# Desarrollo y aplicación de herramientas genómicas para la mejora de especies cucurbitáceas por calidad y resistencia a enfermedades.

TESIS DOCTORAL

Presentada por:

Cristina Esteras Gómez

Dirigida por:

Dra. María Belén Picó Sirvent

Valencia, mayo 2012

# Agradecimientos

Quiero expresar mi agradecimiento de forma muy especial a la directora de esta Tesis, la Dra. Belén Picó, por su constante apoyo durante todos estos años, por tantas enseñanzas recibidas y por la total confianza que ha depositado siempre en mí. Así mismo, me gustaría agradecer al Dr. Fernando Nuez, director del COMAV, el haberme dado la posibilidad de realizar los trabajos de la presente Tesis doctoral en el marco de varios proyectos desarrollados en dicho Instituto.

Desearía dar las gracias también a todos aquellos que han colaborado de una forma u otra a que saliera adelante este trabajo, en especial a Cristina Roig, que tantas cosas me ha enseñado y con la que tantas horas he pasado en el laboratorio, a José Blanca, Antonio Monforte, Joaquín Cañizares, Pello Ziarsolo y los investigadores de otras instituciones con las que se ha colaborado como el IRTA, IFAPA y la Universidad de Nápoles Federico II.

También agradecer todo su apoyo a todos mis compañeros del COMAV, con los que he compartido tantas horas, y en especial a los que "conviven" conmigo en el despacho, a Alicia, Javi y Santi, por sus inestimables consejos siempre que las cosas no salían, su ayuda y sus risas en tantos otros momentos. Y, por supuesto, a todos aquellos que en algún momento u otro han pertenecido al grupo de Cucurbitáceas del COMAV, en especial a Eva y Mercè.

Finalmente, mi más profundo agradecimiento a mi familia, sin la cual nada de esto habría sido posible, y en especial a Juanra, por toda su paciencia en los malos momentos.

**ABSTRACT**

Melon (*Cucumis melo*) and summer squash (*Cucurbita pepo*) are economically important cucurbits cultivated worldwide. New, better-adapted, cultivars, resistant to pests and diseases, and varieties displaying better quality traits are required to optimize their production and to meet market demands. Crop breeding must be carried out efficiently and competitively, using the existing genetic knowledge and the molecular tools available for each species. The development of genomic tools to speed up Cucurbits breeding programs is the main objective of this Doctoral Thesis.

Molecular markers are essential tools for the construction of genetic maps, for an efficient MAS selection, for QTL (Quantitative Trait *Loci*) analysis and mapping and for the development of new breeding populations. Molecular markers are also necessary tools for biodiversity analysis. High-quality markers, such as microsatellites (Simple Sequence Repeats, SSRs) and SNPs (Single Nucleotide Polymorphisms) have been developed and/or validated for both species in this Thesis. The generation of sequence information, required for the development of SSRs and SNPs has dramatically changed in the last 6 years due to the advent of new technological advances. In fact, a fairly complete melon transcriptome has been produced using high-throughput 454 sequencing (NGS, Next Generation Sequencing) and combining the newly produced ESTs with the previously existing Sanger sequences. The melon transcriptome characterization presented is the most complete to date (53.252 unigenes with over 63% of them annotated). The generation of large SSRs and SNPs collections for these species, by mapping the ESTs (Expressed Sequence Tags) against the new version of the transcriptome has been a great achievement starting a new era in melon breeding. The obtained collection of molecular markers (more than 3.000 SSRs and 38.000 SNPs) has allowed the construction of melon maps with increased marker density.

Similar resources have been developed for the first time in *C. pepo*. SSRs identified in the *C. pepo* transcriptome sequencing project has been applied to localize landraces with favourable commercial traits and whose alleles might be exploited in breeding programs. Likewise, the usefulness of both EST-SSRs and genomic SSRs for depicting genetic relationships has been proved. On the other hand, the employment of the first SNP set identified for *Cucurbita* through a massive genotyping platform, GoldenGate, has been a success (90%). The genetic map obtained, with more than 300 SNPs, has allowed to map a large set of QTLs related to early flowering, femaleness tendency, shape, size and fruit color, and other traits of interest in breeding.

Therefore, these works report the validation and application of previously developed markers and the generation of new genomic tools to speed up the breeding process of these crops. The paper-based format of this Thesis confirms the interest of the obtained results.

**RESUMEN**

El melón (*Cucumis melo*) y el calabacín (*Cucurbita pepo*) son especies cucurbitáceas de gran importancia económica a nivel nacional y mundial. Para optimizar su producción se requiere de la obtención de nuevas variedades mejor adaptadas a los sistemas de cultivo, más resistentes frente a nuevas enfermedades o plagas y con mejores características organolépticas, que respondan a las cada vez mayores exigencias del mercado. La mejora debe realizarse de una forma eficiente y competitiva, apoyándose en los crecientes conocimientos genéticos en estas dos especies y en los últimos avances biotecnológicos. El desarrollo de herramientas genómicas con el fin de impulsar la mejora de estos cultivos es el principal objetivo de la presente Tesis doctoral.

