymorphic than transcriptome-derived SSRs and SNP, in our case we have found a considerable level of polymorphism in the markers we tested. Particularly, the combination of both SSR and SNP markers has been highly efficient for genetic fingerprinting. Probably the fact that both types of markers sample different levels of genomic diversity (van Inghelandt et al. 2010) increases the efficiency of fingerprinting when combining both types of markers.

In all cases, accessions sharing a single SSR or SNP profile corresponded to the same eggplant complex (brinjal, gboma, or scarlet eggplant complexes) (Table 3), indicating that both types of markers provide consistent results. The number of accessions having a single genetic profile ranged between two and six for SSRs, between two and seven for SNPs, and only two when combining both SSR and SNP markers. When considering only SSR or SNP markers, accessions from different groups of the scarlet eggplant complex (i.e., *S. anguivi*, the Intermediate *S. anguivi-S. aethiopicum* group, and the four groups of *S. aethiopicum*) often shared a same profile. Our data are in agreement with those of other authors who used AFLPs and SSRs (Sunseri et al. 2010), RAPDs (Aguoru et al. 2015), and SNPs (Acquadro et al., 2017) and also found that accessions of *S. aethiopicum* did not cluster according to the cultivar group. This is an additional indication that these groups, which are distinguished on the basis of morphology (Lester, 1986; Lester and Daunay, 2003; Plazas et al. 2014), are not genetically differentiated.

**Table 3.** Accessions having a shared genetic profile with 11 SSR or 35 SNP markers, or using all of them (11 SNP plus 35 SNP markers).

Shared genetic profiles	Accessions	Species/group
<i>Using 11 SSRs</i>		
GP_SSR1	BBS 168, RNL0367	<i>S. macrocarpon</i>
GP_SSR2	BBS119	<i>S. anguivi</i>
	BBS116	Intermediate <i>S.anguivi-S-aethiopicum</i>
	BBS111	<i>S. aethiopicum</i> gr. Kumba
GP_SSR3	MM1483, RNL0187	<i>S. aethiopicum</i> gr. Aculeatum
GP_SSR4	UPV29803	<i>S. aethiopicum</i> gr. Aculeatum
	BBS148, BBS192	Intermediate <i>S.anguivi-S-aethiopicum</i>
	BBS159, RNL0252	<i>S. aethiopicum</i> gr. Gilo
	BBS110	<i>S. aethiopicum</i> gr. Kumba
GP_SSR5	BBS151, BBS181, RNL0288, RNL0395	<i>S. aethiopicum</i> gr. Gilo
	MM585	<i>S. aethiopicum</i> gr. Kumba
GP_SSR6	BBS107, MM1207	<i>S. aethiopicum</i> gr. Kumba
<i>Using 35 SNPs</i>		
GP_SNP1	BBS117, MM1558	<i>S. macrocarpon</i>
GP_SNP2	BBS180	Intermediate <i>S.anguivi-S-aethiopicum</i>
	BBS140, BBS159, RNL0252	<i>S. aethiopicum</i> gr. Gilo
	MM1483	<i>S. aethiopicum</i> gr. Aculeatum
GP_SNP3	RAREGILO	<i>S. aethiopicum</i> gr. Gilo
	MM1207	<i>S. aethiopicum</i> gr. Kumba
	AN39, IVIA026, UPV29014	<i>S. aethiopicum</i> gr. Gilo
GP_SNP4	BBS111, MM585, PI413783	<i>S. aethiopicum</i> gr. Kumba
	RNL0022	<i>S. aethiopicum</i> gr. Shum
<i>Using 11 SSRs plus 35 SNPs</i>		
GP_SSRSNP1	BBS159, RNL0252	<i>S. aethiopicum</i> gr. Gilo



## SSR and SNP diversity statistics

In general, the diversity statistics for the SSRs had higher average values than those of the SNPs (Tables 4 and 5). This is common because due to its nature, the potential variability of SSRs is larger than that of SNPs (Ellegren, 2004; Kashi and King, 2006; Fischer et al. 2017). In our study, the number of alleles obtained with SSRs ranged between three and seven, with an average value of 5.00 (Table 4), while for SNPs it ranged between two and four with an average value of 2.34 (Table 5). Also, the major allele frequency was generally lower with SSRs, ranging from 0.531 to 0.857, with an average value of 0.689, while for SNPs the values were considerably higher, with a range from 0.510 to 0.989, and an average value of 0.851. These values reveal that, despite a having a great diversity, there is always a major allele with a frequency over 50%, both for SSRs and SNPs. The number of genotypes, as occurs for the alleles, was also higher for SSRs, with a range from four to eight and an average of 5.91 (Table 4), than for SNPs, which presented a range from two to five, with an average of 3.14 (Table 5). As a consequence of the larger number of alleles and lower frequency of the major allele, the PIC values were generally higher for SSRs, with a range from 0.249 to 0.636 and an average of 0.419, than for SNPs, which ranged from 0.021 to 0.386 and an average value of 0.191. Our results are consistent with previous works in eggplants, which find similar values to ours. For example, Vilanova et al. (2012), using genomic SSRs in a collection of 20 accessions of *S. melongena* plus one accession of *S. aethiopicum* and *S. macrocarpon* found similar levels of number of alleles (4.72) and PIC (0.47) than us. Otherwise, other authors found lower values of average PIC rate and number of alleles per

locus. For example, Stàgel et al. (2008) found a mean PIC rate of 0.38 and number of alleles per locus of 3.1 when assessed 11 EST-SSRs in 38 *S. melongena* accessions. Also, Nunome et al. (2009) found lower values for the diversity statistics when they genotyped eight lines of *S. melongena* using 1,054 genomic SSRs (mean PIC value=0.27, mean number of alleles=2.2) and 66 EST-SSRs (mean PIC value= 0.13, mean number of alleles=1.4).

The expected heterozygosity ( $H_e$ ) was, on average, higher for SSRs (0.456) than for SNPs (0.224), and much higher than the observed heterozygosity ( $H_o$ ), which had average values of 0.039 for SSRs and 0.025 for SNPs. The much higher values for  $H_e$  compared to  $H_o$  result in high levels for the fixation index ( $f$ ), with average values of 0.907 for SSRs and 0.837 for SNPs. In fact, for some SSR and SNP markers, the  $f$  value was 1, and so all materials were homozygous for these loci. Our values are similar to those obtained by others with SSRs and SNPs in common eggplant collections (Muñoz-Falcón et al. 2009; Vilanova et al. 2012, 2014, Ge et al. 2013; Augustinos et al. 2016). Although it is known that cultivated brinjal eggplant is fundamentally autogamous (Arumuganathan and Earle 1991; Pessarakli and Dris 2004; Daunay and Hazra 2012) our results also provide evidence that gboma and scarlet eggplants present a reproductive system similar to that of brinjal eggplant.

**Table 4.** Genetic diversity statistics, including major allele frequency, number of genotypes, number of alleles, expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), polymorphic information content (PIC), and coefficient of inbreeding ( $f$ ) for the 11 SSR markers evaluated in a collection of brinjal, gboma and scarlet eggplants accessions.

Marker	Number of alleles	Major allele frequency	Number of genotypes	PIC	$H_e$	$H_o$	$f$
SSR_1	4	0.786	4	0.319	0.354	0.020	0.944
SSR_2	6	0.490	6	0.636	0.679	0.000	1.000
SSR_4	5	0.765	7	0.338	0.379	0.122	0.683
SSR_5	3	0.724	4	0.366	0.422	0.020	0.953
SSR_6	4	0.796	4	0.304	0.339	0.041	0.882
SSR_7	5	0.531	8	0.570	0.625	0.102	0.840
SSR_8	5	0.439	6	0.615	0.672	0.020	0.970
SSR_9	7	0.602	8	0.549	0.587	0.020	0.966
SSR_10	5	0.776	6	0.342	0.373	0.041	0.893
SSR_11	5	0.857	6	0.249	0.259	0.041	0.845
SSR_12	6	0.816	6	0.315	0.326	0.000	1.000
Mean±SE	5.00±0.33	0.689±0.044	5.91±0.44	0.419±0.043	0.456±0.046	0.039±0.012	0.907±0.028

**Table 5.** Genetic diversity statistics, including number of alleles, major allele frequency, number of genotypes, polymorphic information content (PIC), expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), and coefficient of inbreeding (f) for the 35 SNP markers evaluated in a collection of brinjal, gboma and scarlet eggplants accessions.

Marker	Number of alleles	Major allele frequency	Number of genotypes	PIC	$H_e$	$H_o$	f
SNP_38959	2	0.926	3	0.128	0.138	0.021	0.849
SNP_38905	2	0.714	3	0.325	0.408	0.041	0.902
SNP_11564	3	0.723	4	0.386	0.433	0.043	0.904
SNP_19191	3	0.929	4	0.130	0.135	0.020	0.851
SNP_38971	2	0.949	3	0.092	0.097	0.020	0.793
SNP_34715	3	0.959	3	0.078	0.079	0.000	1.000
SNP_00907	3	0.911	3	0.157	0.165	0.000	1.000
SNP_27060	2	0.739	3	0.311	0.386	0.043	0.890
SNP_13379	2	0.949	3	0.092	0.097	0.020	0.793
SNP_23081	2	0.531	3	0.374	0.498	0.041	0.920
SNP_39035	3	0.724	4	0.373	0.425	0.020	0.953
SNP_23613	2	0.854	3	0.218	0.249	0.042	0.836
SNP_15567	2	0.698	3	0.333	0.422	0.021	0.952
SNP_00676	2	0.696	3	0.334	0.423	0.043	0.899
SNP_19562	2	0.724	3	0.320	0.399	0.020	0.950
SNP_10686	4	0.888	5	0.198	0.206	0.061	0.709
SNP_32044	2	0.745	3	0.308	0.380	0.020	0.947
SNP_37940	2	0.745	3	0.308	0.380	0.020	0.947
SNP_30643	2	0.967	3	0.061	0.063	0.022	0.662

SNP_31222	3	0.918	3	0.148	0.153	0.000	1.000
SNP_01600	3	0.750	4	0.325	0.385	0.042	0.894
SNP_23399	2	0.980	2	0.039	0.040	0.000	1.000
SNP_17586	2	0.949	3	0.092	0.097	0.020	0.793
SNP_02438	2	0.969	3	0.058	0.059	0.020	0.662
SNP_38436	3	0.929	4	0.131	0.135	0.020	0.852
SNP_32294	2	0.946	3	0.098	0.103	0.022	0.793
SNP_14499	2	0.989	2	0.021	0.021	0.021	0.000
SNP_14015	2	0.792	3	0.275	0.330	0.042	0.876
SNP_39817	2	0.949	3	0.092	0.097	0.020	0.793
SNP_14306	3	0.878	4	0.202	0.219	0.082	0.634
SNP_39475	2	0.980	2	0.039	0.040	0.000	1.000
SNP_29844	2	0.948	3	0.094	0.099	0.021	0.793
SNP_14718	3	0.949	4	0.096	0.098	0.020	0.796
SNP_30456	2	0.510	2	0.375	0.500	0.000	1.000
SNP_13910	2	0.966	3	0.064	0.066	0.023	0.661
Mean±SE	2.34±0.09	0.851±0.022	3.14±0.11	0.191±0.021	0.224±0.027	0.025±0.003	0.837±0.031

---

## SSR and SNP-based genetic relationships

Genetic identity values among groups within each of the eggplant complexes had high values, both with SSR and SNP markers (Table 6). Not surprisingly, low values of genetic identities have been obtained between *S. incanum* and *S. aethiopicum* groups, as markers were selected for polymorphism between the transcriptomes of two accessions of these taxa (Gramazio et al. 2016). When considering each of the eggplant complexes, the genetic identities within eggplant complexes have been larger than the between eggplant complexes identities. For example for the brinjal eggplant complex, the within complex genetic identity (i.e., between *S. incanum* and *S. melongena*) has been 0.810 for SSRs and 0.588 for SNPs, while the range for genetic identities with other eggplant complex groups has been between 0.045 (*S. melongena* vs. *S. macrocarpon*) and 0.439 (*S. melongena* vs. *S. anguivi*) for SSRs and between 0.107 (*S. incanum* vs. *S. aethiopicum* group Gilo) and 0.565 (*S. melongena* vs. *S. macrocarpon*) for SNPs (Table 6). Amazingly, in this case, the genetic identity between *S. melongena* and *S. macrocarpon* has been largest with SNPs and lowest with SSRs, reflecting that different levels of genetic diversity are sampled by both markers (Ellegren, 2004; Kashi and King, 2006; Fischer et al. 2017). In the case of the gboma eggplant complex, the within complex identity has been very high, with a value of 0.935 for SSRs and 0.911 for SNPs, and it has ranged between 0.045 (*S. melongena* vs. *S. macrocarpon*) and 0.229 (*S. macrocarpon* vs. Intermediate *S. anguivi*-*S. aethiopicum*) for SSRs and between 0.343 (*S. incanum* vs. *S. dasyphyllum*) and 0.753 (*S. macrocarpon* vs. *S. aethiopicum* group Shum). For this complex group the genetic identities

with other complex groups have been much larger with SNPs than with SSRs (Table 6). Finally, for the scarlet eggplant complex the within complex identity has ranged between 0.704 (*S. anguivi* vs. *S. aethiopicum* group Aculeatum) and 0.970 (*S. aethiopicum* group Gilo vs. *S. aethiopicum* group Kumba) for SSRs, and between 0.859 (*S. anguivi* vs. *S. aethiopicum* group Kumba) and 0.976 (*S. aethiopicum* group Gilo vs. *S. aethiopicum* group Kumba) for SNPs; when considering genetic identities with other groups it has ranged between 0.098 (*S. incanum* vs. *S. aethiopicum* group Shum) and 0.439 (*S. melongena* vs. *S. anguivi*) for SSRs and between 0.107 (*S. incanum* vs. *S. aethiopicum* group Gilo) and 0.753 (*S. macrocarpon* vs. *S. aethiopicum* group Shum) (Table 6). In this case, the genetic identities of scarlet eggplant with the brinjal eggplant have been larger than with the gboma eggplant with SSRs, while the contrary occurred with SNPs. Overall, these results provide evidence that important differences exist among SSR and SNP markers for genetic identities among groups. This may have important consequences for the establishment of relationships among eggplant complexes based on genetic profiles, as depending on the markers used the results may be very different.

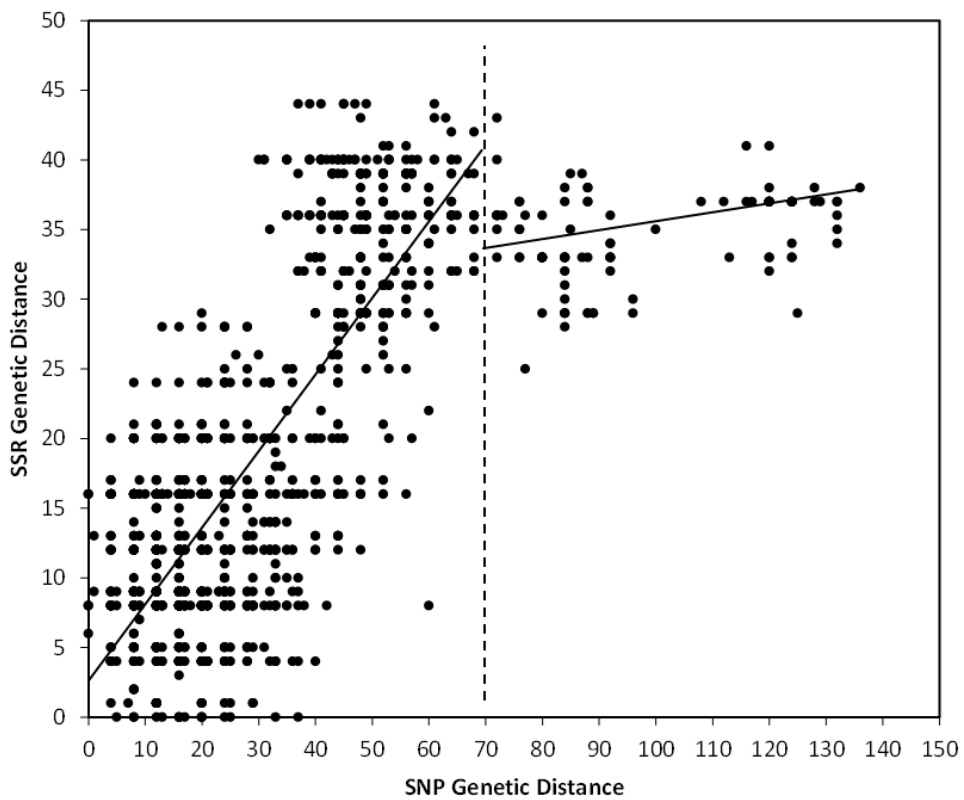
SSR and SNP pair-wise genetic distances among all individual accessions displayed a moderate correlation ( $r^2=0.529$ ;  $b=0.341$ ;  $P<0.0001$ ) (Figure 1). However, it is evident from Figure 1 that there is a difference in the relationship between both markers depending on the value of the SNP genetic distance. In this way, for values of SNP genetic distance below 70, the correlation is much higher ( $r^2=0.592$ ;  $b=0.534$ ;  $P<0.0001$ ) than for values for SNP genetic distance above 70 ( $r^2=0.080$ ;  $b=0.045$ ;  $P=0.0047$ ). This shows that, for the materials used, when genetic

distances based on SNPs are low to intermediate (i.e., within complexes), SNPs and SSRs provide similar levels of information, while when the genetic distances are higher (i.e., among complexes), SNPs provide better resolution, as SSR-based genetic distances seem to reach a saturation. This different performance of both types of markers is very likely due to the much higher rate of mutation of SSRs compared to SNPs (Hamblin et al. 2007; Coates et al. 2009). In this way, when certain levels of phylogenetic distance are reached, the SSR alleles seem to have diverged so much that they are not good to establish relationships based on phylogenetic distance. This has important implications for phylogenetic and germplasm conservation studies.



**Table 6.** Nei (1972) genetic identities among the different groups of the brinjal, gboma and scarlet eggplant complexes based on SSR (above the diagonal) and SNP (below the diagonal) markers. Vertical and horizontal lines separate the different eggplant complexes.

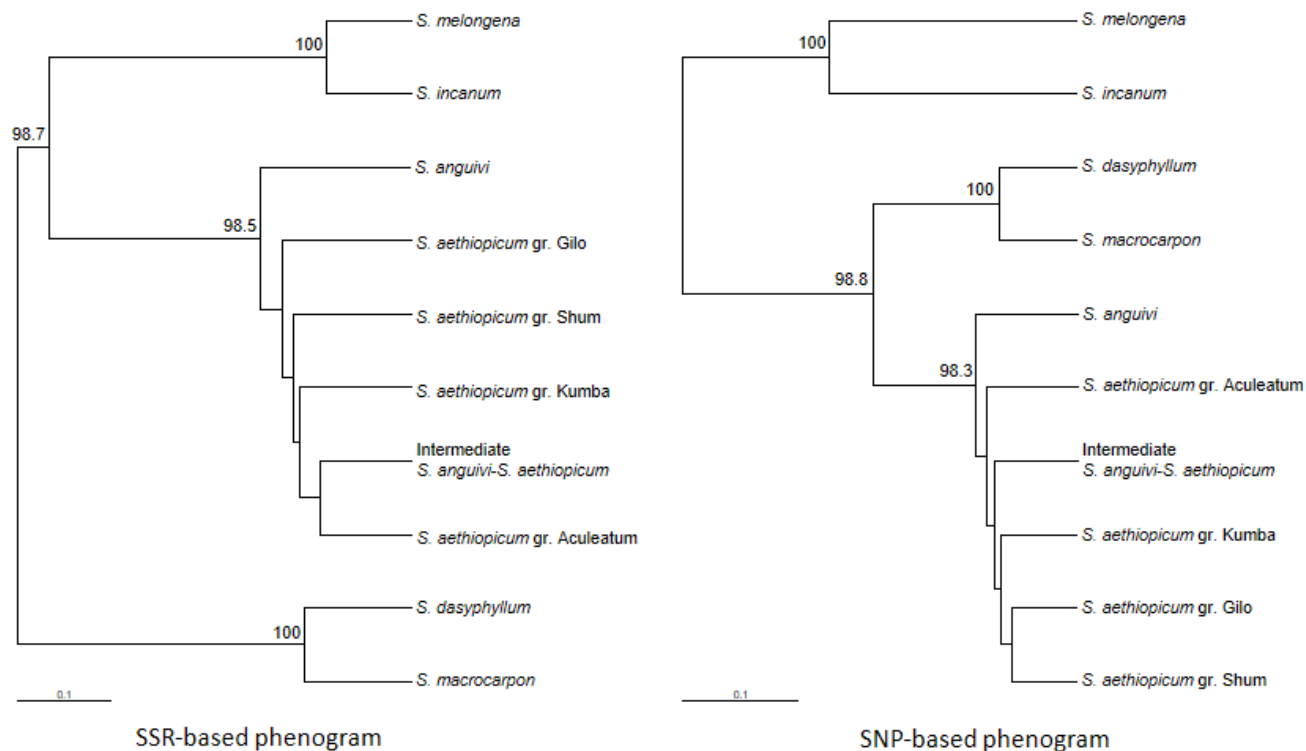
Complex/species/groups	Brinjal eggplant complex		Gboma eggplant complex		Scarlet eggplant complex					
	<i>S. incanum</i>	<i>S. melongena</i>	<i>S. dasyphyllum</i>	<i>S. macrocarpon</i>	<i>S. anguivi</i> ( <i>S. an.</i> )	Intermediate <i>S. an.</i> - <i>S. ae.</i>	<i>S. aethiopicum</i> ( <i>S. ae.</i> )			
							Aculeatum	Gilo	Kumba	Shum
Brinjal eggplant complex										
<i>S. incanum</i>		0.810	0.150	0.152	0.293	0.192	0.145	0.169	0.179	0.098
<i>S. melongena</i>	0.588		0.050	0.045	0.439	0.273	0.217	0.310	0.275	0.195
Gboma eggplant complex										
<i>S. dasyphyllum</i>	0.343	0.500		0.935	0.103	0.132	0.101	0.117	0.103	0.103
<i>S. macrocarpon</i>	0.402	0.565	0.911		0.174	0.229	0.199	0.207	0.196	0.192
Scarlet eggplant complex										
<i>S. anguivi</i>	0.162	0.441	0.694	0.722		0.834	0.704	0.887	0.853	0.750
Intermediate <i>S. an.</i> - <i>S. ae.</i>	0.165	0.434	0.680	0.743	0.924		0.884	0.955	0.957	0.900
<i>S. aethiopicum</i> Aculeatum	0.172	0.470	0.644	0.734	0.878	0.929		0.901	0.925	0.889
<i>S. aethiopicum</i> Gilo	0.107	0.409	0.644	0.712	0.922	0.981	0.954		0.970	0.922
<i>S. aethiopicum</i> Kumba	0.159	0.436	0.636	0.748	0.859	0.948	0.965	0.976		0.940
<i>S. aethiopicum</i> Shum	0.148	0.443	0.652	0.753	0.904	0.953	0.963	0.973	0.970	



**Figure 1.** Relationship between SNP (X-axis) and SSR (Y-axis) pairwise genetic distances among 48 individual accessions of the brinjal, gboma and scarlet eggplant complexes.

## Cluster analyses

The topology of the SSR-based and SNP-based phenograms is different (Figure 2). In other studies, the results obtained on the relationships among the three eggplant complexes depended on the markers used (Sakata et al. 1991; Furini and Wunder, 2004; Levin et al. 2006; Isshiki et al. 2008; Weese and Bohs, 2010; Meyer et al. 2012; Särkinen et al. 2013; Vorontsova et al. 2013, Acquadro et al. 2017). Although in both cases three major clusters are identified, corresponding to each of the three eggplant complexes, in the SSR-based phenogram the gboma eggplant complex cluster is basal to the brinjal and scarlet eggplant complexes, while in the SNP-based phenogram the brinjal eggplant cluster is basal to the gboma and scarlet eggplant complexes. Amazingly, Acquadro et al., (2017), using genomic SNPs found that the scarlet eggplant complex cluster is basal to the brinjal and gboma eggplant complexes. In our case, because markers used here were selected for polymorphism between expressed sequences of one accession of each of *S. incanum* and *S. aethiopicum* (Gramazio et al. 2016), it was expected that the largest distance in the tree should have been between the clusters of brinjal and scarlet eggplants, but this was only true for SNP markers (Figure 2). This provides evidence that SNP markers may provide better resolution than SSRs when evaluating phylogenetic relationships among *Solanum* taxa not belonging to the close primary genepool. When considering the scarlet eggplant complex, in both phenograms the wild *S. anguivi* is basal to the other groups, although the other groups present a different clustering pattern depending on the markers used (Figure 2).



**Figure 2.** UPGMA hierarchical clustering consensus phenograms based on Nei (1972) genetic distances for the brinjal (*S. melongena* and *S. incanum*), gboma (*S. macrocarpon* and *S. dasyphyllum*) and scarlet eggplant (*S. aethiopicum* and *S. anguivi*) complex groups according to SSR (left) and SNP markers (right). Bootstrap values (based on 1000 replications; expressed in percentage) greater than 50% are indicated at the corresponding nodes.

## ***Conclusions***

The transcriptome-derived SSR and SNP markers have been highly polymorphic in the eggplants collection evaluated and proved useful for genetic fingerprinting. As expected, SSRs were on average more informative than SNPs, but a similar number of fingerprints were obtained with 11 SSRs and 35 SNPs. Both SSR and SNP markers confirmed the high fixation index of the eggplant materials and clearly distinguished the three eggplant complexes. However, different results depending on the type of marker were obtained for the relationships among eggplant complexes, indicating that they sample different levels of genetic variation. In this respect, SSRs and SNPs presented a moderate correlation for low to intermediate values of SNP pair-wise genetic distance, but a low correlation for high SNP genetic distances. This suggests that both markers are complementary in the information provided, although SNPs seem more appropriate to evaluate materials genetically distant in the eggplant complexes. This information will be useful for eggplants germplasm management, phylogenetic studies, as well as for genetic fingerprinting and breeding.

## **Acknowledgements**

This work has been funded by European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) and by Spanish Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional (grant AGL2015-64755-R from MINECO/FEDER). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral contract (Programa FPI de la UPV-Subprograma 1/2013 call). Mariola Plazas is grateful to Spanish Ministerio de Economía, Industria y Competitividad for a post-doctoral grant within the Juan de la Cierva-Formación programme (FJCI-2015-24835).







### **Chapter 3: Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes**

Acquadro A<sup>1</sup>, Barchi L<sup>1</sup>, Gramazio P<sup>2</sup>, Portis E<sup>1</sup>, Vilanova S<sup>2</sup>, Comino C<sup>1</sup>, Plazas M<sup>3</sup>, Prohens J<sup>2\*</sup>, Sergio Lanteri<sup>1\*</sup>

<sup>1</sup> University of Turin, DISAFA, Plant Genetics and Breeding, University of Turin, Largo Braccini 2, Grugliasco, Torino, Italy.

<sup>2</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

<sup>3</sup> Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas - Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

\* Corresponding authors: [sergio.lanteri@unito.it](mailto:sergio.lanteri@unito.it), [jprohens@btc.upv.es](mailto:jprohens@btc.upv.es)

**Citation:** Acquadro A, Barchi L, Gramazio P, Portis E, Vilanova S, Comino C, Plazas M, Prohens J, Lanteri S (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLOS ONE* 12(7):e0180774.

<https://doi.org/10.1371/journal.pone.0180774>

## ***Abstract***

Brinjal (*Solanum melongena*), scarlet (*S. aethiopicum*) and gboma (*S. macrocarpon*) eggplants are three Old World domesticates. The genomic DNA of a collection of accessions belonging to the three cultivated species, along with a representation of various wild relatives, was characterized for the presence of single nucleotide polymorphisms (SNPs) using a genotype-by-sequencing approach. A total of 210 million useful reads were produced and were successfully aligned to the reference eggplant genome sequence. Out of the 75,399 polymorphic sites identified among the 76 entries in study, 12,859 were associated with coding sequence. A genetic relationships analysis, supported by the output of the FastSTRUCTURE software, identified four major sub-groups as present in the germplasm panel. The first of these clustered *S. aethiopicum* with its wild ancestor *S. anguivi*; the second, *S. melongena*, its wild progenitor *S. insanum*, and its relatives *S. incanum*, *S. lichtensteinii* and *S. linneanum*; the third, *S. macrocarpon* and its wild ancestor *S. dasyphyllum*; and the fourth, the New World species *S. sisymbriifolium*, *S. torvum* and *S. elaeagnifolium*. By applying a hierarchical FastSTRUCTURE analysis on partitioned data, it was also possible to resolve the ambiguous membership of the accessions of *S. campylacanthum*, *S. violaceum*, *S. lidii*, *S. vespertilio* and *S. tomentosum*, as well as to genetically differentiate the three species of New World Origin. A principal coordinates analysis performed both on the entire germplasm panel and also separately on the entries belonging to sub-groups revealed a clear separation among species, although not between each of the domesticates and their respective wild ancestors. There was no clear differentiation between either distinct

cultivar groups or different geographical provenance. Adopting various approaches to analyze SNP variation provided support for interpretation of results. The genotyping-by-sequencing approach showed to be highly efficient for both quantifying genetic diversity and establishing genetic relationships among and within cultivated eggplants and their wild relatives. The relevance of these results to the evolution of eggplants, as well as to their genetic improvement, is discussed.

**Abbreviations:**

CWR: crop wild relatives

GP1: primary genepool

NGS: next generation sequencing

SNPs: single nucleotide polymorphisms

GBS: genotype-by-sequencing

UPGMA: unweighted pair-group arithmetic mean

PCoA: principal coordinate analysis

COPH: cophenetic values

## ***Introduction***

Eggplant, also known as brinjal eggplant or aubergine (*Solanum melongena* L., Solanaceae,  $2n=2x=24$ ), is cultivated worldwide and is one of the most important vegetable crops, being the second most important solanaceous crop grown for its fruit after tomato (*S. lycopersicum* L.) (Faostat, 2016). The bulk of eggplant production is concentrated in China, India, Iran, Egypt and Turkey, with Italy and Spain representing the most important European Union producers (Faostat, 2016). Because of its importance as a staple vegetable food in many countries from tropical and subtropical regions, eggplant is included with 34 other food crops in the Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (Fowler et al., 2003). Eggplant berries are a source of dietary minerals as well as vitamins and other health-promoting metabolites such as anthocyanins and chlorogenic acid, with nutraceutical and anti-oxidant properties (Stommel and Whitaker, 2003; Kwon et al., 2007; Akanitapichat et al., 2010; Mennella et al., 2010).

Brinjal eggplant selection and breeding over the years have been mainly focused on the improvement of fruit traits (Harlan et al., 1992), such as size, weight, color, and shape (Kashyap et al., 2003; Frary et al., 2006) reduced prickliness, yield potential (Daunay and Lester, 2001), and more recently organoleptic, nutritional and bioactive properties (Plazas et al., 2013a; Zhang et al., 2014; Docimo et al., 2015; Toppino et al., 2016). This has resulted in the development of a large number of eggplant varieties, whose fruit shape varies from flattened to elongated. However, like in many other domesticates, anthropogenic selection has resulted in a drastic reduction of the genetic variation across eggplant genome, due to

both the genetic bottleneck resulting from the sampling process of a limited number of wild plants chosen for domestication (Flint-Garcia et al., 2013) and the migration of a limited number of genotypes from the primary to the secondary centers of domestication (Cericola et al., 2013). The eggplant inter-fertile cultivated species as well as crop wild relatives (CWRs) are a source of variation for many traits of interest and represent an obvious target to aid eggplant improvement; however, to date, their potential use has largely remained unexploited (Rotino et al., 2014).

Unlike tomato (*S. lycopersicum*), potato (*S. tuberosum* L.) and pepper (*Capsicum* spp.), eggplant is native to the Old World and was independently domesticated from *S. insanum* L. in the Indian subcontinent and in China (Cericola et al., 2013; Knapp et al., 2013; Rotino et al., 2014), with a possible additional and independent center of domestication in the Philippines (Meyer et al., 2012a). Besides *S. melongena*, two other eggplant species are commonly grown in sub-Saharan Africa (Daunay et Hazra, 2012), the scarlet eggplant (*S. aethiopicum* L.) and the gboma eggplant (*S. macrocarpon* L.). Both species can be inter-crossed with brinjal eggplant producing hybrids with intermediate fertility (Plazas et al., 2014a). The three cultivated eggplants belong to the *Leptostemonum* clade and to a species-rich subclade composed exclusively of Old World taxa (the Old World clade sensu (Stern et al., 2011; Vorontsova et al., 2013) from Africa, Australia, and Asia (including Eurasia and the Middle East).

Scarlet eggplant (*S. aethiopicum* L.) is an important vegetable in Central and West Africa, but it is also cultivated in the Caribbean and Brazil as well as in some areas of South Italy (Sunseri et al., 2010). It is a hypervariable species and includes hundreds of local varieties (Schippers et al., 2000) clustered in four main cultivar groups: Aculeatum, Gilo,

Kumba and Shum, which are completely inter-fertile (Plazas et al., 2014a). The four cultivar groups are differently exploited, since *Aculeatum* is used as ornamental, Gilo for its fruits, Kumba for both fruits and leaves, while Shum for its leaves (Schippers et al., 2000; Lester and Daunay, 2003). Gboma eggplant (*S. macrocarpon* L.) is less widespread in cultivation, although the species represents a major vegetable in some countries like Benin and in the rain forest regions of Coastal Africa and Congo River (Lester and Jaeger, 1990). It is also a morphologically variable species and it is grown for its fruits, leaves or both (Schippers et al., 2000; Lester and Daunay, 2003). The high variability within both scarlet and gboma eggplants has been recently confirmed by Plazas et al. (2014), whom by applying conventional descriptors as well as the high-throughput Tomato Analyzer phenomics tool characterized a wide set of accessions of both cultivated species as well as from the scarlet eggplant wild ancestor *S. anguivi* Lam., *S. aethiopicum*-*S.anguivi* intermediate forms, and the gboma eggplant wild ancestor *S. dasyphyllum* Schumach. & Thonn. Each of the three cultivated eggplants together with their wild ancestors and the closest wild relatives are commonly referred to as the brinjal, scarlet and gboma eggplant complexes (Daunay and Lester, 2001; Lester and Daunay, 2003; Knapp et al., 2013; Plazas et al., 2014a).

Wild relatives of cultivated eggplants, which are well adapted to grow in a wide range of conditions, from desert to swampy areas and environments with wide ranges of temperatures, are a source of useful traits for eggplants breeding. Unfortunately the latter remain largely unexploited and a limited number of reports on the use of the variation available in the wild species has been reported (Daunay et Hazra, 2012;

Rotino et al., 2014; Liu et al., 2015b) while, to our knowledge, no modern commercial varieties of eggplants carry introgression from wild species.

In brinjal eggplant and related species the delimitation of biologically meaningful genepools is challenging due to limited crossability data reported in literature (Plazas et al., 2016), as well as to the extremely large number of potential genepool members. By taking into account both relatedness, as measured by phylogenetic analyses and available data on crossability, recently Syfert et al. (2016) suggested the inclusion of one species (*S. insanum*) in the primary genepool (GP1), forty-eight species with which eggplant can be inter-crossed with varying degrees of difficulty in GP2, and three wild and weedy species native to the New World in GP3, i.e. *S. sisymbriifolium* Lam., *S. torvum* Sw. and *S. viarum* Dunal, with which only highly sterile hybrids can be obtained through embryo rescue or are not obtainable.

The great advances in next generation sequencing (NGS) technologies, with rapid increases in data volumes and quality combined with reducing costs, have provided breeders with a wide array of genomic tools which facilitate the characterization of germplasm collections and allow to gain a better understanding of how the genome contributes to the diversity detected at phenotypic level (Varshney et al., 2014). Single nucleotide polymorphisms (SNPs) represent the most frequent type of genetic polymorphism and have become the marker of choice for many applications in plant biology, conservation and breeding (Mammadov et al., 2012).

Here we report a genotype by sequencing (GBS) approach based on reducing genome complexity to detect SNPs polymorphisms in a set of seventy-six accessions of species belonging to the brinjal, gboma and

scarlet eggplant complexes, which include taxa included in the *S. melongena* primary, secondary and tertiary genepools. Our main goal was to assess, using a high-throughput genotyping technique, the genetic relationships within and between the genepools of the brinjal eggplant (*S. melongena*) and the two other cultivated eggplants, namely the scarlet (*S. aethiopicum*) and gboma (*S. macrocarpon*) eggplants. Apart from cultivated accessions, we also included in the study accessions of close wild relatives of the three crops, as well more distant species from the tertiary genepool species. The information obtained will be of great relevance for clarifying the relationships among cultivated and wild eggplants and will be useful to breeders using wild species for eggplant breeding



## ***Material and methods***

### **Plant materials**

A total of 76 accessions, including 16 entries of *S. melongena* from Asian and European origin, 30 of *S. aethiopicum* belonging to the four varietal groups (Aculeatum, Gilo, Kumba and Shum) plus intermediate forms between *S. aethiopicum* and *S. anguivi*, five of *S. macrocarpon*, and 25 accessions of 14 wild species were used for the present study (Table 1). Among the 16 entries of brinjal eggplant, two of them are doubled haploids (*S. melongena\_10* and *S. melongena\_12*) obtained by anther culture (Salas et al., 2011). Also, four brinjal eggplant entries come from two original sources (entries *S. melongena\_1* and *S. melongena\_2* from the original source MEL1; and accessions *S. melongena\_6* and *S. melongena\_7* from the original source MEL5) (Table 1). Among the wild relatives are included the putative ancestors of brinjal eggplant (*S. insanum*), scarlet eggplant (*S. anguivi*), and gboma eggplant (*S. dasyphyllum*) (Lester and Niakan, 1986; Bukenya and Carasco, 1994; Ranil et al., 2017), as well as eight other wild species from Old World origin (*S. campylacanthum* Hochst. ex A. Rich, *S. incanum* L., *S. lichtensteinii* Willd., *S. lidii* Sunding, *S. linnaeanum* Hepper & P.-M.L. Jaeger, *S. tomentosum* L., *S. vespertilio* Aiton, and *S. violaceum* Ortega), and three native to the New World (*S. elaeagnifolium* Cav., *S. sisymbriifolium*, and *S. torvum*) (Boyd et al., 1984; Syfert et al., 2016). All these materials are conserved in the germplasm collection maintained at Universitat Politècnica de València (Valencia, Spain).

**Table 1.** Plant materials used including taxon, accession name, accession code used in the present work, country of origin and fruit shape and predominant colour.

Taxon and accession	Code	Country of origin	Fruit shape <sup>a</sup>	Predominant fruit colour <sup>b</sup>
<b><i>S. aethiopicum</i> L. gr. Aculeatum</b>				
MM457	<i>S. aethiopicum</i> aculeatum_1	Japan	1	1.3
UPV29803	<i>S. aethiopicum</i> aculeatum_2	China	1	1.2
RNL0187	<i>S. aethiopicum</i> aculeatum_3	Burkina Faso	1	1.2
MM1483	<i>S. aethiopicum</i> aculeatum_4	Ghana	1	1.3
<b><i>S. aethiopicum</i> L. gr. Gilo</b>				
BBS151A	<i>S. aethiopicum</i> gilo_1	Ivory Coast	7	1.1
IVIA026	<i>S. aethiopicum</i> gilo_2	Unknown	7	1.2
RARE_PLANTS_GILO	<i>S. aethiopicum</i> gilo_3	Unknown	3	1.3
RNL0252	<i>S. aethiopicum</i> gilo_4	Ghana	3	1.2
UPV29014	<i>S. aethiopicum</i> gilo_5	Unknown	5	1.2
RNL0395	<i>S. aethiopicum</i> gilo_6	Liberia	3	1.1
RNL0288	<i>S. aethiopicum</i> gilo_7	Ghana	5	2
BBS181A	<i>S. aethiopicum</i> gilo_8	Ivory Coast	1	1.3
BBS147G	<i>S. aethiopicum</i> gilo_9	Ivory Coast	1	1.3
BBS140B	<i>S. aethiopicum</i> gilo_10	Ivory Coast	3	1.2
BBS159B	<i>S. aethiopicum</i> gilo_11	Ivory Coast	5	1.1
BBS142A	<i>S. aethiopicum</i> gilo_12	Ivory Coast	5	1.2
AN05	<i>S. aethiopicum</i> gilo_13	Angola	3	1.1
<b><i>S. aethiopicum</i> L. gr. Kumba</b>				
INRA_4	<i>S. aethiopicum</i> kumba_1	Senegal	1	1.1
MM1207	<i>S. aethiopicum</i> kumba_2	Mali	1	1.1
BBS111	<i>S. aethiopicum</i> kumba_3	Ivory Coast	1	2
BBS110	<i>S. aethiopicum</i> kumba_4	Ivory Coast	1	1.1
<b><i>S. aethiopicum</i> L. gr. Shum</b>				
RNL0022	<i>S. aethiopicum</i> shum_1	Benin	3	1.3
RNL_0340	<i>S. aethiopicum</i> shum_2	Zimbabwe	1	1.2
<b><i>S. aethiopicum</i> L.-<i>S. anguivi</i> Lam. Intermediate</b>				
BBS116	<i>S. aethiopicum-anguivi</i> _1	Ivory Coast	3	1.3
BBS192E	<i>S. aethiopicum-anguivi</i> _2	Ivory Coast	5	1.2
BBS148D	<i>S. aethiopicum-anguivi</i> _3	Ivory Coast	3	1.1
BBS131C	<i>S. aethiopicum-anguivi</i> _4	Ivory Coast	3	1.1
BBS184	<i>S. aethiopicum-anguivi</i> _5	Ivory Coast	3	1.1
<b><i>S. aethiopicum</i> L.-<i>S. anguivi</i> Lam. Intermediate</b>				
BBS180A	<i>S. aethiopicum-anguivi</i> _6	Ivory Coast	5	1.1
BBS114	<i>S. aethiopicum-anguivi</i> _7	Ivory Coast	5	1.2
<b><i>S. anguivi</i> Lam.</b>				
ANG1	<i>S. anguivi</i> _1	Ivory Coast	3	1.1
ANG2	<i>S. anguivi</i> _2	Ivory Coast	3	1.3

<b><i>S. campylacanthum</i> Hochst. ex A. Rich</b>				
CAM5	<i>S. campylacanthum_1</i>	Tanzania	3	1.2
CAM6	<i>S. campylacanthum_2</i>	Kenya	3	1.2
CAM8	<i>S. campylacanthum_3</i>	Tanzania	3	1.2
<b><i>S. dasyphyllum</i> Schumach. &amp; Thonn.</b>				
DAS1	<i>S. dasyphyllum_1</i>	Uganda	1	1.2
<b><i>S. elaeagnifolium</i> Cav.</b>				
ELE1	<i>S. elaeagnifolium_1</i>	Senegal	3	1.2
ELE2	<i>S. elaeagnifolium_2</i>	Greece	3	1.2
<b><i>S. incanum</i> L.</b>				
MM577	<i>S. incanum_1</i>	Israel	5	1.2
<b><i>S. insanum</i> L.</b>				
INS1	<i>S. insanum_1</i>	Sri Lanka	5	1.2
INS2	<i>S. insanum_2</i>	Sri Lanka	3	1.2
INS3	<i>S. insanum_3</i>	Japan	3	1.2
<b><i>S. lichtensteinii</i> Willd.</b>				
LIC1	<i>S. lichtensteinii_1</i>	South Africa	3	1.3
LIC2	<i>S. lichtensteinii_2</i>	Iran	3	1.1
<b><i>S. lidii</i> Sunding</b>				
LID1	<i>S. lidii_1</i>	Spain	3	1.3
LID2	<i>S. lidii_2</i>	Spain	3	1.3
<b><i>S. linnaeanum</i> Hepper &amp; P.-M.L. Jaeger</b>				
LIN1	<i>S. linnaeanum_1</i>	Spain	3	1.3
LIN3	<i>S. linnaeanum_2</i>	Tunisia	3	1.3
<b><i>S. macrocarpon</i> L.</b>				
MM1558	<i>S. macrocarpon_1</i>	Malaysia	1	2
<b><i>S. macrocarpon</i> L.</b>				
BBS168	<i>S. macrocarpon_2</i>	Ivory Coast	1	2
BBS117	<i>S. macrocarpon_3</i>	Ivory Coast	1	1.3
BBS171B	<i>S. macrocarpon_4</i>	Ivory Coast	1	2
BBS178	<i>S. macrocarpon_5</i>	Ivory Coast	5	1.2
<b><i>S. melongena</i> L.</b>				
MEL1_2	<i>S. melongena_1</i>	Ivory Coast	5	2
MEL1_3	<i>S. melongena_2</i>	Ivory Coast	5	2
MEL2	<i>S. melongena_3</i>	Ivory Coast	5	7
MEL3	<i>S. melongena_4</i>	Ivory Coast	7	1.2
MEL4	<i>S. melongena_5</i>	Sri Lanka	3	7
MEL5_2	<i>S. melongena_6</i>	Sri Lanka	7	7

MEL5_5	<i>S. melongena_7</i>	Sri Lanka	7	7
MEL6	<i>S. melongena_8</i>	Sri Lanka	7	7
AN-S-26	<i>S. melongena_9</i>	Spain	5	7
DH_AN-S-26	<i>S. melongena_10</i>	Spain	5	7
MM1597	<i>S. melongena_11</i>	India	9	1.2
DH_ECAVI	<i>S. melongena_12</i>	Breeding line	7	8
H15	<i>S. melongena_13</i>	Spain	5	7
A0413	<i>S. melongena_14</i>	Unknown	1	2
ASI-S-1	<i>S. melongena_15</i>	China	1	8
IVIA371	<i>S. melongena_16</i>	Spain	5	7

---

***S. sisymbriifolium* Lam.**

SIS1	<i>S. sisymbriifolium_1</i>	Unknown	3	1.2
SIS2	<i>S. sisymbriifolium_2</i>	Unknown	5	1.3

---

***S. tomentosum* L.**

TOM1	<i>S. tomentosum_1</i>	South Africa	3	1.3
------	------------------------	--------------	---	-----

---

***S. torvum* Sw.**

TOR2	<i>S. torvum_1</i>	Sri Lanka	3	1.2
TOR3	<i>S. torvum_2</i>	Unknown	3	1.3

---

***S. vespertilio* Aiton**

VES2	<i>S. vespertilio_1</i>	Spain	3	1.3
------	-------------------------	-------	---	-----

---

***S. violaceum* Ortega**

VIO1	<i>S. violaceum_1</i>	Sri Lanka	3	1.2
------	-----------------------	-----------	---	-----

---

<sup>a</sup> Fruit shape according to the following scale: 1=broader than long; 3=as long as broad; 5=slightly longer than broad; 7=twice as long as broad; 8=three times as long as broad; 9=several times as long as broad.