El desarrollo de marcadores moleculares es esencial para la construcción de mapas genéticos, para la realización de una selección más eficiente, para el análisis y cartografía de QTLs (Quantitative Trait *Loci*) y para el desarrollo de líneas de premejora, además de ser una herramienta fundamental para el análisis de la biodiversidad. En esta Tesis se han desarrollo y/o validado marcadores de alta calidad, de tipo microsatélite (Simple Sequence Repeats, SSRs) y SNPs (Single Nucleotide Polymorphisms), para estas dos especies. La generación de información de secuencia, necesaria para el desarrollo de este tipo de marcadores, ha cambiado en el transcurso de los trabajos presentados en la Tesis, habiéndose abordado finalmente la secuenciación del transcriptoma de melón mediante técnicas de secuenciación de alto rendimiento (NGS, Next Generation Sequencing). La obtención de grandes colecciones de SSRs y SNPs en ambas especies, resultado del ensamblaje de ESTs (Expressed Sequence Tags) procedentes de secuencias Sanger previamente disponibles y de las nuevas secuencias obtenidas por secuenciación masiva, ha supuesto un gran avance para estas especies no

modelo, permitiendo la construcción de mapas más densos en melón y del primer mapa basado en SNPs en calabacín.

En melón, la caracterización del transcriptoma presentada es, con 53.252 unigenes, de los cuales el 63% están anotados funcionalmente con términos GO, la más completa que existe. Este transcriptoma ha proporcionado una valiosa información para la anotación del genoma, actualmente en fase de publicación. Además de la identificación de más de 3.000 SSRs y 38.000 SNPs, ha supuesto el desarrollo de herramientas bioinformáticas para la selección de marcadores polimórficos entre diversos morfotipos de interés, lo que abre camino a una nueva era en la mejora de este cultivo.

La aplicación de los SSRs identificados gracias a la secuenciación del transcriptoma de *C. pepo* ha servido para localizar cultivares tradicionales cuya riqueza alélica podría ser explotada con fines comerciales. Asimismo, ha confirmado la eficacia de los EST-SSRs frente a los de origen genómico para estudios de variabilidad. Por otra parte, la utilización de los primeros SNPs identificados para *Cucurbita* ha supuesto la primera aplicación, con un 90% de éxito, de una plataforma de genotipado masivo, GoldenGate, en este género. El mapa genético basado en más de 300 SNPs ha permitido cartografiar un amplio conjunto de QTLs relacionados con la precocidad de la floración, la tendencia femenina de la planta, la forma, tamaño y coloración del fruto, además de otros caracteres que resultan de gran interés para la mejora.

Por tanto, estos trabajos suponen la validación y la aplicación de marcadores previamente desarrollados y la generación de nuevas herramientas genómicas que supondrán la aceleración del proceso de mejora de estos cultivos. El formato de la Tesis basado en artículos ya publicados por el grupo confirma el interés de los resultados obtenidos.

**RESUM**

El meló (*Cucumis melo*) i la carabasseta (*Cucurbita pepo*) són espècies cucurbitàcies de gran importància econòmica a nivell nacional i mundial. Per a optimitzar la seua producció es requereix de l´obtenció de noves varietats millor adaptades als sistemes de conreu, més resistents a noves malalties o plagues i amb millors característiques organolèptiques, que responguen a les cada vegada majors exigències del mercat. La millora ha de realitzar-se d´una forma eficient i competitiva, basant-se en els creixents coneiximents genètics en aquestes dues espècies i en els últims avançaments biotecnològics. El desenvolupament de ferramentes genòmiques per a impulsar la millora d´aquests conreus és el principal objectiu de la present Tesi doctoral.

El desenvolupament de marcadors moleculars és essencial per a la construcció de mapes genètics, per a la realització d´una selecció més eficient, per a l´anàlisi i cartografia de QTLs (Quantitative Trait *Loci*) i per al desenvolupament de línies de premillora, a més de ser una ferramenta fonamental per a l´anàlisi de la biodiversitat. En aquesta Tesi s´han desenvolupat i/o validat marcadors d´alta qualitat, tipus microsatèl.lit (Simple Sequence Repeats, SSRs) i SNPs (Single Nucleotide Polymorphisms), per ambdues espècies. La generació d´informació de seqüència, necessària per al desenvolupament d´aquest tipus de marcadors, ha canviat en el transcurs dels treballs presentats en la Tesi, havent-se abordat finalment la seqüenciació del transcriptoma de meló mitjançant tècniques de seqüenciació d´alt rendiment (NGS, Next Generation Sequencing). L´obtenció de grans col.leccions de SSRs i SNPs en les dues espècies, resultat de l´acoblament d´ESTs (Expressed Sequence Tags) procedents de seqüències Sanger prèviament disponibles i de les noves seqüències obtingudes per seqüenciació massiva, ha suposat un gran avançament per a aquestes espècies no model, permitint la

construcció de mapes més densos en meló i del primer mapa basat en SNPs en carabasseta.

En meló, la caracterització del transcriptoma presentada és, amb 53.252 unigens, dels quals el 63% estan anotats funcionalment amb termes GO, la més completa que existeix. Això ha proporcionat una valuosa informació per a l´anotació del genoma, actualment en fase de publicació. A més, la identificació de més de 3.000 SSRs i 38.000 SNPs, ha suposat el desenvolupament de ferramentes bioinformàtiques per a la sel.lecció de marcadors polimòrfics entre diversos morfotipus d´interés, amb la qual cosa s´obri camí a una nova era en la millora d´aquest conreu.

L´aplicació dels SSRs identificats gràcies a la seqüenciació del transcriptoma de *C. p*epo ha servit per a localitzar cultivars tradicionals, la riquesa al.lèlica dels quals podria ser explotada amb objectius comercials. A més, ha confirmat l´eficàcia dels EST-SSRs respecte als d´origen genòmic per a estudis de variabilitat. Per una altra banda, la utilització dels primers SNPs identificats per a *Cucurbita* ha suposat la primera aplicació, amb un 90% d´èxit, d´una plataforma de genotipat massiu, GoldenGate, en aquest gènere. El mapa genètic basat en més de 300 SNPs ha permés cartografiar un ampli conjunt de QTLs relacionats amb la precocitat de la floració, la tendència femenina de la planta, la forma, tamany i coloració del fruit, a més d´altres caràcters que resulten de gran interés per a la millora.

Aleshores, aquests treballs suposen la validació i l´aplicació de marcadors prèviament desenvolupats i la generació de noves ferramentes genòmiques que suposaran l´acceleració del procés de millora d´aquests conreus. El format de la Tesi basat en articles ja publicats pel grup confirma l´interés dels resultats obtinguts.

# ÍNDICE

**I.INTRODUCCIÓN GENERAL:**

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools**. En: Cucurbits: Genetics, Genomics and Breeding of Cucurbits. Editado por Wang Y, Behera TK y Kole C. New Hampshire: Science Publishers Inc, Enfield; 2012:140-198.

Deleu W*, Esteras C*, Roig C, González-To M, Fernández-Silva I, Gonzalez-Ibeas D, Blanca J, Aranda MA, Arús P, Nuez F, Monforte AJ, Picó B, Garcia-Mas J: **A set of EST-SNPs for map saturation and cultivar identification in melon.** BMC Plant Biology 2009, 9:90, doi:10.1186/1471-2229-9-90.

(*La contribución de los dos autores fue la misma)

Blanca J, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó B: **Melon transcriptome characterization: Simple Sequence Repeats and Single Nucleotide Polymorphisms discovery for high throughput genotyping across the species.** The Plant Genome 2011, 4(2):118-131.

Formisano G, Roig C, Esteras C, Ercolano MR, Nuez F, Monforte AJ, Picó B: **Genetic diversity of**

Esteras C, Gómez P, Monforte AJ, Blanca J, Vicente-Dólera N, Roig C, Nuez F, Picó B: **High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait *loci* mapping.** BMC Genomics 2012, 13:80, doi:10.1186/1471-2164-13-80.

1. González M, Xu M, Esteras C, Roig C, Monforte AJ, Troadec C, Pujol M, Nuez F, Bendahmane A, Garcia-Mas J, Picó B: **Towards a TILLING platform for functional genomics in Piel de Sapo melons.** BMC Research Notes 2011, 4:289

2. Esteras C, Lunn J, Sulpice R, Blanca J, Garcia-Mas J, Pitrat M, Nuez F, Picó B: **Phenotyping a highly diverse core melon collection to be screened using Ecotilling.** 8 Plant Genomics European Meetings (Plant Gem), Lisbon (Portugal), 7–10 octubre 2009: p214.

3. Esteras C, Pascual L, Saladie M, Dogimont C, Garcia-Mas J, Nuez F, Picó B: **Use of Ecotilling to identify natural allelic variants of melon candidate genes involved in fruit ripening.** 8 Plant Genomics European Meetings (Plant Gem), Lisboa (Portugal), 7-10 octubre 2009: p213.

4. Picó B, Sifres A, Esteras C, Nuez F**: *Cucumis* SSRs markers applied to the study of the genetic diversity in the *Cucurbita* genus.** Cucurbit Genet Coop 2005-2006, 28-29:70-72.

5. Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Picó B, Nuez F, Arus P, Garcia-Mas J, Monforte AJ: **Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.).** Theor Appl Genet 2008, 118:139-150.

6. Esteras C, Sifres A, Nuez F, Picó B: **Variabilidad de *Cucurbita maxima* en su zona de origen: un recurso de interés para la mejora de esta hortaliza.** En Actas de Horticultura 54, VI Congreso Ibérico, XII Nacional de Ciencias Hortícolas. Resúmenes. Editado por Pardo A, Suso

ML y Vázquez N. Logroño (España), 25-29 mayo 2009:1284-1290.

7. Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Esteras C, Picó B: **High throughput SNPs identification in** *C. pepo* **and** *C. moschata* **by Solexa transcriptome resequencing.** 9 Plant Genomics European Meetings (Plant Gem). Istambul (Turkey), 4 -7 mayo 2011.

**PRÓLOGO**

La familia *Cucurbitaceae* es una de las más amplias y diversas. Los últimos estudios indican que contiene un total de 130 géneros y unas 800 especies (Jeffrey, 2005). Engloba diversos cultivos de importancia económica mundial tales como la sandía-*Citrullus lanatus* (Thunb.) Matsum. & Nakai-, el pepino- *Cucumis sativus* L.-, el melón-*Cucumis melo* L.- y el calabacín- *Cucurbita pepo* L. (revisado por Esteras et al., 2012b). En 2010 se produjeron en el mundo más de 89, 57 y 25 millones de toneladas de los 3 cultivos de cucurbitáceas más importantes, sandía, pepino y pepinillos; y melones respectivamente (FAOSTAT, 2012), ocupando el tercer, sexto y décimo puesto de entre los cultivos hortícolas en cuanto a producción mundial. Es difícil obtener datos de superficie y de producción de calabacín por países, ya que la mayoría de ellos incluyen en las estadísticas oficiales calabacines conjuntamente con otras especies de calabazas (fundamentalmente pertenecientes a *C. pepo* y otras dos especies del género *Cucurbita*: *C. moschata* Duchesne y *C. maxima* Duchesne). Según la FAO, se produjeron en 2010 alrededor de 22 millones de toneladas de distintos tipos de calabazas y calabacines, la mayor parte de los cuales correspondieron a calabacín. Tradicionalmente, las calabazas y los calabacines han sido considerados cucurbitáceas menores, pero su importancia ha seguido una tendencia creciente desde la década de los 80, siendo en la actualidad el calabacín la novena hortaliza más cultivada en el mundo y la duodécima en cuanto a producción.

En España, la cucurbitácea más importante es el melón, seguida de la sandía, el pepino y las calabazas y calabacines, con 926, 782, 682 y 364 mil toneladas producidas respectivamente en el año 2010.

El melón es particularmente importante en países de la cuenca mediterránea y del este de Asia. En nuestro país, las regiones productoras más importantes son las provincias de

Ciudad Real, Murcia y Almería (MAPA, 2012). España ocupó el sexto puesto en cuanto a producción mundial en 2010, siendo en 2009 el primer exportador mundial en términos económicos y el segundo en cuanto a toneladas exportadas (FAOSTAT, 2012). El empleo actual de semillas híbridas en este cultivo ha contribuido a un significativo aumento de la producción y, por consiguiente, del impacto económico.

Respecto a la producción de calabacines en España, la mayor parte del cultivo se concentra en la provincia de Almería, mientras que los otros tipos de calabazas, de menor importancia, se producen mayoritariamente en Canarias y la Comunidad Valenciana (MAPA, 2012). En la última década, la superficie cultivada de calabacín ha permanecido más o menos estable. Sin embargo, la producción ha experimentado un incremento importante, sobre todo debido a la mayor importancia relativa del cultivo bajo plástico. En cuanto a la comercialización, la producción de calabacín en España se destina en su mayor parte a la exportación. De hecho, España fue el primer exportador mundial de esta hortaliza en 2009 (FAOSTAT, 2012).

Tanto el melón como el calabacín se encuentran entre las 20 frutas y hortalizas más exportadas en nuestro país (posición 11 y 17, respectivamente), siendo el mercado europeo el de mayor importancia. Por tanto, debido a la importancia de estas especies cucurbitáceas, tanto a nivel nacional como internacional, se hace necesario el desarrollo de nuevas variedades mejor adaptadas a las condiciones existentes o a nuevos ambientes y/o con características deseables para el mercado. Se necesita una mejora genética más eficiente y competitiva, para lo que se requiere el empleo de nuevas tecnologías que ya se han aplicado, o se están aplicando con éxito, en otros cultivos.

Las herramientas que se están convirtiendo en imprescindibles para modernizar la mejora de éstas y otras hortalizas son mapas físicos y genéticos de alta densidad (Shirasawa et al. (2010) y de Boer et al. (2011) en solanáceas; Wang et al. (2011a) en

cereales), nuevas poblaciones más eficientes para el mapeo de características cuantitativas (RILs, Recombinant Inbred Lines; colecciones de ILs, Introgression Lines; o poblaciones MAGIC, Multiparent Advanced Generation Inter-Cross populations) (Gallo et al. (2011) en solanáceas; Qin et al. (2012) en leguminosas; Gu et al. (2012) en cereales; Kover et al. (2009) en la especie modelo *Arabidopsis thaliana*) o plataformas de genética reversa (Eliaset al. (2009), Ibiza et al. (2010) y Okabe et al. (2011) en solanáceas; Uauy et al. (2009) en cereales, Cooper et al. (2008) en leguminosas). La disponibilidad de secuencias ha sido un factor clave para el desarrollo de estas herramientas. La tecnología de secuenciación ha sufrido una revolución en los últimos 6 años. Con el desarrollo de métodos de secuenciación masiva de segunda generación (SGS, Second Generation Sequencing) y su combinación con la metodología de Sanger, se ha ampliado enormemente el rango de cultivos en los que se ha producido masivamente secuencia genómica y transcriptómica. Así pues, la secuenciación de transcriptomas y genomas se ha llevado a cabo en cultivos tan diversos como solanáceas (Mueller et al., 2009; The Potato Genome Sequencing Consortium, 2011), cereales (Nelson et al., 2011; Xu et al., 2011), leguminosas (Kim et al., 2010) o cítricos (Belknap et al., 2011).