<sup>b</sup> Fruit predominant colour when the fruit is physiologically immature according to the following categories, in which the green colour (1) has been subdivided into three subcategories: 1.1=clear green; 1.2=intermediate green; 1.3=dark green; 2=milk white; 3=deep yellow; 4=fire red; 5=scarlet red; 6=lilac grey; 7=purple; 8=purple black; 9=black.

## Library construction and sequencing

DNA was extracted following a modified CTAB method (Doyle and Doyle, 1990) as indicated elsewhere (Acquadro et al., 2009). Library construction (11/2015) was performed as proposed in Peterson et al. (2012) and modified as in Acquadro et al. (2016), by using a HindIII-MseI enzyme combination and adding a final biotin/streptavidin-coated beads based purification step. Quality, quantity and reproducibility of libraries were assessed on a Bioanalyzer instrument (DNA High Sensitivity chip) as well as qPCR. On the basis of the quantitation, DNA libraries were pooled and sequenced on Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA), following the manufacturer protocol using 100SE chemistry.

## Sequence analysis

Raw reads were analyzed with Scythe (<https://github.com/vsbuffalo/scythe>) for filtering out contaminant substrings and Sickle (<https://github.com/najoshi/sickle>), which allows to remove reads with poor quality ends ( $Q < 30$ ). Illumina reads were demultiplexed on the basis of the Illumina TruSeq index. Alignment to the reference eggplant genome (Lanteri et al., 2014; Barchi et al., 2016) was carried out using BWA aligner (Li and Durbin, 2009) (i.e., mem command) with default parameters and avoiding multiple-mapping reads. SNP mining was conducted by adopting a Samtools-based pipeline (Li et al., 2009). Homozygous/heterozygous SNP/Indel calls were considered only with phred-scaled genotype likelihood equal zero. A catalog of candidate high quality SNPs was produced. Relationships among the

genotypes were computed using: i) whole genome, and ii) coding (within exons) SNP/indel datasets. The proportion of heterozygous SNPs for each genotype was estimated by the ratio of total number of heterozygous SNPs and all the detected SNPs (excluding missing SNPs) as well as the ratio of the number of heterozygous SNPs in coding regions and all the detected SNPs in coding regions.

### **Genetic relationships analysis and population structure**

SNP data were coded according to the number of occurring polymorphisms, being assigned a 0 if they showed the homozygous reference type, a 1 if the variant occurred in one chromosome and a 2 if the variant was present in both chromosomes. Genetic similarities between pairs of entries were quantified by the Dice similarity index (Dice, 1945) as  $2m^+/(2m^+ + m^-)$ , where  $m^+$  is the number of matches (1-1 and 2-2) and  $m^-$  is the number of mismatches (0-1, 0-2 and 1-2). Genetic relationships were described by using both the unweighted pair-group arithmetic mean (UPGMA) method with 1,000 bootstraps, and principal coordinate analysis (PCoA) by means of Past 3.14 software (Hammer et al., 2001). A co-phenetic matrix was also produced using the hierarchical cluster system, by means of the COPH (cophenetic values) routine, and correlated with the original distance matrix, in order to test for associations between clusters and the similarity matrix.

*FastSTRUCTURE* (Raj et al., 2014) was used to estimate the number of sub-populations in the panel, applying the admixture model for the ancestry of individuals and correlated allele frequencies. A hierarchical *FastSTRUCTURE* analysis (Emanuelli et al., 2013) was also applied on

accessions which clustered in sub-group 1 and subgroup 4 following UPGMA analysis as well as on the set of all the remaining. The program was run with default setting using simple prior to obtain a reasonable range of values for the number of populations (K), *FastSTRUCTURE* was executed for multiple values of K (K = 1–9). The script chooseK.py (Raj et al., 2014) was then used to infer the most likely number of populations.

## ***Results and discussion***

### **Sequencing and SNPs identification**

A total of 225 million single reads were produced. About 94% of raw reads contained the expected restriction site overhang, along with discriminating inline barcodes. The average number of successfully demultiplexed reads per sample was 2.7 M, with a standard deviation of 1.5 M (S1 Fig). Sequences were trimmed and quality cleaned to 210 million of useful reads (6.2% discarded). The latter were then aligned to the recently produced reference eggplant genome (Lanteri et al., 2014; Barchi et al., 2016) and close to 100% of reads were successfully mapped to single regions (no multiple mapping was permitted). Mapped sequences showed an extensive coverage alongside the 12 chromosomes (data not shown).

In all, 75,399 polymorphic sites were identified among the 76 lines in study. Overall, all the *S. melongena* accessions, together with the three *S. insanum* accessions, showed a reduced level of polymorphism (on average 2.47 and 4.75 % respectively) when aligned to the reference genome. On the other hand the frequency of polymorphic SNPs ranged from 10.62 to 24.32% in the other entries (S1 Table).

*Solanum melongena* is a largely autogamous species (Daunay et Hazra, 2012), thus its low level of heterozygosity (on average 1.66%) is coherent with the expectation that germplasm accessions and non-hybrid varieties should be highly homozygous (Fig 1, S1 Table). Interestingly, the two *S. melongena* varieties (*S. melongena\_10*, *S. melongena\_12*, Fig 1, S1 Table), which are the result of diploidization of haploid plants obtained through anther culture, displayed some heterozygosity (<0.5%). This might be due to somaclonal variation, which is manifested as cytological

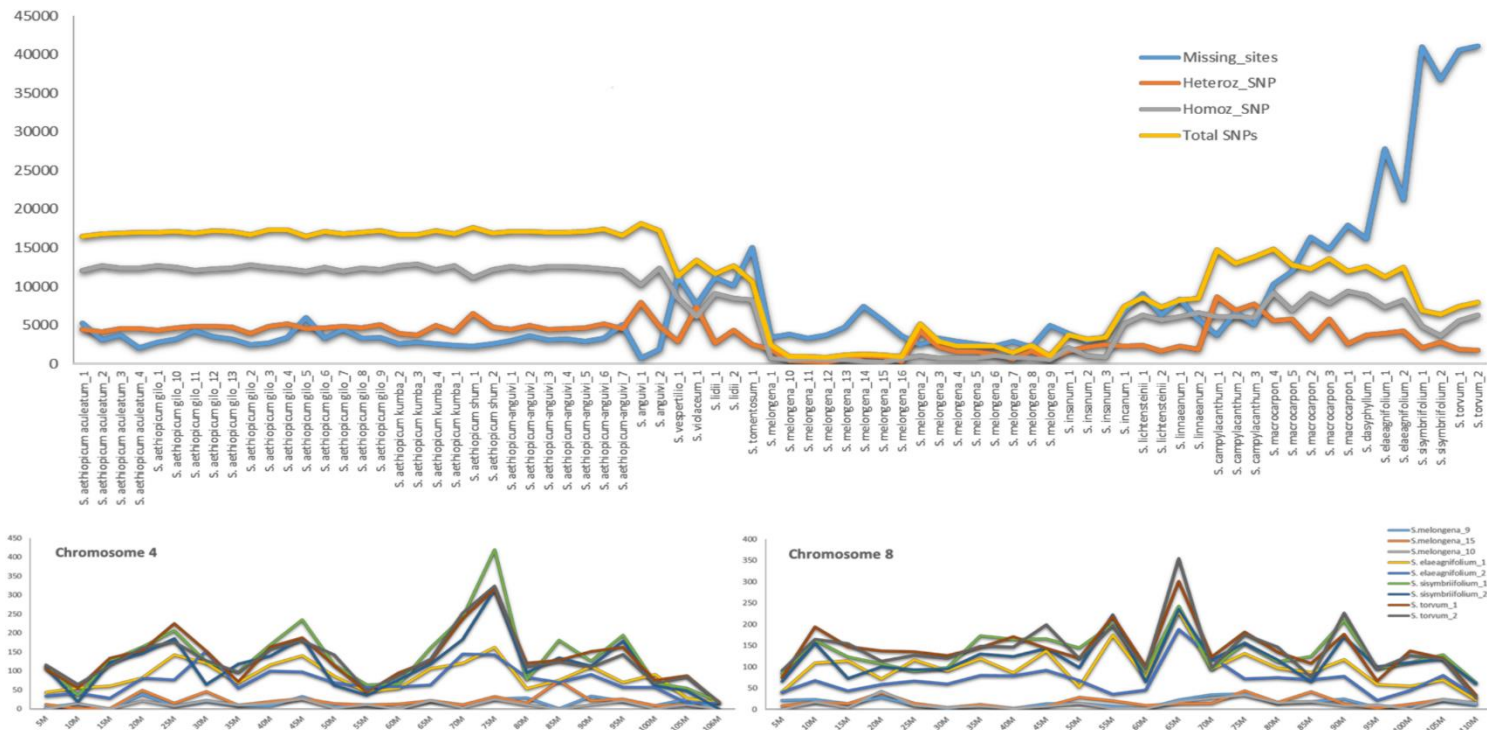


abnormalities, sequence change, and gene activation and silencing which occur through the ‘in vitro’ culture process and that provides evidence that DNA modifications occur more frequently in ‘in vitro’ cultivated than in seed-grown plants (Kaepler et al., 2000). However, it might be also a consequence of SNPs mapping on paralog genes since, similarly to tomato, potato and pepper, also eggplant the genome carries signs of the “T” triplication occurred during Solanaceae evolution (Barchi et al., 2016), or being the results of some mapping artifacts. This would suggest that the heterozygosity detected in the rest of *S. melongena* accessions would be overestimated by almost one third.

The two other cultivated eggplant species (i.e., *S. aethiopicum* and *S. macrocarpon*) showed, on average, higher heterozygosity than *S. melongena*, ranging from 4.52 to 9.53% (Fig 1, S1 Table). This might be a consequence of their higher allogamy and the more limited breeding efforts for stabilizing phenotypic and yield-related traits. Low heterozygosity was also observed in the wild *S. insanum*, *S. lichtensteinii* and *S. linnaeanum* (< 3.5%), while higher values, over 10%, were observed in the wild species *S. campylacanthum*, *S. anguivi* and *S. violaceum*.

Some missing data were observed in *S. melongena* and *S. aethiopicum* (ranging from 2.71% to 9.87%), some accessions showed a medium-high level of missing data (e.g., *S. macrocarpon* 19% on average), while others showed a surprisingly high number of missing data (up to 54.5%, 54.43 and 36.88, in *Solanum torvum*, *S. sisymbriifolium* and *S. elaeagnifolium*, respectively). This might be explained by the fact that these latter species are native to the New World (Boyd et al., 1984; Syfert et al., 2016) and in consequence have a more distant common ancestor,

and greater evolutionary divergence. Missing data were distributed on the different eggplant chromosomes; however, by adopting a five million bases sliding window analysis, some hot spot regions were highlighted (Fig 1). The filtering of the whole SNP dataset for the sites present in CDS regions granted 12,859 SNPs. The latter were used for all the subsequent analyses, since the relative number of missing data was lower in the coding dataset (3% on average) than in the whole dataset (10% on average, Fig 1, S1 Table). As an example the percentage of missing data of South American accessions (*S. elaeagnifolium*, *S. sisymbriifolium* and *S. torvum* – S1 Table) was lowered from about 46.1% to 15.1%, thereby increasing the resolution power of our analyses.



**Figure 1.** SNP numbers and distribution: A) Plot of the number of SNPs (total, homozygous and heterozygous) and missing sites observed within the collection of 76 *Solanum* accessions; B) The distribution of missing sites along two eggplant chromosomes (i.e.: chr4 and chr8). The trend lines track missing sites in three *Solanum melongena* and six accessions of American origin (*S. elaeagnifolium*, *S. sisymbriifolium* and *S. torvum*).

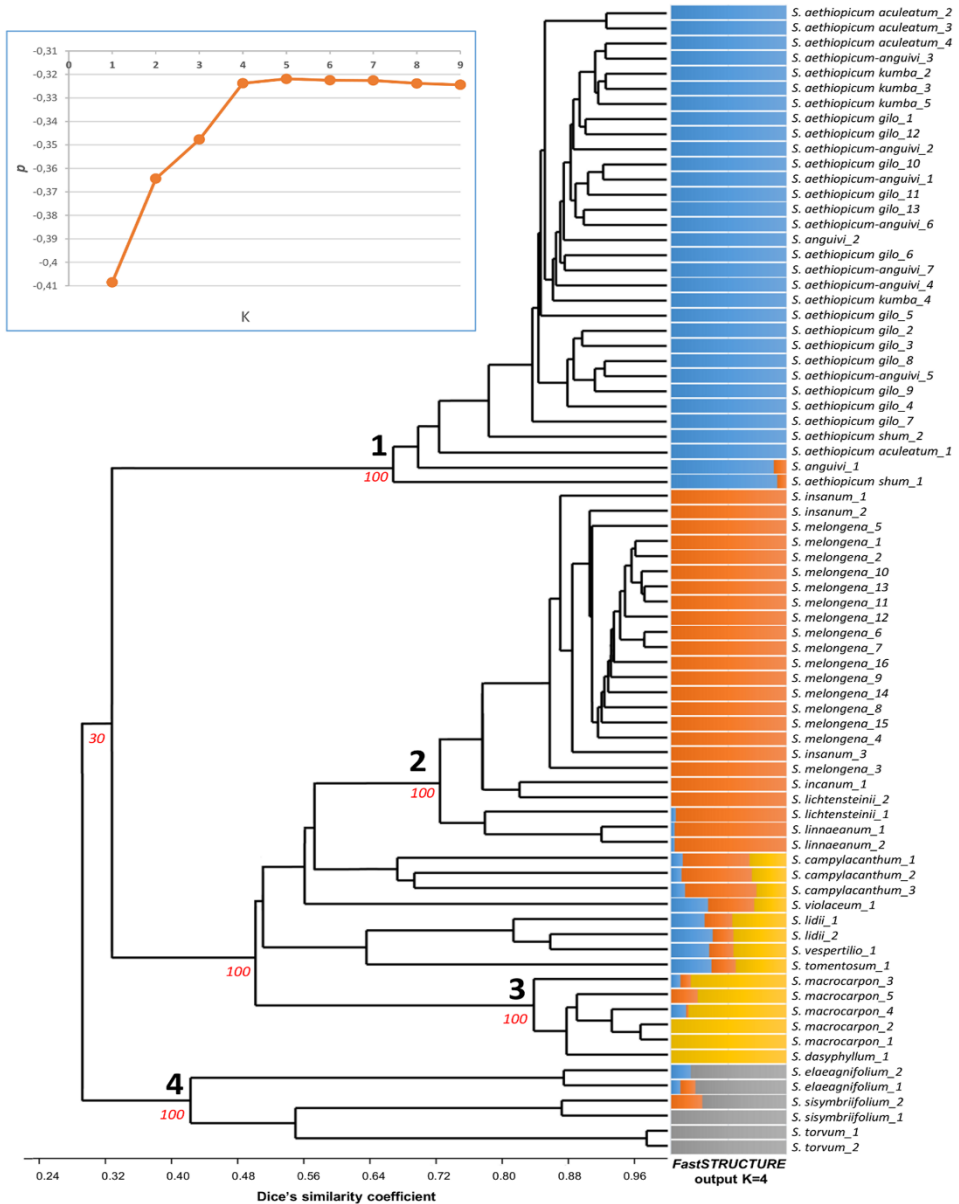
## Genetic relationships analysis and population structure

The UPGMA-based dendrogram and the output of *FastSTRUCTURE* (Raj et al., 2014) analysis (Fig 2) show the genetic relationships between the 76 accessions. Both, as well as the K analysis (Fig 2 box), suggest a population structure comprising four sub-groups. Each entry was fingerprinted and the co-phenetic correlation coefficient (r-value) between the Dice data matrix and the co-phenetic matrix was 0.978, highlighting a very good fit between the dendrogram clusters and the similarity matrices from which they were derived, indicating that the UPGMA method is suitable for the interpretation of our data. The fact that the sister entries *S. melongena\_1* and *S. melongena\_2*, which are derivatives from the original source MEL1 cluster together in the dendrogram, and the same occurs for accessions *S. melongena\_6* and *S. melongena\_7*, which derive from MEL5 provide a confirmation that the analysis is congruent.

According to the level of membership provided by *FastSTRUCTURE* (Raj et al., 2014), sub-group 1 (blue) includes all the accessions of scarlet eggplant (*S. aethiopicum*) and *S. anguivi*, which on the basis of previous studies has been reported to be its wild ancestor (Lester and Niakan, 1986; Lester and Thitai, 1989; Syfert et al., 2016). Sub-group 2 (orange) includes members of the brinjal eggplant complex (Lester and Hasan, 1991; Weese and Bohs, 2010), among which the most genetically related accessions of *S. melongena* and its wild progenitor *S. insanum*, the accession of *S. incanum*, and the two of both *S. lichtensteinii* and *S. linneanum*. Sub-group 3 (yellow) includes the five accessions of gboma eggplant (*S. macrocarpon*) and the one *S. dasyphyllum*, which is its

wild progenitor (Bukenya and Carasco, 1994; Syfert et al., 2016). Sub-group 4 (grey) includes the accessions of the New World species, which form part of the tertiary genepool of brinjal eggplant (Syfert et al., 2016). Finally, the remaining accessions of *S. campylacanthum*, *S. violaceum*, *S. lidii*, *S. vespertilio* and *S. tomentosum* had ambiguous membership and were thus classified as admixed, as their level of membership to a single group was lower than 70% (Fig 2). With the goal to provide insight into the complex relationships of the germplasm used, and to detect additional sub-population structure, a hierarchical *FastSTRUCTURE* analysis was applied by running *STRUCTURE* on partitioned data, i.e. on accessions which clustered in sub-group 1 and sub-group 4 following UPGMA analysis, as well as on the remaining materials (S2 Fig). The hierarchical *FastSTRUCTURE* analysis for the scarlet eggplant complex revealed that the optimal number of populations was obtained at  $K=2$ , and that the accessions of *S. aethiopicum* and *S. anguivi*, included in the UPGMA subgroup 1, share a common genepool. For the brinjal eggplant and gboma eggplant complexes group, the hierarchical *FastSTRUCTURE* analysis suggests that four populations are present. In this set of accessions  $K=2$  separates the brinjal eggplant *S. melongena* and its close relatives *S. insanum*, *S. incanum*, *S. lichtensteinii* and *S. linnaeanum* (Knapp et al., 2013) from the gboma eggplant *S. macrocarpon* and its wild ancestor *S. dasyphyllum* (Bukenya and Carasco, 1994) together with the Canary Islands endemisms *S. lidii* and *S. vespertilio* and the related South African *S. tomentosum* (Levin et al., 2006; Särkinen et al., Vorontsova et al., 2013), while *S. campylacanthum* and *S. violaceum* appear as an admixture (S2 Fig). At  $K=3$  the *S. lidii*, *S. vespertilio*, *S. tomentosum* and *S. violaceum* are separated from the gboma eggplant. Finally, at the optimal  $K=4$ , *S.*

*campylacanthum* accessions group separately, while *S. incanum*, *S. lichtensteinii* and *S. linnaeanum* appear as an admixture of *S. melongena*/*S. insanum* and *S. campylacanthum* (S2 Fig). This might be a consequence of gene flow among them or the result of the recent speciation from a common ancestor or both. These species are phylogenetically closely related but at present are distributed in different geographical areas (Knapp et al., 2013); this suggests that presumably they evolved from a common ancestor for adaptation to different niches, which might difficult gene flow. The hierarchical *FastSTRUCTURE* analysis of the New World species recognized at  $K=2$  two populations, one of which included *S. elaeagnifolium* while the other both *S. sysimbriifolium* and *S. torvum*. However, at the optimal  $K=3$ , the latter was further splitted in two genetically differentiated gene pools, each including one of the two species.



**Figure 2.** The genetic architecture of the full germplasm panel: Consensus UPGMA dendrogram and *FastSTRUCTURE* output at  $K = 4$ . Bootstrap values (%) for the main nodes are reported in red. Each entry is represented by a vertical line representing sub-group 1 (blue) sub-group 2 (orange), sub-group 3 (yellow) and sub-group 4 (grey). The box reports the Probabilities (p) plots derived from the *FastSTRUCTURE* analysis of genotypic data with  $K$  values from 1 to 9.

## PCoA analyses

The whole data set was also subjected to PCoA analysis (Fig 3) which, on the whole, confirmed the grouping of genotypes based on UPGMA and *FastSTRUCTURE* (Raj et al., 2014) clustering. Because a limited number of samples of each of the wild relatives was included in our study, the PCoA analysis did not allow to highlight the within-species diversity as it did in the cultivated species; however, it made possible some additional inferences. The first two axes explained 71.4% of the genetic variation. The first axis, explaining 57.6% of the genetic variation, clearly separated cultivated scarlet eggplant *S. aethiopicum* and its wild ancestor *S. anguivi* from all the other accessions, with no evident separate clustering of the two species. The latter are fully inter-fertile (Lester and Niakan, 1986, 1989) and the identification of intermediate forms (Lester and Jaeger, 1990; Plazas et al., 2016) suggests occurrence of genetic flow between them.

The second axis, explaining 13.8% of the genetic variation, clearly split the entries of *S. sisymbriifolium*, *S. torvum* and *S. elaeagnifolium*, which clustered in the previously described group 4, from the ones of sub-clusters 2 and 3 as well as the entries classified as admixed, i.e. brinjal and gboma eggplants, their respective progenitors *S. insanum* and *S. dasyphyllum* together with other Old World wild species, as well as the entries classified as admixed.

Both *S. sisymbriifolium* and *S. torvum* are native of South and Central America and, together with *S. viarum*, were classified in GP3 by Syfert et al. (2016). They have been also reported to be, within subgenus *Leptostemonum* (Dun.) Bitt., phylogenetically far away from the cultivated



eggplants and the other Old World species (Levin et al., 2006; Särkinen et al., 2013; Vorontsova et al., 2013; Syfert et al., 2016). *Solanum elaeagnifolium* is also a New World species (Boyd et al., 1984) which was not included in the study of Syfert et al. (2016), and whose origin is attributable to GP3 on the basis of the present results.

Both *S. sisymbriifolium* and *S. torvum* are of interest for eggplant breeding, as they are tolerant or resistant to many diseases (Daunay et Hazra, 2012). Their high phylogenetic distance to cultivated eggplants is confirmed by the many ineffective attempts to hybridize them with *S. melongena* (Bletsos et al., 1998; Kumchai et al., 2013; Kouassi et al., 2016; Plazas et al., 2016). No sexual hybrids have ever been reported between *S. melongena* and *S. sisymbriifolium*, while interspecific hybrids obtained through embryo rescue of the cross *S. melongena* x *S. torvum* were highly sterile and no backcrosses have been reported to date (Rotino et al., 2014). Furthermore, although tetraploid somatic hybrids between either *S. sisymbriifolium* or *S. torvum* with *S. melongena* were obtained, they did not produce sexual offspring (Gleddie et al., 1986; Sihachakr et al., 1988).

On the basis of PCoA analysis, the cultivated species which appears genetically closer to the cultivated eggplant is gboma eggplant (*S. macrocarpon*), clustering together with *S. dasyphyllum*, which has been reported by many authors to be its wild ancestor (Lester and Hasan, 1991; Bukenya and Carasco, 1994; Mace et al., 1999; Levin et al., 2006; Vorontsova et al., 2013) (Fig 3). This seems to indicate that gboma eggplant, might be genetically closer to *S. melongena* than the cultivated scarlet eggplant (*S. aethiopicum*), which is included in section *Oliganthes* (Dunn.) Bit. (Lester and Niakan, 1986; Lester et al., 2011). However contrasting results have been reported in literature. Based on chloroplast

DNA RFLPs (Sakata et al., 1991), ISSRs (Isshiki et al., 2008), AFLPs and nrITS sequences (Meyer et al., 2012b) it was previously reported that *S. aethiopicum* is closer to *S. melongena* than *S. macrocarpon*; otherwise Sakata and Lester (Sakata and Lester, 1997), in a study based on chloroplast DNA RFLPs, and Vorontsova et al. [23] using ITS, waxy and trnT-F regions sequences obtained opposite results. Interestingly, Furini and Wunder [66] using AFLPs as well as Levin et al. (2013), Weese and Bohs (2010) and Särkinen et al. (2013) using several nuclear and plastid DNA sequences found that *S. aethiopicum* and *S. macrocarpon* were phylogenetically closer among them than to *S. melongena*. Studies based on the species inter-fertility highlighted that interspecific hybrids between *S. melongena* and *S. aethiopicum* as well as backcrosses could be easily obtained (Daunay et al., 1991; Prohens et al., 2012; Rotino et al., 2014); on the other hand, although hybrids between *S. melongena* and *S. macrocarpon* were obtained (Daunay et al., 1991; Bletsos et al., 1998; Khan et al., 2013), in most cases they were high sterile and only the backcross of a tetraploid hybrid between the two species with *S. melongena* was successful (Khan et al., 2013). The difficulty in obtaining the hybrids between these two species, despite being phylogenetically close (Sakata and Lester, 1997; Vorontsova et al., 2013), might be caused by some chromosomal rearrangement or other hybridization barriers. At last, Kouassi et al. (2016) reported that the backcrosses towards *S. melongena* of the hybrid between *S. dasyphyllum* (wild ancestor of *S. macrocarpon*) and *S. melongena* was successful. A clarification is provided by our data obtained from *FastSTRUCTURE* analysis (Fig 2) which highlights that the three cultivated species belong to clearly separate

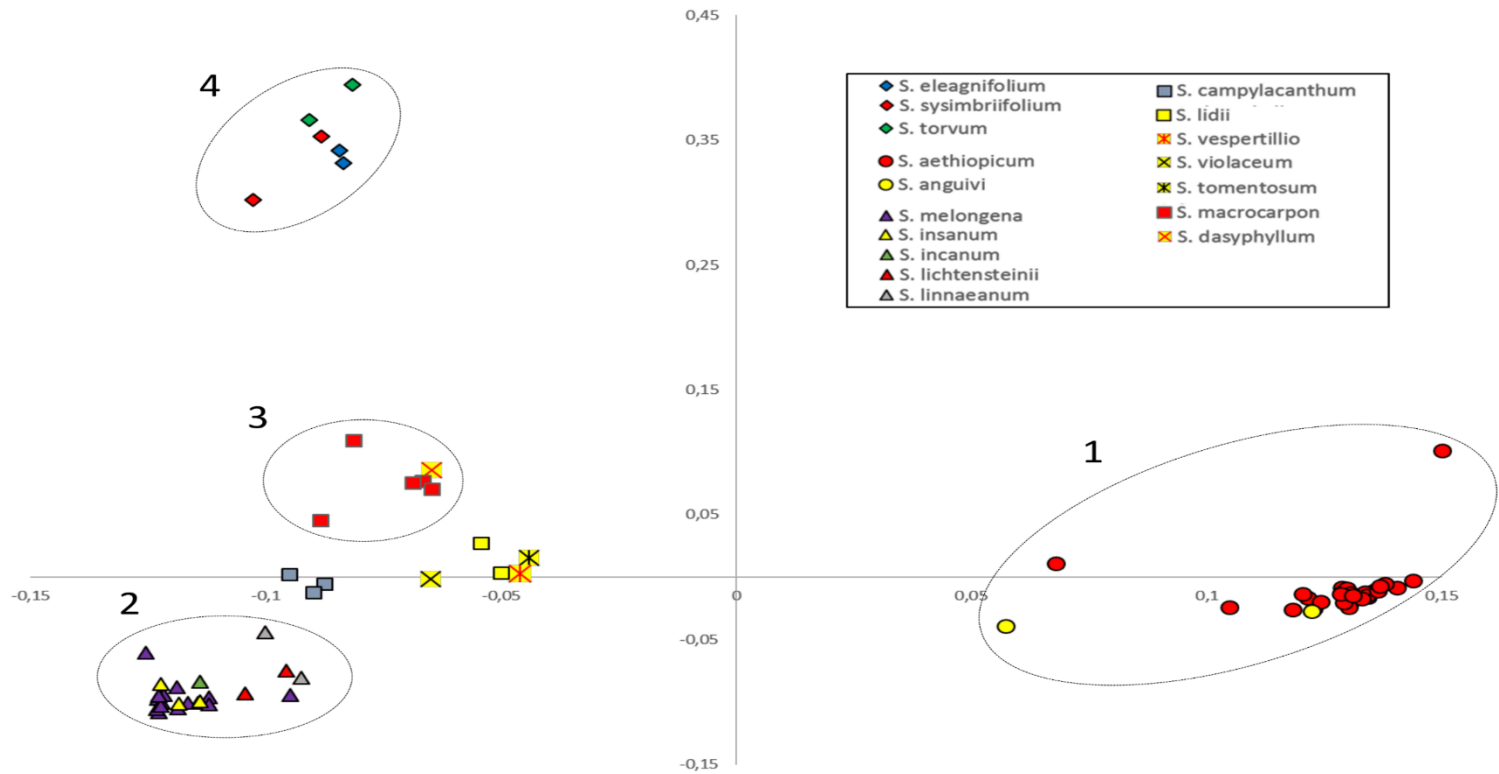
groups, suggesting that *S. macrocarpon* should be excluded from section *Melongena* (Mill.) as proposed by Sakata et al. (1991).

PCoA analysis also showed that *S. campylacanthum*, *S. incanum*, *S. insanum*, *S. lichtensteinii* and *S. linnaeanum*, which form part of the “brinjal eggplant” complex (Lester and Hasan, 1991; Weese and Bohs, 2010), cluster in proximity with eggplant (Fig 3). Among them, *S. campylacanthum* appears to be the most genetically differentiated from the others. This is in agreement with previous AFLP, nuclear and chloroplast DNA sequence results (Mace et al., 1999; Weese and Bohs, 2010; Vorontsova et al., 2013). Indeed, interspecific hybrids were obtained between *S. campylacanthum* and *S. melongena*, but the number of seeded fruits and seeds per fruit was lower in respect to the ones obtained following crosses with other species within the “common eggplant” complex (Lester and Hasan, 1991; Lester and Kang, 1998; Kouassi et al., 2016). *Solanum linnaeanum* and the accession of *S. lichtensteinii* cluster together and close to *S. melongena*. This result confirms that the two species are genetically related (Weese and Bohs, 2010; Vorontsova et al., 2013) and supports the hypothesis that *S. linnaeanum* and *S. lichtensteinii* are of South African origin and share a common ancestor, although the former grows in several tropical and subtropical areas of the world (Knapp et al., 2013; Vorontsova et al., 2013).

*Solanum linnaeanum* and *S. lichtensteinii* produce hybrids with moderate or high fertility when crossed with eggplant (Knapp et al., 2013), which can be also backcrossed with relative ease (Meyer et al., 2012b; Vorontsova et al., 2013; Rotino et al., 2014; Liu et al., 2015; Kouassi et al., 2016). However our data show that they are genetically more distant from *S. melongena* than *S. incanum* or *S. insanum* (Sakata and Lester,

1997; Furini and Wunder, 2004; Weese and Bohs, 2010; Vorontsova et al., 2013; Meyer et al., 2012b). *Solanum incanum* was suggested to be eggplant's pre-domestication ancestor and is being used in eggplant breeding programs as a source of variation for phenolics content and resistance to drought (Knapp et al., 2013). Recent morphological and molecular work has shown that species-level differences exist between *S. incanum* and *S. melongena* and, on the basis of new evidence, *S. insanum* is considered the eggplant wild progenitor (Ranil et al., 2017). The two species are also fully inter-fertile and their hybrid produce many fruits and seeds (Plazas et al., 2016). It is also not surprising that, since frequent genetic flow occurs between both species in the indo-birmanian region (Davidar et al., 2015; Mutegi et al., 2015), in our PCoA analysis the *S. insanum* accessions appear intermingled with the ones of *S. melongena*.

Our data show that the three species *S. lidii*, *S. tomentosum*, *S. vespertilio* cluster into proximity to each other and *S. violaceum* a little more apart (Fig 3). *Solanum lidii* and *S. vespertilio* are endemic to the Canary Islands (Spain) and are genetically similar sister species, which were found to cluster together in previous molecular studies (Anderson et al., 2006; Levin et al., 2006; Prohens et al., 2007a; Särkinen et al., 2013; Vorontsova et al., 2013). In several molecular studies *S. tomentosum* was also found to cluster close to *S. lidii* and *S. vespertilio* (Anderson et al., 2006; Levin et al., 2006; Särkinen et al., 2013; Vorontsova et al., 2013), thus our results confirm that the three species are close relatives. *Solanum violaceum* clusters with these three taxa in both the *FastSTRUCTURE* and PCoA analyses in spite of having a native distribution in India and Southeast Asia (Meyer et al., 2012b).



**Figure 3.** The genetic architecture of the full germplasm panel of 76 *Solanum* accessions. PCoA visualization of the genetic relationships within the full set of accessions

## Within-groups PCoA analyses

In order to gain a better landscape of the genetic relationships among the species in study, PCoA analyses were also separately performed on entries clustering in the sub-groups 1, 2 and 4, following *FastSTRUCTURE* (Raj et al., 2014) analysis (Fig 4 A, B and C). The separate PCoA of entries grouped in sub-group 1 (Fig 4A) confirmed that the different *S. aethiopicum* varietal types are partially intermingled and show a high within varietal type genetic diversity; furthermore, the absence of an evident genetic differentiation with their wild ancestor *S. anguivi* was confirmed. As observed by Sunseri et al. (2010) in a molecular characterization based on AFLP and SSR markers, the different cultivar groups of *S. aethiopicum* were intermingled in the cluster analysis. The four cultivar groups (Aculeatum, Gilo, Kumba, and Shum) are distinguished by simple morphological traits, like fruit size and shape, fruit bitterness, and the presence or absence of prickles and star leaf hairs (Lester and Niakan, 1986; Lester and Daunay, 2003), which allow the differentiation among cultivars based on morphological characterizations (Plazas et al., 2014a). However several of these traits, like prickliness and presence/absence of star leaf hair, seem to have a simple genetic basis in scarlet eggplant (Lester and Thitai; 1989) while, as occurs in common eggplant (Frery et al., 2014; Portis et al., 2014), other traits (fruit size and shape) are under control of a few major genes. The genetic flow occurring between different groups, as a result of spontaneous or artificial hybridization, may thus result in a lack of (or reduced) genetic differentiation. Indeed, in a previous study (Lester and Daunay, 2003), it was reported that the Aculeatum group seems to have been derived from

hybridization between *S. aethiopicum* group Kumba and *S. anguivi*. On the whole, the varietal groups that showed the highest genetic differentiation were Aculeatum, (characterized by the highest anthocyanin content and prickliness in respect to all the others) and Shum, which differed for the mean average values of 8 of 18 morphological traits analysed in a previous study (Plazas et al., 2014a).

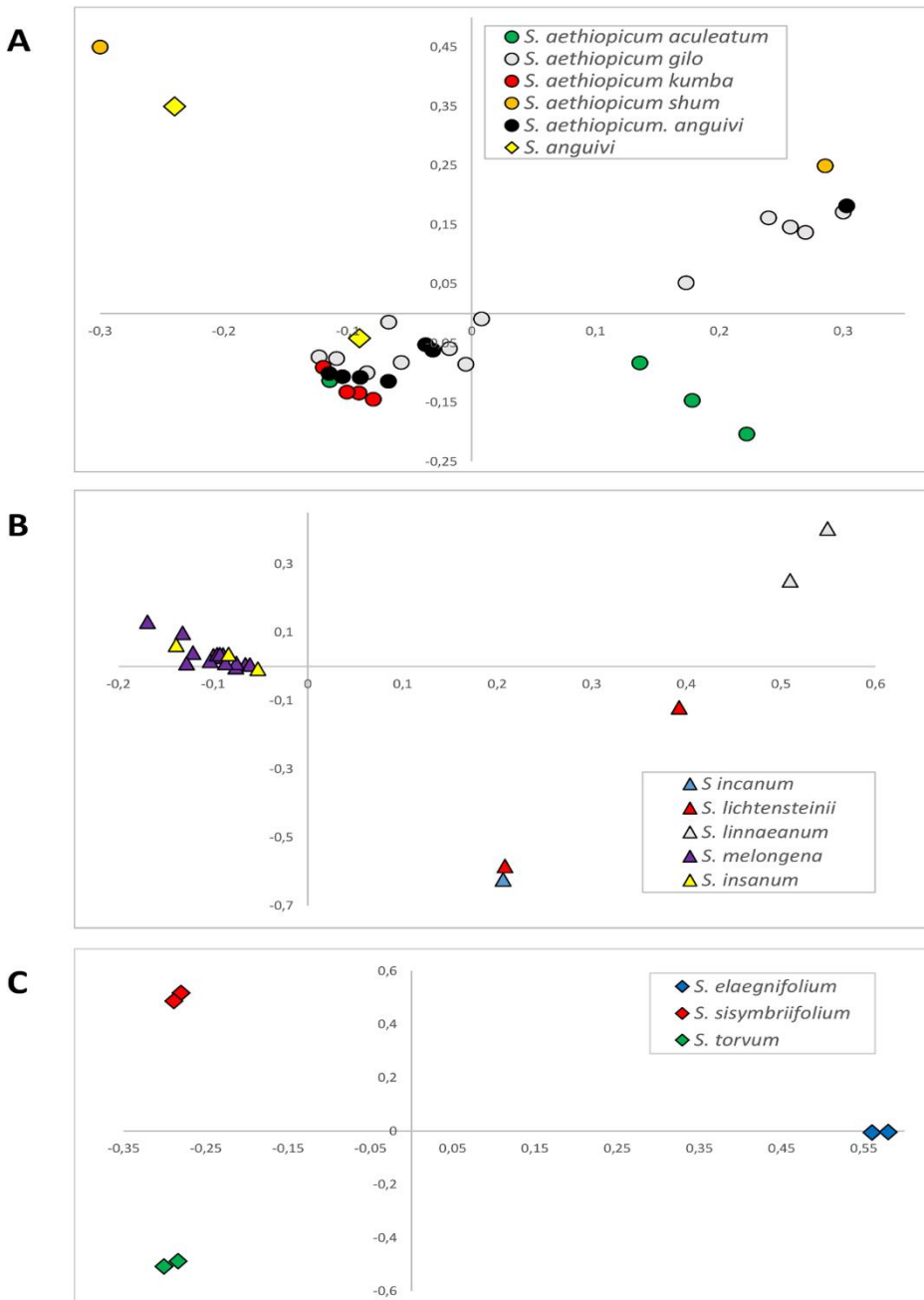
PCoA including accessions of the sub-group 2 (Fig 4B), as expected grouped separately the entries of eggplant and its wild relative *S. insanum* from the close relatives *S. incanum*, *S. lichtensteinii* and *S. linneanum*, the latter being the most genetically differentiated from all the others. The *S. melongena* accessions analysed included types, hailing from Sri Lanka, India and China as well as from Ivory Coast and Spain and producing fruits of different shape and colour. In a previous work (Cericola et al., 2013) 191 eggplant accessions were scored for a set of 19 fruit and plant traits and the analysis of phenotypic data made it possible to classify the genotypes in three main fruit morphological groups producing: (i) elongated fruits, (mean ratio  $fs = \text{fruit length}/\text{fruit maximum diameter}$  around 5.05); (ii) semi-long fruits ( $fs$  from 1.2 to 2) and (iii) round fruits ( $fs$  around 1), which cut across the Oriental and Occidental divide. On the other hand STRUCTURE (Pritchard et al., 2000) analysis based on 24 microsatellite markers (22 genomic ones and two from EST), identified two major sub-groups, which to a large extent mirrored the provenance of the entries. In the present study, in spite of the wide set of polymorphisms detected, the accessions from different origin did not highlight a grouping together trend. This apparent discrepancy can be explained by either the difference in size of the two germplasm sets, but also by the number of markers applied, as the use of a limited number of selected markers might

provide unrealistic estimates of genetic variability in the set of accessions in study.

PCoA including accessions of the sub-group 4 (Fig 4C) highlights that *S. sisymbriifolium*, *S. torvum* and *S. elaeagnifolium* are genetically far away from each other and that their grouping in the sub-group 4 is due to their common high genetic divergence from all the other entries. This is confirmed by previous molecular results that includes *S. torvum* and *S. sisymbriifolium* in different clades within subgenus *Leptostemonum* from the cultivated eggplants (Levin et al., 2006; Särkinen et al., 2013; Vorontsova et al., 2013; Syfert et al, 2016). Furthermore, on the basis of PCoA analysis, the two accessions of *S. torvum* form a group 'per se' in respect to all the others.

Previous phylogenetic studies placed *S. elaeagnifolium* and the rest of species of the *Elaeagnifolium* clade closer to Old World species than either *S. sisymbriifolium* or *S. torvum* (Anderson et al., 2006; Levin et al., 2006; Meyer et al., 2012b; Särkinen et al., 2013; Vorontsova et al., 2013). Recently, crossing data confirm that *S. elaeagnifolium* is closer to eggplant than both *S. sisymbriifolium* and *S. torvum*, as interspecific hybrids have been obtained which present intermediate fertility (Kouassi et al., 2016), and with which it is possible to obtain backcrosses with *S. melongena* (unpublished results).





**Figure 4.** Within-groups PCoA analyses in subgroups of germplasm panel of *Solanum* accessions: visualization of the genetic relationships within sub-group 1 (A; scarlet eggplant complex), sub-group 2 (B; brinjal eggplant complex) and sub-group 4 (C; New World species).