La resecuenciación de genomas y/o transcriptomas de genotipos representativos de la variabilidad de la especie es cada vez más frecuente y permite conocer con detalle la diversidad genética existente. Una de las aplicaciones más inmediatas de la resecuenciación es el desarrollo de amplias colecciones de marcadores de tipo SNP (Single Nucleotide Polymorphisms). Los marcadores de tipo SNP se consideran de calidad porque son altamente reproducibles y codominantes, al igual que los microsatélites o SSRs (Simple Sequence Repeats). La mejora moderna los prefiere a éstos porque su análisis puede ser automatizado, reduciendo enormemente el tiempo y

el coste del genotipado. La optimización del empleo de colecciones masivas de SNPs en plataformas de genotipado masivo de media o gran escala, como las plataformas GoldenGate e Infinium de Illumina o la tecnología iPLEX de Sequenom (Fan et al., 2006; Steermers y Gunderson, 2007; Gabriel et al., 2009) han constituido un gran impulso para la mejora de muchas especies (van Heusden et al. (2011) en solanáceas, Close et al. (2009), Ganal et al. (2010) y Mammadov et al. (2012) en cereales; Hyten et al. (2008) y Deulvot et al. (2010) en leguminosas). La caracterización de la variabilidad a escala genómica, el desarrollo de poblaciones y líneas de premejora, la selección asistida por marcadores y la disección génica de características cuantitativas, QTLs (Quantitative Trait *Loci*), son puntos fundamentales en los programas de mejora que se han visto especialmente impulsados con la disponibilidad de estas herramientas (Grewal et al. (2011), Ma et al. (2012) y Wang et al. (2011b) en cereales; McCord et al. (2011) y Borovsky y Paran (2011) en solanáceas; Sugiyama et al. (2011) y Sahin-Cevik y Moore (2012) en cítricos).

En Cucurbitáceas el desarrollo de las herramientas genómicas y su aplicación a la mejora ha sido más reciente y se ha avanzado de forma distinta en los diversos cultivos. El objetivo general de la presente Tesis doctoral ha sido contribuir al desarrollo y mejora de estas herramientas en cucurbitáceas, fundamentalmente en melón y calabacín, y propiciar su uso en la mejora de estas especies.

La Tesis doctoral que se presenta es una recopilación de trabajos publicados en los que la doctoranda ha intervenido. Los trabajos realizados por la doctoranda durante este proceso de aprendizaje se enmarcan en diversos proyectos de investigación llevados a cabo en el seno del grupo de Mejora Genética de Cucurbitáceas del Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV), perteneciente a la Universitat Politècnica de València (UPV), en estrecha colaboración con el grupo de

Genómica y Bioinformática del mismo instituto, y en colaboración con diversas instituciones como el Centre de Recerca en Agrigenòmica de Barcelona (CSIC-IRTA-UAB-UB), el Instituto de Biología Molecular y Celular de Plantas (IBMCP, centro mixto CSIC-UPV), la Universidad de Nápoles Federico II y el área de Mejora y Biotecnología de cultivos del Instituto de Investigación y Formación Agraria y Pesquera (IFAPA) de Almería.

Las principales actividades realizadas en los dos cultivos se describen a continuación:

**Melón (*C. melo*)**

El grupo en el que se integró la doctoranda al inicio de esta tesis doctoral se encontraba finalizando su participación en un proyecto de desarrollo de herramientas genómicas para su aplicación a la mejora del melón: "MELOGEN: *Desarrollo de herramientas de genómica para la mejora del melón por resistencia a patógenos y calidad del fruto*" (GEN2003-20237-C06, http://www.melogen.upv.es). Como resultado de este proyecto, y de otros anteriores llevados a cabo por los distintos grupos participantes, se había desarrollado un mapa genético de densidad moderada, construido fundamentalmente con marcadores codominantes de tipo SSR (Fernández-Silva et al., 2008), un mapa físico (González et al., 2010), una colección de líneas de introgresión (Eduardo et al., 2005) y una colección de 30.000 ESTs (Expressed Sequence Tags), obtenidos mediante secuenciación por Sanger (González-Ibeas et al., 2007), entre otras herramientas genómicas. La colección de ESTs, que posteriormente se amplió con nuevas genotecas (Clepet et al., 2011), permitió la construcción del primer microarray en melón (Mascarell-Creus et al., 2009) y el avance de los estudios de expresión relacionados con resistencias y calidad del fruto en este cultivo (Dai et al., 2011; González-Ibeas et al.,

2012). Muchas de estas herramientas se integraron y se encuentran accesibles en la base de datos de ICuGI (International Cucurbit Genomics Initiative, http://www.icugi.org/).

Puesto que la colección de ESTs producidos por Sanger se había obtenido a partir de los genotipos parentales del mapa genético de melón, el cultivar T111 de tipo Piel de sapo español, (*C. melo* ssp. *melo* var. *inodorus* H. Jacq) y la variedad exótica coreana Songwhan Charmi (PI161375) (*C. melo* ssp. *melo* var. *conomon* (Thunb) Makino) fue posible identificar, mediante el cribado de la colección, un conjunto de SNPs adecuados para mapado. En el marco de este proyecto, la doctoranda contribuyó al desarrollo del primer mapa de SNPs en melón. Se incrementó significativamente la densidad del mapa existente, mediante el cartografiado de un set de SNPs previamente identificados *in silico*. Además, se comprobó la utilidad de este conjunto de SNPs para la identificación de cultivares (Deleu –Esteras* et al., 2009: "*A set of EST-SNPs for map saturation and cultivar identification in melon*", trabajo incluido en la presente Tesis doctoral. * igual contribución al trabajo). El mapa producido se ha empleado recientemente para obtener un mapa consenso más saturado para la especie (Diaz et al., 2011).

A continuación, el grupo participó en el proyecto trilateral financiado por la acción europea ERA-NET-PG "MELRIP: *Understanding the climacteric vs non-climacteric fruit ripening mechanisms in melon using transcriptomic, metabolomic and reverse genetic approaches*" (GEN2006-27773-C2-2-E), en el que la doctoranda participó activamente en el desarrollo y análisis de herramientas de genética reversa como plataformas de TILLING y de EcoTILLING (González et al., 2011; Esteras et al., 2009a,b trabajos incluidos en los anexos 1, 2 y 3). La población de TILLING fue la primera desarrollada en melón en un fondo genético de tipo Piel de Sapo. Estas poblaciones se encuentran en fase de explotación, habiéndose identificado en las mismas mutantes de interés en genes involucrados en resistencia a enfermedades y en

calidad del fruto. Una segunda población de TILLING, desarrollada por la Unidad de Investigación de Genómica Vegetal del INRA-URGV en un fondo genético de tipo cantalupo, en el marco del mismo proyecto, ha contribuido al estudio de las funciones de genes de interés en melón, y han llevado al clonaje de algunos de los genes involucrados (Boualem et al., 2008; Martin et al., 2009).

Desde 2009, el grupo de Mejora Genética de Cucurbitáceas del COMAV, junto con el grupo de Genómica y Bioinformática del mismo instituto, ha participado en el proyecto financiado por la Fundación Genoma España: MELONOMICS: "*Developing genomics tools in Cucurbitaceae, including the sequencing of the melon genome, and their applications to improve these crops*" (Genoma España, 2009-2011). En el marco de este proyecto se han aplicado los nuevos avances en las tecnologías de secuenciación para la caracterización del transcriptoma de melón (Blanca et al., 2011c: "*Melon transcriptome characterization: Simple Sequence Repeats and Single Nucleotide Polymorphisms discovery for high throughput genotyping across the species*", trabajo incluido en esta Tesis). Las secuencias obtenidas por secuenciación 454 GS FLX Titanium/Roche, a partir de 4 colecciones de cDNA construidas con la línea T111 de Piel de sapo (*C. melo* ssp. *melo* var. *inodorus*), la entrada PI161375 (*C. melo* ssp. *melo* var. *conomon*) y los cultivares Piel de sapo Piñonet (*C. melo* ssp. *melo* var. *inodorus*) y Vedrantais (*C. melo* ssp. *melo* var. *cantalupensis* Naudin), fueron ensambladas con las secuencias ESTs de Sanger ya disponibles (MELOGEN, ICuGI), (Clepet et al., 2011), para obtener una versión más completa del transcriptoma de melón. Este trabajo supuso la primera aplicación de técnicas de secuenciación masiva en melón. Otros estudios empleando técnicas similares se han publicado más recientemente (González-Ibeas et al., 2011; Portnoy et al., 2011).

El transcriptoma obtenido por nuestro grupo, con 53.252 unigenes, representa el más completo de melón hasta la fecha. La identificación de más de 14.000 secuencias con SNVs (Single Nucleotide Variations: SNPs e INDELs) y de más de 3.000 unigenes con motivos SSRs ha supuesto una nueva fuente de marcadores de calidad inestimable (Blanca et al., 2011c). Un total de 3.593 SSRs, 38.587 SNPs y 5.795 INDELs se detectaron en el conjunto de unigenes, obteniéndose un porcentaje de validación en una muestra de los mismos cercano al 90%.