## ***Conclusions***

One of the most exciting developments in the past decade has been the application of powerful and ultra-rapid nucleic acid sequencing techniques to the study of genetic relationships and phylogeny of crop species (Patwardhan et al., 2014). As previously reported by Bajaj et al. (2015) in chickpea, our results demonstrate that the high-throughput genotyping of numerous genome-wide SNP markers represents a highly and more effective approach, in respect to the ones based on limited sets of genome-wide markers or a small set of gene sequences, for understanding the extent of natural allelic diversity and genetic relationships among and within wild and cultivated species belonging to eggplant complexes. The high number of detected polymorphisms were analysed by *FastSTRUCTURE* (Emanuelli et al., 2013; Raj et al., 2014), UPGMA and PCoA analyses and the three approaches showed to be complementary in the interpretation of data. On the whole, we confirm a wide genetic base and broad molecular diversity among wild and cultivated species within and among the three cultivated eggplant complexes and the New World eggplant CWRs. Thanks to a reduced complexity genome sequencing approach, we were able to fingerprint all accessions in the study and gathered information which may efficiently guide further exploration of the diversity and relationships in the large *Solanum* subgenus *Leptostemonum* group. The approach used and data obtained lay the foundation also to address the evaluation of gene flow among inter-fertile sympatric taxa (Davidar et al., 2015), recent speciation and domestication processes of cultivated eggplants. In addition, the large number of markers distributed across the genome may also contribute to facilitate the transfer of target genomic regions controlling useful

agronomic traits, such as biotic and abiotic stress tolerance or fruit quality traits, from related species into the genetic background of cultivated eggplants.

## *Acknowledgements*

This work has been funded in part by European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) and by Spanish Ministerio de Economía, Industria y Competitividad and Fondo Europeo de Desarrollo Regional (grant AGL2015-64755-R from MINECO/FEDER). Funding has also been received from the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives", which is supported by the Government of Norway. This last project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew and implemented in partnership with national and international gene banks and plant breeding institutes around the world. For further information see the project website: <http://www.cwrdiversity.org/>. Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral (Programa FPI de la UPV-Subprograma 1/2013 call) contract. Mariola Plazas is grateful to Spanish Ministerio de Economía, Industria y Competitividad for a post-doctoral grant within the Santiago Grisolíá Programme (FCJI-2015-24835). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

***Additional files***

**S1 Table.** SNPs detected in the genome and in CDSs. In both cases, number and percentage of: (i) missing sites; (ii) detected SNPs , the percentage is evaluated as the ratio between detected SNPs/Genomic or CDS total SNPs-missing sites; (iii) heterozygous SNPs , the percentage is evaluated as the ratio between heterozygous SNPs/ Genomic or CDS total SNPs-missing sites; (iv) homozygous SNPs , the percentage is evaluated as the ratio between homoyigous SNPs/ Genomic or CDS total SNPs-missing sites.

<https://doi.org/10.1371/journal.pone.0180774.s001>

**S1 Fig.** Distribution of sequenced reads, after quality cleaning and trimming procedures, across a germplasm panel of 76 *Solanum* accessions (in million reads).

<https://doi.org/10.1371/journal.pone.0180774.s002>

**S2 Fig.** *FastSTRUCTURE* output at K=4 from full germplasm panel together with outputs of separate analyses performed with subsets of taxa. Asterisks indicate the best K choice based on the  $\Delta K$  method.

<https://doi.org/10.1371/journal.pone.0180774.s003>



## **Chapter 4: Location of the chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant**

Gramazio P<sup>1</sup>, Prohens J<sup>1\*</sup>, Plazas M<sup>1</sup>, Andújar I<sup>1</sup>, Herraiz FJ<sup>1</sup>, Castillo E<sup>1</sup>, Knapp S<sup>2</sup>, Meyer RS<sup>3</sup>, Vilanova S<sup>1</sup>

<sup>1</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

<sup>2</sup> Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom.

<sup>3</sup> Center for Genomics and Systems Biology, New York University, 12 Waverly Place, New York, USA.

\* Corresponding author: [jprohens@btc.upv.es](mailto:jprohens@btc.upv.es)

**Citation:** Gramazio P, Prohens J, Plazas M, Andújar I, Herraiz FJ, Castillo E, Knapp S, Meyer RS, Vilanova S (2014). Location of the chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. *BMC Plant Biology* 14:350.

<https://doi.org/10.1186/s12870-014-0350-z>

## ***Abstract***

**Background:** Eggplant is a powerful source of polyphenols which seems to play a key role in the prevention of several human diseases, such as cancer and diabetes. Chlorogenic acid is the polyphenol most present in eggplant, comprising between the 70% and 90% of the total polyphenol content. Introduction of the high chlorogenic acid content of wild relatives, such as *S. incanum*, into eggplant varieties will be of great interest. A potential side effect of the increased level polyphenols could be a decrease on apparent quality due to browning caused by the polyphenol oxidase enzymes mediated oxidation of polyphenols. We report the development of a new interspecific *S. melongena* × *S. incanum* linkage map based on a first backcross generation (BC1) towards the cultivated *S. melongena* as a tool for introgressing *S. incanum* alleles involved in the biosynthesis of chlorogenic acid in the genetic background of *S. melongena*.

**Results:** The interspecific genetic linkage map of eggplant developed in this work anchor the most informative previously published genetic maps of eggplant using common markers. The 91 BC1 plants of the mapping population were genotyped with 42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs and one morphological polymorphic markers. Segregation marker data resulted in a map encompassing 1085 cM distributed in 12 linkage groups. Based on the synteny with tomato, the candidate genes involved in the core chlorogenic acid synthesis pathway in eggplant (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*) as well as five polyphenol oxidase (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*) were mapped. Except for *4CL* and *HCT* chlorogenic acid genes were not linked. On the contrary, all *PPO* genes clustered



together. Candidate genes important in domestication such as fruit shape (*OVATE*, *SISUNI*) and prickliness were also located.

**Conclusions:** The achievements in location of candidate genes will allow the search of favorable alleles employing marker-assisted selection in order to develop new varieties with higher chlorogenic content alongside a lower polyphenol oxidase activity. This will result into an enhanced product showing a lower fruit flesh browning with improved human health properties.

**Keywords:** chlorogenic acid, genetic map, polyphenol oxidases, *Solanum incanum*, *Solanum melongena*, synteny

**Abbreviations:**

HCA: hydroxycinnamic acid

CGA: chlorogenic acid

PPOs: polyphenol oxidase enzymes

SOL: *Solanum* orthologous gene sets

HRM: high resolution melting

*PAL*: phenylalanine ammonia lyase

*C4H*: cinnamate 4-hydroxylase

*4CL*: 4-hydroxycinnamoyl-CoA ligase

*HCT*: hydroxycinnamoyl-coA shikimate/quinic acid hydroxycinnamoyl transferase

*C3'H*: *p*-coumaroyl ester 3'-hydroxylase

*HQT*: hydroxycinnamoyl CoA quinic acid hydroxycinnamoyl transferase

## ***Introduction***

Eggplant (*Solanum melongena* L., Solanaceae;  $2n = 2x = 24$ ) ranks third in the genus *Solanum*, after potato and tomato, in total production and economic importance and is the most important Solanaceae crop native to the Old World (Daunay and Lester, 2001). The most nutritionally important bioactive constituents of the eggplant fruit are phenolics, which are responsible of the high antioxidant activity of eggplant (Hanson et al., 2006; Lo Scalzo et al., 2010; Mennella et al., 2012; Plazas et al., 2013a, 2013b). The most abundant phenolics of eggplant are hydroxycinnamic acid (HCA) conjugates, which are synthesized by converting phenylalanine to cinnamic acid. Among HCA conjugates, chlorogenic acid (5-*O*-caffeoyl-quinic acid; CGA) constitutes between 70% to over 95% of the total phenolics content (Whitaker and Stommel, 2003; Stommel and Whitaker, 2003; Singh et al., 2009; Prohens et al., 2013). Growing interest in this compound is due to its many beneficial properties for the treatment for various metabolic and cardiovascular diseases and ailments. Several *in vitro* and *in vivo* experiments have shown that CGA has anti-oxidant, anti-inflammatory, analgesic, antipyretic, neuroprotective, cardioprotective, anti-carcinogenic, anti-microbial, hypotensive, anti-obesity and anti-diabetic activity (Plazas et al., 2013b; Dos Santos et al., 2006; Cho et al., 2010). Moreover, CGA is highly stable at high temperatures, and its bioavailability in eggplant increases, as compared to the raw product, after cooking (Lo Scalzo et al., 2010).

Great diversity in the content of total phenolics and CGA has been observed in eggplant, due both to genetic and environmental factors (Whitaker and Stommel, 2003; Stommel and Whitaker, 2003; Hanson et

al., 2006; Luthria et al., 2010; Mennella et al., 2012; Prohens et al., 2013). Some close wild relatives of cultivated eggplant, such as *S. incanum* (Lester and Hasan, 1991; Knapp et al., 2013), have high levels of CGA (Stommel and Whitaker, 2003; Ma et al., 2010; Prohens et al., 2013). *Solanum incanum* is native to northern Africa and the Middle East to Pakistan (Knapp et al., 2013), and is a cross-compatible with *S. melongena* (Daunay and Lester, 2001; Prohens et al., 2013). Therefore, *S. incanum* shows promise for use in breeding programs for developing new eggplant varieties with increased phenolic content (Plazas et al., 2013b).

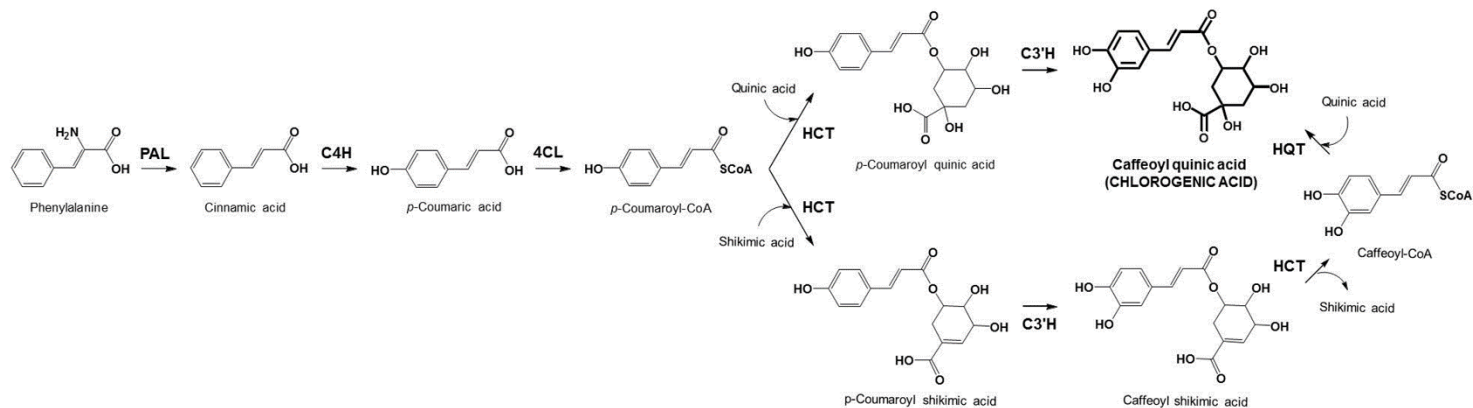
Raising the total phenolics content, however, may cause a negative effect on apparent quality of the fruit. When eggplant fruit flesh is cut, phenolics, mostly stored in vacuoles, become available to polyphenol oxidase enzymes (PPOs), which are present in chloroplasts. PPOs catalyse the oxidation of phenolics to quinones, which in turn, react non-enzymatically with oxygen in the air to give brown compounds, thus causing browning of fruit flesh (Ramirez et al., 2003). Several authors have found differences in PPO activity between varieties of eggplant, which can lead to differences in the degree of browning in fruit flesh between varieties with similar content of total phenolics (Doğan et al., 2002; Shetty et al., 2011; Concellón et al., 2012). Molecular breeding for high CGA content and low PPO activity could contribute to developing improved cultivars with higher bioactive properties through a combination of high antioxidant activity and presenting a low degree of browning. For this purpose, a candidate gene approach shows promise, given that the genes involved in the CGA synthesis pathway, which include *phenylalanine ammonia lyase*, *PAL*; *cinnamate 4-hydroxylase*, *C4H*; *4-hydroxycinnamoyl-CoA ligase*, *4CL*; *hydroxycinnamoyl-coA*

*shikimate/quinic acid hydroxycinnamoyl transferase, HCT*; *p-coumaroyl ester 3'-hydroxylase, C3'H*; and, *hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase, HQT*, (Figure 1), in addition to the PPO genes, are known (Niggeweg et al., 2004; Comino et al., 2007, 2009; Menin et al., 2010; Shetty et al., 2011).

Understanding of eggplant genome organization, which is of great relevance for molecular breeding, has lagged behind that for other solanaceous crops such as potato, tomato and pepper. Several linkage maps for eggplant have been developed. Nunome et al. (2001) developed a first intraspecific linkage map in eggplant using RAPD and AFLP markers. Two improved versions of the Nunome et al. (2001) map were developed by adding SSR markers (Nunome et al., 2003a, 2009). Doganlar et al. (2002) developed the first interspecific map using RFLP markers resulting from crossing *S. melongena* and *S. linnaeanum*. The resolution of this map was further improved by adding COSII and AFLP markers (Wu et al., 2009; Doganlar et al., 2014). Barchi et al. (2010) also developed an intraspecific mostly AFLP and SSR marker. Finally, an intraspecific saturated integrated map of *S. melongena* was developed by Fukuoka et al. (2012) from two F<sub>2</sub> populations in which SSR and SNP markers were mapped. Of the markers used by Fukuoka et al. (2012), many were obtained from *Solanum* orthologous (SOL) gene sets from a multiple alignment between the unigenes of eggplant, tomato and potato.

Here we report the development of a new interspecific *S. melongena* × *S. incanum* linkage map with the aim of locating, and in the future introgressing, *S. incanum* alleles involved in the biosynthesis of CGA in the genetic background of *S. melongena*. In order to devise molecular tools for minimizing browning associated with high CGA

levels, PPO genes were also targeted. This new map is anchored to the tomato genetic map and previous eggplant maps, which will facilitate molecular breeding in eggplant for high CGA content and reduced browning as well as other morphological traits of importance in eggplant breeding.

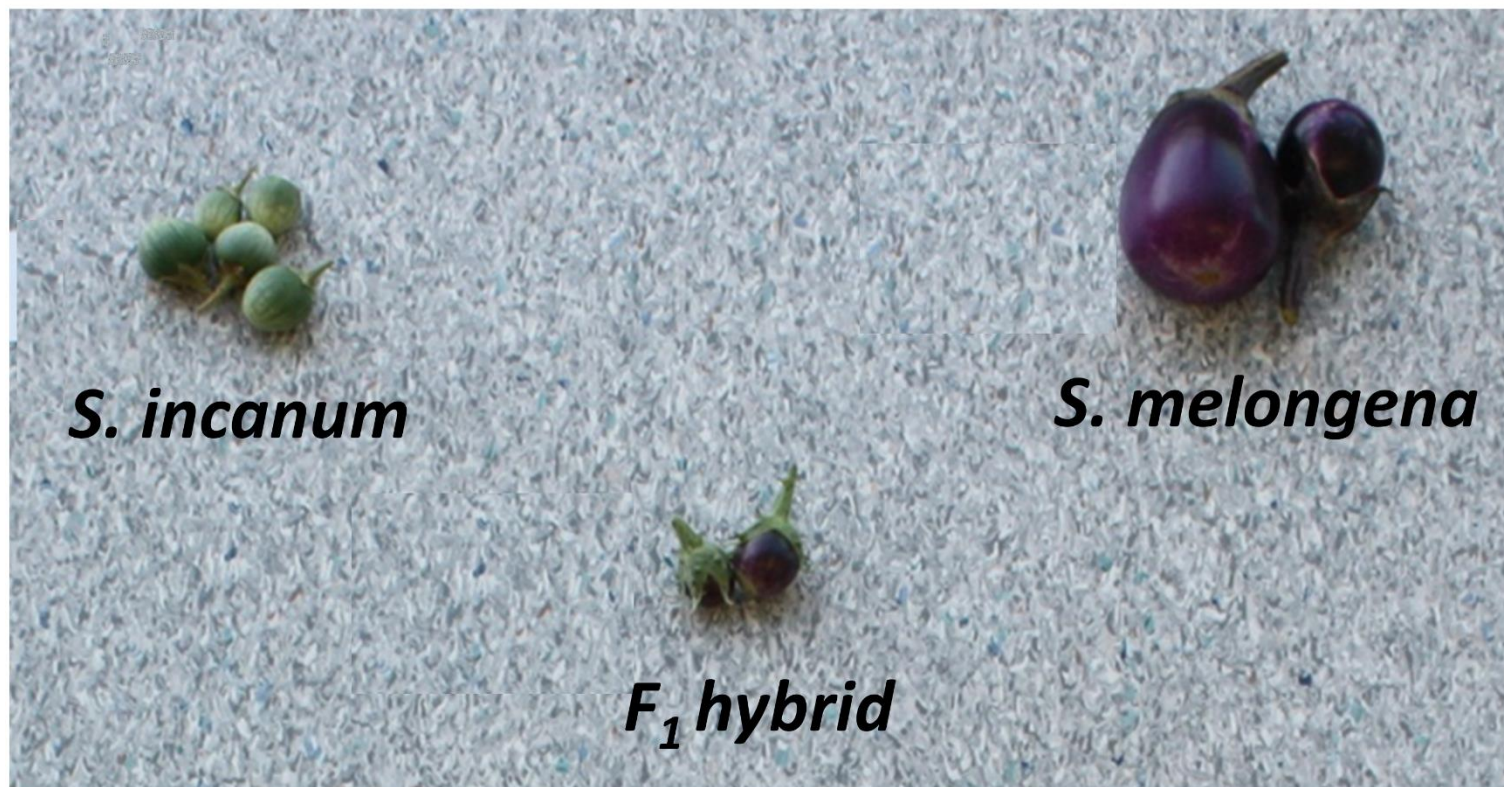


**Figure 1.** Biochemical pathway for chlorogenic acid (CGA) synthesis in eggplant. Enzymes involved in the CGA pathway are indicated: *PAL*, phenylalanine ammonia lyase; *C4H*, cinnamate 4-hydroxylase; *4CL*, 4-hydroxycinnamoyl-CoA ligase; *HCT*, hydroxycinnamoyl-coA shikimate/quinic hydroxycinnamoyl transferase; *C3'H*, p-coumaroyl ester 3'-hydroxylase; *HQT*, hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase (Comino et al., 2007; 2009; Menin et al., 2010; Niggeweg et al., 2010).

## ***Material and methods***

### **Plant materials**

Parentals for developing the interspecific genetic linkage map (denominated SMIBC, making reference to *S. melongena* and *S. incanum* [SMI] and backcross [BC] mapping population) were *S. melongena* accession AN-S-26 and *S. incanum* accession MM577 (Figure 2). The F<sub>1</sub> hybrid was obtained using AN-S-26 as a female parent and MM577 as male parent. After that, the F<sub>1</sub> hybrid was backcrossed (as a female) to the *S. melongena* AN-S-26 parental to obtain a BC<sub>1</sub> population. BC<sub>1</sub> seeds were germinated and 91 BC<sub>1</sub> plants were randomly selected for the development of the genetic linkage map.



**Figure 2.** Parents and F1 hybrid used to generate the BC1 mapping population used to develop the SMIBC genetic map.



## Development of the SMIBC genetic map

### *DNA extraction*

Total genomic DNA was isolated from leaves of the parental, F1 and BC1 plants according to the CTAB procedure (Doyle and Doyle, 1987). DNA quality was checked in 1% agarose gels. DNA concentrations were measured with a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA).

### *COSII markers*

A total of 123 Universal Primers of COSII markers (Wu et al., 2006), previously mapped in an interspecific cross between *S. linnaeanum* and *S. melongena* (Wu et al., 2009), were screened in parental and F1 plants, and in the BC1 mapping population. Amplicons were analyzed on an ABI PRISM 3100-Avant DNA genetic analyzer (Applied Biosystems, Foster City, California). Amplicon sequences were examined with Sequence Scanner v1.0 (Applied Biosystems) software and aligned using the CAPS Designer software (<http://solgenomics.net/>) in order to find polymorphisms between the parental species. Restriction enzymes were used to cut in the polymorphic regions using the CAPS method (Cleaved Amplified Polymorphic Sequence). Alternatively, the High Resolution Melting (HRM) technique on a LightCycler 480 Real-Time PCR (Roche Diagnostics, Meylan, France) was used.

## *SSR markers*

Three sources of SSR markers were used: a) EST-SSRs developed ‘*in silico*’ specifically for this study, b) genomic SSRs developed by Nunome et al. (2009), and c) genomic SSRs developed using a genomic library (Vilanova et al., 2012). Development of new SSR markers *in silico* was performed using 16,000 eggplant unigene sequences obtained from the VegMarks database (<http://vegmarks.nivot.affrc.go.jp/>).

The identification of perfect di- and tri-nucleotide motifs with a repeat of  $\geq 6$  times were identified using the software SciRoKo (Kofler et al., 2007). This allowed us the identification of 254 ESTs-SSRs, which were subjected to a BLASTN search of the SGN Cornell unigene database (Bombarely et al., 2011) (<http://solgenomics.net/>) against the tomato genome database. SSRs were validated in a 2-3% agarose gel when the differences in the bands sizes bands allowed polymorphisms in samples to be distinguished by eye. When this was not possible, the detection of polymorphisms was carried out by: a) capillary electrophoresis in polyacrylamide gel on an LI-COR 4300 DNA Analysis System combining fragment size and fluorophores IRD700 and IRD800 in a multiplex reaction, or b) by labeling forward primers with the fluorochromes FAM, VIC, NED, and PET. PCR products were diluted in formamide and analyzed on an automated DNA sequencer ABI PRISM 3100-Avant with a GeneScan-600LIZ (Applied Biosystems) size standard. The data were analyzed using the GeneScan software to obtain the electropherograms and polymorphisms were analyzed with Genotyper DNA Fragment Analysis software.

*AFLP markers*

Genomic DNA was digested with a two enzymes combination, EcoRI and MseI (Vos et al., 1995). Then EcoRI and MseI adapters were ligated using an AFLP Core Reagent (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. A total of 12 AFLP primer combinations, generated by four EcoRI primers (EcoAGC, EcoACT, EcoACA, EcoACG) combined with three MseI primers (MseCAA, MseCTA, MseCAC) were used. A pre-amplification PCR reaction was executed with primers based on the adapter sequences with one additional selective nucleotide at the 3' end of each primer (EcoA + MseC). Using an aliquot of the PCR product, a selective amplification reaction was performed with Eco and Mse primers combination with three additional selective nucleotides at the 3' end of each primer. Each 5' end of the EcoRI primers was labelled with different fluorescent dyes (PET, FAM, NED, and VIC). AFLP fragment analysis was performed on an automatic capillary eletrophoresis sequencer ABI PRISM 3100-Avant (Applied Biosystems, Foster City, CA). The data were analysed using GeneScan (Applied Biosystems) and Genotyper DNA Fragment Analysis (Applied Biosystems) software as above. Only AFLP bands that were present in in *S. incanum* and absent in *S. melongena* parental lines were scored.

*Sequence search for candidate genes and in silico comparison*

Orthologous sequences of candidate genes involved in CGA synthesis pathway were obtained from tomato and potato, by BLASTN searching in the SGN database. These sequences were used to find

orthologous unigenes in an eggplant local unigene database, developed with the BioEdit software (Hall, 1999) (<http://www.mbio.ncsu.edu/bioedit/>). This local database contains the 16,245 unigenes published by Fukuoka et al. (2010), available in the VegMarks webpage (<http://vegmarks.nivot.affrc.go.jp>). In addition, an eggplant local contigs database was made employing contigs developed by Barchi et al. (2011). The sequences of PPO genes published by Shetty et al. (2011) were obtained from the NCBI database (<http://www.ncbi.nlm.nih.gov/>).

Intron/exon structure of the unigenes was detected by comparison with tomato sequences. Subsequently primers were designed with Primer3 software (Rozen and Skaletsky, 1999) to amplify preferentially introns in order to have more probabilities to find polymorphism. The amplification of genes was done according to the following protocol: denaturalization at 94° C for 5 min, 35 cycles at 94° C for 30 s, annealing at 55° C for 1 min, extension at 72° C for 2 min, and a final extension at 72° C for 10 min. In case of nonspecific amplifications, additional tests with higher annealing temperatures and/or lower MgCl<sub>2</sub> concentrations were carried out.

The amplification products were purified and sequenced with an automatic ABI PRISM3100-Avant sequencer. Results were analysed with Sequence Scanner v1.0 (Applied Biosystems) software. The program Blast2Seq (NCBI) was used to compare the parental sequences to detect SNPs. SNPs found were transformed in CAPS markers with CAPS Designer software. In the cases in which no enzymes were available, primers to detect SNPs with High Resolution Melting (HRM) were developed. Synteny was studied between the SMIBC and the Tomato-EXPEN 2000 map (Fulton et al., 2002).

### *Additional markers*

The sequences and markers developed for *SISUNI* and *OVATE*, which are candidate genes involved in fruit shape (Huang et al., 2013), were obtained as described above. The morphological marker *PRICKLINESS* was developed by phenotyping the plants and taking data on the presence or absence of stem prickles.

### **Linkage analysis and map construction**

The mapping population was genotyped and  $\chi^2$  test ( $\chi^2$  value  $\leq \chi^2$   $\alpha=0.05$ ) were performed to check if individual markers segregated following Mendelian ratios. Linkage analysis was carried out using Joinmap v4.0 software (Van Ooijen, 2006) to construct the map with settings LOD threshold of 3.0 and maximum recombination fraction  $\theta=0.4$ . Kosambi mapping function (Kosambi et al., 1944) was used to convert recombination units into genetic distance (cM).

### **Macro-synteny between SMIBC and other genetic linkage maps**

The synteny of SMIBC map with other eggplant genetic linkage maps was established using shared markers (Nunome et al., 2009; Wu et al., 2009; Barchi et al., 2012; Fukuoka et al., 2012). The macro-synteny of SMIBC with the Tomato EXPEN-2000 (Fulton et al., 2002) was carried out mainly using COSII markers developed by Wu et al. (2006). Some SMIBC SSR markers were positioned based on reciprocal best-hit relationships using BLASTN search against the “Tomato gene models

CDS (ITAG release 2.30)” sequence database. Synteny was visualized using the Circos software (Krzywinski et al., 2009).

## **Results**

### **Genetic map construction**

The mapping population was genotyped with 243 molecular markers comprising 42 COSII, 99 SSRs, 88 AFLPs, nine CAPS, four SNPs and the morphological marker *PRICKLINESS*. Genotypic data generated a genetic linkage map that spans 1085 cM distributed in 12 major and three minor linkage groups (Figure 5). Synteny with maps of Wu et al. (2009), Fukuoka et al. (2012), Barchi et al. (2012) and Tomato-EXPEN 2000 (Fulton et al., 2002) anchored the three minor linkage group to the corresponding major linkage groups (E05, E10, E11). The linkage groups ranged in length between 58.6 cM (E05) and 132.9 cM (E01) (Table 1). The average genome-wide density was 4.46 cM, with linkage group E01 having the lowest average density (5.77 cM inter-locus separation), and E08 showing the highest density (3.19 cM inter-locus separation). The number of loci per linkage group was highest in E06 (27) and lowest in E04 (16). Segregation distortion was observed for 22.6% of the markers (Table 1). The linkage groups with greater distortion were E02, E03 and E09 with around half of their markers skewed. A clear distortion in favour of *S. incanum* was found in E03 and E09 whereas in E02 and E06, alleles for *S. melongena* were more abundant. In order to develop a strong framework map, only markers joined at  $\text{LOD} > 3$  were selected and those that had a lower LOD were discarded to avoid errors in positioning.

**Table 1.** Statistics of framework of SMIBC map. The table shows the length in cM, the number of markers of each type, the average density in cM, the gaps larger than 15 cM and the skewed segregation for each linkage group. Linkage groups are designated as E01- E12

Statistics	Linkage groups												
	E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	E12	Total
<b>Length (cM)</b>	132.9	78.7	94.3	76	58.6	111.9	101.7	78.3	96.1	96.7	79.7	80.1	1085.0
<b>Number of markers</b>	23	18	21	16	18	27	21	20	22	19	19	19	243
<b>COSII</b>	5	7	8	2	2	4	3	1	1	5	3	1	42
<b>SSRs</b>	8	7	4	4	7	13	8	9	12	6	9	12	99
<b>AFLPs</b>	9	3	7	10	9	8	9	5	8	7	7	6	88
<b>CAPS</b>	1	1	1	0	0	0	0	4	1	1	0	0	9
<b>SNPs</b>	0	0	1	0	0	1	1	1	0	0	0	0	4
<b>Morphological</b>	0	0	0	0	0	1	0	0	0	0	0	0	1
<b>Average density (cM)</b>	5.77	4.37	4.49	4.75	3.25	4.14	4.84	3.19	4.36	5.09	4.19	4.21	4.46
<b>Gaps (&gt;15 cM)</b>	2	0	1	2	1	1	0	1	2	2	1	3	16
<b>Skewed markers (<math>P &lt; 0,05</math>)</b>	0	9	11	0	3	5	2	1	11	7	2	4	55
<b>Percentage skewed markers</b>	0	50.0	52.3	0	16.6	18.5	9.5	5.0	50.0	36.8	10.5	21.0	22.6



*COSII analysis*

A total of 35 (28.5%) out of the 123 COSII developed by Wu et al. (2009) were polymorphic in our mapping population (Table 1). Seven other COSII markers were identified comparing the sequences of SSRs mapped with the tomato genome database (Sol Genomic Network). Six of them were EST-SSRs, obtained *in silico*, and one was a genomic SSR marker (Table 2). COSII markers allowed us to establish synteny with Wu et al. [29] eggplant map and with the Tomato EXPEN-2000 map (Fulton et al., 2002) (Figures 3 and 4, Additional file 1: Figure S1) as well as with other members of the Asterid clade (Bremer et al., 2003).

**Table 2.** COSII markers identified in the SMIBC genetic map from CSM and EST-SSR markers based on sequence homology found after a BLASTN search on the SGN Cornell marker database.

SMIBC LG	Current marker name	Previous marker name	Type of SSR
E03	C2_At3g17970	CSM44	genomic SSR
E04	C2_At1g09920	SmFL32K22A	EST-SSR
E05	C2_At3g24490	LS502D19A	EST-SSR
E06	C2_At5g61910	PLA03P07F	EST-SSR
E06	C2_At5g48630	ROT01I04F	EST-SSR
E08	C2_At5g45680	SmFL04A12A	EST-SSR
E12	C2_At4g32930	LS004H02A	EST-SSR

### *SSRs analysis*

A total of 99 SSR markers of different sources were mapped in the BC1 population (Table 1). One hundred and twenty-eight of the 254 EST-SSRs, obtained *in silico* using eggplant unigenes from the VegMarks database, showed homology with tomato unigenes (Sol Genomics Network). Of these, forty-seven were selected based on the theoretical position and screened for segregation in our parental (*S. incanum* and *S. melongena*) plants. Twenty of them (42.5%) showed polymorphism and were mapped. In addition, 71 genomic SSR markers from Nunome et al. (2009), with an average of 4-5 per linkage group, were tested. Of these, 53 (74.6%) showed polymorphism and were positioned on the map. Most of them were also used by Fukuoka et al. (2012) in the LWA2010 genetic integrated map, allowing the comparison between the two maps (Figures 3 and 4). Finally, 33 genomic SSRs (CSM markers) developed by Vilanova et al. (2012) could be mapped. Seven of them were also used in Barchi et al. (2012) in their genetic linkage map, enabling us to establish synteny (Figures 3 and 4).

### *AFLPs analysis*

A total of 116 AFLP polymorphic bands were produced from 12 AFLP primer combination combined with three MseI primers. A total of 88 AFLP markers were identified and mapped (Table 1), with an average of 9.6 polymorphic bands per primer pair. Scoring only AFLP bands present in *S. incanum* and absent in *S. melongena* in a backcrossing

population where *S. incanum* is a donor parent had the advantage that AFLPs could be scored in a codominant manner.

### **Mapping of CGA pathway genes**

The six genes (Figure 1) involved in the core CGA synthesis pathway (Niggeweg et al., 2004; Comino et al., 2007, 2009; Menin et al., 2010) were amplified and positioned in the SMIBC genetic map (Figure 5) based on the syntenic position with the Tomato EXPEN-2000 genetic linkage map (Fulton et al., 2002) (Additional files 1,2: Figure S1 Table S1).

#### *PAL (phenylalanine ammonia lyase)*

The search for phenylalanine ammonia lyase in SOL database give several orthologous gene sequence in tomato clustered in the same region of chromosome nine. A Blast of these orthologs in our local database allow to identify eggplant unigene OVS02A18A obtained by Fukuoka et al. (2010). A reciprocal Blast of OVS02A18A in SOL database shows the high homology with [SGN:Solyc09g007890.1.1] whose length is approximately 2.3 Kb and consist of two exons and one intron. In *S. melongena* and *S. incanum* parents a SNP (T/C) was found, after the amplification and sequencing of intron region, which was validated by restriction enzyme using a CAPS method (Additional file 3: Table S2). After genotyping the BC1 population the gene was mapped into the linkage group E09 at 15.2 cM from the linkage group end (Figure 5). The tomato orthologous gene is also positioned in the upper part on

chromosome 9, between markers [SGN:CLED-9-D21] (15.0 cM) and [SGN:C2\_At2g37025] (15.3 cM) and (Figure 4).

#### *C4H (cinnamate 4-hydroxylase)*

A search in the SOL database for the *C4H* gene in tomato yielded no results, however a *C4H* ortholog was found in potato [SGN:PGSC0003DMG402030469]. This gene in potato is approximately 3.7 kb and comprises three exons and two introns. Using this sequence it was possible to find the eggplant unigene SmFL27M04A and develop primers that amplify a region comprising the first intron. The amplicon in parents was sequenced and a SNP (A/G) was located and validated with high resolution melting (HRM) technique (Additional file 3: Table S2). The gene could be mapped on the bottom of linkage group E06 at 108.5 cM from the linkage group end (Figure 5). Synteny analysis reveals high co-linearity between linkage group 6 of eggplant and tomato. After a BLAST search using eggplant unigene SmFL27M04A, we found the tomato ortholog on linkage group 6 that corresponded to unigene [SGN-U590064]. Although this tomato gene, positioned in the Tomato EXPEN-2000 map between markers C2\_At3g51630 (92.5 cM) and T1789 (95.0 cM) (Figure 3), was not annotated as *C4H*, it is certainly orthologous to *C4H*.

#### *4CL (4-hydroxycinnamoyl-CoA ligase)*

Using a tomato *4CL* sequence [SGN:Solyc03g117870.2], which was 3.6 Kb with five exons and four introns, it was possible to find the eggplant unigene SmFL38N19A. Analysis of the amplified sequences of

introns three and four, allowed us to detect a polymorphism (A/G) that was transformed in a CAP marker (Additional file 3: Table S2). The gene could be mapped in the lower part of linkage group E03 at 90.9 cM from the linkage group end (Figure 5). The orthologous gene in tomato is positioned in linkage group 3 between COSII markers [SGN:C2\_At1g09760] (133.3 cM) and [SGN:C2\_At1g16180] (133.5 cM) (Figure 3).

*HCT (hydroxycinnamoyl-coA shikimate/quininate hydroxycinnamoyl transferase)*

The eggplant unigene ROT01O23W was identified using a tomato ortholog [SGN:Solyc03g117600.2]. The gene in tomato (5.3 Kb) consists of three exons and two introns. The first intron was located in the 5' UTR and there was insufficient data available for primers to be designed. The second intron was large (3.5 Kb) and we were unable to amplify it. No polymorphism between the two parental sequences was found after the amplification of first and the second exons. After a BLAST search with ROT01O23W unigene against database of contigs developed by Barchi et al. (2011) we found a positive contig (22573:15433\_PStI\_67/3\_NODE\_1\_L378; 15220\_PStI\_305E40\_NODE\_1\_L282). On the basis of the contig sequence obtained, primers that partially amplify the second intron were developed. After the analysis of the sequenced amplicon, a SNP (T/A) was found and validated using HRM (Additional file 3: Table S2). The gene was mapped into the linkage group E03 at 89.6 cM very close to the *4CL* gene (90.9 cM) (Figure 5). In the Tomato EXPEN-2000 map both genes also appear close together (Figure 3), separated only by 159.8 Kb in the tomato physical map.

### *C3'H (p-coumaroyl ester 3'-hydroxylase)*

An eggplant unigene YFR01I20A was identified using a tomato orthologous sequence [SGN:Solyc01g096670.2] which was approximately 3 Kb and contains three exons and two introns. Primers were developed to amplify the second intron of the gene where an Indel (TT) was found (Additional file 3: Table S2). Using this polymorphism as a CAP marker, the gene could be mapped in linkage group E01 at 85.9 cM from the linkage group end (Figure 5). Synteny study reveals that the linkage group 1 in eggplant and tomato are collinear except for minor position changes. The orthologous gene in tomato is located between the markers [SGN:T0852] (93.0 cM) and [SGN:cLPT-1-k7] (93.5 cM) (Figure 3).

### *HQT (hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase)*

Using tomato ortholog sequence, the eggplant unigene YFR01H03A was identified. The sequence of the *HQT* gene in tomato [SGN:Solyc07g005760.2) was 3.7 Kb and shows two exons and one large intron of 2.1 kb. We tried to amplify the intron of the eggplant ortholog, but we were not able to obtain a clear band. The amplification of the two exons areas allowed us to detect a SNP (A/G) that was validated by HRM (Additional file 3: Table S2). The gene was located in the upper part of linkage group E07 at a distance of 3.0 cM from the first marker (Figure 5). The tomato orthologous gene is also positioned in the upper part of linkage group 7 between the markers [SGN:U176363] (0.2 cM) and [SGN:TG131] (2.0 cM) (Figure 4). Study of synteny based on 5 anchor points reveals

certain collinearity between the two linkage groups although more anchor markers would be desirable.

### Mapping of PPO genes

From the alignment of the PPO sequences published by Shetty et al. (2011), primers were designed in order to amplify six PPO genes in *S. melongena* and *S. incanum* parental plants. In order to shorten the names, in this paper the SmePPO genes described by Shetty et al. (2011) are here simply termed PPO. All of them were selectively amplified and sequenced except for *PPO6*, which was amplified in *S. melongena* but not in *S. incanum*. SNPs polymorphisms were found in the other five PPO genes. CAPS could be developed for *PPO1* (C/A), *PPO2* (C/G), *PPO4* (G/A) and *PPO5* (G/A and T/G) while the SNP of *PPO3* (G/A) was validated by HRM (Additional file 3: Table S2). As we expected, based on synteny with Tomato EXPEN-2000 map (Fulton et al., 2002) (Figure 4), all eggplant PPOs were mapped in SMIBC in the same genomic region in the linkage group E08, where *PPO1* and *PPO3* are situated at a distance of 42.2 cM, *PPO2* and *PPO4* at a distance of 42.4 cM, and *PPO5* at 37.3 cM from the linkage group end (Figure 5).

Synteny reveals that PPO orthologous genes in tomato are located in an area of 95.5 Kb, comprising the markers [SGN:TG624] (36.70 cM) and [SGN:CIET-8-E2] (38.0 cM) (Figure 4). Several comparisons between eggplant and tomato PPO were made to correctly assign orthologous PPO genes between the two species, but a clear identification was not reached, probably due to high sequence similarity among the PPO genes of each species.





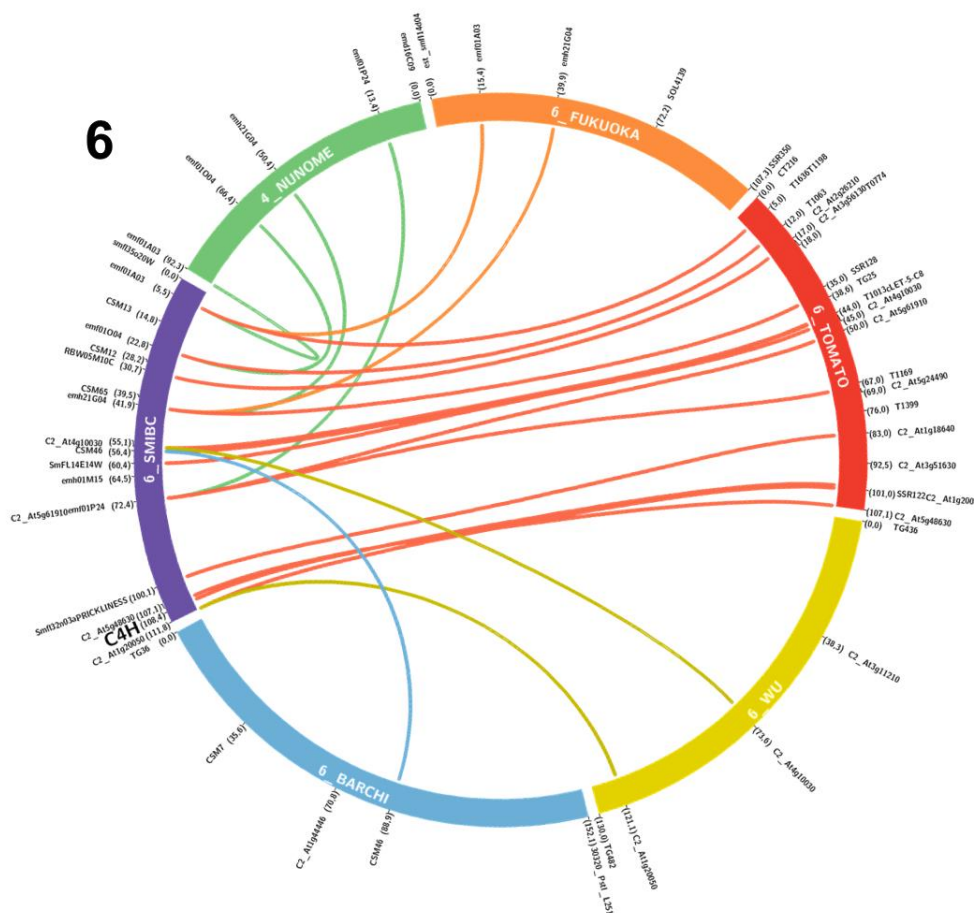




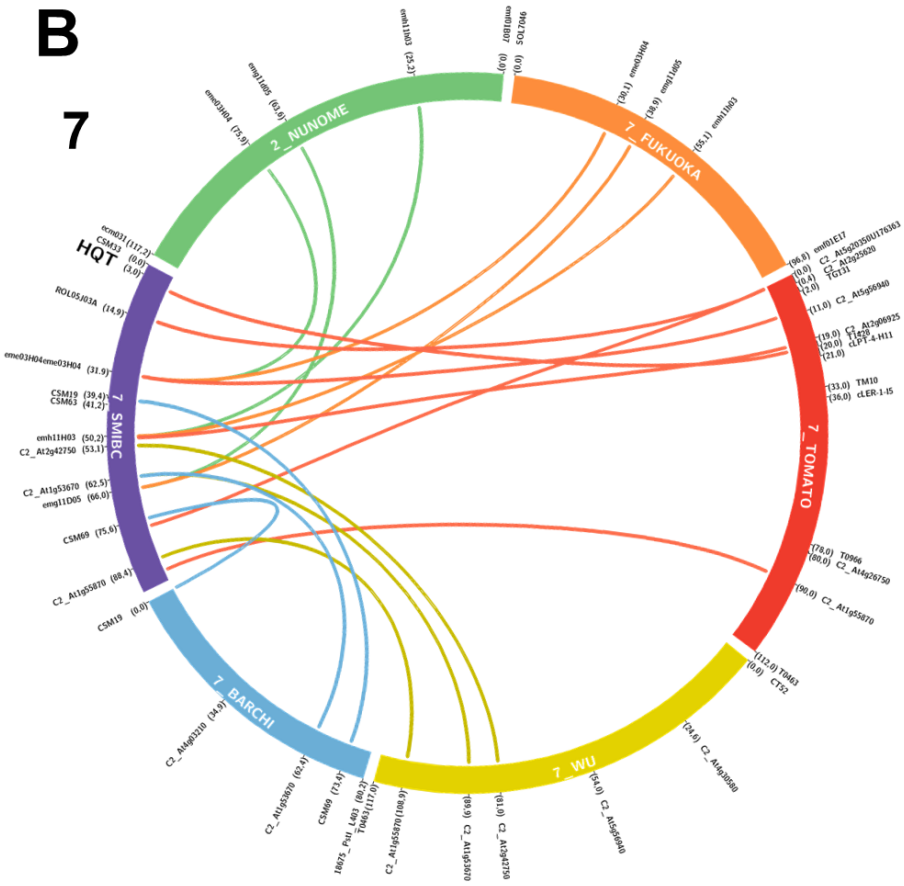


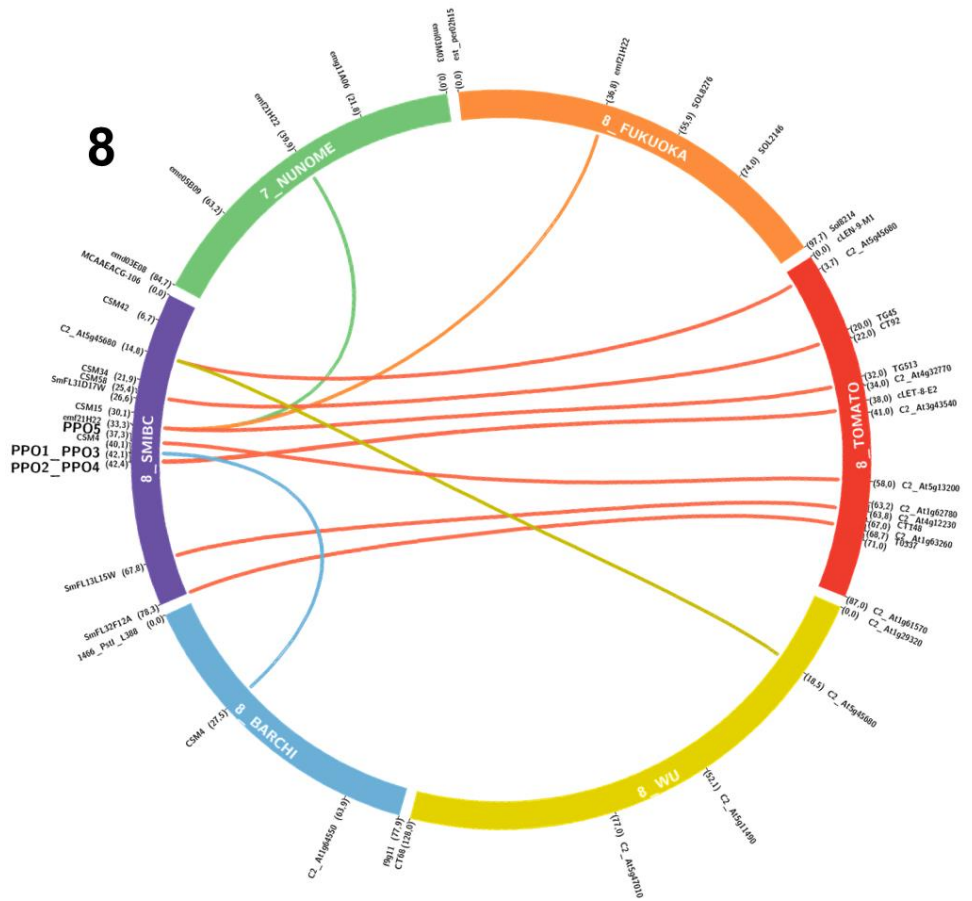






**Figure 3.** Macro-synteny between SMIBC interspecific eggplant map, tomato and eggplant maps (linkage groups E01-E06). Different colours were used for linkage groups and links to distinguish and anchor the maps. SMIBC interspecific eggplant map was depicted in purple, tomato EXPEN-2000 map [34] in red, Barchi et al. [33] eggplant map in blue, and Fukuoka et al. [32] eggplant map in orange, Nunome et al. [27] eggplant map in green, and Wu et al. [29] eggplant map in yellow. Inside of linkage groups in white are shown the corresponding map and number of each linkage group. On the external part of circular ideograms are indicated the markers name and their position. The candidate genes for chlorogenic acid (CGA) synthesis pathway and polyphenol oxidases (PPOs) are shown in bold letters.



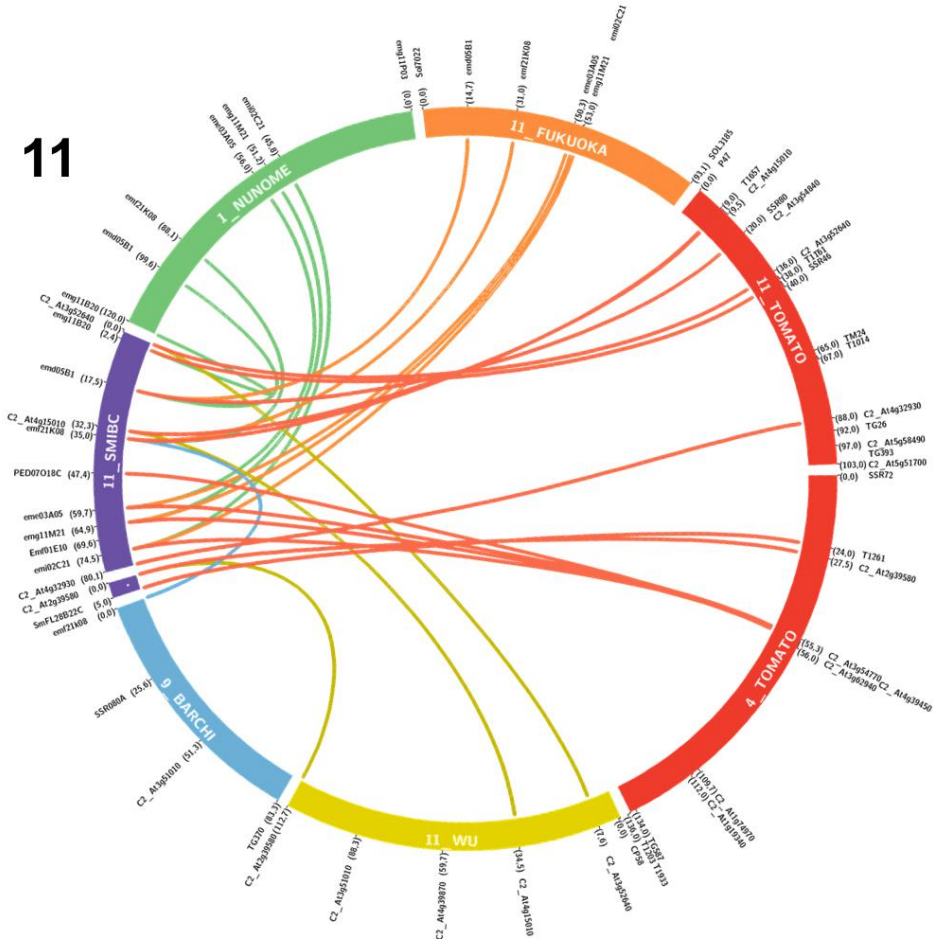




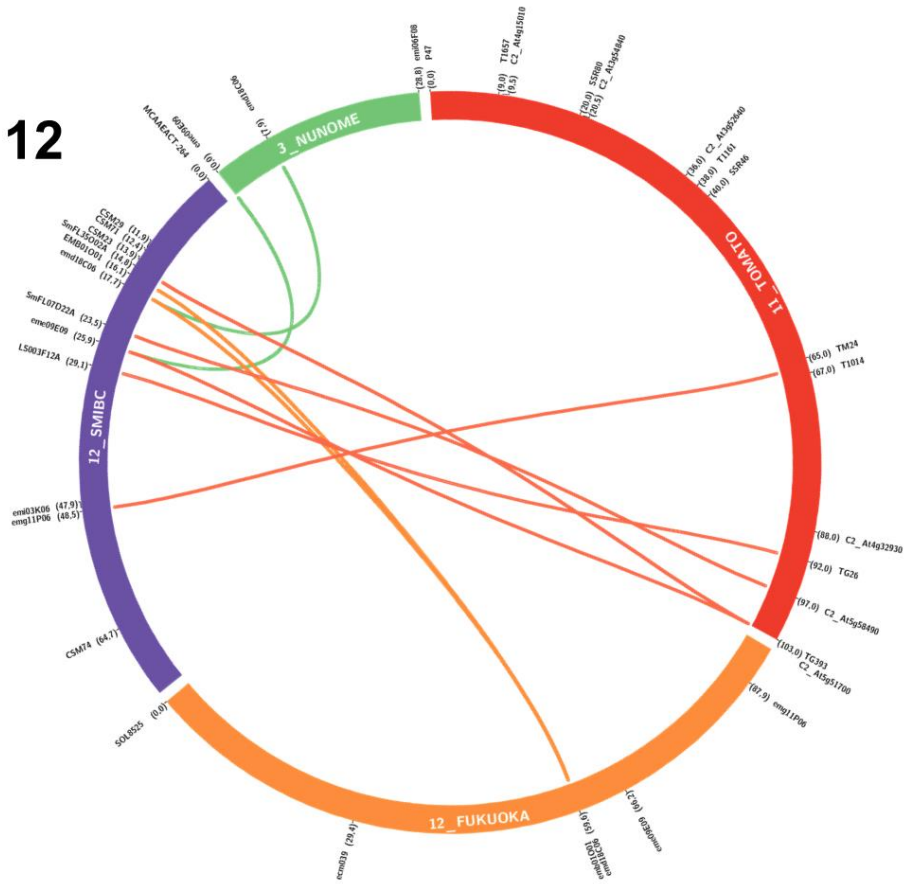




11



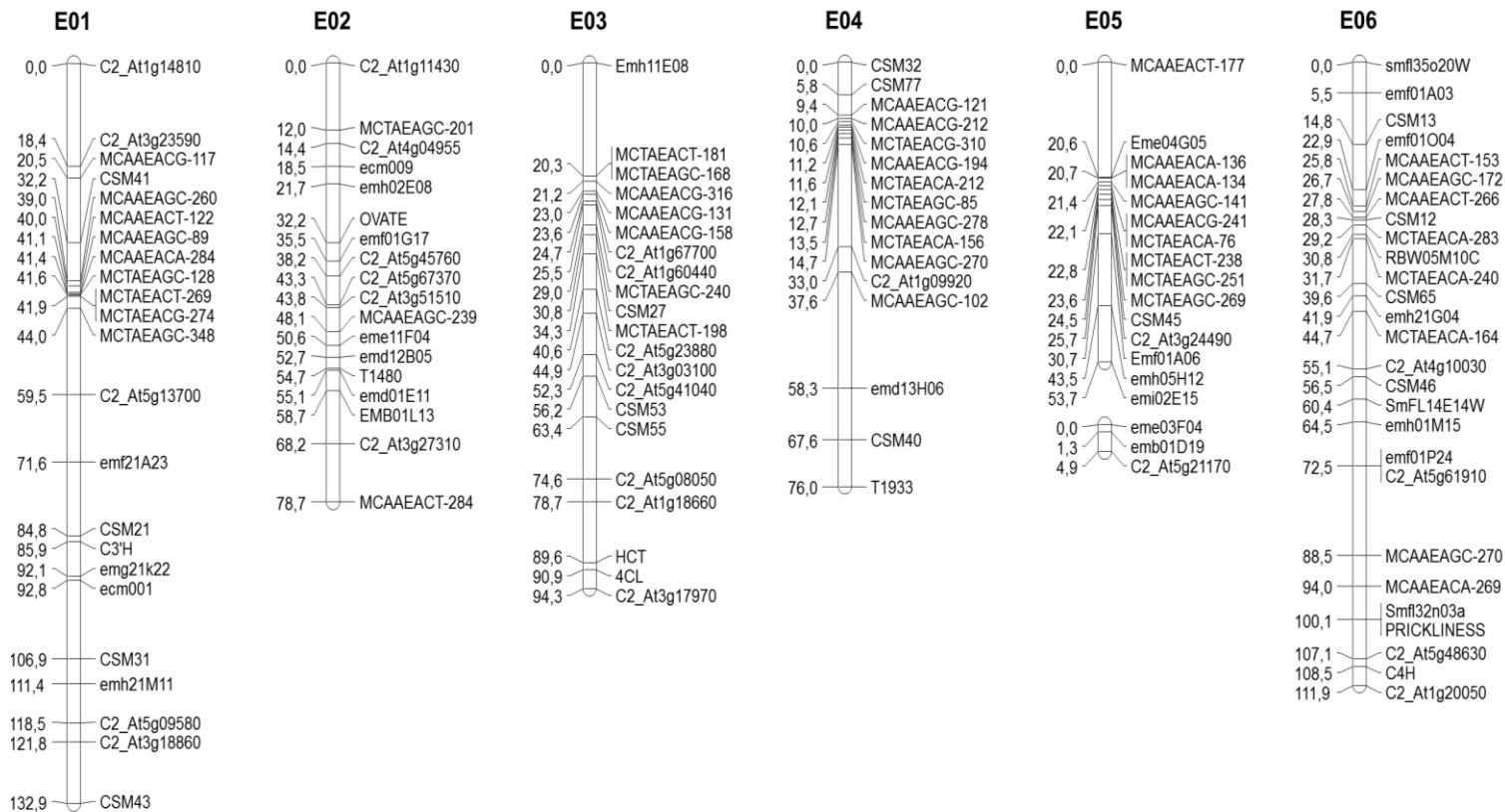
12

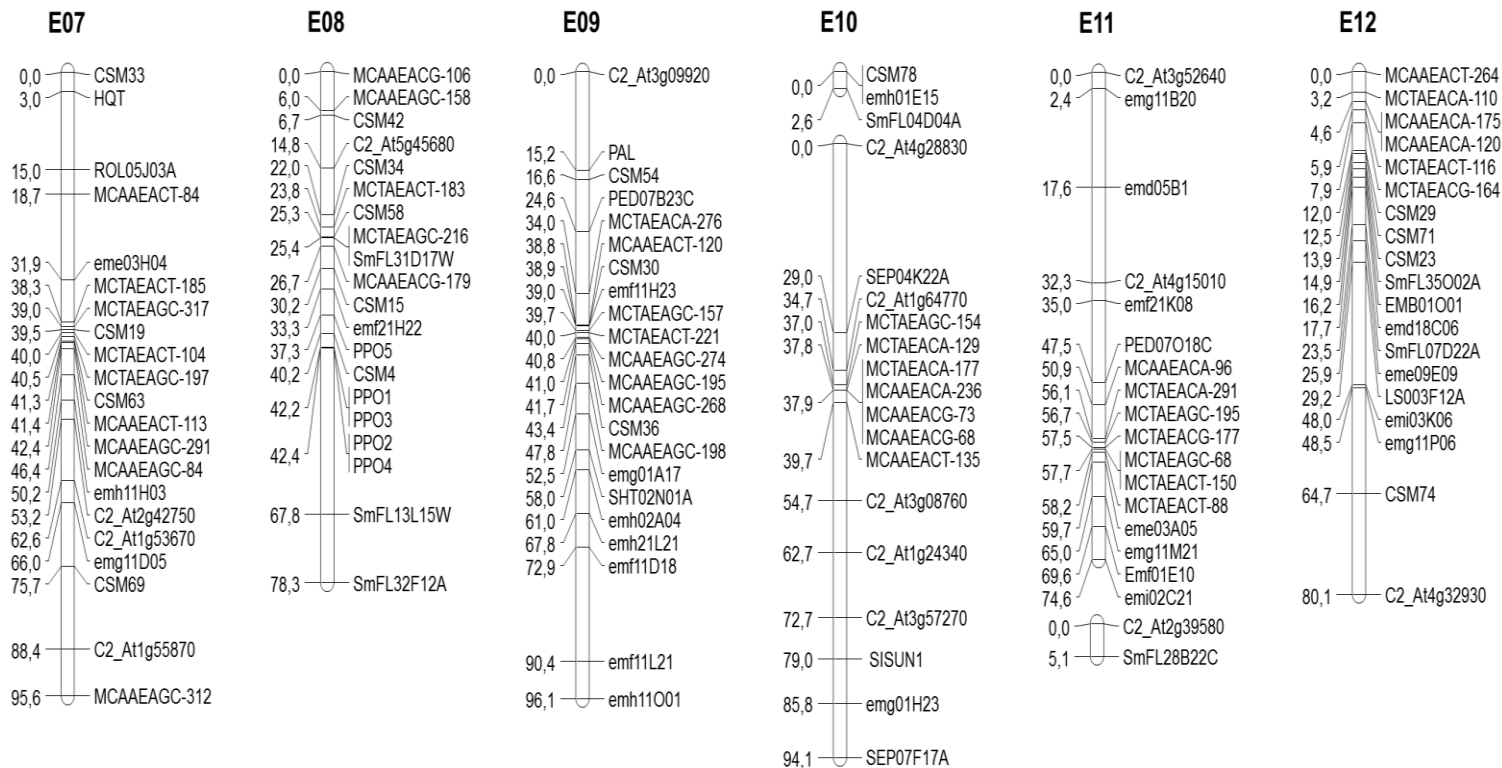


**Figure 4.** Macro-synteny between SMIBC interspecific eggplant map, tomato and eggplant maps (linkage groups E07-E12). Different colours were used for linkage groups and links to distinguish and anchor the maps. SMIBC interspecific eggplant map was depicted in purple, tomato EXPEN-2000 map (Fulton et al., 2002) in red, Barchi et al. (2012) eggplant map in blue, and Fukuoka et al. (2012) eggplant map in orange, Nunome et al. (2009) eggplant map in green, and Wu et al. (2009) eggplant map in yellow. The corresponding number of each linkage group and map are shown inside the same. Markers name and their position are shown on the external part of circular ideograms. The candidate genes for chlorogenic acid (CGA) synthesis pathway and polyphenol oxidases (PPOs) are shown in bold letters.

## Mapping of other genes and traits of agronomic importance

The sequence of the *OVATE* gene [SGN:Solyc02g085500.2], which determines the conversion from round to pear-shaped fruit in tomato, corresponded to eggplant unigene SmFL28E15A. The gene was mapped into the linkage E02 at 32.2 cM from the linkage group end (Figure 5). In Wu et al. (2009) map, *OVATE* gene was mapped at 79.0 cM from the linkage group end (Figure 3), and in tomato the orthologous gene is positioned into linkage group 2 at 89.50 cM (Figure 3). The *SISUNI* gene in tomato [SGN:SGN-U569959], which controls elongated and pointed fruit shape, is positioned in chromosome 10, near the marker [SGN:C2\_At3g10140] (52.80 cM) and shows a high identity with the eggplant contig (15541:35662\_PstI\_305E40\_NODE\_1\_L250; 36203\_PStI\_67/3\_NODE\_1\_L230) (Barchi et al., 2011). The gene was mapped in SMIBC onto linkage group 10 at 79.0 cM from the linkage group end (Figure 5). The morphological marker *PRICKLINESS* could be mapped in linkage group E06 at 100.1 cM from the linkage group end (Figure 5).





**Figure 5.** The interspecific genetic linkage map SMIBC. Linkage groups were denominated E01 to E12 in agreement with the denomination of linkage groups in other eggplant genetic maps (Doganlar et al., 2002a; Barchi et al., 2012; Fukuoka et al., 2012). The map distances, given in cM, are shown on the left side of the linkage groups. Marker names are shown on the right side.

## Synteny and orthologous candidate genes with other maps

Synteny using common molecular markers was established in order to develop a genetic linkage map taking advantage of the information provided by previous eggplant genetic linkage maps (Figures 3 and 4, Additional file 2: Table S1). The reference maps were the F<sub>2</sub> intraspecific maps developed by Nunome et al. (2009), the integrated map LWA2010 derived from two F<sub>2</sub> linkage maps (LW2010 and AL2010) developed by Fukuoka et al. (2012), the interspecific map *S. linnaeanum* (MM195) x *S. melongena* (MM738) developed by Wu et al. (2009) and the F<sub>2</sub> intraspecific maps developed by Barchi et al. (2012). In general a good conservation of marker location between SMIBC and the previous four eggplant maps was observed with the exception of six markers (Table 4). These markers show inconsistencies in position with Fukuoka et al. (2012) map. In addition, macro-synteny between SMIBC and Tomato EXPEN-2000 map was established (Figures 3 and 4, Additional files 1,2: Figure S1 Table S1). The syntenic relationship between these two maps was highly collinear, except for a few small parts of the genomes. In this way, a small part of the eggplant linkage group E03 was syntenic to tomato chromosome 5 (T05), and the same occurred between E05 and T05 and T12, E10 and T10 and T05, E11 and T11 and T04, and E12 with T11. These results are in agreement with the synteny observed by Wu et al. (2009) among eggplant and tomato.

The highest number of anchoring points was observed with the Tomato EXPEN-2000 map, that shared 130 markers in common with SMIBC, varying from 15 (E02 and E06) to 4 (E04) per linkage genetic group, followed by the eggplant maps of Wu et al. (2009) (42 anchoring



points), Nunome et al. (2009) (37), Fukuoka et al. (2012) (32), and Barchi et al. (2012) (12) for a total of 253 anchoring points. E02 was the linkage group with the highest number of links (38), while E04 had the fewest connections (8) (Table 3).

**Table 3.** Statistics of macro-synteny between SMIBC map, Tomato EXPEN-2000 map developed by Fulton et al. (2002) and eggplant genetic maps developed by Barchi et al. (2012), Fukuoka et al.(2012), Nunome et al. [27], and Wu et al. (2009). Number of shared markers with eggplant linkage groups, indicated as E01-E12, are indicated.

<b>Linkage groups</b>													
<b>Genetic maps</b>	<b>E01</b>	<b>E02</b>	<b>E03</b>	<b>E04</b>	<b>E05</b>	<b>E06</b>	<b>E07</b>	<b>E08</b>	<b>E09</b>	<b>E10</b>	<b>E11</b>	<b>E12</b>	<b>Total</b>
<b>Tomato EXPEN-2000</b>	12	15	12	4	8	15	8	11	12	12	12	9	130
<b>Barchi <i>et al.</i> (2012)</b>	1	1	1	0	2	1	3	1	1	0	1	0	12
<b>Fukuoka <i>et al.</i> (2012)</b>	3	7	1	1	3	2	3	1	4	2	5	0	32
<b>Nunome <i>et al.</i> (2009)</b>	3	7	1	1	5	4	3	1	3	1	6	2	37
<b>Wu <i>et al.</i> (2009)</b>	5	8	7	2	1	2	3	1	1	5	3	4	42
<b>Total</b>	24	38	22	8	19	24	20	15	21	20	27	15	253

**Table 4.** Markers with conflictive position according to the synteny between SMIBC interspecific eggplant map and Fukuoka et al. [32] LWA2010 eggplant genetic map. Linkage group (LG) in which the markers are positioned in SMIBC and LWA2010 maps are shown

<b>LG SMIBC map</b>	<b>Marker name</b>	<b>LG LWA2010 map</b>
<b>E01</b>	emf21A23	E08
<b>E06</b>	emh01M15	E03
<b>E09</b>	emh02A04	E11
<b>E09</b>	emh21L21	E07
<b>E09</b>	emh11O01	E03
E11	Emf01E10	E01

## ***Discussion***

Although notable efforts have been made recently to better understand the structure and organization of the eggplant genome, the available genomic information is still very limited when compared to other major Solanaceae crops such as tomato, potato and pepper. Our interspecific map spread 1085 cM, and the 15 linkage groups we have found could be traced back to 12 chromosomes through the use of markers shared with other previously published maps. The SMIBC was the first map that connects all the most informative previously published genetic maps for eggplant, allowing us to exploit the information generated by these maps (Figures 3 and 4). A total of 130 anchoring points connect SMIBC and Tomato EXPEN-2000, many of which are COSII, enabling us to establish synteny between the two genomes (Figures 3 and 4, Additional files 1,2: Figure S1 Table S1). This allowed us to confirm the position of all candidate genes according to orthologous gene location in tomato. The syntenic relationships between these two maps are in agreement with those observed by Doganlar *et al.* (2002), Wu *et al.* (2009), and Fukuoka *et al.* (2012) and provide information on the genome evolution of both crops.

A general problem in the construction of genetic maps is that different types of molecular markers tend to map in a specific regions of the genome (Jeuken *et al.*, 2001; Fukuoka *et al.*, 2012; Shirasawa *et al.*, 2013; Nunome *et al.*, 2009). Our use of a wide variety of molecular markers including those used in previous eggplant genetic maps (Nunome *et al.*, 2009; Wu *et al.*, 2009; Barchi *et al.*, 2012; Fukuoka *et al.*, 2012) allowed us to achieve better representation of the eggplant genome, which

is important for effective molecular breeding and synteny studies in the future.

Distorted markers occurred in almost all linkage groups, except for E01 and E04. Segregation distortion in mapping populations is a well-documented phenomenon in different crops (Alheit et al., 2011; Ren et al., 2014). Similar distortion levels to those we observed have been reported by Doganlar et al. (2002) in their interspecific map between *S. melongena* and *S. linnaeanum*, with 16% of markers being skewed towards the one or the other parent.

CGA is the major phenolic compound present in eggplant, which is one of the crops with highest content in hydroxycinnamic acids (Whitaker and Stommel, 2003; Stommel and Whitaker, 2003; Singh et al., 2009; Plazas et al., 2013b). Until now, attempts to improve the CGA content in eggplant have been conducted in a classical way through hybridization and selection of materials with high CGA content (Plazas et al., 2013a; Plazas et al., 2013b; Prohens et al., 2013). The results of these programs have been positive, but due to moderate heritability ( $H^2$ ) of the CGA content trait and total phenolics in general, genetic advances have been limited (Prohens et al., 2007b, 2013; Raigón et al., 2008). Although these studies indicate that phenotypic selection to obtain materials with higher contents of CGA is possible, it would be desirable to apply marker-assisted selection (MAS) to improve the efficiency of selection for this trait.

The candidate gene (CG) approach has been successful in both animal and plant genetics (Byrne and McMullen, 1996). The involvement of CGs in CGA biosynthesis pathway has been studied in plant species. The PAL gene catalyses one of the first steps of the phenylpropanoid

pathway, that produces hundreds of specialized metabolic products including lignins, flavonoids, alkaloids and many other important phenolics in plants (Vogt, 2010). In *Populus trichocarpa*, five *PAL* genes have been described that are involved in wood formation (Shi et al., 2013), and in raspberry (*Rubus idaeus*) two *PAL* genes have been identified: *RiPAL1* associated with early fruit ripening events while *RiPAL2* involved with larger flower and fruit development (Kumar and Ellis, 2001). Candidate genes involved in phenylpropanoid pathway have also been mapped in other plants like apple (Chagné et al., 2012) and artichoke (Portis et al., 2009, 2012; Moglia et al., 2013). These are examples showing that homologs can have different functions, and paralogs can play largely different roles in CGA or lignin synthesis. Even within the Solanaceae, different genes in the pathway are strongly associated with the abundance of CGA. *HQT* seems to be the most important contributor to CGA synthesis in tomato and *Nicotiana benthamiana* (Niggeweg et al., 2004), while *HCT* and *C3'H* are the most relevant in potato (Rommens et al., 2008; André et al., 2009).

In this work we were able to locate all the genes involved in the core CGA pathway in eggplant. The positions of the genes are in agreement with those expected based on synteny with tomato. With the exception of *HCT* and *4CL*, which are co-situated at a genetic distance of 1.3 cM in the linkage group E03, all of the rest of the genes of the biosynthesis pathway of CGA are situated in separate linkage groups. This has important implications for breeding and for strategies based on pyramiding of favourable alleles introgressed from *S. incanum* in the genetic background of eggplant.

A potential problem for developing materials with higher CGA content is their oxidation mediated by polyphenol oxidase (PPO), which leads to the browning of the fruit flesh, reducing apparent fruit quality (Concellón et al., 2012; Plazas et al., 2013b). A positive phenotypic linear correlation was found in eggplant between the degree of fruit flesh browning and total content in phenolic and CGA content (Prohens et al., 2007b; Plazas et al., 2013a).

Several studies in other crops associated quantitative trait loci (QTL) for enzymatic browning with PPO alleles and these alleles were then used as markers for this trait (Urbany et al., 2011; Di Guardo et al., 2013). On the other hand, silencing of PPO genes result in a reduced browning reaction as has been shown in potato and apple (Coetzer et al., 2001; Murata et al., 2001).

The studies of PPO genes suggest that in plants they form a gene family with high homology, usually have no introns and normally are present in several copies. For instance, tomato has seven PPOs, potato has six PPOs and soybean has eleven PPOs (Newman et al., 2011; Urbany et al., 2011; Tran et al., 2012; Chai et al., 2013). These features have complicated the search of polymorphisms between PPOs because amplification of a single isoform is difficult. In this study we were able to detect polymorphism in all PPO genes with the exception of *PPO6* (amplified in *S. melongena* but not in *S. incanum*) leading us to speculate that in *S. incanum* this PPO gene is absent or presents substantial changes with respect to *S. melongena*. As in tomato (Newma et al., 2011), eggplant PPOs were mapped very closely to each other in linkage group E08; this confirms the existence of a cluster also in eggplant (Figure 5), although a correct assignment using synteny was not possible between eggplant and

tomato orthologs, due to high sequence similarity among the PPO genes of each species. In eggplant PPO genes identity varying between 72% and 95% at nucleotide level and between 62% and 92% at amino acid level (Shetty et al., 2011).

Since both consumers and industry prefer eggplant varieties with white flesh and a low degree of browning, future varieties with high phenolic content should also be bred to have limited browning (Prohens et al., 2007). Accordingly, the search for allelic variants for increasing CGA content should be conducted in concert with a search for allelic variants decreasing PPOs. In eggplant, several authors have found differences in PPO activity between varieties (Shetty et al., 2011; Concellón et al., 2012) which suggest mining the biodiversity of this crop and its relatives can lead to discovering desirable alleles useful to produce varieties with low browning in fruit flesh regardless of the total phenolic abundance. In addition, the study of selection sweeps at these loci due to the culturally distinctive organoleptic preferences, would be of interest to understand the history of evolution of eggplant. A reduction in browning is quite probably a significant domestication trait in eggplant and many other species.

In the present study an interspecific genetic linkage map was developed to better understanding the genetics of CGA content and fruit flesh browning in eggplant through the mapping of candidate genes involved in these processes. This was assisted by the use of synteny of the orthologous genes in tomato. The use of *S. incanum* in this work and in breeding programs is not accidental, as it is a close relative of eggplant (Lester and Hasan, 1991; Knapp et al., 2013) and produces fully fertile hybrids with regular meiosis [7,14]. *Solanum incanum* has characteristics of great interest for plant breeders, such as bacterial and fungal resistance,



including to *Fusarium oxysporum*, and tolerance to various abiotic stresses, especially drought (Yamakawa et al., 1979; Prohens et al., 2013). Recent discoveries have shown that *S. incanum* is a powerful source of variation for phenolic content, as its total phenolic content is several times higher than that of *S. melongena* (Stommel and Whitaker, 2003; Prohens et al., 2013). *Solanum incanum* also exhibits absence of anthocyanins, presence of prickles and subovoid green fruits (Prohens et al., 2013)).

Building an interspecific genetic map from the backcross generation of two species that differ in many traits has two important advantages. First of all, in this manner the problem of low intraspecific polymorphism in eggplant (Doganlar et al., 2002b; Stàgel et al., 2008; Wu et al., 2009; Fukuoka et al., 2012) is partially overcome. Secondly, this BC1 population can be used as the starting point to develop a set of introgression lines of *S. melongena*, with *S. incanum* as the donor parent. Once obtained, these lines will be useful to understand and dissect candidate gene roles in CGA accumulation and browning, to detect QTLs and genes involved in the synthesis of other polyphenols present in *S. incanum* (mainly *N*-(*E*)-caffeoylputrescine, 3-*O*-Malonyl-5-*O*-(*E*)-caffeoylquinic acid and 5-*O*-Malonyl-4-*O*-(*E*)-caffeoylquinic acid) (Prohens et al., 2013), as well as many other morphological and physiological traits of interest for breeding or for crop evolution and domestication studies. Candidate genes for traits thought to be important in domestication such as fruit shape and *PRICKLINESS* are mapped in SMIBC in agreement with their positions based on previous QTL studies in eggplant and in syntenic regions of tomato (Frary et al., 2014; Portis et al., 2014), also highlighting the common occurrence of parallelism in loci causing domestication phenotypes between species.

## ***Conclusion***

Despite its economic importance few genomic and biotechnological resources exist for eggplant. Eggplant contains high quantities of phenolics, in particular of CGA, a secondary metabolite with many properties beneficial for human-health and which in the last years is attracting a lot of attention from the scientific community. We report the development of a new interspecific map in eggplant using different types of markers. We have mapped the six genes involved in CGA synthesis pathway on five linkage groups, and determined that the PPO genes cluster together on LG8. This new map is anchored to the tomato genetic map as well as to four previous genetic maps of eggplant, and therefore constitutes an important tool for exploiting synteny and for molecular breeding in eggplant. The molecular tools we have developed will be of interest for eggplant breeding programmes, in particular those aimed at using marker assisted selection for improved CGA content combined with low browning effects. These tools and results, together with the recently published eggplant genome sequence (Hirakawa et al., 2014) will greatly facilitate the development of a new generation of eggplant varieties with dramatically enhanced bioactive properties that can significantly contribute to better human nutrition.

## ***Acknowledgements***

This research has been partially funded by Ministerio de Economía y Competitividad and FEDER (grant AGL2012-34213). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral grant.





## **Chapter 5: Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background**

Gramazio P<sup>1,\*</sup>, Prohens J<sup>1</sup>, Plazas M<sup>2</sup>, Mangino G<sup>1</sup>, Herraiz FJ<sup>1</sup>, Vilanova S<sup>1</sup>

<sup>1</sup> Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

<sup>2</sup> Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas - Universitat Politècnica de València, Camino de Vera 14, Valencia, Spain.

\* Corresponding author: [piegra@upv.es](mailto:piegra@upv.es)

**Citation:** Gramazio P<sup>1,\*</sup>, Prohens J<sup>1</sup>, Plazas M<sup>2</sup>, Mangino G<sup>1</sup>, Herraiz FJ<sup>1</sup>, Vilanova S<sup>1</sup> (2017). Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. *Frontiers in Plant Science* 8:1477.

<https://doi.org/10.3389/fpls.2017.01477>

## ***Abstract***

Advanced backcrosses (ABs) and introgression lines (ILs) of eggplant (*Solanum melongena*) can speed up genetics and genomics studies and breeding in this crop. We have developed the first full set of ABs and ILs in eggplant using *Solanum incanum*, a wild eggplant that has a relatively high tolerance to drought, as a donor parent. The development of these ABs and IL eggplant populations had a low efficiency in the early stages, because of the lack of molecular markers and genomic tools. However, this dramatically improved after performing genotyping-by-sequencing in the first round of selfing, followed by high-resolution-melting single nucleotide polymorphism genotyping in subsequent selection steps. A set of 73 selected ABs covered 99% of the *S. incanum* genome, while 25 fixed immortal ILs, each carrying a single introgressed fragment in homozygosis, altogether spanned 61.7% of the *S. incanum* genome. The introgressed size fragment in the ILs contained between 0.1% and 10.9% of the *S. incanum* genome, with a mean value of 4.3%. Sixty-eight candidate genes involved in drought tolerance were identified in the set of ILs. This first set of ABs and ILs of eggplant will be extremely useful for the genetic dissection of traits of interest for eggplant, and represents an elite material for introduction into the breeding pipelines for developing new eggplant cultivars adapted to the challenges posed by the climate-change scenario.

**Keywords:** *Solanum melongena*, *Solanum incanum*, advanced backcrosses, introgression lines, drought tolerance, high-throughput genotyping

**Abbreviations:**

AB: advanced backcross

COSII: conserved ortholog set of genes II

HRM: high resolution melting

IL: introgression line

QTL: quantitative trait locus

RIL: recombinant inbred line

SNP: single nucleotide polymorphism

SSR: simple sequence repeat

UPV: Universitat Politècnica de València

## ***Introduction***

Eggplant (*Solanum melongena* L.,  $2n = 2x = 24$ ) is a major vegetable crop of increasing importance during the last few years; while other major crops in the genus *Solanum* and in the family Solanaceae, like tomato, potato, pepper, tobacco and petunia, have been widely studied from the genetic and genomic points of view, developments in eggplant genomics have lagged behind. In fact, the first draft of the eggplant genome was published only in 2014 (Hirakawa *et al.*, 2014), while in the other major solanaceous crops high-quality genome sequences were available much earlier (Potato Genome Sequencing Consortium 2011; Tomato Genome Consortium 2012). While other eggplant genomic tools are slowly being published, like the first transcriptome (Yang *et al.*, 2014), genetic and association maps (Barchi *et al.*, 2012; Fukuoka *et al.*, 2012; Gramazio *et al.*, 2014; Portis *et al.*, 2015) and quantitative trait locus (QTL) studies (Miyatake *et al.*, 2016; Rinaldi *et al.*, 2016; Toppino *et al.*, 2016), there is a general lack of eggplant experimental populations, such as recombinant inbred lines (RILs), advanced backcrosses (ABs), introgression lines (ILs) or multiparent advanced generation intercross (MAGIC) populations.

However, although a great phenotypic diversity can be found in the common eggplant (Kaushik *et al.*, 2016; Portis *et al.*, 2015), its genetic base is quite narrow due to the genetic bottlenecks undergone during the domestication and subsequent crop evolution processes (Muñoz-Falcón *et al.*, 2009; Meyer *et al.*, 2012a). This low genetic diversity could be a serious constraint for the development of new highly performing cultivars, which have to be able to contribute to increasing yield under the



environmental changes resulting from a climate-change scenario (Tilman *et al.*, 2011; Ray *et al.*, 2013).

Although many resistances and tolerances to biotic and abiotic stresses have been described in eggplant wild relatives (Rotino *et al.*, 2014), just a few attempts have been made to transfer them to the cultivated eggplant (Toppino *et al.*, 2008; Liu *et al.* 2015). Recent efforts in enhancing wild relatives for the adaptation of eggplant to climate change have resulted in the development of over 50 combinations of hybrids and 30 first backcross (BC1) generations (Kouassi *et al.*, 2016; Plazas *et al.*, 2016). These materials were obtained from six *S. melongena* parents and 14 wild species from different gene pools, including the primary gene pool (GP1, which includes relatives that can be easily crossed with the cultivated species to produce a highly fertile progeny), secondary gene pool (GP2, which includes species that can be crossed with the cultivated species, although with some difficulties due to reproductive barriers, and that may give a low fertility progeny) and the tertiary gene pool (GP3, which is made of phylogenetically distant species where pre- and post-zygotic barriers require specific breeding techniques, like embryo rescue, for hybridization and the progeny is often sterile).

The development of ABs and ILs in the eggplant gene pool would make a major contribution to enhancing the use of crop wild relatives in eggplant breeding (Prohens *et al.*, 2017). ABs consist of materials obtained through a backcross breeding scheme containing fixed or non-fixed single or multiple introgressions from a donor in the genetic background of a recurrent parent (Fulton *et al.*, 1997). ABs represent powerful tools that allow more precise mapping of QTLs compared to balance populations (e.g. F2, BC1, RIL) and an elite material for breeding (Tanksley *et al.*,

1996). IL collections are a special type of AB population, consisting of fixed lines with single introgressed fragments from a donor parent (Eshed and Zamir 1995; Tian *et al.*, 2006). ILs are immortal populations making them more suitable for direct incorporation into breeding pipelines. The general aim for developing sets of ILs is to represent the whole genome of a donor parent, almost always an exotic or wild genotype, in the genetic background of a recipient cultivated parent through a set of lines carrying overlapping homozygous chromosome fragments of the former in the latter (Eshed and Zamir 1994; Zamir 2001; Eduardo *et al.*, 2005). Several studies have demonstrated the higher efficiency of ILs and near isogenic lines (NILs) in QTL estimation than other segregating populations, such as F<sub>2</sub>, double haploid lines and RILs (Wehrhahn and Allard 1965; Eshed and Zamir 1995; Kearsley and Farquhar 1998; Lebreton *et al.*, 1998; Alonso-Blanco *et al.*, 2006; Yin *et al.*, 2016).

Although the availability of IL populations in crop species is still limited, in the last two decades many have been developed, mainly in major cultivated crops such as maize, wheat, rice, barley, soybean and melon (Koester *et al.*, 1993; Concibido *et al.*, 2003; Von Korff *et al.*, 2004; Eduardo *et al.*, 2005; Liu *et al.*, 2006; Pestsova *et al.*, 2006; Tian *et al.*, 2006). The availability of a large number of IL populations in tomato has promoted and facilitated studies of the genetics of morphology (Chitwood *et al.*, 2013), physiology (Ranjan *et al.*, 2016), biotic resistance (Verlaan *et al.*, 2013), abiotic tolerance (Albacete *et al.*, 2014), chemical composition (Ikeda *et al.*, 2017), enzyme activity (Caruso *et al.*, 2016), metabolic profiling (Schauer *et al.*, 2006), transcriptional profiles (Baxter *et al.*, 2005) and many more traits.

In the case of eggplant, up until now no sets of ABs or ILs covering the whole genome of a relative have been obtained. In order to obtain the first set of introgression materials of this type in eggplant, we chose *Solanum incanum* L. as a donor parent. This wild species belongs to the secondary gene pool of common eggplant (Syfert *et al.*, 2016), and is naturally distributed in the desertic and dryland areas from western Pakistan, Afghanistan and Iran across the Middle East and northern Africa (Ranil *et al.*, 2016; Vorontsova and Knapp 2016). *S. incanum* is reported as having considerable drought tolerance (Lester and Hasan 1991; Daunay 2008; Knapp *et al.*, 2013), which has been confirmed under experimental conditions (Savarino 2017). Drought tolerance is a trait of major interest for adaptation to climate change, where new drought-tolerant varieties are required (Dempewolf *et al.*, 2014). In addition, *S. incanum* presents resistance to some fungal diseases, such as *Fusarium oxysporum* and *Phomopsis vexans* (Yamakawa and Mochizuki 1979; Collonnier *et al.*, 2001b). Also, *S. incanum* exhibits a high content of bioactive phenolic compounds, mostly of chlorogenic acid type (Ma *et al.*, 2011; Stommel and Whitaker 2003; Prohens *et al.*, 2013; Meyer *et al.*, 2015). These favorable features led us to use *S. incanum* for the construction of an interspecific genetic map with *S. melongena*, which is linked to four eggplant and one tomato genetic maps (Gramazio *et al.*, 2014). Furthermore, the transcriptome of *S. incanum* was sequenced to create a large set of molecular markers directly applicable in breeding programs and for a better understanding of the target genes involved in metabolic pathways relevant for eggplant breeding (Gramazio *et al.*, 2016). Also, recently, the entire genome of *S. incanum* has been sequenced (Gramazio, unpublished data).

Here, we present the development of the first full set of ABs and ILs in the eggplant genetic background, from the first generations, where little genomic information was available and low–medium throughput genotyping techniques were used, to the last ones, where more markers were available and affordable high-throughput genotyping-by-sequencing (GBS) helped in the characterization of the ABs and in obtaining a set of ILs with introgressions containing a single fragment from the *S. incanum* donor.