En el marco de este mismo proyecto se ha llevado a cabo la secuenciación del genoma de melón (Garcia-Mas et al., 2012, enviado para su publicación) para cuya anotación se ha empleado el transcriptoma desarrollado. Los polimorfismos detectados anteriormente se han situado en el genoma, lo que incrementa su utilidad. Además, un conjunto de 768 SNPs se han implementado en una plataforma de genotipado masivo de tipo GoldenGate, empleándose para incrementar la densidad del mapa genético y para anclar el genoma al mismo. Toda esta información, en cuyo desarrollo ha participado la doctoranda, se encuentra en la actualidad organizada en la página http://www.melonomics.upv.es, que será de acceso libre en cuanto se publique la secuencia del genoma. Por otra parte, las herramientas desarrolladas en el marco de los proyectos descritos ya se están aplicando con diversos fines en el grupo de Cucurbitáceas del COMAV, concretamente en el marco de dos proyectos en los que participa la doctoranda: el proyecto INIA "*Recolección, multiplicación, conservación y caracterización de variedades emblemáticas de melón de Castilla-La Mancha*" (RF2008-00003-C02-02), en el que los microsatelites se emplean para la conservación de variedades locales de interés, y el proyecto trilateral de la convocatoria europea PLANT KBBE, SAFQIM "*Sugars and fruit quality in melon"*(PIM2010PKB-00691), en el que los SNPs se están empleando para el desarrollo de líneas de introgresión

derivadas del cruce de cultivares de interés comercial, Piel de sapo y Vedrantais, y exóticos con características de interés (*dudaim* y *makuwa*).

El avance en la producción de herramientas en melón se asemeja a la situación de pepino, en el que la secuenciación del genoma ya está finalizada (Huang et al., 2009), o sandía, en la que se encuentra muy avanzada (Guo et al., 2010; ICuGI). La disponibilidad de estos recursos genómicos ha contribuido a la construcción de mapas altamente saturados (Ren et al., 2012). En pepino, recientemente ha sido publicado un mapa consenso altamente saturado con casi 1.400 marcadores, de los que un 84% son SSRs (Zhang et al., 2012). Además, nuevas secuencias producidas mediante RNA-seq han mejorado recientemente la anotación del genoma (Li et al., 2011).

**Calabacín (*C. pepo*)**

La producción de herramientas genómicas ha sido mucho más lenta en el género *Cucurbita*. Hasta el año 2008 en que Gong et al. publicaron la primera colección importante de SSRs (unos 400) a partir de dos genotecas genómicas enriquecidas, de un cultivar del morfotipo Pumpkin (*C. pepo* ssp. *pepo* morfotipo Pumpkin) y de un cultivar perteneciente a la especie relacionada *C. moschata*, muy pocos marcadores de este tipo estaban disponibles en el género. Previamente al desarrollo de esta colección, se trataba de poner a punto en *Cucurbita* SSRs desarrollados en el género *Cucumis* (melón o pepino). Sin embargo, la tasa de transferibilidad tanto en gSSRs (genómicos) como en EST-SSRs resultaba muy baja (revisado en Esteras et al., 2012b). En este sentido, la doctoranda colaboró en diversos proyectos INIA "*Desarrollo de una colección nuclear de germoplasma español de calabacín (C. pepo) y especies relacionadas*" (RF2004-00003-00-00) y "*Control de la erosión genética en bancos de germoplasma: el caso de la colección de germoplasma español de calabazas y calabacines del COMAV*" (RF03-

003) transfiriendo marcadores SSRs de *Cucumis* a *Cucurbita* (Picó et al., 2005-2006; Fernández-Silva et al., 2008, trabajos incluidos en los anexos 4 y 5) o aplicando marcadores dominantes y menos reproducibles como AFLPs (Amplified Fragment Length Polymorphisms) (Esteras et al., 2009c, trabajo incluido en el anexo 6). La necesidad de desarrollar marcadores codominantes específicos para este género requería información de secuencia, que no ha estado disponible para *Cucurbita* hasta la publicación del primer transcriptoma de *C. pepo* por Blanca et al. (2011b), pues hasta entonces tan sólo 1.000 secuencias ESTs de la colección del NCBI pertenecían a este género. Así pues, la continuación del trabajo de la doctoranda en *Cucurbita* spp. partió de la información de secuencia generada por el grupo de Mejora de Cucurbitáceas y de Bioinformática y Genómica del COMAV en el proyecto INIA "*Mejora competitiva del calabacín por características de calidad de fruto y resistencia a enfermedades: Desarrollo de RILs, marcadores codominantes SSR, CAPS y SNP y construcción de un mapa genético*" (RTA2008-00035-C02-02). La secuenciación del transcriptoma *via* 454 GS FLX Titanium a partir de dos genotecas multi-tejido construidas con la entrada MU-CU-16 (*C. pepo* ssp. *pepo* morfotipo Zucchini) y UPV-196 (*C. pepo* ssp. *ovifera* morfotipo Scallop), condujo a la identificación de 1.935 SSRs, 19.980 SNPs y 1.174 INDELs (Blanca et al., 2011b). Los trabajos incluidos en la presente Tesis referentes a *C. pepo* incluyen, por una parte, la validación de parte de esta colección de EST-SSRs y su aplicación en un estudio de variabilidad en comparación con los SSRs genómicos previamente desarrollados por Gong et al. (2008) (Formisano et al., 2011: "*Genetic diversity of Spanish Cucurbita pepo landraces: an unexploited resource for summer squash breeding*") y, por otra parte, la aplicación de 384 EST-SNPs para la construcción del primer mapa genético en el género basado en este tipo de marcadores (Esteras et al., 2012a: "*High-throughput SNP genotyping in Cucurbita pepo for map construction and*