## ***Materials and Methods***

### **Plant material**

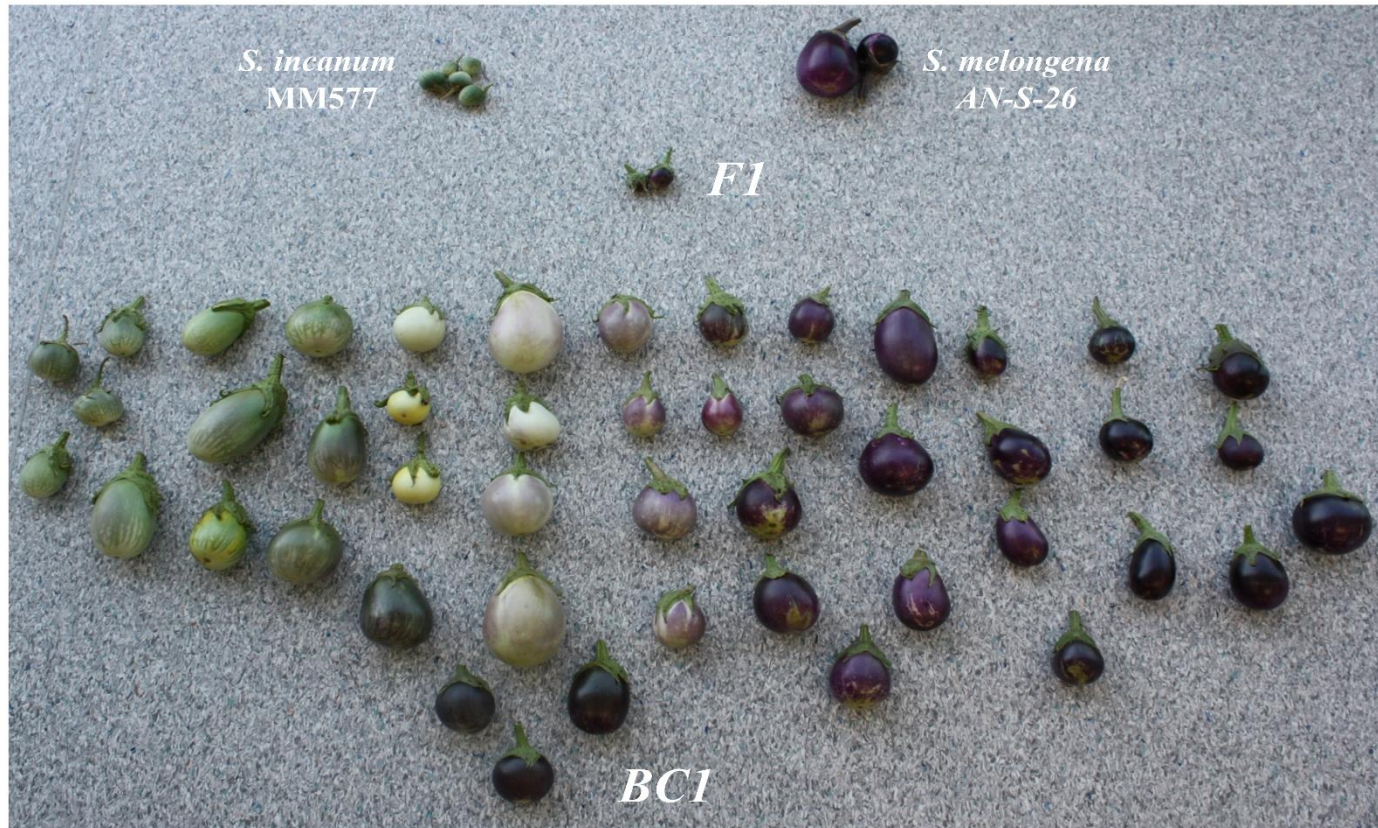
The donor parent selected for the development of the set of ABs and ILs was *S. incanum* accession number MM577, collected in the wild in Israel and provided by the INRA Avignon (France) germplasm bank. MM577 plants are prickly, with no anthocyanin pigmentation and produce small green rounded fruits (Figure 1). This MM577 plant has been used in several genetic and genomic studies, as it has been characterized for morphological traits and phenolic acids content (Stommel and Whitaker 2003), for androgenic competence through anther culture (Salas *et al.*, 2011), as a rootstock for eggplant yield and its effects on fruit quality (Gisbert *et al.*, 2011), for phylogeographic studies (Meyer *et al.*, 2015), for breeding for adaptation to climate change (<http://eggplantprebreeding.upv.es/>), as a parental of an interspecific genetic map (Gramazio *et al.*, 2014), for transcriptome sequencing (Gramazio *et al.*, 2016) and for whole genome resequencing (Gramazio, unpublished data).

The recipient parent was *S. melongena* accession number AN-S-26, a Spanish local variety from Andalusia obtained from the Universitat Politècnica de València (UPV), Valencia, Spain, germplasm bank. The AN-S-26 plant is non-prickly, with anthocyanin pigmentation and large obovoid dark purple fruits (Figure 1). The morphological characterization and phenolics content of *S. incanum* MM577 and *S. melongena* AN-S-26 revealed important differences in plant and fruit morphology and phenolic concentration between them (Prohens *et al.*, 2013). The genetic characterization of both parents with simple sequence repeat (SSR) and

conserved ortholog set of genes II (COSII) markers also revealed a high level of genetic polymorphism between them (Gramazio et al., 2014). Both accessions numbers are maintained at the UPV germplasm genebank.

Seeds were germinated in Petri dishes and were subsequently transferred to seedling trays. These were kept in a climatic chamber under a photoperiod and temperature regime of 16 h light (25 °C): 8 h dark (18 °C). The plantlets were then transplanted to a pollinator-free glasshouse situated in the campus of the UPV, Valencia, Spain (GPS coordinates: latitude, 39° 28' 55" N; longitude, 0° 20' 11" W; 7 metres above sea level). Plants were grown in 15 L pots filled with coconut fiber, irrigated and fertilized using a drip irrigation system, and pruned and trained with vertical strings. Phytosanitary treatments against spider mites and whiteflies were performed when necessary. Plants were hand pollinated and crosses were performed between the recurrent parent and the selected descendants of a family (hereafter referred to as progeny). *S. melongena* AN-S-26 was used mainly as the male parent, except for when the F1 and BC1 generations were obtained, where AN-S-26 was the female parent.

The hand pollinations were carried out as follows: the flowers of the female parent were emasculated before anthesis (i.e. before flower opening), then pollen was deposited on the stigma by gently rubbing it with a glass slide covered with the pollen of the male parent. Pollinated flowers were covered with a paper or mesh bag in order to avoid undesired cross-pollination, and tagged with information on the cross or selfing performed and the date.



**Figure 1.** Parents (*Solanum incanum* MM577 and *S. melongena* AN-S-26) of the AB and IL sets, their hybrid (F1) and the segregating first backcross (BC1) generation to the recipient parent *S. melongena* AN-S-26.

## Breeding scheme

The breeding scheme used to develop the ABs and ILs is depicted in Figure 2. The F1 hybrid was obtained in 2008 using the recipient parent as a female. The hybrid was completely fertile and in the following year was backcrossed to the AN-S-26 parent to generate the BC1 population.

The BC1 population, which consisted of 91 plants, was used to build the interspecific genetic map (Gramazio *et al.*, 2014). A total of 12 selected BC1 plants (hereafter referred to as families) were backcrossed to generate the BC2 population. A subset of 50 selected BC2 plants, four to five per BC1 family, were grown in the greenhouse and backcrossed to obtain the BC3 population.

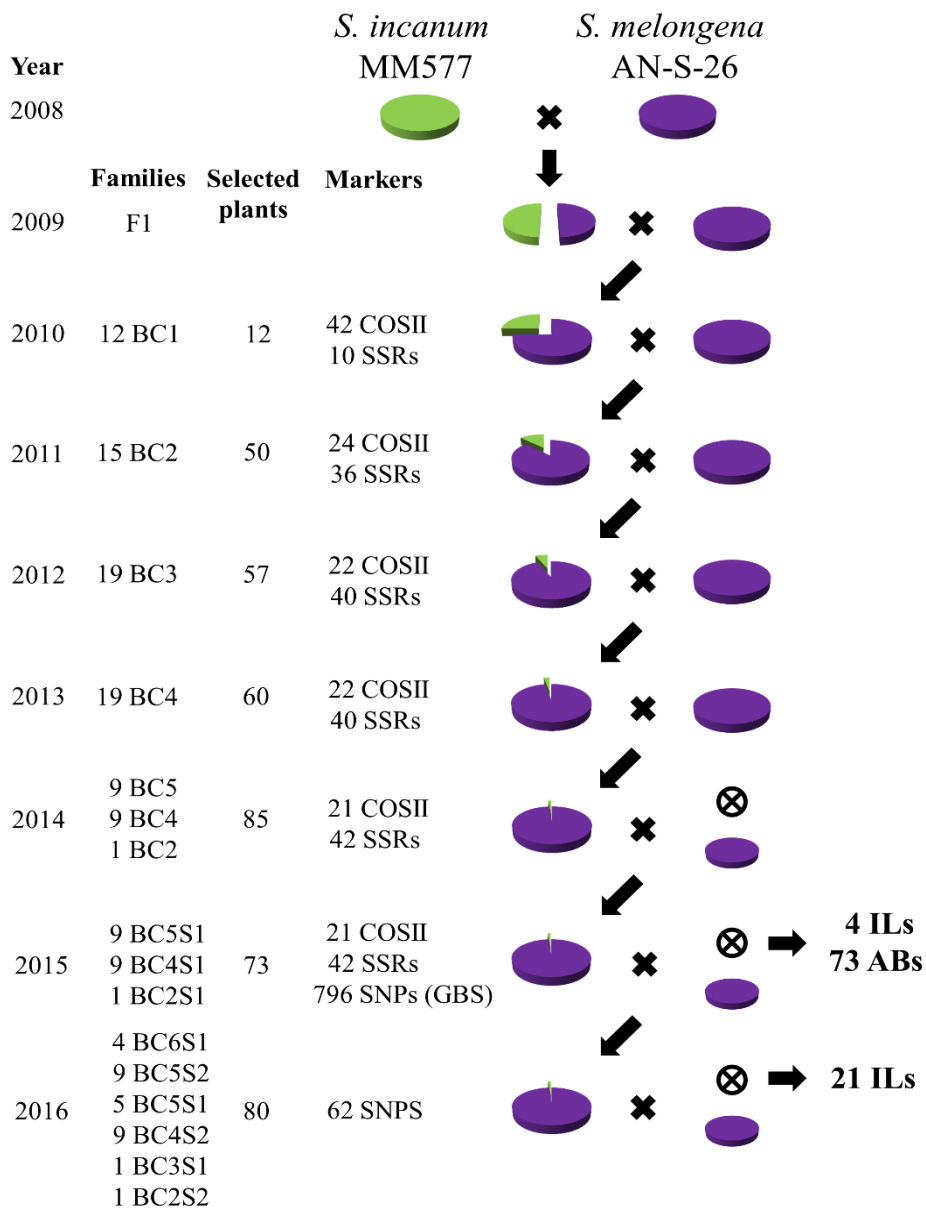
Twenty plants of each BC3 family were genotyped and three of each family were selected for the next round of backcross. Fifteen to twenty BC4 plants were genotyped at the seedling stage and 60 of them were selected for producing the next backcross generation. Some BC4 families did not set fruit; thus in 2014, in addition to the BC5 plants, some BC4 and BC2 plants, which had not given any progeny, were genotyped and selfed for a total of 85 plants.

After a preliminary manual genotyping of the progenies obtained, and driven by the recent publication of the first draft of the eggplant genome (Hirakawa *et al.*, 2014), a high-throughput GBS was performed for the selected selfed genotypes. A total of 73 genotyped plants, including BC5S1, BC4S1 and BC2S1, were grown in the greenhouse and selfed to produce seeds in order to obtain the set of ABs and a first subset of ILs with single homozygous introgressions. In addition, the plants that



presented two or more homozygous introgressions were backcrossed in order to obtain single homozygous introgressions.

A last round of selection for obtaining fixed ILs, including BC6S1, BC5S2, BC5S1, BC4S2, BC3S1 and BC2S2 families, was made with a new set of single nucleotide polymorphisms (SNPs) obtained by the GBS method based on the physical position. An additional 21 ILs with a single homozygous introgression, were identified. A total of 80 selected plants, around three to four per family, were grown in the greenhouse. Depending on the introgressions carried, they were selfed to produce seeds for the new ILs identified and to reduce the residual heterozygosity, or backcrossed to reduce the number of homozygous and off-target introgressions.



**Figure 2.** Breeding scheme used to develop the first set of eggplant ABs and ILs in the genetic background of *Solanum melongena* AN-S-26 (purple circle) by using *S. incanum* MM577 (green circle) as a donor. For each year (one generation per year), the number of families, selected plants, molecular markers assessed and type of cross are shown, the backcrosses with the recurrent parents are represented by an × and a purple circle, and the selfing rounds by an encircled ×.

## DNA extraction

Total genomic DNA was isolated from young leaves of plantlets at the 3-4 true leaves stage according to the CTAB method (Doyle and Doyle 1987) ,with slight modifications. The extracted DNA was dissolved in Milli-Q water and its quality was checked in a 0.8% agarose gel. DNA concentration was measured with a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA) and adjusted to 30 ng/μL for PCR and high resolution melting (HRM) amplification (Wittwer *et al* 2003) and to 100 ng/μL for GBS.

## Genotyping methods

At the same time as the ABs and ILs were developed, a genetic linkage map was built from a cross between the parents, and was anchored to four eggplant and one tomato genetic maps (Gramazio *et al.*, 2014). This map comprised 42 COSII, 99 SSRs, 88 amplified fragment length polymorphisms (AFLPs), 9 cleaved amplified polymorphic sequences (CAPS), 4 SNPs and one morphological polymorphic markers, and encompassed 1085 cM distributed in 12 linkage groups. A subset of these markers was used to assist in the selection of the plants during the development of the ABs and ILs.

Forty-two universal primers of COSII markers (Wu *et al.*, 2006), which were previously mapped in an interspecific genetic map between *S. linnaeanum* and *S. melongena* (Wu *et al.*, 2009), and 10 SSRs, developed by Nunome *et al.*, (2009), which corresponded to the first set of markers mapped, were used to genotype the 91 BC1 plants and to select the 12 BC1

used to generate the BC2 population. The 12 plants selected presented all the markers from the donor for the target chromosome, each one different from the other 11 BC1 plants. The 240 BC2 progeny were genotyped at the seedling stage with 60 markers, 24 COSII and 36 SSRs, some of which were developed by our group (Vilanova *et al.*, 2012). The details of the marker types, polymorphism and segregation distortion were reported in Gramazio *et al.*, (2016). The new SSRs substituted for some of the COSII markers for reaching new genetic positions uncovered by the latter. For the BC2 population, not the whole background of each plant was screened, but just the target chromosome of each family. So, for example, the BC2 plants derived from the BC1 plant selected for chromosome 1 were genotyped only for chromosome 1 and not for chromosomes 2 to 12.

This strategy was employed to reduce the time and resources required to assess all the markers in all the chromosomes, and to perform genotyping at the seedling stage in order to grow only the selected plants to the reproductive stage. In fact, the approach was to develop, as a first step and as long as possible, a set of chromosome substitution lines in order to reduce the number of plants to be genotyped, grown and backcrossed. The set of markers used for genotyping the BC3, BC4, BC5 and BC5S1 generations was practically the same, except for a few changes where SSRs replaced some COSII markers.

The selected plants from the first round of selfing were also genotyped with a GBS platform. A total of 796 SNPs was selected to improve the estimation of recombination breakpoints and to detect non-target introgressions that were dragged during the ABs and ILs development. From the whole collection of SNPs detected by GBS, 62 of them were selected, from four to eight markers per chromosome, and

adapted for HRM analysis to genotype the plants involved in the second round of selfing. The number of SNPs assessed for each progeny was variable, depending on the size and the type (target or not-target) of the introgressions. The size and the position of the introgressions in megabases (Mb) and the percentage of coverage were calculated according to the physical position of the new eggplant genome assembly developed by the Italian Eggplant Genome Sequencing Consortium (Barchi *et al.*, 2016). Primers pairs were designed using the Primer3 tool (v. 0.4.0, <http://bioinfo.ut.ee/primer3-0.4.0/primer3/>) with a mean amplicon range of 80 to 120 bp.

#### *COSII marker analysis*

COSII primer pairs were amplified by PCR in a 12  $\mu$ L reaction volume comprising 7.21  $\mu$ L water, 1.2  $\mu$ L 1 $\times$  PCR buffer, 0.6  $\mu$ L MgCl<sub>2</sub> (50 mM), 0.24  $\mu$ L dNTPs (10 mM), 0.3  $\mu$ L of each primer (10  $\mu$ M), 0.15  $\mu$ L *Taq* DNA Polymerase (5U/ $\mu$ L) and 2  $\mu$ L DNA template (20 ng/ $\mu$ L), under the following cycling conditions: denaturation at 95 °C for 3 min; followed by 30 cycles of 30 s at 95 °C, 30 s at 65 °C and of 30 s at 72 °C; with a final extension at 72 °C for 5 min. Then, restriction enzymes were used to cut in the polymorphic regions and the different size bands were visualized in 2–3% agarose gel.

#### *SSR marker analysis*

The amplification of SSRs was performed by PCR in a final volume of 12  $\mu$ L: 7.21  $\mu$ L water, 1.2  $\mu$ L 1 $\times$  PCR buffer, 0.6  $\mu$ L MgCl<sub>2</sub> (50 mM), 0.24  $\mu$ L dNTPs (10 mM), 0.3  $\mu$ L reverse primer (10  $\mu$ M), 0.06  $\mu$ L

forward primer with M13 tail (10  $\mu$ M), 0.24  $\mu$ L fluorochrome (FAM, VIC, NED and PET) (10  $\mu$ M), 0.15  $\mu$ L *Taq* DNA Polymerase (5U/ $\mu$ L) and 2  $\mu$ L DNA template (20 ng/ $\mu$ L). The PCR amplification program consisted of: a first step of denaturation at 95 °C for 3 min; followed by 30 cycles of 30 s at 95 °C, 30 s at 65 °C and of 30 s at 72 °C; and a last step of extension at 72 °C for 5 min. Subsequently, the PCR products were diluted in formamide and analyzed with an automated ABI PRISM 3100-*Avant* DNA sequencer, with a GeneScan 600LIZ (Applied Biosystems, California, USA) size standard. The fragments were analyzed using GeneScan software (Applied Biosystems) to obtain the electropherograms and polymorphisms were analyzed with Genotyper DNA Fragment Analysis software (Applied Biosystems).

#### *SNP marker analysis*

SNPs were validated in a LightCycler 480 Real-Time PCR (Roche, Basel, Switzerland). The reactions were performed in a 10  $\mu$ L volume comprising 5  $\mu$ L Master Mix (2 $\times$ ), 0.8  $\mu$ L MgCl<sub>2</sub> (25 mM), 0.25  $\mu$ L each primer, 1.7  $\mu$ L water and 2  $\mu$ L DNA (30 ng/ $\mu$ L) with the following touchdown PCR program: denaturation at 95 °C for 10 min; followed by 55 cycles of 10 s at 95 °C, 15 s at 65 °C (decreasing 1 °C each cycle until 55 °C) and 15 s at 72 °C; and a final melting step of 1 min at 95 °C, 1 min at 40 °C and 1 s at 60 °C with the temperature being raised by 0.02 °C/s until 95 °C was reached.

*GBS analysis*

GBS was performed in 83 plants comprising 73 pre-selected plants from different generations (BC5S1, BC4S1 and BC2S1), 2 samples of each parent, 2 additional samples of doubled haploids of the recurrent parent and 2 samples of the hybrid. After DNA extraction and dilution, the samples were sent to the Cornell University genomic facilities in August 2015 for library preparation (a PstI enzyme was used) and sequencing. The raw data were demultiplexed and the Illumina barcode removed using the software 

GBS	Barcode	Splitter
-----	---------	----------

 (<https://sourceforge.net/projects/gbsbarcode/>). Subsequently, the sequences were filtered at a 25 Phred quality score and the Illumina adapters removed with the Fastq-mcf program (Aronesty 2013). A quality control before and after the trimming was performed with the FastQC tool (Andrew 2010) and general statistics were obtained using Fastq-stats (Aronesty 2013). Then, the clean reads were mapped against the reference published genome (Hirakawa *et al.*, 2014) and against another eggplant genome with improved quality, kindly provided by the Italian Eggplant Genome Sequencing Consortium (Barchi *et al.*, 2016). Due to the better assembly of the latter (12 pseudomolecules versus 33,873 scaffolds of the former), we decided to use it for the further analysis. The mapping was performed using BWA-MEM software (Li 2013) with default settings and the SNP calling using the FreeBayes program (Garrison and Marth 2012), ignoring those variants that presented less coverage than 15, a mapping quality of 20 or a base quality of 20. In order to select the most reliable SNPs for genotyping the plants with HRM, different filters from the Seq Crumbs tools ([https://bioinf.comav.upv.es/seq\\_crumbs/](https://bioinf.comav.upv.es/seq_crumbs/)) were applied to

the VCF file for maximizing the polymorphism validation. The filters CS60 and CL60 identified polymorphisms that were at a distance of less than 60 nucleotides to another polymorphism or to the alignment edge, respectively. The filter HV0.05 determined whether the region had more than 5 polymorphisms per 100 bases, while the filter VKS discriminated INDELs from SNPs. GBS raw data were deposited in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra/>) repository under the accession number SAMN07249022.

### *Positioning candidate genes for drought tolerance*

Candidate genes involved in pathways strictly related to drought tolerance were gathered from the current scientific literature (Bolger *et al.*, 2014; Zhang *et al.*, 2014; Krannich *et al* 2015; Landi *et al.*, 2016). Orthologous genes of these candidates were searched for in tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) through EnsemblPlants (<http://plants.ensembl.org/index.html/>) and NCBI (National Center for Biotechnology Information; <https://www.ncbi.nlm.nih.gov/>) databases, and the corresponding Fasta sequences were retrieved. Subsequently, a Blast search (with a cut-off value of 1e-60) was performed with the Fasta sequences to position the candidate genes on the eggplant genome (Barchi *et al.*, 2016) and to assign them to the corresponding ILs.



## Results

### Development of the ABs and ILs

A total of 91 BC1 plants were obtained by crossing the F1 hybrid *S. melongena* AN-S-26 × *S. incanum* MM577 to the recurrent parent (AN-S-26), generating a highly diverse population in plant morphology and in fruit color, shape, size and prickles (Figure 1). Based on the information of the 52 molecular markers assessed, twelve BC1 plants were selected. Each of them was heterozygous for all the scored markers of the target chromosome and had a reduced percentage of donor parent genetic background in the rest of their chromosomes. The BC2 progeny, obtained from the 12 BC1 plants, were genotyped with a mean of four to five molecular markers (24 COSII and 36 SSRs). This allowed selection of 15 BC2 families, which were taken to the greenhouse in order to obtain the BC3 population. For chromosomes 5, 7 and 8, no individual plants displayed a chromosome with all scored markers completely heterozygous; thus, two complementary individuals for each of them were selected to cover all the markers for the target chromosomes. As in the previous year (2011), the selected plants were transplanted to the greenhouse at the end of October with the aim of obtaining two cycles of backcross per year (winter–spring and summer–fall). For several reasons, including seed dormancy, slow growth, poor flowering with low-quantity pollen during winter and parthenocarpy, just one cycle per year was achievable.

The BC3 progeny were genotyped with 22 COSII and 38 SSRs. Four chromosomes (1, 2, 4 and 10) were not completely heterozygous for the markers used in any of the selected plants; hence, the number of

progeny rose to nineteen. The genotyping of the BC4 plants (22 COSII and 42 SSRs) allowed selection of the progeny for the last round of backcross. Unfortunately, nine BC4 families (two from chromosome 1, two from chromosome 2, one from chromosome 6, two from chromosome 7, one from chromosome 10 and one from chromosome 11) did not produce fruit due to poor fruit flowering and setting. Thus, in order to avoid loss of these introgressions, new BC4 families for those from which we could not obtain seeds were genotyped again, together with the progenies of the BC5 families. Moreover, for chromosome 6, the few remaining seeds of the BC4 and BC3 generations did not germinate and the BC2 progeny were used to continue the process.

The first round of selfing was performed for nine BC5, nine BC4 and one BC2 families. In addition, the BC4 and BC2 families were backcrossed for a backup, due to the fact that the BC2 and most of the BC4 generations that were re-genotyped had few remaining seeds. The progenies obtained from selfing were manually genotyped with 24 COSII and 44 SSRs. Subsequently, 73 plants that were pre-selected from the BC5S1, BC4S1 and BC2S1 families were additionally sequenced through GBS. The selfed offspring of these 73 plants constituted the ABs set.

The genotyping by GBS allowed a very accurate detection of introgression length, precisely defining the recombination breakpoints and the double recombination events undetected due to the low–medium density molecular marker assisted selection used in the previous generations. The sequencing of the 83 genotypes (73 pre-selected plants from BC5S1, BC4S1 and BC2S1, 2 samples of each parent, 2 additional samples of doubled haploids of the recurrent parent and 2 samples of the F1 hybrid) generated a total of 201,181,531 reads, with a mean of 288

2,423,873 reads per genotype. After the filtering steps, more than 85% clean reads were aligned to the 12 pseudomolecules of the new eggplant genome developed by the Italian Eggplant Genome Sequencing Consortium (Barchi *et al.*, 2016). The SNP calling allowed identification of a total of 58,401 polymorphisms. Several filters were applied to the VCF file, using Seq\_crumbs software ([https://bioinf.comav.upv.es/ngs\\_crumbs/](https://bioinf.comav.upv.es/ngs_crumbs/)), in order to select the most suitable variants for assessment using the HRM technique. Although all polymorphisms matched the quality criteria, a subset of 796 was selected for the genotyping.

The mean number of SNPs per chromosome was 66, with a range of 38 (chromosomes 9 and 11) to 92 (chromosome 3). Using the subset of 796 filtered SNPs, increasing the number of markers for genotyping over elevenfold, more than 90% of introgressions were confirmed by GBS, as well as the segregation of the heterozygous regions previously analyzed by manual genotyping. A total of 36% of the introgressions was totally homozygous for the donor parent, 21.3% were still totally heterozygous, and 33.3% of them displayed a combination of homozygous and heterozygous regions in the first round of selfing (Figure 3).

Double recombinations, which were practically undetected in the previous manual genotyping, were identified in the regions with the highest density of molecular markers. In fact, the distribution of SNPs was very uneven along the chromosomes, being generally greater in the euchromatin and less in the centromeric regions. The development of the backcross generations without screening the genetic background of the selected progenies until the first round of selfing led to the dragging of non-target introgressions. In fact, 84% of the genotyped plants presented

at least one non-target introgression, with the most common case being two non-targeted introgressions per plant (37.3%), followed by one (26.6%) and three (8%) non-target introgressions per plant. Obviously, the plants selected for chromosome 6 were those that presented the highest number of non-targeted introgressions (from seven to eleven) due to the fact that they were backcrossed just twice instead of being subjected to four or five rounds of backcrosses of the others.

However, in most cases, the non-target introgressions of the 73 ABs progenitors (i.e. selected BC5S1, BC4S1 and BC2S1 plants) were smaller than 5 Mb and positioned on the edges of the chromosomes, while few others were larger, especially those of chromosome 6, and located in the centromeric regions where generally the recombination is low. Most of the 159 non-target introgressions in the 73 ABs progenitors were heterozygous (60.3%), which meant an easier elimination by segregation through selfing, while 52 non-target introgressions (32.7%) were homozygous for the donor parent, and 11 (7%) were both heterozygous and homozygous in different percentages. The negative selection against the homozygous non-target introgressions is more complicated compared to the heterozygous ones, requiring at least an additional round of backcross before the elimination by segregation in a subsequent selfing generation.

However, genetic characterization by GBS revealed the high coverage of the donor parent across the ABs progenitors (Table 1). In fact, thanks also to the non-target introgressions, the representation of *S. incanum* was around 99% in the AB materials, while the other 1% belonged to non-polymorphic regions between the two parentals. This means that no region of the donor parental had been lost during the

backcross process and all the *S. incanum* genome had been introgressed into the ABs, although some materials will need more rounds of backcrossing and selfing to produce ILs with just one introgression in homozygosis.

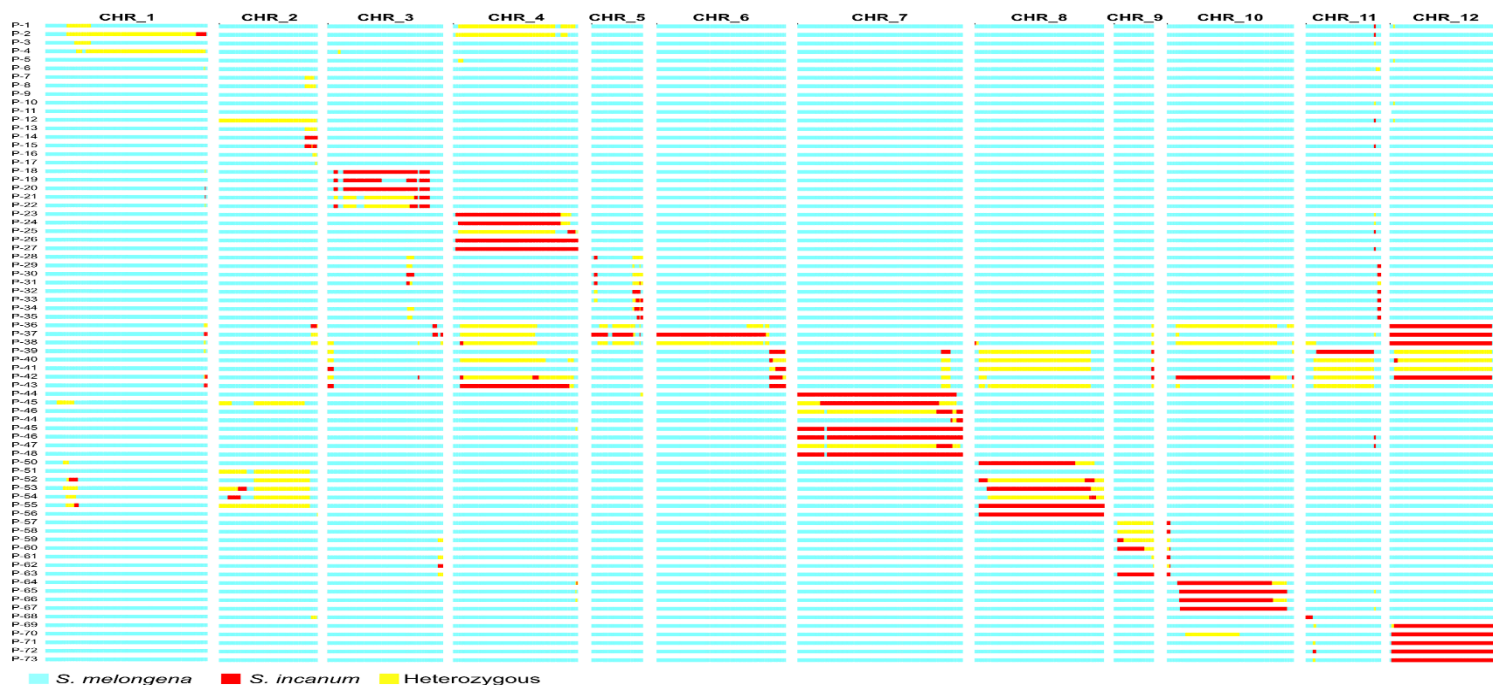
The first round of selfing and the high-throughput genotyping by GBS allowed identification of the first four fixed ILs (SMI\_2-1, SMI\_7-1, SMI\_7-3, SMI\_8-1), with two of them (SMI\_7-1 and SMI\_8-1) covering most of the entire chromosome physical size (Table 2 and Figure 4). Based on the genotype of the target introgressions (homozygous, heterozygous or both), and the number and the genotype of the non-target introgressions, selected plants of the ABs were selfed or backcrossed in order to obtain new ILs. The progenies obtained were screened with a set of 62 SNPs derived from the GBS and converted for assessment using the HRM technique, the mean marker density being higher in the euchromatin regions and lower in the centromeres. The design of the new markers was based on fixing the single-introgression ILs and reducing the number of non-target introgressions. Hence, the number of SNPs was different for each chromosome and depended on the size of the latter, resulting in four to eight SNPs per chromosome.

The genotyping of an additional generation of selfing from the ABs allowed identification of 21 more fixed single-introgression homozygous ILs, raising the total set of ILs to 25 different lines (Table 2). On average, the ILs contained 4.3% of the donor parent genome, ranging from 0.1% to 10.9%, altogether covering 61.7% of the *S. incanum* genome. The mean size of the introgressions was 49.5 Mb (ranging from 2 to 125 Mb), encompassing between 1.8% and 97.2% of the target chromosome length. Chromosome 2 presented the largest number of ILs (4), while chromosome

6 had none. The most covered chromosome was 8, with 97.2% of the total length, and the least covered among those having at least one IL was chromosome 10, with 1.8%.

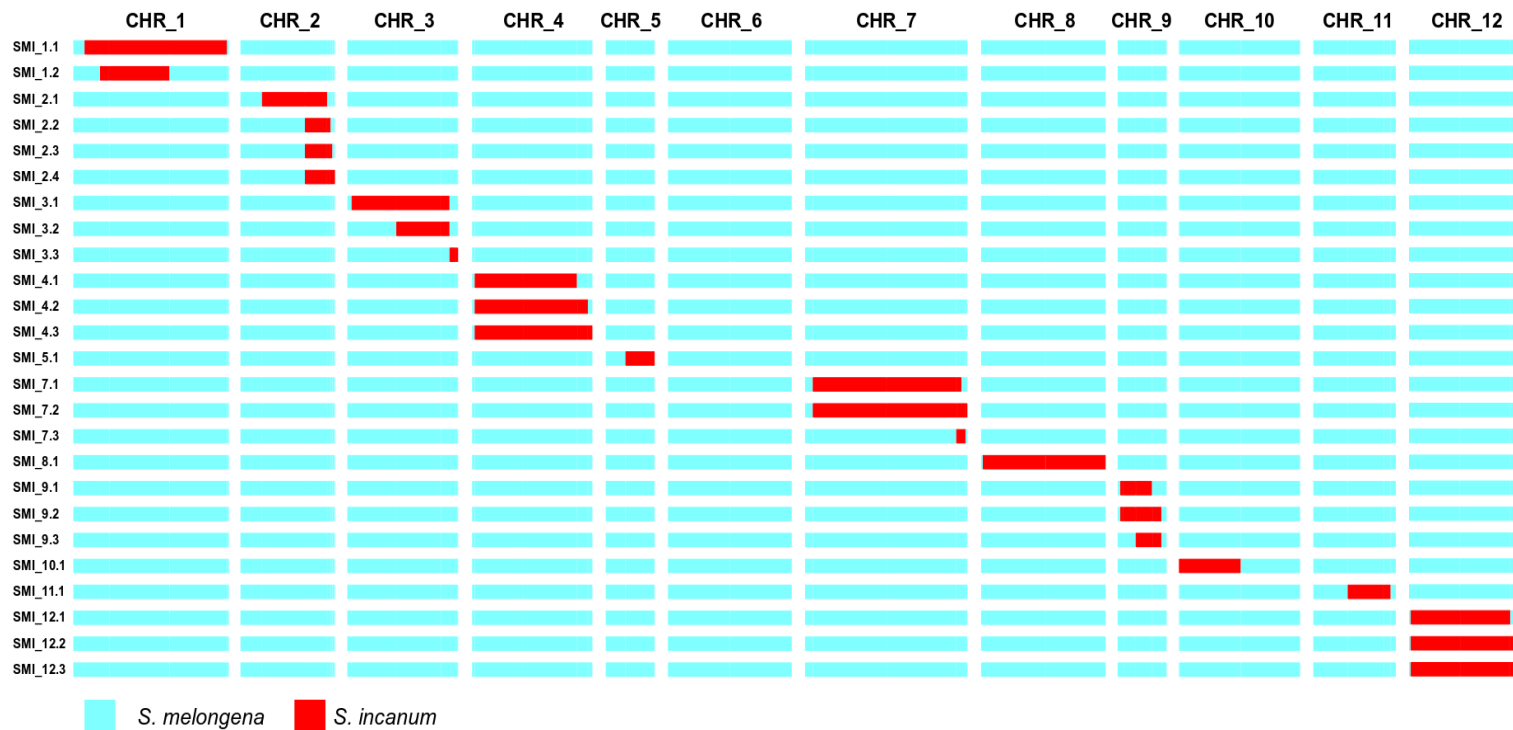
**Table 1.** Statistics of *Solanum incanum* MM577 genome coverage in the 73 progenitors (corresponding to selected plants of the BC5S1, BC4S1 and BC2S1 generations) of the advanced backcrosses (ABs) set, after the genotyping by genotyping-by-sequencing (GBS). *Solanum melongena* AN-S-26 regions were assigned when no heterozygous or *S. incanum* introgressions were present, while heterozygous regions were assigned when no donor parent introgressions were present at the same time. All other regions were considered as *S. incanum* regions.

Chrom.	Chrom. size (Mb)	<i>S. incanum</i> size (Mb)	<i>S. incanum</i> (%)	Heter. regions size (Mb)	Heterozygous (%)	<i>S. melongena</i> size (Mb)	<i>S. melongena</i> (%)
1	136	20.8	15.3	109.7	80.7	5.5	4.0
2	83	25.1	30.2	57.9	69.8	-	-
3		92.4	95.2	4.6	4.8	-	-
4	105	105	100.0	-	-	-	-
5	43	43	100.0	-	-	-	-
6	108	108	100.0	-	-	-	-
7	142	142	100.0	-	-	-	-
8	109	106.8	97.9	2.2	2.1	-	-
9	43	43	100.0	-	-	-	-
10	106	95.4	90.1	5.6	5.2	5.0	4.7
11	72	72	100.0	-	-	-	-
12	100	100	100.0	-	-	-	-
Total	1144	953.5	83.4	180.0	15.7	10.5	0.9



**Figure 3.** Graphical genotypes of the selected plants genotyped by genotyping-by-sequencing (GBS) after four to five (two for P-36 to P-43) rounds of backcross (typically five, but two or four for some genotypes) and one round of selfing, using *Solanum incanum* MM577 as the donor parent in the genetic background of *S. melongena* AN-S-26. The first row indicates chromosome number and the first column the plant code. Homozygous introgressions of the donor parent are depicted in red, while the genetic background of the recipient parent is depicted in blue and the heterozygous regions in yellow.





**Figure 4.** Graphical genotypes of ILs developed using *Solanum incanum* MM577 as the donor parent in the genetic background of *S. melongena* AN-S-26. The first row indicates the chromosome number and the first column the IL code. Homozygous introgressions of the donor parent are depicted in red, while the genetic background of the recipient parent is depicted in blue.

## Candidate genes for drought tolerance

Due to the lack of genomic and transcriptomic studies in the eggplant genepool, candidate genes from different crops related to drought tolerance were retrieved from the current scientific literature in order to detect the potential genomic region in eggplant and in the recently developed ILs. A total of 142 candidate genes gave a positive Blast hit (cut-off value of  $1e-60$ ) in the eggplant genome, most of them involved in the abscisic acid biosynthesis, catabolism and signaling network, in ethylene synthesis and signaling, in osmolyte biosynthesis and cell protection, in aquaporins and detoxification regulation (Supplementary material 1). The chromosomes containing most candidate genes for drought tolerance were chromosome 12 with 28 genes, chromosome 5 with 20 genes and chromosome 10 with 18 genes. The *S. incanum* alleles of these candidate genes were present in the AB materials, while 68 of them were present in the IL set. For the IL set, SMI\_12.3 was the line carrying the most candidate genes (18), followed by SMI\_12.2 (14 genes), and SMI\_5.1, SMI\_8.1 and SMI\_12.1 (11 genes each) (Table 2).

**Table 2.** Statistics of the *S. incanum* MM577 introgression lines (ILs) in the genetic background of *S. melongena*

AN-S-26 using a 12 pseudomolecule eggplant genome (Italian Eggplant Genome Sequencing Consortium; Barchi *et al.*, 2016), and a number of candidate genes for drought tolerance. The positions and size of the ILs are represented in megabases (Mb) according to the physical position in each chromosome (Chrom.).

ILs	Chrom.	Donor parent (%)	IL size (Mb)	IL position (Mb)	Chrom. IL size (%)	Chrom. total ILs size (Mb)	Chrom. total ILs size (%)	Drought-tolerance related genes <sup>a</sup>
SMI_1.1	1	9.9	114	19–133	83.8	114	83.8	-
SMI_1.2		0.7	9	27–36	6.5			-
SMI_2.1	2	3.2	37	38–75	44.5	43	51.8	1
SMI_2.2		0.1	2	75–77	2.4			-
SMI_2.3		0.4	5	75–80	6.0			2
SMI_2.4		0.5	6	75–81	7.2			3
SMI_3.1	3	6.9	79	7–86	81.4	82	84.5	7
SMI_3.2		0.6	8	78–86	8.2			2
SMI_3.3		0.2	3	93–96	3.0			1
SMI_4.1	4	7.0	81	4–85	75.2	101	96.1	3
SMI_4.2		8.2	94	4–98	89.5			4
SMI_4.3		8.8	101	4–105	96.1			6
SMI_5.1	5	0.6	8	35–43	18.6	8	18.6	11
SMI_7.1	7	10.5	121	14–135	85.2	125	88.0	4
SMI_7.2		10.9	125	14–139	88.0			7
SMI_7.3		0.8	10	129–139	7.0			4
SMI_8.1	8	9.2	106	3–109	97.2	106	97.2	11
SMI_9.1	9	1.8	21	5–26	48.3	29	64.4	3
SMI_9.2		2.5	29	5–34	64.4			3
SMI_9.3		0.7	9	26–34	20.9			1
SMI_10.1	10	0.1	2	0–2	1.8	2	1.8	-
SMI_11.1	11	0.2	3	60–63	4.1	3	4.1	-
SMI_12.1	12	7.2	83	3–86	83.0	93	93.0	11
SMI_12.2		7.7	89	3–92	89.0			14
SMI_12.3		8.1	93	3–96	93.0			18
Mean		4.3	49.5		48.2	64	62.1	5.8
Total						706		68

## Discussion

*S. incanum* is a close relative of eggplant and its hybrids and backcrosses with *S. melongena* are fully fertile (Knapp *et al.*, 2013; Kouassi *et al.*, 2016; Plazas *et al.*, 2016).

The development of the ABs and ILs of *S. incanum* in the genetic background of *S. melongena*, in its early stages, was characterized by a lack of information available for genotyping the progenies efficiently. This led, at the beginning of the process, to the use of markers for low–medium throughput genotyping from information obtained in other eggplant studies and to the generation of new molecular markers more suitable for our materials. Although in general some molecular markers can be useful and polymorphic across species (Saha *et al.*, 2014; Barbará *et al.*, 2007), their identification and amplification can be very challenging. In fact, the markers used for genotyping the BC1 generation were developed in other eggplant segregating populations (Nunome *et al.*, 2009; Wu *et al.*, 2009), although they were progressively substituted by markers identified in our population. From the initial set of markers, just one fifth were polymorphic in our BC1 population, which led to a low saturation of some chromosomal regions, which reduced the efficiency for selecting the most appropriate genotypes.

Our strategy for this first step was to select 12 BC1 plants, each one with one chromosome completely heterozygous for the markers from the donor, and at the same time with a higher percentage of the recurrent parent background in the other 11 chromosomes. This approach, consisting of the development of a set of chromosome substitution lines, accelerated the process of recovering the recipient parent background using a small set of

selected plants, avoiding the tracking many families until the first round of selfing.

Unfortunately, this strategy did not last long, because, during the selection of the BC2 generation, no plants with complete selected chromosomes heterozygous for the markers screened were found in the progeny of three out of the twelve BC1 plants. The same drawback occurred also in further selection steps in the BC3 and BC5 generations. In addition, to reduce the time and resources needed during selection, we genotyped the plants only for the target chromosome instead of for the whole set of chromosomes. The strategy was to develop ILs without screening the genetic background, for the sake of time and resource saving, until AB generations or even selfing rounds were performed also by other authors (Eichten *et al.*, 2011; Pea *et al.*, 2013). In fact, no affordable high-throughput genotyping platforms or SNP arrays were available in eggplant to genotype a large number of plants with a wide set of SNPs in a very short time, as in the major crops like maize (Ganal *et al.*, 2011), wheat (Wang *et al.*, 2014), tomato (Sim *et al.*, 2012a) and potato (Vos *et al.*, 2015); so, in our case, all the markers were assessed manually using low–medium throughput genotyping. Moreover, more than half of the markers assessed were SSRs (24 COSII and 36 SSRs), which are very robust and reliable (Varshney *et al.*, 2005), but compared to SNPs are generally more difficult to automate and more time-consuming and expensive, due to the need of detection through an agarose or polyacrylamide gel or capillary sequencing (Jones *et al.*, 2007). Some authors, Yang *et al.*, (2011), calculated that assessing SNPs through a high-throughput genotyping platform can be 100-fold faster and 75% less expensive than SSR gel-based or capillary methods.

All these combined circumstances made the genotyping and selection steps the bottlenecks in the AB and IL development process, more than the interspecific crossing of two genetically distant species, which probably discouraged attempts to start sets of ILs with eggplant genepool materials. In addition, in part due to the excessive amount of time needed for the genotyping and selection steps, and also due to difficulties in obtaining seeded fruit set during the winter–spring season (e.g. slow plant growth, poor flowering with reduced production of pollen, low pollen viability and parthenocarpy), just one cycle of backcrossing per year was feasible. This meant a very long process of eight years has been required to obtain a first set of ILs.

For other crops, where two or even three growing cycles per year are possible, the entire set of ILs could be obtained in three or four years (Eduardo *et al.*, 2005; Barrantes *et al.*, 2014). The strategy to bypassing the screening of the non-targeted chromosomes of the selected plants, even if it has the advantage of reducing the time and the cost for genotyping, increases the number of rounds of backcrosses required to recover the almost totality of the recipient parent background (Frisch *et al.*, 1999). In fact, although generally five rounds of backcross are recommended to recover more than 98% of the recurrent parent, due to the fact that the donor genome proportion is halved with each backcross generation, an adequately high percentage of the recipient parent can be recovered after three generations of marker-assisted backcrossing (Tanksley *et al.*, 1989; Hospital *et al.*, 1992; Hospital and Charcosset 1997), although this depends on the number of plants screened and the number of markers adequately distributed along the genome (Frisch and Melchinger 2005).

High-throughput genotyping in the first generations of

backcrossing not only accelerates the process of obtaining ILs, but also provides early detection of non-target introgressions and easy elimination of them during the backcross process. When the genetic background is not screened until advanced generations of backcrossing or even after the final collection of ILs, the presence of non-target introgressions is almost certain, as happened in several studies (Shen *et al.*, 2001; Eichten *et al.*, 2011; Pea *et al.*, 2013; Arbelaez *et al.*, 2015). However, non-target introgressions can go unnoticed, until a massive genotyping is performed, even when a low-density genotyping is performed during the IL development process, due to double recombination events undetected by two adjacent markers (Xu *et al.*, 2010; Schmalenbach *et al.*, 2011; Zhang *et al.*, 2011).

Our case was similar to the first scenario. In fact, the genetic background of the progenies was not screened until the first round of selfing when a GBS was performed. The majority of the plants genotyped contained at least one non-target introgression (84%), although most of them were small and regularly placed in certain regions, which mostly corresponded with low-recombination regions. This finding allowed the characterization of the ABs, but caused indirectly a further delay in the development of a set of nearly isogenic lines due to the necessity to clean the genetic background of the non-target introgressions.

Given the high representation of the donor genome in the AB materials, the removal of non-target introgressions will allow obtainment of a complete set of single-introgression fixed ILs where more than 95% of the donor genome is represented. Also, double recombinants were identified by GBS both in target and non-target introgressions. The medium-density genotyping applied during the population development

did not allow detect of all the double-crossing-over events, mainly in the euchromatin regions where the recombination rate is higher (Wang *et al.*, 2006; Gaut *et al.*, 2007). In fact, although the selection of the markers before GBS was based on the genetic map, which tends to represent more the euchromatic regions and less the centromeric regions (Chen *et al.*, 2006, Zhang *et al.*, 2014), some high-recombination regions were not properly saturated due to the fact that the map was already under development. Then, after the GBS, for a more efficient genotyping design, the selection of the new SNPs was based on a combination of the genetic and physical position, plus the necessary markers to screen the non-target introgressions.

Selection based just on genetic maps could under-represent telomeric regions and drag large centromeric regions (Wang *et al.*, 2000), as happened in our case, while a selection based exclusively on physical maps tends to reduce the adequate marker density of high-recombination regions, running the risk of not detecting double recombinants and of over-representing centromeric regions (Barrantes *et al.*, 2014), which is not cost-effective due to the low-recombination ratio. The most efficient approach for reducing the time for the IL population development, avoiding most of the non-target introgressions and double recombinants, and detecting regions undetected by low–medium genotyping, could be to perform high-throughput genotyping in the first cycles of backcrossing. In this way, most of the ILs can be developed just in four generations, as was the case for Barrantes *et al.*, (2014), who developed a set of tomato ILs with *Solanum pimpinellifolium* introgressions, where most of the lines derived from a BC3S1 generation. Even though a few years ago the cost of massive genotyping techniques was high and prohibitive for many labs,



currently, the decreasing cost of sequencing and the emergence of new genotyping technologies, like GBS and similar (Campbell *et al.*, 2015), have made the task of genotyping more affordable and precise. This makes possible the development of experimental populations for non-model crops for which genomic resources are limited (Sallam *et al.*, 2015).

The ABs and ILs developed may be of great interest not only for dissecting the genetic basis of domestication traits, but also for developing varieties of eggplant with new traits, including complex ones like drought resistance, introgressed from *S. incanum*. Drought is considered by many experts the stress that presents the strongest negative impact on yield and crop productivity (Krannich *et al.*, 2015). Plants have developed different strategies, involving several different genes and pathways, to mitigate the negative effects of drought (Reguera *et al.*, 2012). In this respect, *S. incanum*, being distributed from northern Africa to the Middle East, is able to survive in desert and dryland areas (Ranil *et al.*, 2016; Vorontsova and Knapp 2016) and, therefore, is a potential source of variation for tolerance to drought, as demonstrated by Savarino (2017). However, almost nothing is known about the mechanisms and strategies used by *S. incanum* to tolerate this abiotic stress. For the major crops, like rice and wheat (Tang *et al.*, 2016; Valluru *et al.*, 2016), several mechanisms have been described involving many genes and pathways required for tolerance to drought. In this study, we have identified 142 eggplant genome regions that are orthologs of candidate genes related to drought tolerance in other crops; while all these genome regions are represented in the AB materials, 68 of these regions are present in the ILs newly developed using a donor parent (*S. incanum* MM577) that is much more resistant to drought than the recipient parent (*S. melongena* AN-S-26) (Savarino 2017). Some of these

ILs are very promising, carrying several orthologs related to drought tolerance from different crops. Monitoring the expression of these regions during a screening test for drought tolerance using the AB materials and the subset of ILs could pave the way for understanding the regulatory mechanisms for drought tolerance in eggplant. At the same time, it could facilitate the development of new eggplant varieties in a scenario when drought and desertification are increasingly a serious constrain for crop production.

A preliminary drought screening test realized by Savarino (2017), using intermediate backcross materials during the IL development, showed significant drought-tolerance levels among the parents, hybrid and the backcross materials. Although the results were promising, the screening was exploratory and a further analysis is required. In addition, Sotomayor (2016) realized a phenotypic characterization of 62 AB materials for morphological traits related to fruit, flower, prickles and chlorogenic acid content. A great variation among the materials was found for almost all the traits (Supplementary material 2). Once again, these results are preliminary and more extensive analyses are being undertaken to discover QTLs and dissect the genetic basis of important agronomic traits. In the meantime, and without any further delay, the collection of ABs and ILs is available for the scientific community.

In conclusion, in this study, using the available genomic information (reduced at the beginning and more comprehensive at the later stages), we have developed the first set of ABs and ILs in eggplant that cover the whole genome of a wild donor species (*S. incanum*). These materials will have a tremendous impact in a wide range of studies in this genepool, which was lagging behind compared to other important crops of

the same genus, like tomato and potato. In fact, precision QTL mapping in the ABs and ILs will allow confirmation of existing QTLs (Buerstmayr *et al.*, 2011; Sayed *et al.*, 2012) and detection of new ones, which has been complicated until now by the lack of an experimental population for this crop. In addition, new genetic diversity has been introduced in a crop that had suffered a reduction in the genetic base due to repeated genetic bottlenecks during the domestication process (Muñoz-Falcón *et al.*, 2009; Meyer *et al.*, 2012a). The ABs and ILs developed could be used directly by breeders in their breeding pipelines, as most of them do not display undesirable wild traits because they just have a small percentage of the donor parent genome. We have provide further evidence that the implementation of high-throughput genotyping technologies can have an extraordinary impact on the development and precision of IL development, reducing the time and, probably, the cost of obtaining this segregating population (Schmalenbach *et al.*, 2011; Arbelaez *et al.*, 2015). Also, the ABs and ILs developed up to now will allow isolation of improved versions of the IL population, which will further contribute to the enhancement of eggplant breeding.

## *Acknowledgements*

This work has been funded in part by the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives", which is supported by the Government of Norway. This project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew and implemented in partnership with national and international gene banks and plant breeding institutes around the world. For further information see the project website: <http://www.cwrdiversity.org/>. Funding has also been received from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) and from Spanish Ministerio de Economía, Industria y Competitividad and Fondo Europeo de Desarrollo Regional (grant AGL2015-64755-R from MINECO/FEDER). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral (Programa FPI de la UPV-Subprograma 1/2013 call) contract. Mariola Plazas is grateful to Ministerio de Economía, Industria y Competitividad for a post-doctoral grant within the Juan de la Cierva programme (FCJI-2015-24835). Giulio Mangino is grateful to Conselleria d'Educació, Investigació, Cultura i Esport de la Generalitat Valenciana for a pre-doctoral grant within the Santiago Grisolí programme (GRISOLIAP/2016/012). Authors also thank the Italian Eggplant Genome Sequencing Consortium for providing access to an improved version of the eggplant genome.





# **GENERAL DISCUSSION**





Genetic and genomic resources are crucial to develop new improved varieties that have to deal with the expanding human population in a changing climate scenario and under the constant threat of new biotic stresses. By 2050, the amount of crop production required to feed a burgeoning population might need to be doubled compared to the levels of 2010 (Tilman et al., 2011). This unprecedented challenge will have to be addressed in conjunction with a forecast of constantly deterioration and aridification of cultivation environments, with the intensification of the onset of pests and diseases and the exacerbation of abiotic stresses (drought, high temperatures, salinity) due to the substantial impact of climate change (Bebber et al., 2013, Fita et al., 2015).

In addition, domestication processes and modern plant breeding have reduced the genetic diversity of cultivated species, which is generally quite narrow and might be insufficient to achieve the dramatic increase of yields required in agriculture in the future decades (van de Wouw et al., 2010; Meyer and Purugganan, 2013; Dempewolf et al., 2014). To overcome this challenging scenario, plant breeders are pointing at crop wild relatives (CWRs) as sources of novel beneficial variation for resistance or tolerance to stresses, using new approaches like "introgressomics" that we proposed to exploring and exploiting novel genetic diversity from germplasm collections and CWRs. In the last fifteen years, many breakthroughs have revolutionized plant breeding. Undoubtedly, the advent of next-generation sequencing (NGS) and subsequent platforms has been one of the most relevant events that changed the way to improve crops. In addition, the continuous decrease of the sequencing costs has made this revolution affordable to almost all

research groups and in the last few years it is bringing high benefits also for non-model organisms and neglected crops.

In fact, before 2014 common eggplant had few genetic and genomic tools and resources available to assist breeders in their objectives and to face old and new threats and challenges. When this Ph.D. thesis has begun, in September 2013, eggplant breeders could just rely on a few molecular markers, identified mainly from intraspecific crosses through genetic mapping, which most of the times were not sufficiently informative in other germplasm backgrounds. Furthermore, the use of cultivated relatives (*S. aethiopicum* and *S. macrocarpon*) and CWRs for common eggplant breeding had been insignificant (Rotino et al., 2014).

For all these reasons, this Ph.D. thesis was strategically designed as a multi-pronged and interdisciplinary project, combining disciplines like genetics, genomics, plant breeding, to provide genetic and genomic tools for different stakeholders to foster the development of new and resilient varieties exploiting the valuable genetic diversity of CWRs.

### ***Genomic information available in the eggplant genepool***

As remarked in our review (Gramazio et al., 2017), few genomic studies have been performed until the last years in eggplant genepool despite the economic importance of common eggplant. We decided to contribute with genomic information from the sequencing and the de novo assembly of two related species of common eggplant with great interest for breeding, *S. incanum* and *S. aethiopicum* (Cappelli et al., 1995; Collonnier et al., 2001a,b; Rizza et al., 2002; Stommel and Whitaker, 2003; Gisbert et al., 2011; Prohens et al., 2013, Plazas et al., 2014a,b).

In order to capture a higher representation of the whole transcriptome and increase the diversity of transcripts, RNA was extracted from different tissues (young leaves, floral buds, and young fruits) and mixed. More than 100 million raw reads per species were obtained from sequencing and deposited in the SRA-NCBI database (bioproject SRP063088). Those sequences constituted the first genomic resources in both species. The de novo assembly generated more than 100,000 transcripts per species and 83,905 unigenes for *S. incanum* and 87,084 unigenes for *S. aethiopicum*, showing similar values with the most recent published transcriptomes. Compared to the available transcriptomes of *S. melongena* and *S. torvum* (Yang et al., 2014), assembled in 34,174 and 38,185 unigenes respectively, our assemblies exhibited a higher representation of unigenes/transcripts, due to a better coverage and higher sequencing data.

The unigenes were structurally and functionally annotated based on comparison with public databases and using bioinformatics tools. Open reading frame (ORF) and introns regions were detected for the structural annotation, providing valuable information about gene structure and primers design, while protein-coding genes, GO terms, EC numbers and KEGG biological pathways were assigned for the functional annotation. In general, the unigenes annotated were between 30,000 and 35,000, very similar to *S. melongena* and *S. torvum* and to tomato and other *Solanum* species, which confirmed the high quality of our assembly and annotation (Tomato Genome Consortium, 2012, Yang et al., 2014).

In this study, we have proved how the new NGS technologies can enhance and increase the development of high-quality genomic information in non-model species with a limited amount of resources. The

genomic information generated will pave the way for a broad range of studies, from identification of candidate genes for traits of interest to biological pathways, as well as, domestication and comparative mapping with related species to enhance and promote breeding in the eggplant genepool.

### ***Molecular markers identification***

Before the advent of NGS technologies, the development of molecular markers was not a straightforward task. Several types of markers have been developed during the last decades (e.g., isozymes, RFLPs, RAPDs, ISSRs, AFLPs, etc.), but for most of them, their characteristics were too far from what an excellent molecular marker has to be. In fact, an ideal marker should be codominant, polymorphic, highly informative, robust, reliable, abundant, ubiquitous, inexpensive and easy to automate (Varshney et al. 2005). Two more recent types of markers, SNPs and SSRs, present the majority of these features and have practically replaced the rest of markers. In addition, they can be easily identified in high-throughput sequencing and genotyping projects.

Prior to the start of this thesis, the markers of this type publicly available for the eggplant community were collections of SSRs developed from genomic libraries for intraspecific genetic mapping (Nunome et al., 2003b; Stàgel et al., 2008; Nunome et al., 2009; Vilanova et al., 2012), whose development required a quite high degree of expertise and a significant investment of time and resources. On the contrary, few SNPs were available, which generally are more easy to automate and less

expensive than SSRs (Jones et al., 2007; Yan et al., 2010), and almost no markers available in eggplant relatives and CWRs.

We designed a strategy to identify a large set of SSRs in two eggplant relatives, *S. incanum* and *S. aethiopicum*, and test their usefulness and versatility across other species of eggplant genepool. In addition, we planned to detect a comprehensive set of SNPs from transcriptome in both species and test part of them in the same eggplant related materials of the SSRs. In that way, we intended to make a comparison between the two marker types to test which one was more suitable for eggplant related fingerprinting and for establishing relationships within the eggplant genepool (the latter will be discussed in the following paragraph).

Apart from that, we implemented a modified genotyping-by-sequencing approach to discover polymorphic sites for characterizing a collection of 76 accessions of cultivated and eggplant CWRs.

After sequencing and assembly the transcriptomes and performing a comprehensive annotation, 976 and 1,278 SSRs were identified in *S. incanum* and *S. aethiopicum*, respectively. In order to know their physical position and their position in the transcripts (ORFs, 3'-UTR and 5'-UTR), the unigenes that contained the microsatellites, were blasted against the reference genome.

On the other hand, the SNP calling was performed for *S. incanum* and *S. aethiopicum* and for the *S. melongena* and *S. torvum* transcriptomes already available (Yang et al., 2014). A total of 12,396 and 164,127 intraspecific SNVs (SNPs and INDEL) were identified for *S. incanum* and *S. aethiopicum* and 2,780 and 25,147 intraspecific SNVs for the *S. melongena* and *S. torvum*, respectively. Subsequently, comparisons between two species or three species at the same time were performed to

identify highly informative interspecific polymorphisms suitable for genotyping assays, either manually or with high-throughput platforms. Hence, apart from generating valuable genomic information, the sequencing of the transcriptome of a species provide a large set of molecular markers that are very helpful in marker-assisted selection (MAS) breeding.

Regarding the high-throughput genotyping of 76 accessions of cultivated and eggplant CWRs, we were able to identify 75,399 polymorphic sites, 12,859 of them were associated with the coding sequence. Apart of discovering a large set of interspecific markers, we demonstrated that the massive genotyping represents a more effective approach in terms of accuracy and time and resource consumed with respect to using a small set of markers assessed manually for understanding genetic diversity and establish relationships among and within the species of the eggplant genepool.

For both approaches, transcriptome sequencing and high-throughput genotyping, we confirmed their great efficiency for identification large sets of molecular markers in non-model species like eggplant in order to perform different kinds of studies, from gene annotation to phylogeny, or to use them for MAS selection.

### ***Diversity evaluation and the establishment of relationships in eggplant genepool***

Several authors using small sets of different types of markers (chloroplast DNA RFLPs, ISSRs, AFLPs, nrITS sequences, etc) have examined the complex relationships among the species belonging to the eggplant genepool (Sakata et al. 1991; Sakata and Lester, 1997; Furini and

Wunder, 2004; Levin et al. 2006; Isshiki et al. 2008; Weese and Bohs, 2010; Tumbilen et al. 2011; Meyer et al. 2012; Vorontsova et al. 2013; Särkinen et al. 2013). The wide discrepancy in some results led us to design a study screening a large set of markers in phylogenetically distant eggplant materials. However, before performing the study, we aimed at choosing an optimal marker for this specific case.

SSRs and SNPs, as indicated above, have practically replaced the markers mentioned above in genetic and genomic studies; thus, we considered them for the selection of the most suitable marker for genetic fingerprinting, diversity evaluation and establishment of relationships in eggplant. From the set of markers identified in the transcriptome sequencing study of *S. incanum* and *S. aethiopicum*, we selected a subset of 11 SSRs and 35 SNPs, all of them polymorphic between the two species. Those markers were assessed in a collection of 48 accessions belonging to the three cultivated eggplant species (*S. melongena*, *S. aethiopicum* and *S. macrocarpon*) and their respective wild relatives (*S. incanum*, *S. anguivi* and *S. dasyphyllum*).

Apart from our main objective, we could confirm the reliability and transferability of markers obtained *in silico* from transcriptomes to related species for genetic fingerprinting. Regarding the comparison of SSRs and SNPs, different results were obtained in the establishment of the relationships among the cultivated eggplants, suggesting each marker type sampled different levels of genetic variation. In this respect, we observed a moderate correlation between SSR and SNP genetic distances ( $r^2=0.592$ ) for low and intermediate values of pair-wise SNP genetic distances, but a lower correlation ( $r^2=0.080$ ) for high SNP genetic distances. These results indicated that for close phylogenetically accessions, both markers

provided a similar level of information, while for materials are more distant, SNPs provide a better resolution since SSRs reached a saturation for higher genetic distance.

In the light of these results, we decided to select the SNP as marker type for the relationships and diversity study, apart from for its lower price and greater ease to automate (Yan et al., 2010). Using a modified GBS (genotyping-by-sequencing) approach we performed a massive genotyping of 76 accessions belonging to 17 species from the primary (*S. insanum*, *S. incanum*), secondary (*S. aethiopicum*, *S. macrocarpon*, *S. anguivi*, *S. dasyphyllum*, *S. campylacanthum*, *S. lichtensteinii*, *S. lidii*, *S. linnaeanum*, *S. tomentosum*, *S. vespertilio*, *S. violaceum*) and tertiary genepool (*S. eleagnifolium*, *S. sisymbriifolium* and *S. torvum*) of common eggplant (*S. melongena*, inclusive).

The sequencing produced a 225 million total raw reads, with an average of 2.7 million reads per accession that allowed identifying 75,399 polymorphisms among the accessions, 12,859 of them in CDS regions. Genetic relationships analysis and population structure suggested an assignment of the species to four main sub-groups. Sub-group 1 included the accessions belonging to scarlet eggplant (*S. aethiopicum*) and *S. anguivi*, which it is considered its wild ancestor (Lester and Niakan, 1986; Lester and Thitai, 1989; Syfert et al., 2016). Sub-group 2 consisted of accessions of the common eggplant complex, the accessions of its wild relative (*S. insanum*) (Knapp et al., 2013) and of *S. incanum*, *S. lichtensteinii* and *S. linnaeanum* accessions. Sub-group 3 included the accessions of gboma eggplant (*S. macrocarpon*) and its wild relative (*S. dasyphyllum*) (Bukanya and Carasco, 1994; Syfert et al., 2016) and finally, sub-group 4 comprised the New World species from the tertiary genepool



(*S. eleagnifolium*, *S. sisymbriifolium* and *S. torvum*). The remaining accessions of *S. campylacanthum*, *S. violaceum*, *S. lidii*, *S. vespertilio*, *S. tomentosum* were classified as admixed.

Furthermore, to disclose the relationships in each sub-group, an additional hierarchical structure analysis was performed using partitioned data (detailed results in Acquadro et al., 2017). This clustering was also confirmed by a PCoA analysis, where the first two axes explained 71.4% of the genetic variation (57.6% the first axis and 13.8% the second one). The most relevant result was the closer genetical proximity of the African cultivated gboma eggplant to the common eggplant rather than the scarlet eggplant. The results in literature, based on different types of markers and with few markers assessed, are contrasting. Some of them using chloroplast DNA RFLPs, ISSRs, AFLPs and nrITS sequences (Sakata et al., 1991, Furini et al., 2004; Meyers et al., 2012) reported that *S. aethiopicum* is closer to *S. melongena* than *S. macrocarpon*, others using several nuclear and plastid DNA sequences (Furini and Wunder, 2004; Levin et al., 2006; Weese and Bohs, 2010; Särkinen et al. 2013) affirmed that *S. aethiopicum* and *S. macrocarpon* are phylogenetically closer among them than to *S. melongena*. On the other hand, Sakata and Lester (1997) and Vorontsova et al. (2013) assessing chloroplast DNA RFLPs and ITS, waxy and trnT-F regions obtained results similar to ours.

On the contrary, studies based on the species inter-fertility affirmed that interspecific hybrids between common and gboma eggplant are more challenging to obtain and with higher level of sterility compared to common and scarlet eggplant (Daunay et al., 1991; Bletsos et al., 1998, Prohens et al., 2012; Khan et al., 2013; Rotino et al., 2014), probably due to some chromosomal rearrangement or hybridization barriers.

Interestingly, our group achieved an interspecific hybrid between *S. melongena* and *S. dasyphyllum*, the wild ancestor of *S. macrocarpon*, as well as three generations of backcross towards *S. melongena* (Kouassi et al., 2016).

These results suggested that SNP is the most suitable and convenient marker type for establishing relationships and estimate the genetic diversity among the species belonging to the eggplant gene pool. In fact, we demonstrate that massive genotyping with a large set of SNPs represented a more effective and accurate approach with respect to the ones based on limited sets of markers. We established an exhaustive and detailed evaluation of the natural allelic diversity and genetic relationships among and within wild and cultivated species belonging to eggplant complexes using three different approaches (hierarchical population structure, UPGMA-based dendrogram, and PCoA analysis). The relevance of these findings is essential and crucial both for eggplant breeding as well as for basic research and for developing new genetic tools, like interspecific genetic maps and experimental population.

### ***Development of genetic and genomic tools using CWRs***

Despite being one of the most important vegetables in term of production and for food security in many regions, for eggplant there is a general lack of genetic tools and experimental populations. In the last few years, some efforts have been done in this direction, although the gap with other important solanaceous vegetable crops, like tomato, potato or pepper, is still quite large. Genomic tools play a vital role in understanding the genetics of complex traits, as well as, in developing experimental

populations. Apart from performing genetic studies like QTLs identification, the development of experimental population is one of the most efficient ways to introgress genes from CWRs, which is crucial in a species like common eggplant with a narrow genetic diversity.

An interesting eggplant CWR is *S. incanum*, which is reported as having considerable drought tolerance and resistance to some fungal diseases (Yamakawa and Mochizuki 1979; Lester and Hasan 1991; Collonnier et al., 2001a; Daunay 2008; Knapp et al., 2013). In addition, *S. incanum* exhibits a high content of bioactive phenolic compounds, mostly of chlorogenic acid type (Ma et al., 2011; Stommel and Whitaker 2003; Prohens et al., 2013; Meyer et al., 2015). Thus, in order to develop new improved eggplant varieties and dissect the genetics of some interesting traits, we built an interspecific linkage genetic map and developed an introgression line population between common eggplant and *S. incanum*.

The mapping population consisted of 91 BC1 individuals obtained backcrossing the interspecific towards the *S. melongena* parent. A total of 243 molecular markers were successfully mapped consisting of 42 COSII, 99 SSRs, 88 AFLPs, 9 CAPS, 4 SNPs and one morphological polymorphic markers encompassed 1085 cM distributed in 12 linkage groups. The wide variety of molecular markers employed avoided the excess of clustering in some genomic regions (Nunome et al., 2009; Wu et al., 2009; Barchi et al., 2012; Fukuoka et al., 2012), typical of some marker types, and allowed achieving a better representation of the eggplant genome. Our interspecific map, named SMICB, was linked to the four most informative previous eggplant genetic maps by using shared markers and allowing to exploit the information generated by these maps (Fulton et al., 2002; Nunome et al., 2009; Wu et al., 2009; Barchi et al., 2012; Fukuoka et al., 2012). In

addition, macro-synteny between SMIBC and Tomato EXPEN-2000 map was established with 130 anchoring points. Based on the synteny with tomato, the candidate genes involved in the core chlorogenic acid synthesis pathway in eggplant (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*), five polyphenol oxidase genes (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*), as well as two genes involved in fruit shape (*OVATE*, *SISUNI*), were mapped. Chlorogenic acid (5-O-caffeoyl-quinic acid) is the most abundant phenolic compound of eggplant, between 70% to over 95% of the total phenolics content (Whitaker and Stommel, 2003; Stommel and Whitaker, 2003; Singh et al., 2009; Prohens et al., 2013). Chlorogenic acid have shown antioxidant, anti-inflammatory, analgesic, antipyretic, neuroprotective, cardioprotective, anti-carcinogenic, anti-microbial, hypotensive, anti-obesity and anti-diabetic activity (Dos Santos et al., 2006; Cho et al., 2010; Plazas et al., 2013a). On the other hand, polyphenol oxidase genes catalyze the oxidation of phenolics causing the browning of the fruit flesh (Ramirez et al., 2003).

The location of these genes is of great relevance in eggplant breeding since the identification of favorable alleles by marker-assisted selection could contribute to developing improved varieties with higher chlorogenic content alongside a lower polyphenol oxidase activity. These new varieties will present higher bioactive properties through a combination of high antioxidant activity and alongside a low degree of browning.

SMIBC map was extremely helpful in the early stages of the development of the introgression line (IL) population using *S. incanum* as donor parent in the genetic background of *S. melongena*. In fact, during the five rounds of backcrossing, we performed a low–medium throughput

genotyping using the markers mapped in SMIBC, with a range of 52 to 63 markers per round. In addition, for the sake of time and resource saving, we just genotyped the target chromosome of each plant and not the whole background. At that time, no affordable high-throughput genotyping platforms or SNP arrays were available in eggplant to genotype a large number of plants with a wide set of SNPs in a very short time as in other species (Wang et al., 2014; Vos et al., 2015). That strategy increased the number of rounds of backcrosses required since other authors, in species with more genetic and genomic tools, just performed three rounds to achieve the same level of recipient parent recovered (Tanksley et al., 1989; Hospitalet al., 1992; Hospital and Charcosset, 1997). In addition, we dragged, along the backcrossing process, many non-target introgressions, as happened in several studies (Shen et al., 2001; Eichten et al., 2011; Pea et al., 2013; Arbelaez et al., 2015). In fact, when massive genotyping is performed in the first generations of backcrossing the process of obtaining ILs is accelerated and simultaneously it provided an early detection of non-target introgressions and easy elimination of them during the backcross process (Barrantes et al., 2014).

A high-throughput genotyping by GBS was performed in the first round of selfing detecting 796 very robust SNPs, as well as, recombination breakpoints, not-targeting introgressions, and double recombination events. The GBS improved dramatically the precision of the length of the introgressions and giving a whole picture of the genetic background.

From the whole set of SNPs detected by GBS, 62 of them were selected and adapted for manually genotyping by HRM (high-resolution melting) platform to continue to process of fixing material through rounds of selfing. The selection of SNPs was made using a combination of genetic

and physical position, plus the necessary markers to screen the non-target introgressions. In that way, we avoided an under-representation of telomeric regions and drag large centromeric regions, as it usual in selection based just on genetic maps (Wang et al., 2000).

During this thesis, we obtained 73 selected ABs (advanced backcrossed) materials covered 99% of the *S. incanum* genome and 25 fixed immortal ILs, each carrying a single introgressed fragment in homozygosis, spanned 61.7% of the *S. incanum* genome. In those materials 142 eggplant genome regions were identified as orthologs of candidate genes related to drought tolerance in other crops (Bolger et al., 2014; Zhang N. et al., 2014; Krannich et al., 2015; Landi et al., 2016), with 68 of them being present in the fixed ILs.

Although the IL development process is not concluded because additional fixed lines are being refined and some large fixed ILs are being broken into smaller fragments, the materials developed so far are currently being tested for QTLs identification. Furthermore, a preliminary drought-screening test has been performed in the framework of a MSc thesis in which some promising materials have been identified (Savarino, 2017).

This IL population has been the first developed in eggplant and will be extremely useful for the genetic dissection of traits of interest for eggplant. In addition, it is a valuable resource for eggplant breeding since they are advanced materials, they do not display undesirable wild traits, and can be directly introduced into the breeding pipelines for developing new eggplant cultivars adapted to the challenges posed by the climate-change scenario.

### ***Future perspectives***

The information, knowledge, and tools developed in this multidisciplinary and multi-pronged thesis provide a basis for further studies and projects in eggplant and in non-model species.

This study paves the way for a new generation of improved varieties with higher bioactive properties, through a combination of high antioxidant activity and low degree of browning, as well as, new resilient varieties resistant and tolerant to biotic and abiotic stresses. The systematic "introgressomics" approach that we proposed would greatly broaden the narrow genetic diversity in common eggplant employing CWRs.

In this respect, in the framework of the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives" managed by the Global Crop Diversity Trust, the team in which I developed my thesis has obtained so far over 90 interspecific hybrid combinations with among six common eggplant and 14 eggplant CWRs, over 50 BC1 combinations using 12 CWRs and more than 30 BC2 combinations with 10 CWRs.

Taking advantage of the experience gained in this thesis in the development of experimental populations has led us to generate more tools to introgress wild genes in the common eggplant background. Currently, we are developing three other IL populations with species belonging to the primary (*Solanum insanum*), secondary (*S. dasyphyllum*) and tertiary (*S. elaeagnifolium*) eggplant genepools (Knapp et al., 2013; Syfert et al., 2016). These species are of great interest for many agronomic traits as phenolic compounds and content, tolerance to abiotic stresses or for biotic resistance (Plazas et al., 2014a,b; Ranil et al., 2017).

In addition, we are also developing a multi-parent experimental population (MAGIC) in eggplant, employing seven common eggplants and one *S. incanum* accession. According to our knowledge, this will be the first eggplant MAGIC population. MAGIC populations have been extremely useful in model species to identify QTLs with high precision and to dissect complex quantitative traits (Cavanagh et al., 2008; Pascual et al., 2014). In order to speed up this process, we have re-sequenced and structurally and functionally annotated the whole genome of the eight parents and identified millions of markers.

In conclusion, this thesis, apart from providing highly relevant information and tools for eggplant genetic, genomics and breeding enhancement, is a starting point for a new era in eggplant breeding where a new generation of improved and resilient varieties will be developed assisted by the cutting-edge genetic and genomic tools available. This will contribute to cope with future and unpredictable challenges in eggplant breeding.







# CONCLUSIONS



- ❖ The currently available genomic resources in eggplant genepool are still very limited and insufficient to assist future breeding efforts and challenges. In this thesis, using different approaches, we have developed and provided genomic knowledge, information and tools that will be useful for the eggplant research community and for eggplant breeders to exploit efficiently the natural genetic diversity in cultivated and eggplant relatives.
  
- ❖ In order to introgress valuable wild genes in breeding pipelines and based on our views of using eggplant CWRs in breeding, we have proposed a new approach, called 'introgressiomics'. This approach, which can be applied in general to all crops, consists of a systematic and mass scale development of materials using crop wild relatives as a source of genetic diversity. In this thesis, following this approach, we developed advanced materials that are readily available for a wide range of breeding purposes.
  
- ❖ The transcriptome sequencing of the wild *Solanum incanum*, which is report as highly tolerant to drought and with high content in phenolics, and the African cultivated *S. aethiopicum*, which is usually used as rootstock for common eggplant for being tolerant to many soil diseases, has revealed as a very efficient and pragmatic way to generate information in non-model species. In fact, with a modest amount of resources, the sequencing of a transcriptome of this two species with a great interest for eggplant breeding, provided a detailed structural and functional information about the genes, as well as the

identification of tens of thousand molecular markers (SSRs, SNPs and INDELs).

- ❖ We were able to confirm that transcriptome-derived SSR and SNP markers are useful for genetic fingerprinting, diversity evaluation and the establishment of relationships in eggplants, as they sample different levels of genetic variation. Both markers are complementary in the information provided, although SNPs seem more appropriate to evaluate materials genetically distant in the eggplant complexes.
- ❖ High-throughput genotyping of numerous genome-wide SNP markers represents a highly and more effective approach with respect to the ones based on limited sets of markers for understanding the extent of natural allelic diversity and establish genetic relationships among and within wild and cultivated species belonging to eggplant complexes. In this way, we have been able to solve relationships among and within the three cultivated eggplant groups.
- ❖ We have developed an interspecific genetic map linked to the four most informative previous eggplant genetic maps and tomato EXPEN-2000 genetic map, allowing exploiting the information generated by these maps. Based on the synteny with tomato, we were able to locate in our map the candidate genes involved in the core chlorogenic acid synthesis pathway in eggplant (*PAL*, *C4H*, *4CL*, *HCT*, *C3'H*, *HQT*), five polyphenol oxidase genes (*PPO1*, *PPO2*, *PPO3*, *PPO4*, *PPO5*), as well as two genes involved in fruit shape (*OVATE*, *SISUNI*). In addition, our map has been essential in the early

stages of marker-assisted selection in the development of the introgression line population also developed in this thesis.

- ❖ Finally, in this thesis, we developed the first introgression line population in the eggplant genepool using *S. incanum* as donor parent. Specifically, the materials developed consisted of a set of 73 selected advance backcrosses covering 99% of the *S. incanum* genome and 25 fixed immortal introgression lines (ILs) spanning 61.7% of the donor genome. Furthermore, 68 orthologous candidate genes involved in pathways related to drought tolerance were in silico assigned to the ILs. Regarding the ILs development, we observed that the implementation of high-throughput genotyping technologies has an extraordinary impact in the development and precision of IL populations, reducing the time and the resources and increasing the precision of the introgression detection.
  
- ❖ The multidisciplinary and multi-pronged focus of this thesis has allowed generating knowledge, information, tools, resources, approaches and advanced materials to enhance eggplant breeding. In this way, a new generation of advanced and resilient eggplant varieties to face with the future challenges can be efficiently developed using the valuable genetic diversity of CWRs and the precision and efficiency of high-throughput techniques. We consider that the information gathered in this work could be useful also for other non-model species that has lagged behind in the exploitation of the dramatic advanced provided by the genetic and genomics tools and next-generation sequencing platforms.





## **REFERENCES**



- Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., Costa de Oliveira, A., Cseke, L.J., Dempewolf, H., De Pace, C., et al. 2016. Global agricultural intensification during climate change: A role for genomics. *Plant Biotechnol. J.* 14, 1095–1098.
- Acquadro, A., Lanteri, S., Scaglione, D., Arens, P., Vosman, B., Portis, E., 2009. Genetic mapping and annotation of genomic microsatellites isolated from globe artichoke. *Theor. Appl. Genet.* 118, 1573–1587.
- Acquadro, A., Barchi, L., Portis, E., Carrasquilla-Garcia, N., Tirone, M., Lanteri, S., Comino, C., 2016. RAD2seq: An efficient protocol for plant genotyping by Sequencing. In: *Acta Horticulturae*. pp. 1–7.
- Aflitos, S., Schijlen, E., De Jong, H., De Ridder, D., Smit, S., Finkers, R., Wang, J., Zhang, G., Li, N., Mao, L., et al. 2014. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *Plant J.* 80, 136–148.
- Akanitapichat, P., Phraibung, K., Nuchklang, K., Prompitakkul, S., 2010. Antioxidant and hepatoprotective activities of five eggplant varieties. *Food Chem. Toxicol.* 48, 3017–3021.
- Albacete, A., Andújar, C., Dodd, I., Giuffrida, F., Hichri, I., Lutts, S., Thompson, A., Asins, M., 2015. Rootstock-mediated variation in tomato vegetative growth under drought, salinity and soil impedance stresses. *Acta Hort.* 1086, 141–146.
- Alexander, L.J., 1963. Transfer of a dominant type of resistance to 4 known Ohio pathogenic strains of *Tobacco Mosaic Virus* (TMV), From *Lycopersicon Peruvianum* To *L. Esculentum*. In: *Phytopathology* 53, 869.
- Alfares, W., Bouguennec, A., Balfourier, F., Gay, G., Bergès, H., Vautrin, S., Sourdille, P., Bernard, M., Feuillet, C., 2009. Fine mapping and marker development for the crossability gene *SKr* on chromosome 5BS of hexaploid wheat (*Triticum aestivum* L.). *Genetics* 183, 469–481.
- Alheit, K.V., Reif, J.C., Maurer, H.P., Hahn, V., Weissmann, E.A., Miedaner, T., Würschum, T., 2011. Detection of segregation distortion loci in triticale (x *Triticosecale* Wittmack) based on a high-density DArT marker consensus genetic linkage map. *BMC*

- Alonso-Blanco, C., Peeters, A.J.M., Koornneef, M., Lister, C., Dean, C., Van Bosch, N., Den Pot, J., Kuiper, M.T.R., 1998. Development of an AFLP based linkage Map of Ler, Col and Cvi *arabidopsis thaliana* ecotypes and construction of a Ler/Cvi recombinant inbred line population. *Plant J.* 14, 259–271.
- Alonso-Blanco, C., Koornneef, M., van Ooijen, J.W., 2006. QTL analysis. *Methods Mol. Biol.* 323, 79–99.
- Alseekh, S., Ofner, I., Pleban, T., Tripodi, P., Di Dato, F., Cammareri, M., Mohammad, A., Grandillo, S., Fernie, A.R., Zamir, D., 2013. Resolution by recombination: Breaking up *Solanum pennellii* introgressions. *Trends Plant Sci.* 18, 536–538.
- Amar, M.H., Biswas, M.K., Zhang, Z., Guo, W.W., 2011. Exploitation of SSR, SRAP and CAPS-SNP markers for genetic diversity of Citrus germplasm collection. *Sci. Hortic.* 128, 220–227.
- Anderson, G., Bernardello, G.L.B., Weese, T., Santos-Guerra, A., Santos-Guerra, A., 2006. Phylogeny and biogeography of the Canarian *Solanum vespertilio* and *S. lidii* (Solanaceae). *An. del Jardín Botánico Madrid* 63, 159–167.
- André, C.M., Schafleitner, R., Legay, S., Lefèvre, I., Aliaga, C.A.A., Nomberto, G., Hoffmann, L., Hausman, J.F., Larondelle, Y., Evers, D., 2009. Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. *Phytochemistry* 70, 1107–1116.
- Andrew, A., 2010. FastQC: A quality control tool for high throughput tequence.
- Anis, M., Baksh, S., Iqbal, M., 1994. Cytogenetic studies on the F1 hybrid *Solanum incanum* \* *S. melongena* var. american wonder. *Cytologia* 59, 433–436.
- Arbelaez, J.D., Moreno, L.T., Singh, N., Tung, C.W., Maron, L.G., Ospina, Y., Martinez, C.P., Grenier, C., Lorieux, M., McCouch, S., 2015. Development and GBS-genotyping of introgression lines (ILs) using two wild species of rice, *O. meridionalis* and *O. rufipogon*, in a common recurrent parent, *O. sativa* cv. Curinga. *Mol. Breed.* 35, 81.
- Aronesty, E., 2013. Comparison of sequencing utility programs. *Open*

- Bioinforma. J. 7, 1–8.
- Arumuganathan, Earle, 1991. Nuclear DNA content of some important plant species. *Plant Mol. Biol.* 9, 208–218.
- Ashrafi, H., Hill, T., Stoffel, K., Kozik, A., Yao, J., Chin-Wo, S.R., Van Deynze, A., 2012. De novo assembly of the pepper transcriptome (*Capsicum annuum*): A benchmark for in silico discovery of SNPs, SSRs and candidate genes. *BMC Genomics* 13, 571.
- Bagnaresi, P., Sala, T., Irdani, T., Scotto, C., Lamontanara, A., Beretta, M., Rotino, G.L., Sestili, S., Cattivelli, L., Sabatini, E., 2013. *Solanum torvum* responses to the root-knot nematode *Meloidogyne incognita*. *BMC Genomics* 14, 540.
- Bajaj, D., Das, S., Badoni, S., Kumar, V., Singh, M., Bansal, K.C., Tyagi, A.K., Parida, S.K., 2015. Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci. Rep.* 5, 357–373.
- Bao, Y., Xu, S., Jing, X., Meng, L., Qin, Z., 2015. De novo assembly and characterization of *Oryza officinalis* leaf transcriptome by using RNA-seq. *Biomed Res. Int.* 2015, 1–7.
- Barbará, T., Palma-Silva, C., Paggi, G.M., Bered, F., Fay, M.F., Lexer, C., 2007. Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Mol. Ecol.* 16, 3759–3767.
- Barchi, L., Delledonne, M., Lanteri, S., Dal Molin, A., Minio, A., Ferrarini, A., Venturini, L., Avanzato, C., Toppino, L., Sala, T., et al. 2016. A high quality eggplant genome sequence: a new tool for the analysis of the Solanaceae family evolution and for the molecular deciphering of complex traits. In: Kölliker, R., Boller, B.C. (Eds.), *Plant Breeding: The Art of Bringing Science to Life - 20th Eucarpia General Congress*. Agroscope, Institute for Sustainability Sciences ISS, Zurich, p. 122.
- Barchi, L., Lanteri, S., Portis, E., Stàgel, A., Valè, G., Toppino, L., Rotino, G.L., 2010. Segregation distortion and linkage analysis in eggplant (*Solanum melongena* L.). *Genome* 53, 805–15.
- Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Vale, G., Toppino, L., Rotino, G.L., 2011. Identification of SNP and SSR markers in

- eggplant using RAD tag sequencing. *BMC Genomics* 12, 304.
- Barchi, L., Lanteri, S., Portis, E., Valè, G., Volante, A., Pulcini, L., Ciriaci, T., Acciarri, N., Barbierato, V., Toppino, L., Rotino, G.L., 2012. A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. *PLoS One* 7.
- Bari, A., Street, K., Mackay, M., Endresen, D.T.F., de Pauw, E., Amri, A., 2012. Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genet. Resour. Crop Evol.* 59, 1465–1481.
- Barrantes, W., Fernández-del-Carmen, A., López-Casado, G., González-Sánchez, M.Á., Fernández-Muñoz, R., Granell, A., Monforte, A.J., 2014. Highly efficient genomics-assisted development of a library of introgression lines of *Solanum pimpinellifolium*. *Mol. Breed.* 34, 1817–1831.
- Baute, G.J., Dempewolf, H., Rieseberg, L.H., 2015. Using genomic approaches to unlock the potential of CWR for crop adaptation to climate change. In: Redden, R., Yadav, S., Maxted, N., Dulloo, M., Guarino, L., Smith, P. (Eds.), *Crop wild relatives and climate change*. Wiley, Hoboken, pp. 268–280.
- Baxter, C.J., Sabar, M., Quick, W.P., Sweetlove, L.J., 2005. Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. *J. Exp. Bot.* 56, 1591–1604.
- Bebber, D.P., Ramotowski, M.A.T., Gurr, S.J., 2013. Crop pests and pathogens move polewards in a warming world. *Nat. Clim. Chang.* 3, 985–988.
- Bedő, Z., Láng, L., 2015. Wheat breeding: current status and bottlenecks. In: *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics*. Springer International Publishing, Cham, pp. 77–101.
- Behera, T.K., Singh, N., 2002. Inter-specific crosses between eggplant (*Solanum melongena* L.) with related *Solanum* species. *Sci. Hortic.* 95, 165–172.
- Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Nekrasov, V., 2013. Plant genome editing made easy: Targeted mutagenesis in model and crop

- plants using the CRISPR/Cas system. *Plant Methods* 9, 39.
- Bessey, C.E., 1906. Crop improvement by utilizing wild species. *J. Hered.* 2, 112–118.
- Beyer, P., Morell, M., Mackay, I., Powell, W., 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11, 215–221.
- Blanca, J., Cañizares, J., Roig, C., Ziarsolo, P., Nuez, F., Picó, B., 2011. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* 12, 104.
- Bletsos, F.A., Roupakias, D.G., Tsaktsira, M.L., Scaltsoyjanis, A.B., Thanassouloupoulos, C.C., 1998. Interspecific hybrids between three eggplant (*Solanum melongena* L.) cultivars and two wild species (*Solanum torvum* Sw. and *Solanum sisymbriifolium* Lam.). *Plant Breed.* 117, 159–164.
- Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., et al. 2014. The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* 46, 1034–1038.
- Bombarely, A., Menda, N., Teclé, I.Y., Buels, R.M., Strickler, S., Fischer-York, T., Pujar, A., Leto, J., Gosselin, J., Mueller, L.A., 2011. The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res.* 39, D1149–D1155.
- Bors, R.H., Sullivan, J.A., 2005. Interspecific hybridization of *Fragaria vesca* subspecies with *F. nilgerrensis*, *F. nubicola*, *F. pentaphylla*, and *F. viridis*. *J. Am. Soc. Hortic. Sci.* 130, 418–423.
- Boyd, J.W., Murray, D.S., Tyrll, R.J., 1984. Silverleaf nightshade, *Solarium elaeagnifolium*, origin, distribution, and relation to man. *Econ. Bot.* 38, 210–217.
- Brown, T.A., 2007. *Genomes 3*. Garland Science Pub.
- Brozynska, M., Furtado, A., Henry, R.J., 2016. Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085.
- Buckler, E.S., Ilut, D.C., Wang, X., Kretschmar, T., Gore, M.A.,

- Mitchell, S.E., 2016. rAmpSeq: Using repetitive sequences for robust genotyping. bioRxiv 2016, 096628.
- Buerstmayr, M., Lemmens, M., Steiner, B., Buerstmayr, H., 2011. Advanced backcross QTL mapping of resistance to *Fusarium* head blight and plant morphological traits in a *Triticum macha* x *T. aestivum* population. Theor. Appl. Genet. 123, 293–306.
- Bukenya, Z., Carasco, J., 1994. Biosystematic study of *Solanum macrocarpon*-*S. dasyphyllum* complex in Uganda and relations with *S. linnaeanum*. East African Agric. For. J. 59, 187–204.
- Byrne, P., McMullen, M., 2010. Defining genes for agricultural traits: QTL analysis and the candidate gene approach. Proc. Natl. Acad. Sci. Probe, 24–27.
- Campbell, N.R., Harmon, S.A., Narum, S.R., 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. 15, 855–867.
- Campi, M., Nuvolari, A., 2015. Intellectual property protection in plant varieties: A worldwide index. Res. Policy 44, 951–964.
- Cappelli, C., Stravato, V.M., Rotino, G.L., Buonaurio, R., 1995. Sources of resistance among *Solanum* spp. to an Italian isolate of *Fusarium oxysporum* f. sp. *melongenae*. In: Proceedings of the 9th EUCARPIA Meeting on Genetics Andbreeding of Capsicum and Eggplant. pp. 221–224.
- Cardi, T., 2016. Cisgenesis and genome editing: Combining concepts and efforts for a smarter use of genetic resources in crop breeding. Plant Breed. 135, 139–147.
- Caruso, G., Gomez, L.D., Ferriello, F., Andolfi, A., Borgonuovo, C., Evidente, A., Simister, R., McQueen-Mason, S.J., Carputo, D., Frusciante, L., Ercolano, M.R., 2016. Exploring tomato *Solanum pennellii* introgression lines for residual biomass and enzymatic digestibility traits. BMC Genet. 17, 56.
- Castañeda-Álvarez, N.P., Khoury, C.K., Achicanoy, H.A., Bernau, V., Dempewolf, H., Eastwood, R.J., Guarino, L., Harker, R.H., Jarvis, A., Maxted, et al. 2016. Global conservation priorities for crop wild relatives. Nat. Plants 2, 16022.



- Cattivelli, L., Rizza, F., Badeck, F.W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., Marè, C., Tondelli, A., Stanca, A.M., 2008. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. *F. Crop. Res.* 105, 1–14
- Centro Internacional de Agricultura Tropical, 2017. A global database for the distributions of crop wild relatives (<https://www.gbif.org/dataset/07044577-bd82-4089-9f3af4a9d2170b2e>).
- Cericola, F., Portis, E., Toppino, L., Barchi, L., Acciarri, N., Ciriaci, T., Sala, T., Rotino, G.L., Lanteri, S., 2013. The Population Structure and Diversity of Eggplant from Asia and the Mediterranean Basin. *PLoS One* 8, e73702.
- Cericola, F., Portis, E., Lanteri, S., Toppino, L., Barchi, L., Acciarri, N., Pulcini, L., Sala, T., Rotino, G., 2014. Linkage disequilibrium and genome-wide association analysis for anthocyanin pigmentation and fruit color in eggplant. *BMC Genomics* 15, 896.
- Chagné, D., Krieger, C., Rassam, M., Sullivan, M., Fraser, J., André, C., Pindo, M., Troglio, M., Gardiner, S.E., Henry, R.A., 2012. QTL and candidate gene mapping for polyphenolic composition in apple fruit. *BMC Plant Biol.* 12, 12.
- Chai, C., Lin, Y., Shen, D., Wu, Y., Li, H., Dou, D., 2013. Identification and functional characterization of the Soybean GmaPPO12 promoter conferring *Phytophthora sojae* induced expression. *PLoS One* 8, e67670.
- Chen, H., Wang, L., Wang, S., Liu, C., Blair, M.W., Cheng, X., 2015. Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST-SSR markers. *PLoS One* 10, e0120273.
- Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., Town, C.D., 2017. Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 89, 789–804.
- Chitwood, D.H., Kumar, R., Headland, L.R., Ranjan, A., Covington, M.F., Ichihashi, Y., Fulop, D., Jiménez-Gómez, J.M., Peng, J., Maloof, J.N., Sinha, N.R., 2013. A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. *Plant Cell* 25, 2465–81.

- Cho, A.S., Jeon, S.M., Kim, M.J., Yeo, J., Seo, K.I., Choi, M.S., Lee, M.K., 2010. Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced-obese mice. *Food Chem. Toxicol.* 48, 937–943.
- Coetzer, C., Corsini, D., Love, S., Pavek, J., Tumer, N., 2001. Control of enzymatic browning in potato (*Solanum tuberosum* L.) by sense and antisense RNA from tomato polyphenol oxidase. *J. Agric. Food Chem.* 49, 652–657.
- Collonnier, C., Fock, I., Kashyap, V., Rotino, G.L., Daunay, M.C., Lian, Y., Mariska, I.K., Rajam, M. V., Servaes, A., Ducreux, G., Sihachakr, D., 2001a. Applications of biotechnology in eggplant. *Plant Cell. Tissue Organ Cult.* 65, 91–107.
- Collonnier, C., Mulya, K., Fock, I., Mariska, I., Servaes, A., Vedel, F., Siljak-Yakovlev, S., Souvannavong, V., Ducreux, G., Sihachakr, D., 2001b. Source of resistance against *Ralstonia solanacearum* in fertile somatic hybrids of eggplant (*Solanum melongena* L.) with *Solanum aethiopicum* L. *Plant Sci.* 160, 301–313.
- Collonnier, C., Fock, I., Mariska, I., Servaes, A., Vedel, F., Siljak-Yakovlev, S., Souvannavong, V., Sihachakr, D., 2003. GISH confirmation of somatic hybrids between *Solanum melongena* and *S. torvum*: Assessment of resistance to both fungal and bacterial wilts. *Plant Physiol. Biochem.* 41, 459–470.
- Comino, C., Lanteri, S., Portis, E., Acquadro, A., Romani, A., Hehn, A., Larbat, R., Bourgaud, F., 2007. Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. *BMC Plant Biol.* 7, 14.
- Comino, C., Hehn, A., Moglia, A., Menin, B., Bourgaud, F., Lanteri, S., Portis, E., 2009. The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. *BMC Plant Biol.* 9, 30.
- Concellón, A., Zaro, M.J., Chaves, A.R., Vicente, A.R., 2012. Changes in quality and phenolic antioxidants in dark purple American eggplant (*Solanum melongena* L. cv. Lucía) as affected by storage at 0° C and 10° C. *Postharvest Biol. Technol.* 66, 35–41.
- Concibido, V., Vallee, B. La, Mclaird, P., Pineda, N., 2003. Introgression

- of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theor. Appl. Genet.* 106, 575–582.
- Conesa, A., Gotz, S., 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 619832.
- Cowling, W.A., Buirchell, B.J., Falk, D.E., 2009. A model for incorporating novel alleles from the primary gene pool into elite crop breeding programs while reselecting major genes for domestication or adaptation. *Crop Pasture Sci.* 60, 1009–1015.
- Daunay, M.C., Lester, R.N., Laterrot, H., 1991. The use of wild species for the genetic improvement of Brinjal eggplant (*Solanum melongena*) and tomato (*Lycopersicon esculentum*). In: Hawkes, J., Nee, M., Estrada, N. (Eds.), *Solanaceae III: Taxonomy, Chemistry, Evolution*. Royal Botanic Gardens: Kew, UK, London, pp. 389–413.
- Daunay, M., Lester, R., 2001. Eggplant. In: Charrier, A., Jacquot, M., Hamon, S., Nicolas, D. (Eds.), *Tropical Plant Breeding*. CIRAD, SciencePublishers Inc, Plymouth, UK., pp. 199–221.
- Daunay, M., Lester, R., Ano, G., 2001. Cultivated eggplants. *Trop. plant Breed.* 200–225.
- Daunay, M.-C., 2008. Eggplant. In: Prohens, J., Nuez, F. (Eds.), *Vegetables II: Handbook of Plant Breeding*. Springer, New York, pp. 163–220.
- Daunay, M., Hazra, P., 2012. Eggplant. In: Peter, K., Hazra, P. (Eds.), *Handbook of Vegetables*. Studium Press LLC, Houston, TX, USA, pp. 257–322.
- Davey, J.W., Blaxter, M.L., 2010. RADSeq: Next-generation population genetics. *Brief. Funct. Genomics* 9, 416–423.
- Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510.
- Davidar, P., Snow, A.A., Rajkumar, M., Pasquet, R., Daunay, M.C., Mutegi, E., 2015. The potential for crop to wild hybridization in eggplant (*Solanum melongena*; Solanaceae) in Southern India. *Am. J. Bot.* 102, 129–139.

- De Storme, N., Mason, A., 2014. Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* 2014, 10–33.
- Dempewolf, H., Hodgins, K.A., Rummell, S.E., Ellstrand, N.C., Rieseberg, L.H., 2012. Reproductive isolation during domestication. *Plant Cell* 24, 2710–2717.
- Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J. V., Toll, J., 2014. Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecol. Sustain. Food Syst.* 38, 369–377.
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., Guarino, L., 2017. Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082.
- Dhaliwal, H.S., 1992. Unilateral incompatibility. In: Kalloo, G., Chowdhury, J. (Eds.), *Distant Hybridization of Crop Plants*. Springer, Berlin, Heidelberg, Berlin, pp. 32–46.
- Di Guardo, M., Tadiello, A., Farneti, B., Lorenz, G., Masuero, D., Vrhovsek, U., Costa, G., Velasco, R., Costa, F., 2013. A multidisciplinary approach providing new insight into fruit flesh browning physiology in Apple (*Malus x domestica* Borkh.). *PLoS One* 8, e78004.
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.
- Díez, M., Nuez, F., 2008. Tomato. In: Prohens, J., Nuez, F. (Eds.), *Vegetables II: Fabaceae, Liliaceae, Solanaceae, and Umbelliferae*. Springer, New York, NY, pp. 249–323.
- Ding, Q., Li, J., Wang, F., Zhang, Y., Li, H., Zhang, J., Gao, J., 2015. Characterization and development of EST-SSRs by deep transcriptome sequencing in chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Int. J. Genomics* 473028.
- Docimo, T., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., Toppino, L., Rotino, G.L., Mennella, G., Tucci, M., 2015. Phenylpropanoids accumulation in eggplant fruit: Characterization of biosynthetic genes and regulation by a MYB transcription factor. *Front. Plant Sci.* 6, 1233.

- Dodsworth, S., Chase, M.W., Särkinen, T., Knapp, S., Leitch, A.R., 2016. Using genomic repeats for phylogenomics: A case study in wild tomatoes (*Solanum* section *Lycopersicon*: Solanaceae). *Biol. J. Linn. Soc.* 117, 96–105.
- Doğan, M., Arslan, O., Doğan, S., 2002. Substrate specificity, heat inactivation and inhibition of polyphenol oxidase from different aubergine cultivars. *Int. J. Food Sci. Technol.* 37, 415–423.
- Doganlar, S., Frary, A., Daunay, M.C., Lester, R.N., Tanksley, S.D., 2002a. Conservation of gene function in the Solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* 161, 1713–1726.
- Doganlar, S., Frary, A., Daunay, M.C., Lester, R.N., Tanksley, S.D., 2002b. A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics* 161, 1697–1711.
- Doganlar, S., Frary, A., Daunay, M.C., Huvenaars, K., Mank, R., Frary, A., 2014. High resolution map of eggplant (*Solanum melongena*) reveals extensive chromosome rearrangement in domesticated members of the Solanaceae. *Euphytica* 198, 231–241.
- Dos Santos, M.D., Almeida, M.C., Lopes, N.P., De Souza, G.E.P., 2006. Evaluation of the anti-inflammatory, analgesic and antipyretic activities of the natural polyphenol chlorogenic acid. *Biol. Pharm. Bull.* 29, 2236–2240.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Doyle, J.J., Doyle, J.L., 1990. Isolation of Plant DNA from fresh tissue. *Focus* 12, 13–15.
- Durand, J., Bodenes, C., Chancerel, E., Frigerio, J.M., Vendramin, G., Sebastiani, F., Buonamici, A., Gailing, O., Koelewijn, H.P., Villani, F., 2010. A fast and cost-effective approach to develop and map EST-SSR markers: Oak as a case study. *BMC Genomics* 11, 570.
- Dwivedi, S.L., Upadhyaya, H.D., Stalker, H.T., Blair, M.W., Bertioli, D.J., Nielsen, S., Ortiz, R., 2008. Enhancing crop gene pools with beneficial traits using wild relatives. *Plant Breed. Rev.* 30, 179–230.
- Eduardo, I., Arús, P., Monforte, A.J., 2005. Development of a genomic

- library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI161375. *Theor. Appl. Genet.* 112, 139–148.
- Eichten, S.R., Foerster, J.M., de Leon, N., Kai, Y., Yeh, C.-T., Liu, S., Jeddelloh, J.A., Schnable, P.S., Kaeppeler, S.M., Springer, N.M., 2011. B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. *Plant Physiol.* 156, 1679–1690.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, e19379.
- Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troglio, M., Myles, S., Martinez-Zapater, J.M., Zyprian, E., Moreira, F.M., Grando, M.S., 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol.* 13, 39.
- Erayman, M., Sandhu, D., Sidhu, D., Dilbirligi, M., Baenziger, P.S., Gill, K.S., 2004. Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Res.* 32, 3546–3565.
- 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.
- Fan, J.B., Chee, M.S., Gunderson, K.L., 2006. Highly parallel genomic assays. *Nat. Rev. Genet.* 7, 632–644.
- FAO, 2015. Food and Agriculture Organization of the United Nations (<http://faostat.fao.org>).
- FAO, 2017. Food and Agriculture Organization of the United Nations (<http://faostat.fao.org>).
- FAO, 2016. Food and Agriculture Organization of the United Nations (<http://faostat.fao.org>).
- Faure, D., Joly, D., 2015. Next-generation sequencing as a powerful motor

- for advances in the biological and environmental sciences. *Genetica* 143, 129–132.
- Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K.R., Ylitalo-Ward, H., Bowen, B.W., Toonen, R.J., Goetze, E., Karl, S.A., 2013. Microsatellites for next-generation ecologists: A post-sequencing bioinformatics pipeline. *PLoS One* 8, e55990.
- Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S., Eversole, K., 2011. Crop genome sequencing: Lessons and rationales. *Trends Plant Sci.* 16, 77–88
- Filippi, C. V., Aguirre, N., Rivas, J.G., Zubrzycki, J., Puebla, A., Cordes, D., Moreno, M. V, Fusari, C.M., Alvarez, D., Heinz, R.A., et al. 2015. Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biol.* 15, 52.
- Fita, A., Rodríguez-Burruezo, A., Boscaiu, M., Prohens, J., Vicente, O., 2015. Breeding and domesticating crops adapted to drought and salinity: A new paradigm for increasing food production. *Front. Plant Sci.* 6, 978.
- Flint-Garcia, S.A., 2013. Genetics and consequences of crop domestication. *J. Agric. Food Chem.* 61, 8267-8276.
- Fowler, C., Moore, G.K., Hawtin, G., 2003. The International Treaty on Plant Genetic Resources for Food and Agriculture: a primer for the future harvest centres of the CGIAR.
- Frary, A., Doganlar, S., Daunay, M.C., Tanksley, S.D., 2003. QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. *Theor. Appl. Genet.* 107, 359–370.
- Frary, A., Doganlar, S., Daunay, M.C., 2006. Eggplant. In: Kole, C. (Ed.), *Genome Mapping and Molecular Breeding in Plants, Volume V: Vegetables II.* Springer, Heidelberg, Germany., pp. 287–313.
- Frary, A., Frary, A., Daunay, M.C., Huvenaars, K., Mank, R., Doğanlar, S., 2014. QTL hotspots in eggplant (*Solanum melongena*) detected with a high resolution map and CIM analysis. *Euphytica* 197, 211–228.
- Friebe, B., Jiang, J., Raupp, W.J., McIntosh, R.A., Gill, B.S., 1996.

Characterization of wheat-alien translocations conferring resistance to diseases and pests: Current status. *Euphytica* 91, 59–87.

Friebe, B., Qi, L., Liu, C., Liu, W., Gill, B.S., 2012. Registration of a hard red winter wheat genetic stock homozygous for facilitating alien introgression for crop improvement. *J. Plant Regist.* 6, 121–123.

Frisch, M., Bohn, M., Melchinger, A.E., 1999. Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Sci.* 39, 1295–1301.

Frisch, M., Melchinger, A., 2005. Selection theory for marker-assisted backcrossing. *Genetics* 2, 909–917.

Fukuoka, H., Miyatake, K., Nunome, T., Negoro, S., Shirasawa, K., Isobe, S., Asamizu, E., Yamaguchi, H., Ohshima, A., 2012. Development of gene-based markers and construction of an integrated linkage map in eggplant by using *Solanum* orthologous (SOL) gene sets. *Theor. Appl. Genet.* 125, 47–56.

Fukuoka, H., Yamaguchi, H., Nunome, T., Negoro, S., Miyatake, K., Ohshima, A., 2010. Accumulation, functional annotation, and comparative analysis of expressed sequence tags in eggplant (*Solanum melongena* L.), the third pole of the genus *Solanum* species after tomato and potato. *Gene* 450, 76–84.

Fulton, T.M., Beck-Bunn, T., Emmatty, D., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., Zamir, D., Tanksley, S.D., 1997. QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. *Theor. Appl. Genet.* 95, 881–894.

Fulton, T.M., 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14, 1457–1467.

Furini, A., Wunder, J., 2004. Analysis of eggplant (*Solanum melongena*)-related germplasm: Morphological and AFLP data contribute to phylogenetic interpretations and germplasm utilization. *Theor. Appl. Genet.* 108, 197–208.

Gabriel, S., Ziaugra, L., Tabbaa, D., 2009. SNP genotyping using the sequenom massARRAY iPLEX Platform. *Curr. Protoc. Hum. Genet.* 2009, 2–12



- Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.M., Hansen, M., Joets, J., 2011. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6, e28334.
- Gandhi, G.R., Ignacimuthu, S., Paulraj, M.G., Sasikumar, P., 2011. Antihyperglycemic activity and antidiabetic effect of methyl caffeate isolated from *Solanum torvum* Swartz. fruit in streptozotocin induced diabetic rats. Eur. J. Pharmacol. 670, 623–631.
- Gao, C., Xin, P., Cheng, C., Tang, Q., Chen, P., Wang, C., Zang, G., Zhao, L., 2014. Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers. PLoS One 9, e110638.
- Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv Prepr. arXiv 1207, 3907.
- Ge, H., Liu, Y., Jiang, M., Zhang, J., Han, H., Chen, H., 2013. Analysis of genetic diversity and structure of eggplant populations (*Solanum melongena* L.) in China using simple sequence repeat markers. Sci. Hortic. 162, 71–75.
- Gisbert, C., Prohens, J., Raigón, M.D., Stommel, J.R., Nuez, F., 2011. Eggplant relatives as sources of variation for developing new rootstocks: Effects of grafting on eggplant yield and fruit apparent quality and composition. Sci. Hortic. 128, 14–22.
- Gleddie, S., Keller, W.A., Setterfield, G., 1986. Production and characterization of somatic hybrids between *Solanum melongena* L. and *S. sisymbriifolium* Lam. Theor. Appl. Genet. 71, 613–621.
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: The challenge of feeding 9 billion people. Science 327, 812–818.
- Gonzaga, Z.J., Aslam, K., Septiningsih, E.M., Collard, B.C.Y., 2015. Evaluation of SSR and SNP markers for molecular breeding in rice. Plant Breed. Biotechnol. 3, 139–152.
- Gonzalez-Ibeas, D., Blanca, J., Roig, C., Gonzalez-To, M., Pico, B., Truniger, V., Gomez, P., Deleu, W., Cano-Delgado, A., Arus, P., et

- al. 2007. MELOGEN: an EST database for melon functional genomics. *BMC Genomics* 8, 306.
- Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351.
- Gousset, C., Collonnier, C., Mulya, K., Mariska, I., Rotino, G.L., Besse, P., Servaes, A., Sihachakr, D., 2005. *Solanum torvum*, as a useful source of resistance against bacterial and fungal diseases for improvement of eggplant (*S. melongena* L.). *Plant Sci.* 168, 319–327.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
- Gramazio, P., Prohens, J., Plazas, M., Andújar, I., Herraiz, F., Castillo, E., Knapp, S., Meyer, R.S., Vilanova, S., 2014. Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. *BMC Plant Biol.* 14, 350.
- Gramazio, P., Blanca, J., Ziarsolo, P., Herraiz, F.J., Plazas, M., Prohens, J., Vilanova, S., 2016. Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common eggplant (*Solanum melongena*) with interest for breeding. *BMC Genomics* 17, 300.
- Groves, R.A., Hagel, J.M., Zhang, Y., Kilpatrick, K., Levy, A., Marsolais, F., Lewinsohn, E., Sensen, C.W., Facchini, P.J., 2015. Transcriptome profiling of khat (*Catha edulis*) and *Ephedra sinica* reveals gene candidates potentially involved in amphetamine-type alkaloid biosynthesis. *PLoS One* 10, e0119701.
- Guerrero, R.F., Posto, A.L., Moyle, L.C., Hahn, M.W., 2016. Genome-wide patterns of regulatory divergence revealed by introgression lines. *Evolution.* 70, 696–706.
- Gupta, P.K., Rustgi, S., Mir, R.R., 2008. Array-based high-throughput DNA markers for crop improvement. *Heredity* 101, 5–18.
- Gupta, M., Mason, A.S., Batley, J., Bharti, S., Banga, S., Banga, S.S., 2016. Molecular-cytogenetic characterization of C-genome

- chromosome substitution lines in *Brassica juncea* (L.) Czern and Coss. *Theor. Appl. Genet.* 129, 1153–1166.
- Gur, A., Zamir, D., 2004. Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol.* 2, e245.
- Gur, A., Zamir, D., 2015. Mendelizing all Components of a Pyramid of Three Yield QTL in Tomato. *Front. Plant Sci.* 6, 1096.
- Guri, A., Sink, K.C., 1988. Interspecific somatic hybrid plants between eggplant (*Solanum melongena*) and *Solanum torvum*. *Theor. Appl. Genet.* 76, 490–496.
- Haghighi, K.R., Ascher, P.D., 1988. Fertile, intermediate hybrids between *Phaseolus vulgaris* and *P. acutifolius* from congruity backcrossing. *Sex. Plant Reprod.* 1, 51–58.
- Hajjar, R., Hodgkin, T., 2007. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* 156, 1–13.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Hammer, K., 1984. Das Domestikationssyndrom. *Die Kult.* 32, 11–34.
- Hammer, Ø., Harper, D., Ryan, P., 2008. PAST-palaeontological statistics, ver. 1.89. Paleontological Museum, University of Oslo, Norway.
- Hanson, P.M., Yang, R.Y., Tsou, S., Ledesma, D., Engle, L., Lee, T.C., 2006. Diversity in eggplant (*Solanum melongena*) for superoxide scavenging activity, total phenolics, and ascorbic acid. *J. Food Compos. Anal.* 19, 594–600.
- Harlan, J.R., de Wet, J.M.J., 1971. Toward a rational classification of cultivated plants. *Taxon* 20, 509.
- Harlan, J.R., 1992. Crops and man. American Society of Agronomy 1992.
- Hartung, F., Schiemann, J., 2014. Precise plant breeding using new genome editing techniques: Opportunities, safety and regulation in the EU. *Plant J.* 78, 742–752.

- Herzog, E., Falke, K.C., Presterl, T., Scheuermann, D., Ouzunova, M., Frisch, M., 2014. Selection strategies for the development of maize introgression populations. *PLoS One* 9, e92429.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., Yamaguchi, H., Sato, S., Isobe, S., Tabata, S., Fukuoka, H., 2014. Draft genome sequence of eggplant (*Solanum melongena* L.): The representative *Solanum* species indigenous to the old world. *DNA Res.* 21, 649–660.
- Hogenboom, N.G., 1984. Incongruity: Non-Functioning of intercellular and intracellular partner relationships through non-matching information. In: Linskens, H., Heslop-Harrison, J. (Eds.), *Cellular Interactions*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 640–654.
- Hong, C.P., Piao, Z.Y., Kang, T.W., Batley, J., Yang, T., Hur, Y., 2007. Genomic distribution of simple sequence repeats in *Brassica rapa*. *Mol. Cells* 23, 349–356.
- Hospital, F., Chevalet, C., Mulsant, P., 1992. Using markers in gene introgression breeding programs. *Genetics* 132, 1199–1210.
- Hospital, F., Charcosset, A., 1997. Marker-assisted introgression of quantitative trait loci. *Genetics* 148, 1469–1485.
- Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Huang, Z., Van Houten, J., Gonzalez, G., Xiao, H., Van Der Knaap, E., 2013. Genome-wide identification, phylogeny and expression analysis of *SUN*, *OFP* and *YABBY* gene family in tomato. *Mol. Genet. Genomics* 288, 111–129.
- Hurtado, M., Vilanova, S., Plazas, M., Gramazio, P., Herraiz, F.J., Andújar, I., Prohens, J., 2013. Phenomics of fruit shape in eggplant (*Solanum melongena* L.) using tomato analyzer software. *Sci. Hortic.* 164, 625–632.
- Ikeda, H., Shibuya, T., Nishiyama, M., Nakata, Y., Kanayama, Y., 2017. Physiological mechanisms accounting for the lower incidence of

- blossom-end rot in tomato introgression line IL8-3 fruit. *Hortic. J.* 86, 327–333.
- Iseli, C., Jongeneel, C.V., Bucher, P., 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In: *International Conference on Intelligent Systems for Molecular Biology*. pp. 138–148.
- Isshiki, S., Iwata, N., Khan, M.M.R., 2008. ISSR variations in eggplant (*Solanum melongena* L.) and related *Solanum* species. *Sci. Hortic.* 117, 186–190.
- Jacobsen, E., Schouten, H.J., 2007. Cisgenesis strongly improves introgression breeding and induced translocation breeding of plants. *Trends Biotechnol.* 25, 219–223.
- Jaiswal, B.S., 2012. *Solanum torvum*: A review of its traditional uses, phytochemistry and pharmacology. *Int. J. Pharma Bio Sci.* 3, 104–111.
- Jarvis, D.I., Hodgkin, T., 1999. Wild relatives and crop cultivars: Detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. *Mol. Ecol.* 8, S159–S173.
- Jarvis, A., Lane, A., Hijmans, R.J., 2008. The effect of climate change on crop wild relatives. *Agric. Ecosyst. Environ.* 126, 13–23.
- Jeuken, M., Van Wijk, R., Peleman, J., Lindhout, P., 2001. An integrated interspecific AFLP map of lettuce (*Lactuca*) based on two *L. sativa* × *L. saligna* F2 populations. *Theor. Appl. Genet.* 103, 638–647.
- Jo, K.R., Kim, C.J., Kim, S.J., Kim, T.Y., Bergervoet, M., Jongsma, M.A., Visser, R.G., Jacobsen, E., Vossen, J.H., 2014. Development of late blight resistant potatoes by cisgene stacking. *BMC Biotechnol.* 14, 50.
- Johnson, A., Veilleux, R., 2001. Somatic hybridization and application in plant breeding. In: *Plant Breeding Reviews*. John Wiley & Sons, Inc., Oxford, UK, pp. 167–225.
- Jones, T.A., 2003. The restoration gene pool concept: Beyond the native versus non-native debate. *Restor. Ecol.* 11, 281–290.
- Jones, E.S., Sullivan, H., Bhatramakki, D., Smith, J.S.C., 2007. A comparison of simple sequence repeat and single nucleotide

- polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). *Theor. Appl. Genet.* 115, 361–371.
- Kaeppler, S.M., Kaeppler, H.F., Rhee, Y., 2000. Epigenetic aspects of somaclonal variation in plants. *Plant Mol. Biol.* 43, 179–188.
- Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., Dhawan, A.K., 2011. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* 177, 309–334.
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Kantar, M.B., Sosa, C.C., Khoury, C.K., Castañeda-Álvarez, N.P., Achicanoy, H.A., Bernau, V., Kane, N.C., Marek, L., Seiler, G., Rieseberg, L.H., 2015. Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.). *Front. Plant Sci.* 6, 841.
- Kashyap, V., Vinod Kumar, S., Collonnier, C., Fusari, F., Haicour, R., Rotino, G.L., Sihachakr, D., Rajam, M. V., 2003. Biotechnology of eggplant. *Sci. Hortic.* 97, 1–25.
- Kaushik, P., Prohens, J., Vilanova, S., Gramazio, P., Plazas, M., 2016. Phenotyping of eggplant wild relatives and interspecific hybrids with conventional and phenomics descriptors provides insight for their potential utilization in breeding. *Front. Plant Sci.* 7, 677.
- Kearsey, M.J., Farquhar, a G.L., 1998. QTL analysis in plants; where are we now? *Heredity* 80, 137–142.
- Khan, M.M.R., Hasnunnahar, M., Isshiki, S., 2013. Production of amphidiploids of the hybrids between *Solanum macrocarpon* and eggplant. *HortScience* 48, 422–424.
- Khan, M.M.R., Hasnunnahar, M., Iwayoshi, M., Ogura-Tsujita, Y., Isshiki, S., 2015. Pollen degeneration in three functional male-sterile lines of eggplant with the wild *Solanum* cytoplasms. *Hortic. Environ. Biotechnol.* 56, 350–357.
- Khazaei, H., Street, K., Bari, A., Mackay, M., Stoddard, F.L., 2013. The FIGS (Focused Identification of Germplasm Strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS One* 8, e63107.

- Khlestkina, E.K., 2014. Current applications of wheat and wheat–alien precise genetic stocks. *Mol. Breed.* 34, 273–281.
- Khoury, C., Laliberté, B., Guarino, L., 2010. Trends in ex situ conservation of plant genetic resources: A review of global crop and regional conservation strategies. *Genet. Resour. Crop Evol.* 57, 625–639.
- Khush, G.S., Brar, D.S., 1992. Overcoming the Barriers in Hybridization. In: *Distant Hybridization of Crop Plants*. Springer, Berlin, Heidelberg, pp. 47–61.
- Kilian, B., Mammen, K., Millet, E., Sharma, R., Graner, A., Salamini, F., Hammer, K., Özkan, H., 2011. *Aegilops*. In: *Wild Crop Relatives: Genomic and Breeding Resources*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–76.
- Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L.S., Paterson, A.H., 2016. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* 242, 14–22
- King, J., Grewal, S., Yang, C., Hubbart, S., Scholefield, D., Ashling, S., Edwards, K.J., Allen, A.M., Burrige, A., Bloor, C., et al. 2017. A step change in the transfer of interspecific variation into wheat from *Amblyopyrum muticum*. *Plant Biotechnol. J.* 15, 217–226.
- Knapp, S., Vorontsova, M.S., Prohens, J., 2013. Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): New understanding of species names in a complex group. *PLoS One* 8, e57039.
- Knox, J., Hess, T., Daccache, A., Wheeler, T., 2012. Climate change impacts on crop productivity in Africa and South Asia. *Environ. Res. Lett.* 7, 34032.
- Koester, R.P., Sisco, P.H., Stuber, C.W., 1993. Identification of quantitative trait loci controlling days to flowering and plant height in two near isogenic lines of maize. *Crop Sci.* 33, 1209–1216.
- Kofler, R., Schlötterer, C., Lelley, T., 2007. SciRoKo: A new tool for whole genome microsatellite search and investigation. *Bioinformatics* 23, 1683–1685.
- Kole, C., Muthamilarasan, M., Henry, R., Edwards, D., Sharma, R., Abberton, M., Batley, J., Bentley, A., Blakeney, M., Bryant, J., 2015. Application of genomics-assisted breeding for generation of climate resilient crops: Progress and prospects. *Front. Plant Sci.* 6, 563.

- Kosambi, D., 1944. The estimation of map distances from recombination values. *Ann Eugen* 12, 172–175.
- Kouassi, B., Prohens, J., Gramazio, P., Kouassi, A.B., Vilanova, S., Galán-Ávila, A., Herraiz, F.J., Kouassi, A., Seguí-Simarro, J.M., Plazas, M., 2016. Development of backcross generations and new interspecific hybrid combinations for introgression breeding in eggplant (*Solanum melongena*). *Sci. Hortic.* 213, 199–207.
- Krannich, C.T., Maletzki, L., Kurowsky, C., Horn, R., 2015. Network candidate genes in breeding for drought tolerant crops. *Int. J. Mol. Sci.* 16, 16378-16400.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circo: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645.
- Kumar, A., Ellis, B.E., 2001. The phenylalanine ammonia-lyase gene family in raspberry. Structure, expression, and evolution. *Plant Physiol.* 127, 230–239.
- Kumar, S., Banks, T.W., Cloutier, S., 2012. SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genomics* 2012.
- Kumar, B., Kumar, U., Yadav, H.K., 2015. Identification of EST–SSRs and molecular diversity analysis in *Mentha piperita*. *Crop J.* 3, 335–342.
- Kumchai, J., Wei, Y.C., Lee, C.Y., Chen, F.C., Chin, S.W., 2013. Production of interspecific hybrids between commercial cultivars of the eggplant (*Solanum melongena* L.) and its wild relative *S. torvum*. *Genet. Mol. Res.* 12, 755–764.
- Kwon, Y.I., Apostolidis, E., Shetty, K., 2008. In vitro studies of eggplant (*Solanum melongena*) phenolics as inhibitors of key enzymes relevant for type 2 diabetes and hypertension. *Bioresour. Technol.* 99, 2981–2988.
- Landi, S., Nurcato, R., De Lillo, A., Lentini, M., Grillo, S., Esposito, S., 2016. Glucose-6-phosphate dehydrogenase plays a central role in the response of tomato (*Solanum lycopersicum*) plants to short and long-term drought. *Plant Physiol. Biochem.* 105, 79–89.
- Lanteri, S., Barchi, L., Toppino, L., Sala, T., Acquadro, A., Portis, E.,



- Rinaldi, R., Scaglione, D., Dal Molin, A., Minio, A., 2014. The first eggplant (*Solanum melongena* L.) high quality genome draft. In: Programme and Abstracts-Applied Vegetable Genomics. p. 24.
- Lebeau, A., Gouy, M., Daunay, M.C., Wicker, E., Chiroleu, F., Prior, P., Frary, A., Dintinger, J., 2013. Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theor. Appl. Genet.* 126, 143–158.
- Lebreton, C.H., Visscher, P.M., Haley, C.S., Semikhodskii, A., Quarrie, S.A., 1998. A nonparametric bootstrap method for testing close linkage vs. pleiotropy of coincident quantitative trait loci. *Genetics* 150, 931–943.
- Lee, S., Abecasis, G.R., Boehnke, M., Lin, X., 2014. Rare-variant association analysis: Study designs and statistical tests. *Am. J. Hum. Genet.* 95, 5–23.
- Lester, R.N., Niakan, L., 1986. Origin and domestication of the scarlet eggplant, *Solanum aethiopicum*, from *S. anguivi* in Africa. In: D'Arcy, W.G. (Ed.), *Solanaceae: Biology and Systematics*. Columbia University Press, New York, NY, pp. 433–456.
- Lester, R.N., Thitai, G.N.W., 1989. Inheritance in *Solanum aethiopicum*, the scarlet eggplant. *Euphytica* 40, 67–74.
- Lester, R.N., Jaeger, P.M.L., Bleijendaal-Spierings, B.H.M., Bleijendaal, H.P.O., Holloway, H.L.O., 1990. African eggplants-a review of collecting in West Africa. *Plant Genet. Resour. Newsl.* 17–26.
- Lester, R.N., Hasan, S.M.Z., 1991. Origin and domestication of the brinjal eggplant, *Solanum melongena*, from *S. incanum*, in Africa and Asia. In: Hawkes, J.G., Lester, R.N., Nee, M.H., Estrada, N. (Eds.), *Solanaceae III: Taxonomy, Chemistry, Evolution*. Royal Botanic Gardens, Kew, London, U.K., pp. 369–388.
- Lester, R.N., Kang, J.H., 1998. Embryo and endosperm function and failure in *Solanum* species and hybrids. *Ann. Bot.* 82, 445–453.
- Lester, R.N., Daunay, M.C., 2003. Diversity of African vegetable *Solanum* species and its implications for a better understanding of plant domestication. In: Rudolf Mansfeld and Plant Genetic Resources. *Schriften Zu Genetischen Ressourcen*. pp. 137–152.
- Lester, R.N., Jaeger, P.M.L., Child, A., 2011. *Solanum* in Africa.

Birmingham: Celia Lester.

- Levin, R.A., Myers, N.R., Bohs, L., 2006. Phylogenetic relationships among the ‘Spiny Solanums’ (*Solanum* subgenus *Leptostemonum*, Solanaceae). *Am. J. Bot.* 93, 157–169.
- Li, B., Dewey, C.N., 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr.* 1303, 3997.
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J., 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20, 265–272.
- Li, Y.C., Korol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21, 991–1007.
- Lippman, Z.B., Semel, Y., Zamir, D., 2007. An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr. Opin. Genet. Dev.* 17, 545–552.
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., Arenas-Huertero, C., Chua, N.H., 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345.
- Liu, J., Wang, H., Chua, N., 2015a. Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* 13, 319–328.
- Liu, J., Zheng, Z., Zhou, X., Feng, C., Zhuang, Y., 2015b. Improving the resistance of eggplant (*Solanum melongena*) to *Verticillium* wilt using wild species *Solanum linnaeanum*. *Euphytica* 201, 463–469.

- Liu, S., Zhou, R., Dong, Y., Li, P., Jia, J., 2006. Development, utilization of introgression lines using a synthetic wheat as donor. *Theor. Appl. Genet.* 112, 1360–1373.
- Lo Scalzo, R., Fibiani, M., Mennella, G., Rotino, G.L., Dal Sasso, M., Culici, M., Spallino, A., Braga, P.C., 2010. Thermal treatment of eggplant (*Solanum melongena* L.) increases the antioxidant content and the inhibitory effect on human neutrophil burst. *J. Agric. Food Chem.* 58, 3371–3379.
- Lu, X., Kim, H., Zhong, S., Chen, H., Hu, Z., Zhou, B., 2014. De novo transcriptome assembly for rudimentary leaves in Litchi chinensis Sonn. and identification of differentially expressed genes in response to reactive oxygen species. *BMC Genomics* 15, 805.
- Lusser, M., Paris, i C., Plan, D., Rodríguez-Cerezo, E., 2011. New plant breeding techniques: State-of-the-art and prospects for commercial development. Reference Report by the Joint Research Centre of the European Commission.
- Lusser, M., 2014. Workshop on public-private partnerships in plant breeding. In: Directorate Growth & Innovation and JRC-Seville, Joint Research Centre.
- Luthria, D., Singh, A.P., Wilson, T., Vorsa, N., Banuelos, G.S., Vinyard, B.T., 2010. Influence of conventional and organic agricultural practices on the phenolic content in eggplant pulp: Plant-to-plant variation. *Food Chem.* 121, 406–411.
- M. Perez-de-Castro, A., Vilanova, S., Canizares, J., Pascual, L., M. Blanca, J., J. Diez, M., Prohens, J., Pico, B., 2012. Application of Genomic Tools in Plant Breeding. *Curr. Genomics* 13, 179–195.
- Ma, C., Dastmalchi, K., Whitaker, B.D., Kennelly, E.J., 2011. Two new antioxidant malonated caffeoylquinic acid isomers in fruits of wild eggplant relatives. *J. Agric. Food Chem.* 59, 9645–9651.
- Mace, E.S., Lester, R.N., Gebhardt, C.G., 1999. AFLP analysis of genetic relationships among the cultivated eggplant, *Solanum melongena* L., and wild relatives (Solanaceae). *Theor. Appl. Genet.* 99, 626–633.
- Mammadov, J., Aggarwal, R., Buyyarapu, R., Kumpatla, S., 2012. SNP markers and their impact on plant breeding. *Int. J. Plant Genomics*.
- Martín, A., Alvarez, J.B., Martín, L.M., Barro, F., Ballesteros, J., 1999.

The development of Tritordeum: a novel cereal for food processing. *J. Cereal Sci.* 30, 85–95.

- Maundu, P., Achigan-Dako, E., Morimoto, Y., 2009. Biodiversity of African vegetables. In: Shackleton, C., Pasquini, M., Drescher, A. (Eds.), *African Indigenous Vegetables in Urban Agriculture*. Routledge, pp. 65–104.
- Mavromatis, K., Land, M.L., Brettin, T.S., Quest, D.J., Copeland, A., Clum, A., Goodwin, L., Woyke, T., Lapidus, A., Klenk, H.P., et al., 2012. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS One* 7, e48837.
- Maxted, N., Ford-Lloyd, B. V., Jury, S., Kell, S., Scholten, M., 2006. Towards a definition of a crop wild relative. *Biodivers. Conserv.* 15, 2673–2685.
- Maxted, N., Dulloo, E., Ford-Lloyd, B. V., Iriondo, J.M., Jarvis, A., 2008. Gap analysis: A tool for complementary genetic conservation assessment. *Divers. Distrib.* 14, 1018–1030.
- Maxted, N., Kell, S., 2009. Establishment of a global network for the in situ conservation of crop wild relatives: Status and needs. *FAO Commission on Genetic Resources for Food and Agriculture*, Rome, Italy, 266.
- McIntosh, R.A., 1992. Pre-emptive breeding to control wheat rusts. *Euphytica* 63, 103–113.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- McKhann, H.I., Camilleri, C., Bérard, A., Bataillon, T., David, J.L., Reboud, X., Le Corre, V., Caloustian, C., Gut, I.G., Brunel, D., 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* 38, 193–202.
- Menda, N., Strickler, S.R., Edwards, J.D., Bombarely, A., Dunham, D.M., Martin, G.B., Mejia, L., Hutton, S.F., Havey, M.J., Maxwell, D.P., Mueller, L.A., 2014. Analysis of wild-species introgressions in

- tomato inbreds uncovers ancestral origins. *BMC Plant Biol.* 14, 287.
- Menin, B., Comino, C., Moglia, A., Dolzhenko, Y., Portis, E., Lanteri, S., 2010. Identification and mapping of genes related to caffeoylquinic acid synthesis in *Cynara cardunculus* L. *Plant Sci.* 179, 338–347.
- Mennella, G., Rotino, G.L., Fibiani, M., D’Alessandro, A., Franceses, G., Toppino, L., Cavallanti, F., Acciarri, N., Scalzo, R.L.O., 2010. Characterization of health-related compounds in eggplant (*Solanum Melongena* L.) lines derived from introgression of allied species. *J. Agric. Food Chem.* 58, 7597–7603.
- Mennella, G., Lo Scalzo, R., Fibiani, M., D’Alessandro, A., Francese, G., Toppino, L., Acciarri, N., de Almeida, A.E., Rotino, G.L., 2012. Chemical and bioactive quality traits during fruit ripening in eggplant (*S. melongena* L.) and allied species. *J. Agric. Food Chem.* 60, 11821–11831.
- Metzgar, D., Bytof, J., Wills, C., 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.* 10, 72–80.
- Meyer, R.S., Duval, A.E., Jensen, H.R., 2012a. Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48.
- Meyer, R.S., Karol, K.G., Little, D.P., Nee, M.H., Litt, A., 2012b. Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. *Mol. Phylogenet. Evol.* 63, 685–701.
- Meyer, R.S., Purugganan, M.D., 2013. Evolution of crop species: Genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852.
- Meyer, R.S., 2015. Encouraging metadata curation in the Diversity Seek initiative. *Nat. Plants* 1, 15099.
- Meyer, R.S., Whitaker, B.D., Little, D.P., Wu, S.B., Kennelly, E.J., Long, C.L., Litt, A., 2015. Parallel reductions in phenolic constituents resulting from the domestication of eggplant. *Phytochemistry* 115, 194–206.
- Miceli, A., Sabatino, L., Moncada, A., Vetrano, F., D’Anna, F., 2014. Nursery and field evaluation of eggplant grafted onto unrooted cuttings of *Solanum torvum* Sw. *Sci. Hortic. (Amsterdam)*. 178, 203–

- Miyatake, K., Saito, T., Negoro, S., Yamaguchi, H., Nunome, T., Ohyama, A., Fukuoka, H., 2012. Development of selective markers linked to a major QTL for parthenocarpy in eggplant (*Solanum melongena* L.). *Theor. Appl. Genet.* 124, 1403–1413.
- Miyatake, K., Saito, T., Negoro, S., Yamaguchi, H., Nunome, T., Ohyama, A., Fukuoka, H., 2016. Detailed mapping of a resistance locus against *Fusarium* wilt in cultivated eggplant (*Solanum melongena*). *Theor. Appl. Genet.* 129, 357–367.
- Mizuno, H., Kawahara, Y., Wu, J., Katayose, Y., Kanamori, H., Ikawa, H., Itoh, T., Sasaki, T., Matsumoto, T., 2011. Asymmetric distribution of gene expression in the centromeric region of rice chromosome 5. *Front. Plant Sci.* 2, 16.
- Moglia, A., Comino, C., Menin, B., Portis, E., Acquadro, A., Beekwilder, J., Hehn, A., Bourgaud, F., Lanteri, S., 2013. Caffeoylquinic acids biosynthesis and accumulation in *Cynara cardunculus*: State of the art. In: *Acta Horticulturae*. pp. 401–406.
- Mohan, M., Kamble, S., Gadhi, P., Kasture, S., 2010. Protective effect of *Solanum torvum* on doxorubicin-induced nephrotoxicity in rats. *Food Chem. Toxicol.* 48, 436–440.
- Moore, G., 2015. Strategic pre-breeding for wheat improvement. *Nat. Plants* 1, 15018.
- Morgulis, A., Gertz, E.M., Schäffer, A.A., Agarwala, R., 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* 13, 1028–1040.
- Morrell, P.L., Buckler, E.S., Ross-Ibarra, J., 2011. Crop genomics: advances and applications. *Nat. Rev. Genet.* 13, 85–96.
- Mott, R., 1997. EST\_GENOME: A program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* 13, 477–478.
- Muñoz-Falcón, J.E., Prohens, J., Vilanova, S., Nuez, F., 2009. Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. *Ann. Appl. Biol.* 154, 453–465.

- Muñoz, L.C., Blair, M.W., Duque, M.C., Tohme, J., Roca, W., 2004. Introgression in common bean x tepary bean interspecific congruity-backcross lines as measured by AFLP markers. *Crop Sci.* 44, 637–645.
- Murata, M., Nishimura, M., Murai, N., Haruta, M., Homma, S., Itoh, Y., 2001. A transgenic apple callus showing reduced polyphenol oxidase activity and lower browning potential. *Biosci. Biotechnol. Biochem.* 65, 383–388.
- Mutegi, E., Snow, A.A., Rajkumar, M., Pasquet, R., Ponniah, H., Daunay, M.-C., Davidar, P., 2015. Genetic diversity and population structure of wild/weedy eggplant (*Solanum insanum*, Solanaceae) in southern India: implications for conservation. *Am. J. Bot.* 102, 140–148.
- Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G., Stahl, F., 2013. Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* 24, 22–30.
- Newman, S.M., Tantasawat, P., Steffens, J.C., 2011. Tomato polyphenol oxidase B is spatially and temporally regulated during development and in response to ethylene. *Molecules* 16, 493–517.
- NGS\_Crumbs software ([https://bioinf.comav.upv.es/ngs\\_crumbs](https://bioinf.comav.upv.es/ngs_crumbs)).
- Niederhuth, C.E., Patharkar, O.R., Walker, J.C., 2013. Transcriptional profiling of the *Arabidopsis* abscission mutant *hae hsl2* by RNA-Seq. *BMC Genomics* 14, 37.
- Niggeweg, R., Michael, A., Martin, C., 2004. Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat. Biotechnol.* 22, 746–754.
- Nilsson, N.O., Hallden, C., Hansen, M., Hjerdin, A., Sall, T., 1997. Comparing the Distribution of Rapd and RFLP Markers in a High Density Linkage Map of Sugar Beet. *Genome* 40, 644–651.
- Nunome, T., Ishiguro, K., Yoshida, T., Hirai, M., 2001. Mapping of fruit shape and color development traits in eggplant (*Solanum melongena* L.) based on RAPD and AFLP markers. *Breed. Sci.* 51, 19–26.
- Nunome, T., Suwabe, K., Iketani, H., Hirai, M., 2003. Identification and characterization of microsatellites in eggplant. *Plant Breed.* 122, 256–262.

- Nunome, T., Suwabe, K., Ohya, A., Fukuoka, H., 2003. Characterization of trinucleotide microsatellites in eggplant. *Breed. Sci.* 53, 77–83.
- Nunome, T., Negoro, S., Kono, I., Kanamori, H., Miyatake, K., Yamaguchi, H., Ohya, A., Fukuoka, H., 2009. Development of SSR markers derived from SSR-enriched genomic library of eggplant (*Solanum melongena* L.). *Theor. Appl. Genet.* 119, 1143–1153.
- Ozsolak, F., Milos, P.M., 2011. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98.
- Pariasca-Tanaka, J., Lorieux, M., He, C., McCouch, S., Thomson, M.J., Wissuwa, M., 2015. Development of a SNP genotyping panel for detecting polymorphisms in *Oryza glaberrima/O. sativa* interspecific crosses. *Euphytica* 201, 67–78.
- Pascual, L., Albert, E., Sauvage, C., Duangjit, J., Bouchet, J.P., Bitton, F., Desplat, N., Brunel, D., Le Paslier, M.C., Ranc, N., et al. 2016. Dissecting quantitative trait variation in the resequencing era: Complementarity of bi-parental, multi-parental and association panels. *Plant Sci.* 242, 120–130.
- Pascual, L., Desplat, N., Huang, B.E., Desgroux, A., Bruguier, L., Bouchet, J.P., Le, Q.H., Chauchard, B., Verschave, P., Causse, M., 2015. Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol. J.* 13, 565–577.
- Patwardhan, A., Ray, S., Roy, A., 2014. Molecular markers in phylogenetic studies-a review. *J. Phylogenetics Evol. Biol.* 2, 131.
- Pea, G., Aung, H.H., Frascaroli, E., Landi, P., Pè, M.E., 2013. Extensive genomic characterization of a set of near-isogenic lines for heterotic QTL in maize (*Zea mays* L.). *BMC Genomics* 14, 61.
- Peleg, Z., Fahima, T., Krugman, T., Abbo, S., Yakir, D., Korol, A.B., Saranga, Y., 2009. Genomic dissection of drought resistance in durum wheat × wild emmer wheat recombinant inbred line population. *Plant, Cell Environ.* 32, 758–779.
- Peleman, J.D., Van Der Voort, J.R., 2003. Breeding by design. *Trends Plant Sci.* 8, 330–334.
- Pestsova, E.G., Börner, A., Röder, M.S., 2006. Development and QTL



- assessment of *Triticum aestivum*-*Aegilops tauschii* introgression lines. *Theor. Appl. Genet.* 112, 634–647.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7, e37135.
- Pico, B., Herraiz, J., Ruiz, J.J., Nuez, F., 2002. Widening the genetic basis of virus resistance in tomato. *Sci. Hortic.* 94, 73–89.
- Pinosio, S., González-Martínez, S.C., Bagnoli, F., Cattonaro, F., Grivet, D., Marroni, F., Lorenzo, Z., Pausas, J.G., Verdú, M., Vendramin, G.G., 2014. First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus halepensis* Mill. *Mol. Ecol. Resour.* 14, 846–856.
- Plazas, M., Andújar, I., Vilanova, S., Hurtado, M., Gramazio, P., Herraiz, F.J., Prohens, J., 2013a. Breeding for Chlorogenic Acid Content in Eggplant: Interest and Prospects. *Not. Bot. Horti Agrobot. Cluj-Napoca* 41, 26–35.
- Plazas, M., López-Gresa, M.P., Vilanova, S., Torres, C., Hurtado, M., Gramazio, P., Andújar, I., Herráiz, F.J., Bellés, J.M., Prohens, J., 2013b. Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: Fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning. *J. Agric. Food Chem.* 61, 8871–8879.
- Plazas, M., Andújar, I., Vilanova, S., Gramazio, P., Herraiz, F.J., Prohens, J., 2014a. Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (*Solanum aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplant complexes. *Front. Plant Sci.* 5, 318.
- Plazas, M., Prohens, J., Cuñat, A.N., Vilanova, S., Gramazio, P., Herraiz, F.J., Andújar, I., 2014b. Reducing capacity, chlorogenic acid content and biological activity in a collection of scarlet (*Solanum aethiopicum*) and Gboma (*S. macrocarpon*) eggplants. *Int. J. Mol. Sci.* 15, 17221–17241.
- Plazas, M., Vilanova, S., Gramazio, P., Rodríguez-Burruezo, A., Fita, A., Herraiz, F.J., Ranil, R., Fonseka, R., Niran, L., Fonseka, H., et al. 2016. Interspecific hybridization between eggplant and wild relatives

- from different genepools. *J. Am. Soc. Hortic. Sci.* 141, 34–44.
- Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J.P., Hyvonen, J., 2013. Advances in plant gene-targeted and functional markers: A review. *Plant Methods* 9, 6.
- Porch, T., Beaver, J., Debouck, D., Jackson, S., Kelly, J., Dempewolf, H., 2013. Use of Wild Relatives and Closely Related Species to Adapt Common Bean to Climate Change. *Agronomy* 3, 433–461.
- Portis, E., Mauromicale, G., Mauro, R., Acquadro, A., Scaglione, D., Lanteri, S., 2009. Construction of a reference molecular linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus*). *Theor. Appl. Genet.* 120, 59–70.
- Portis, E., Scaglione, D., Acquadro, A., Mauromicale, G., Mauro, R., Knapp, S.J., Lanteri, S., 2012. Genetic mapping and identification of QTL for earliness in the globe artichoke/cultivated cardoon complex. *BMC Res. Notes* 5, 252.
- Portis, E., Barchi, L., Toppino, L., Lanteri, S., Acciarri, N., Felicioni, N., Fusari, F., Barbierato, V., Cericola, F., Valè, G., Rotino, G.L., 2014. QTL mapping in eggplant reveals clusters of yield-related loci and orthology with the tomato genome. *PLoS One* 9, e89499.
- Portis, E., Cericola, F., Barchi, L., Toppino, L., Acciarri, N., Pulcini, L., Sala, T., Lanteri, S., Rotino, G.L., 2015. Association mapping for fruit, plant and leaf morphology traits in eggplant. *PLoS One* 10.
- Prakash, S., Ahuja, I., Upreti, H.C., Dinesh Kumar, V., Bhat, S.R., Kirti, P.B., Chopra, V.L., 2001. Expression of male sterility in alloplasmic *Brassica juncea* with *Erucastrum canariense* cytoplasm and the development of a fertility restoration system. *Plant Breed.* 120, 479–482.
- Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Prohens, J., Anderson, G.J., Herraiz, F.J., Bernardello, G., Santos-Guerra, A., Crawford, D., Nuez, F., 2007a. Genetic diversity and conservation of two endangered eggplant relatives (*Solanum vesperilio* Aiton and *Solanum lidii* Sunding) endemic to the Canary Islands. *Genet. Resour. Crop Evol.* 54, 451–464.
- Prohens, J., Rodríguez-Burruezo, A., Raigón, M.D., Nuez, F., 2007b. Total

- phenolic concentration and browning susceptibility in a collection of different varietal types and hybrids of eggplant: Implications for breeding for higher nutritional quality and reduced browning. *J. Am. Soc. Hortic. Sci.* 132, 638–646.
- Prohens, J., Plazas, M., Raigón, M.D., Seguí-Simarro, J.M., Stommel, J.R., Vilanova, S., 2012. Characterization of interspecific hybrids and first backcross generations from crosses between two cultivated eggplants (*Solanum melongena* and *S. aethiopicum* Kumba group) and implications for eggplant breeding. *Euphytica* 186, 517–538.
- Prohens, J., Whitaker, B.D., Plazas, M., Vilanova, S., Hurtado, M., Blasco, M., Gramazio, P., Stommel, J.R., 2013. Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). *Ann. Appl. Biol.* 162, 242–257.
- Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M.J., Fita, A., Herraiz, F.J., Rodríguez-Burruezo, A., Soler, S., et al. 2017. Introgressomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213, 158.
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., et al. 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci.* 111, 5135–5140.
- Quinlan, A., Hall, I., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rafalski, J.A., 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci.* 162, 329–333.
- Raigón, M.D., Prohens, J., Muñoz-Falcón, J.E., Nuez, F., 2008. Comparison of eggplant landraces and commercial varieties for fruit content of phenolics, minerals, dry matter and protein. *J. Food Compos. Anal.* 21, 370–376.
- Raj, A., Stephens, M., Pritchard, J.K., 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589.
- Ramesh, K.R., Hemalatha, R., Vijayendra, C.A., Arshi, U.Z.S., Dushyant, S.B., Dinesh, K.B., 2016. Transcriptome analysis of *Solanum*

*melongena* L. (eggplant) fruit to identify putative allergens and their epitopes. *Gene* 576, 64–71.

- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L., 2010. A Gap analysis methodology for collecting crop gene pools: A case study with *Phaseolus* beans. *PLoS One* 5, e13497.
- Ramirez, E.C., Whitaker, J.R., Virador, V.M., 2003. Polyphenol oxidase. In: Whitaker, J.R., Voragen, A.G.J., Wong, D.W.S. (Eds.), *Handbook of Food Enzymology*. Marcel Dekker, Inc., New York, pp. 509–524.
- Ramkumar, G., Madhav, M.S., Rama Devi, S.J.S., Umakanth, B., Pandey, M.K., Prasad, M.S., Sundaram, R.M., Viraktamath, B.C., Ravindra Babu, V., 2016. Identification and validation of novel alleles of rice blast resistant gene Pi54, and analysis of their nucleotide diversity in landraces and wild *Oryza* species. *Euphytica* 209, 725–737.
- Ramos, A.M., Crooijmans, R.P.M.A., Affara, N.A., Amaral, A.J., Archibald, A.L., Beever, J.E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* 4, e6524.
- Ranil, R.H.G., Niran, H.M.L., Plazas, M., Fonseka, R.M., Fonseka, H.H., Vilanova, S., Andújar, I., Gramazio, P., Fita, A., Prohens, J., 2015. Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. *Sci. Hortic.* 193, 174–181.
- Ranil, R.H.G., Prohens, J., Aubriot, X., Niran, H.M.L., Plazas, M., Fonseka, R.M., Vilanova, S., Fonseka, H.H., Gramazio, P., Knapp, S., 2017. *Solanum insanum* L. (subgenus *Leptostemonum* Bitter, Solanaceae), the neglected wild progenitor of eggplant (*S. melongena* L.): A review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. *Genet. Resour. Crop Evol.* 64, 1707–1722.
- Ranjan, A., Ichihashi, Y., Farhi, M., Zumstein, K., Townsley, B., David-Schwartz, R., Sinha, N.R., 2014. De novo assembly and characterization of the transcriptome of the parasitic weed dodder identifies genes associated with plant parasitism. *Plant Physiol.* 166, 1186–1199.
- Ranjan, A., Budke, J.M., Rowland, S.D., Chitwood, D.H., Kumar, R.,

- Carriedo, L., Ichihashi, Y., Zumstein, K., Maloof, J.N., Sinha, N.R., 2016. eQTL regulating transcript levels Associated with diverse biological processes in tomato. bioRxiv 40592.
- Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. PLoS One 8, e66428.
- Redden, R., Yadav, S., Maxted, N., Dulloo, M., Guarino, L., Smith, P., 2015. Crop Wild Relatives and Climate Change.
- Ren, Y., McGregor, C., Zhang, Y., Gong, G., Zhang, H., Guo, S., Sun, H., Cai, W., Zhang, J., Xu, Y., 2014. An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). BMC Plant Biol. 14, 33.
- Rieseberg, L., Carney, S., 1998. Plant hybridization. New Phytol. 140, 599–624.
- Rieseberg, L.H., Arias, D.M., Ungerer, M.C., Linder, C.R., Sinervo, B., 1996. The effects of mating design on introgression between chromosomally divergent sunflower species. Theor. Appl. Genet. 93, 633–644.
- Rinaldi, R., Van Deynze, A., Portis, E., Rotino, G.L., Toppino, L., Hill, T., Ashrafi, H., Barchi, L., Lanteri, S., 2016. New Insights on Eggplant/Tomato/Pepper Synteny and Identification of Eggplant and Pepper Orthologous QTL. Front. Plant Sci. 7, 1031.
- Rizza, F., Mennella, G., Collonnier, C., Sihachakr, D., Kashyap, V., Rajam, M., Prestera, M., Rotino, G., 2002. Androgenic dihaploids from somatic hybrids between *Solanum melongena* and *S. aethiopicum* group gilo as a source of resistance to *Fusarium oxysporum* f. sp. melongenae. Plant Cell Rep. 20, 1022–1032.
- Robbins, M.D., Sim, S.C., Yang, W., Van Deynze, A., van der Knaap, E., Joobeur, T., Francis, D.M., 2011. Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. J. Exp. Bot. 62, 1831–1845.
- Rommens, C.M., Richael, C.M., Yan, H., Navarre, D.A., Ye, J., Krucker, M., Swords, K., 2008. Engineered native pathways for high

- kaempferol and caffeoylquinic acid production in potato. *Plant Biotechnol. J.* 6, 870–886.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., et al. 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl. Acad. Sci.* 111, 3268–3273.
- Rotino, G.L., Sala, T., Toppino, L., 2014. Eggplant In: Kumar, J., Pratap, A., (Eds.), *Alien Gene Transfer in Crop Plants*. Springer New York, New York, NY.
- Rozen, S., Skaletsky, H., 1999. Primer3 on the WWW for general users and for biologist programmers. *Bioinforma. Methods Protoc.* 132, 365–386.
- Sabatino, L., Palazzolo, E., D’Anna, F., 2013. Grafting suitability of Sicilian eggplant ecotypes onto *Solanum torvum*: Fruit composition, production and phenology. *J. Food, Agric. Environ.* 11, 1195–1200.
- Sakata, Y., Nishio, T., Matthews, P.J., 1991. Chloroplast DNA analysis of eggplant (*Solanum melongena*) and related species for their taxonomic affinity. *Euphytica* 55, 21–26.
- Sakata, Y., Lester, R.N., 1997. Chloroplast DNA diversity in brinjal eggplant (*Solanum melongena* L.) and related species. *Euphytica* 97, 295–301.
- Salamini, F., Ozkan, H., Brandolini, A., Schäfer-Pregl, R., Martin, W., 2002. Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* 3, 429–441.
- Salas, P., Prohens, J., Seguí-Simarro, J.M., 2011. Evaluation of androgenic competence through anther culture in common eggplant and related species. *Euphytica* 182, 261–274.
- Salgon, S., Jourda, C., Sauvage, C., Daunay, M.-C., Reynaud, B., Wicker, E., Dintinger, J., 2017. Eggplant resistance to the *Ralstonia solanacearum* species complex involves both broad-spectrum and strain-specific quantitative trait loci. *Front. Plant Sci.* 8, 828.
- Salinas, M., Capel, C., Alba, J.M., Mora, B., Cuartero, J., Fernández-Muñoz, R., Lozano, R., Capel, J., 2013. Genetic mapping of two QTL from the wild tomato *Solanum pimpinellifolium* L. controlling

- resistance against two-spotted spider mite (*Tetranychus urticae* Koch). *Theor. Appl. Genet.* 126, 83–92.
- Sallam, A.H., Endelman, J.B., Jannink, J., Smith, K.P., 2015. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *Plant Genome* 8.
- Särkinen, T., Bohs, L., Olmstead, R.G., Knapp, S., 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): A dated 1000-tip tree. *BMC Evol. Biol.* 13, 214.
- Sarma, D., Knott, D.R., 1966. The transfer of leaf rust resistance from *Agropyron* to *Triticum* by irradiation. *Can. J. Genet. Cytol.* 8, 137–143.
- Savage, J.A., Haines, D.F., Holbrook, N.M., 2015. The making of giant pumpkins: How selective breeding changed the phloem of *Cucurbita maxima* from source to sink. *Plant, Cell Environ.* 38, 1543–1554.
- Savarino, F., 2017. Sviluppo di popolazioni biparentali e multiparentali in melanzana e prove di resistenza alla siccità. Tuscia University.
- Sayed, M. a, Schumann, H., Pillen, K., Naz, A., Leon, J., 2012. AB-QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (*Hordeum vulgare* L.). *BMC Genet.* 13, 61.
- Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., Pleban, T., Perez-Melis, A., Bruedigam, C., Kopka, J., et al. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24, 447–454.
- Scheben, A., Batley, J., Edwards, D., 2017. Genotyping-by-sequencing approaches to characterize crop genomes: Choosing the right tool for the right application. *Plant Biotechnol. J.* 15, 149-161.
- Schippers, R., 2000. African indigenous vegetables: An overview of the cultivated species. University of Greenwich, Natural Resources Institute.
- Schlautman, B., Fajardo, D., Bougie, T., Wiesman, E., Polashock, J., Vorsa, N., Steffan, S., Zalapa, J., 2015. Development and Validation of 697 Novel Polymorphic Genomic and EST-SSR Markers in the American Cranberry (*Vaccinium macrocarpon* Ait.). *Molecules* 20,

2001–2013.

- Schmalenbach, I., March, T.J., Bringezu, T., Waugh, R., Pillen, K., 2011. High-resolution genotyping of wild barley introgression lines and fine-mapping of the threshability locus *thresh-1* using the Illumina GoldenGate assay. *G3 Genes, Genomes, Genet.* 1, 187–196.
- Schwarz, D., Roupheal, Y., Colla, G., Venema, J., 2010. Grafting as a tool to improve tolerance of vegetables to abiotic stresses: Thermal stress, water stress and organic pollutants. *Sci. Hortic.* 127, 162–171.
- Sevon, P., Toivonen, H.T.T., Onkamo, P., 2005. Gene mapping by pattern discovery. In: *Data Mining in Bioinformatics*. Springer-Verlag, London, pp. 105–126.
- Sharma, D.R., Kaur, R., Kumar, K., 1996. Embryo rescue in plants—a review. *Euphytica* 89, 325–337.
- Shen, L., Courtois, B., McNally, K.L., Robin, S., Li, Z., 2001. Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. *Theor. Appl. Genet.* 103, 75–83.
- Shetty, S.M., Chandrashekar, A., Venkatesh, Y.P., 2011. Eggplant polyphenol oxidase multigene family: Cloning, phylogeny, expression analyses and immunolocalization in response to wounding. *Phytochemistry* 72, 2275–2287.
- Shi, R., Shuford, C.M., Wang, J.P., Sun, Y.H., Yang, Z., Chen, H.C., Tunlaya-Anukit, S., Li, Q., Liu, J., Muddiman, D.C., 2013. Regulation of phenylalanine ammonia-lyase (PAL) gene family in wood forming tissue of *Populus trichocarpa*. *Planta* 238, 487–497.
- Shirasawa, K., Asamizu, E., Fukuoka, H., Ohyama, A., Sato, S., Nakamura, Y., Tabata, S., Sasamoto, S., Wada, T., Kishida, Y., et al. 2010. An interspecific linkage map of SSR and intronic polymorphism markers in tomato. *Theor. Appl. Genet.* 121, 731–739.
- Shirasawa, K., Ishii, K., Kim, C., Ban, T., Suzuki, M., Ito, T., Muranaka, T., Kobayashi, M., Nagata, N., Isobe, S., 2013. Development of Capsicum EST–SSR markers for species identification and in silico mapping onto the tomato genome sequence. *Mol. Breed.* 31, 101–110.
- Shivanna, K.R., Bahadur, B., 2015. Efficacy of biotechnological approaches to raise wide sexual hybrids. In: *Plant Biology and*



- Biotechnology: Volume II: Plant Genomics and Biotechnology. Springer India, New Delhi, pp. 347–362.
- Sihachakr, D., Haicour, R., Chaput, M.-H., Barrientos, E., Ducreux, G., Rossignol, L., 1989. Somatic hybrid plants produced by electrofusion between *Solanum melongena* L. and *Solanum torvum* Sw. Theor. Appl. Genet. 77, 1–6.
- Sihachakr, D., Haicour, R., Serraf, I., Barrientos, E., Herbreteau, C., Ducreux, G., Rossignol, L., Souvannavong, V., 1988. Electrofusion for the production of somatic hybrid plants of *Solanum melongena* L. and *Solanum khasianum* C.B. Clark. Plant Sci. 57, 215–223.
- Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S., Francis, D.M., 2012a. Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS One 7, e40563.
- Sim, S.C., van Deynze, A., Stoffel, K., Douches, D.S., Zarka, D., Ganal, M.W., Chetelat, R.T., Hutton, S.F., Scott, J.W., Gardner, R.G., et al. 2012b. High-Density SNP Genotyping of Tomato (*Solanum lycopersicum* L.) Reveals Patterns of Genetic Variation Due to Breeding. PLoS One 7, e45520.
- Singh, A.P., Luthria, D., Wilson, T., Vorsa, N., Singh, V., Banuelos, G.S., Pasakdee, S., 2009. Polyphenols content and antioxidant capacity of eggplant pulp. Food Chem. 114, 955–961.
- Smith, P.G., 1944. Embryo culture of a tomato species hybrid. Proc. Am. Soc. Hortic. Sci. 44, 413–416.
- Sotomayor, I., 2016. Desarrollo y fenotipado de dos generaciones para la construcción de líneas de introgresión de *S. incanum* en el fondo genético de la berenjena (*S. melongena*). Universitat Politècnica de València.
- St.Laurent, G., Wahlestedt, C., Kapranov, P., 2015. The Landscape of long noncoding RNA classification. Trends Genet. 31, 249–251.
- Stàgel, A., Portis, E., Toppino, L., Rotino, G.L., Lanteri, S., 2008. Gene-based microsatellite development for mapping and phylogeny studies in eggplant. BMC Genomics 9, 357.
- Stern, S., Agra, M., Bohs, L., 2011. Molecular delimitation of clades

- within New World species of the spiny solanums (*Solanum* subg. *Leptostemonum*). *Taxon* 5, 1429–1441.
- Stommel, J.R., Whitaker, B.D., 2003. Phenolic Acid Content and composition of eggplant fruit in a germplasm core subset. *J. Amer. Soc. Hort. Sci.* 128, 704–710.
- Street, K., Bari, A., Mackay, M., Amri, A., 2016. How the focused identification of germplasm strategy (FIGS) is used to mine plant genetic resources for adaptive traits. In: Maxted, N., Dulloo, M., Ford-Lloyd, B. (Eds.), *Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement*. CABI, Wallingford, pp. 54–65.
- Sunseri, F., Sciancalepore, A., Martelli, G., Acciarri, N., L. Rotino, G., Valentino, D., Tamietti, G., 2003. Development of RAPD-AFLP map of eggplant and improvement of tolerance to *Verticillium* wilt. In: *Acta Horticulturae*. pp. 107–115.
- Sunseri, F., Polignano, G.B., Alba, V., Lotti, C., Bisignano, V., Mennella, G., D'alessandro, A., Bacchi, M., Riccardi, P., Fiore, M.C., Ricciardi, L., 2010. Genetic diversity and characterization of African eggplant germplasm collection. *African J. Plant Sci.* 4, 231–241.
- Syfert, M.M., Castañeda-Álvarez, N.P., Khoury, C.K., Särkinen, T., Sosa, C.C., Achicanoy, H.A., Bernau, V., Prohens, J., Daunay, M.C., Knapp, S., 2016. Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. *Am. J. Bot.* 103, 635–651.
- Takeda, H., Sato, A., Nishihara, E., Arao, T., 2007. Reduction of cadmium concentration in eggplant (*Solanum melongena*) fruits by grafting with *Solanum torvum* rootstock. *Japanese J. Soil Sci. Plant Nutr.* 2007.
- Tang, N., Ma, S., Zong, W., Yang, N., Lv, Y., Yan, C., Guo, Z., Li, J., Li, X., Xiang, Y., et al. 2016. MODD mediates deactivation and degradation of OsbZIP46 to negatively regulate ABA signaling and drought resistance in rice. *Plant Cell* 28, 2161–2177.
- Tanksley, S.D., Young, N.D., Paterson, A.H., Bonierbale, M.W., 1989. RFLP mapping in plant breeding: New tools for an old science. *Bio/Technology* 7, 257–264.

- Tanksley, S.D., Ganai, M.W., Prince, J.P., De Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., et al. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132, 1141–1160.
- Tanksley, S.D., Nelson, J.C., 1996. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92, 191–203.
- Tanksley, S.D., McCouch, S.R., 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277, 1063–1066.
- Thapa, S.P., Miyao, E.M., Michael Davis, R., Coaker, G., 2015. Identification of QTLs controlling resistance to *Pseudomonas syringae* pv. tomato race 1 strains from the wild tomato, *Solanum habrochaites* LA1777. *Theor. Appl. Genet.* 128, 681–692.
- The Angiosperm Phylogeny Group, 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141, 399–436.
- The eggplant genome database (<http://eggplant.kazusa.or.jp>).
- The GOterm database (<http://geneontology.org>).
- The IGV software (<http://www.broadinstitute.org/igv/>).
- The ITAG2.4 database (<ftp://ftp.sgn.cornell.edu>).
- The lncRNA (Long Noncoding RNA) database (<http://www.lncrnadb.org>).
- The NCBI (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov>).
- The Primer3 software (<http://bioinfo.ut.ee/primer3-0.4.0>).
- The Swiss-Prot database (<http://web.expasy.org>).
- The TAIR database: The Arabidopsis Information Resource (<https://www.arabidopsis.org/>).
- The UniRef90 database (<http://www.ebi.ac.uk/uniprot>).
- Thomson, M.J., Alfred, J., Dangl, J., Kamoun, S., McCouch, S., Baird, N., Etter, P., Atwood, T., Currey, M., Shiver, A., et al. 2014. High-

- Throughput SNP Genotyping to Accelerate Crop Improvement. *Plant Breed. Biotechnol.* 2, 195–212.
- Tian, F., De, J.L., Fu, Q., Zuo, F.Z., Yong, C.F., Xiang, K.W., Chuan, Q.S., 2006. Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. *Theor. Appl. Genet.* 112, 570–580.
- Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* 108, 20260–20264.
- Tomato Genome Consortium, 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641.
- Toppino, L., Barchi, L., Lo Scalzo, R., Palazzolo, E., Francese, G., Fibiani, M., D'Alessandro, A., Papa, V., Laudicina, V.A., Sabatino, L., et al. 2016. Mapping Quantitative Trait Loci Affecting Biochemical and Morphological Fruit Properties in Eggplant (*Solanum melongena* L.). *Front. Plant Sci.* 7, 1–16.
- Toppino, L., Valè, G., Rotino, G.L., 2008. Inheritance of *Fusarium* wilt resistance introgressed from *Solanum aethiopicum* Gilo and Aculeatum groups into cultivated eggplant (*S. melongena*) and development of associated PCR-based markers. *Mol. Breed.* 22, 237–250.
- Tran, L., Taylor, J., Constabel, C., 2012. The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion. *BMC Genomics* 13, 395.
- Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Trethowan, R.M., Mujeeb-Kazi, A., 2008. Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. *Crop Sci.* 48, 1255–1265.
- Urbany, C., Stich, B., Schmidt, L., Simon, L., Berding, H., Junghans, H., Niehoff, K.H., Braun, A., Tacke, E., Hofferbert, H.R., et al. 2011. Association genetics in *Solanum tuberosum* provides new insights into potato tuber bruising and enzymatic tissue discoloration. *BMC Genomics* 12, 7.

- Valkoun, J.J., 2001. Wheat pre-breeding using wild progenitors. In: *Euphytica*. Kluwer Academic Publishers, pp. 17–23.
- van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., Visser, B., 2010. Genetic erosion in crops: Concept, research results and challenges. *Plant Genet. Resour.* 8, 1–15.
- van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426.
- Van Ooijen, J.W., 2006. JoinMap 4. *Softw. Calc. Genet. Link. maps Exp. Popul.* 33.
- Varshney, R.K., Graner, A., Sorrells, M.E., 2005. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol.* 23, 48–55.
- Varshney, R.K., Terauchi, R., McCouch, S.R., 2014. Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12, e1001883.
- Verlaan, M.G., Szinay, D., Hutton, S.F., De Jong, H., Kormelink, R., Visser, R.G.F., Scott, J.W., Bai, Y., 2011. Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene Ty-1. *Plant J.* 68, 1093–1103.
- Vilanova, S., Manzur, J.P., Prohens, J., 2012. Development and characterization of genomic simple sequence repeat markers in eggplant and their application to the study of diversity and relationships in a collection of different cultivar types and origins. *Mol. Breed.* 30, 647–660.
- Villegas, D., Casadesús, J., Atienza, S.G., Martos, V., Maalouf, F., Karam, F., Aranjuelo, I., Nogués, S., 2010. Tritordeum, wheat and triticale yield components under multi-local mediterranean drought conditions. *F. Crop. Res.* 116, 68–74.
- Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N.P., Guarino, L., Eastwood, R., Lén, B., Maxted, N., 2013. A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265–275.
- Viquez-Zamora, M., Vosman, B., Geest, H. van de, Bovy, A., Visser, R.G.,

- Finkers, R., Heusden, A.W. van, Miller, J., Tanksley, S., Agarwal, M., et al. 2013. Tomato breeding in the genomics era: Insights from a SNP array. *BMC Genomics* 2013 141 14, 437–448.
- Vogt, T., 2010. Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20.
- Vorontsova, M.S., Stern, S., Bohs, L., Knapp, S., 2013. African spiny *Solanum* (subgenus leptostemonum, solanaceae): A thorny phylogenetic tangle. *Bot. J. Linn. Soc.* 173, 176–193.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van De Lee, T., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414.
- Wall, J.R., 1970. Experimental introgression in the genus *Phaseolus*. Effect of mating systems on interspecific gene flow. *Evolution* 24, 356–366.
- Warburton, M.L., Crossa, J., Franco, J., Kazi, M., Trethowan, R., Rajaram, S., Pfeiffer, W., Zhang, P., Dreisigacker, S., Ginkel, M. van, 2006. Bringing wild relatives back into the family: Recovering genetic diversity in CIMMYT improved wheat germplasm. *Euphytica* 149, 289–301.
- Warschefsky, E., Varma Penmetsa, R., Cook, D.R., Von Wettberg, E.J.B., 2014. Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Bot.* 101, 1791–1800.
- Waterhouse, P.M., Helliwell, C.A., 2003. Exploring plant genomes by RNA-induced gene silencing. *Nat. Rev. Genet.* 4, 29–38.
- Webb, E.C., 1992. Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes. Academic Press.
- Weese, T.L., Bohs, L., 2010. Eggplant origins: Out of Africa, into the Orient. *Taxon* 59, 49–56.
- Wendler, N., Mascher, M., Himmelbach, A., Johnston, P., Pickering, R., Stein, N., 2015. Bulbosum to go: A toolbox to utilize *Hordeum vulgare/bulbosum* introgressions for breeding and beyond. *Mol. Plant* 8, 1507–1519.

- Whitaker, B.D., Stommel, J.R., 2003. Distribution of hydroxycinnamic acid conjugates in fruit of commercial eggplant (*Solanum melongena* L.) cultivars. *J. Agric. Food Chem.* 51, 3448–3454.
- Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., Pryor, R.J., 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin. Chem.* 49, 853–860.
- Wu, F., Mueller, L.A., Crouzillat, D., Pétiard, V., Tanksley, S.D., 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in the euasterid plant clade. *Genetics* 174, 1407–1420.
- Wu, F., Eannetta, N.T., Xu, Y., Tanksley, S.D., 2009. A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theor. Appl. Genet.* 118, 927–935.
- Wulff, B.B.H., Moscou, M.J., 2014. Strategies for transferring resistance into wheat: From wide crosses to GM cassettes. *Front. Plant Sci.* 5, 692.
- Xiao, M., Zhang, Y., Chen, X., Lee, E.J., Barber, C.J.S., Chakrabarty, R., Desgagné-Penix, I., Haslam, T.M., Kim, Y.B., Liu, E., et al. 2013. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* 166, 122–134.
- Potato Genome Sequencing Consortium. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195.
- Yamaguchi, H., Fukuoka, H., Arao, T., Ohyama, A., Nunome, T., Miyatake, K., Negoro, S., 2010. Gene expression analysis in cadmium-stressed roots of a low cadmium-accumulating solanaceous plant, *Solanum torvum*. *J. Exp. Bot.* 61, 423–437.
- Yamakawa, K., Mochizuki, H., 1979. Nature and inheritance of *Fusarium* wilt resistance in eggplant cultivars and related wild *Solanum* species. *Bull. Veg. Orn. Crop. Res. Stn.* 6, 19–27.
- Yan, J., Yang, X., Shah, T., Sánchez-Villeda, H., Li, J., Warburton, M., Zhou, Y., Crouch, J.H., Xu, Y., 2010. High-throughput SNP genotyping with the Goldengate assay in Maize. *Mol. Breed.* 25, 441–451.

- Yang, L., Jue, D., Li, W., Zhang, R., Chen, M., Yang, Q., 2013. Identification of MiRNA from Eggplant (*Solanum melongena* L.) by Small RNA Deep Sequencing and Their Response to *Verticillium dahliae* Infection. PLoS One 8, e72840.
- Yang, X., Cheng, Y.F., Deng, C., Ma, Y., Wang, Z.W., Chen, X.H., Xue, L.B., 2014. Comparative transcriptome analysis of eggplant (*Solanum melongena* L.) and turkey berry (*Solanum torvum* Sw.): Phylogenomics and disease resistance analysis. BMC Genomics 15, 412.
- YiKui, W., YanYan, F., Hong, W., WenJia, L., Yan, L., 2014. Pollen development and anther callus induction in hybrid F1 by *Solanum torvum* SW.(♂)× *Solanum melongena* L.(♀). J. South. Agric. 45, 1967–1971.
- Yuan, Y., Bayer, P.E., Batley, J., Edwards, D., 2017. Improvements in Genomic Technologies: Application to Crop Genomics. Trends Biotechnol. 2017.
- Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. 99, 193–208.
- Zamir, D., Ekstein-Michelson, I., Zakay, Y., Navot, N., Zeidan, M., Sarfatti, M., Eshed, Y., Harel, E., Pleban, T., Van-Oss, H., et al. 1994. Mapping and introgression of a tomato yellow leaf curl virus tolerance gene, TY-1. Theor. Appl. Genet. 88, 141–146.
- Zamir, D., 2001. Improving plant breeding with exotic genetic libraries. Nat. Rev. Genet. 2, 983–989.
- Zeng, S., Xiao, G., Guo, J., Fei, Z., Xu, Y., Roe, B.A., Wang, Y., 2010. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. BMC Genomics 11, 94.
- Zenkteler, M., 1990. In Vitro Fertilization and Wide Hybridization in Higher Plants. CRC. Crit. Rev. Plant Sci. 9, 267–279.
- Zhang, J., Liu, W., Han, H., Song, L., Bai, L., Gao, Z., Zhang, Y., Yang, X., Li, X., Gao, A., 2015. De novo transcriptome sequencing of *Agropyron cristatum* to identify available gene resources for the



- enhancement of wheat. *Genomics* 106, 129–136.
- Zhang, S., Shi, Y., Cheng, N., Du, H., Fan, W., Wang, C., 2015. De novo characterization of fall dormant and nondormant alfalfa (*Medicago sativa* L.) leaf transcriptome and identification of candidate genes related to fall dormancy. *PLoS One* 10, e0122170.
- Zhang, Y., Hu, Z., Chu, G., Huang, C., Tian, S., Zhao, Z., Chen, G., 2014. Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). *J. Agric. Food Chem.* 62, 2906–2912.
- Zhou, C., He, X., Li, F., Weng, Q., Yu, X., Wang, Y., Li, M., Shi, J., Gan, S., 2014. Development of 240 novel EST-SSRs in *Eucalyptus* L'Hérit. *Mol. Breed.* 33, 221–225.
- Zhou, X., Bao, S., Liu, J., Zhuang, Y., 2016. De Novo Sequencing and Analysis of the Transcriptome of the Wild Eggplant Species *Solanum Aculeatissimum* in Response to *Verticillium dahliae*. *Plant Mol. Biol. Report.* 34, 1193–1203.