*quantitative trait loci mapping*"). Los mapas previos estaban basados principalmente en marcadores dominantes y poco transferibles a otras poblaciones (Lee et al., 1995 y Brown y Myers, 2002), basados en el cruce interespecífico *C. pepo* x *C. moschata*; Zraidi et al., 2007), o basados en cruces intraespecíficos e inter-subespecíficos dentro de *C. pepo*), y sólo el publicado por Gong et al., (2008) incluía gSSRs, algunos de los cuales se anclan al mapa presentado aquí (Esteras et al., 2012a). La aplicación de estos marcadores de alta calidad no sólo ha impulsado los estudios de diversidad en la especie, incluyendo tipos comerciales y tradicionales, sino que ha permitido la puesta a punto de técnicas de genotipado de alto rendimiento como la plataforma GoldenGate de Illumina y un primer análisis de QTLs (Esteras et al., 2012a).

Posteriormente, el grupo ha llevado a cabo una resecuenciación del transcriptoma en *Cucurbita* para incrementar el número de genes y alelos representados, pero en este caso mediante tecnología Solexa/Illumina, empleando dos colecciones de cDNA, una con entradas pertenecientes a *C. pepo* ssp. *pepo* y otra con tipos de *C. moschata,* especie relacionada que constituye un puente genético para la transferencia de caracteres de interés de tipos silvestres hacia los tipos cultivados de *C. pepo* (Blanca et al., 2011a, incluido en el anexo 7).

Aunque en fases más iniciales, en *Cucurbita* spp. también comienzan a existir estudios de expresión como los publicados recientemente por Manzano et al. (2010), que clonan y caracterizan dos genes involucrados en el desarrollo sexual de las flores de *C. pepo*, Martínez et al. (2011); que estudian la expresión de CpACS1 y CpACO1 en relación a la producción de etileno; Guo et al. (2011), que clonan y estudian la expresión de un gen relacionado con la cobertura de la semilla, u Obrero et al. (2011), en el que se seleccionan genes de referencia para análisis de expresión mediante PCR cuantitativa (qPCR). El reciente avance en el desarrollo de herramientas transcriptómicas para este

género y la disponibilidad del mapa de ligamiento obtenido, gracias a las colecciones de marcadores generadas, impulsará seguro una mejora más eficiente de esta hortaliza. Para facilitar la utilización de esta información se ha generado la base de datos Cucurbigene (http://cucurbigene.net/), desarrollada en colaboración por ambos grupos de Cucurbitáceas y Bioinformática del COMAV.

El trabajo realizado en el marco de esta Tesis doctoral ha sido esencial para el desarrollo y uso de herramientas genómicas en melón y calabaza, permitiendo una mejora más eficiente de ambas especies. Además de los trabajos indicados anteriormente, se incluye una introducción inicial correspondiente con un capítulo del libro "Genetics, Genomics and Breeding of Cucurbits" titulado "Genetic diversity studies in Cucurbits using molecular tools" (Esteras et al., 2012b) en el que se presenta una revisión acerca de la taxonomía y los diversos estudios de variabilidad, tanto morfológica como molecular mediante el empleo de diferentes sistemas de marcadores, en la familia *Cucurbitaceae*.

* Algunos de los trabajos que conforman esta Tesis presentan archivos adicionales (figuras, tablas, archivos de secuencias) que no están impresos y únicamente están en versión digital (CD).

**Referencias**

Belknap WR, Wang Yi; Huo NX, Wu JJ, Rockhold DR, Gu YQ, Stover E: **Characterizing the citrus cultivar Carrizo genome through 454 shotgun sequencing.** Genome 2011, 54(12):1005-1015, doi: 10.1139/G11-070.

Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Esteras C, Picó B: **High throughput SNPs identification in *C. pepo* and *C. moschata* by Solexa transcriptome resequencing.** 9 Plant Genomics European Meetings (Plant Gem). Istambul (Turkey), 4 -7 mayo 2011a.

Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Picó B: **Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (*Cucurbitaceae*).** BMC Genomics 2011b, 12:104.

Blanca J, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó B: **Melon transcriptome characterization. SSRs and SNPs discovery for high throughput genotyping across the species.** The Plant Genome 2011c, 4(2):118-131.

Borovsky Y, Paran I: **Characterization of fs10.1, a major QTL controlling fruit elongation in *Capsicum*.** Theor Appl Genet 2011, 123(4):657-665, doi: 10.1007/s00122-011-1615-7.

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari MA, Collin F, Flowers JM, Pitrat M, Purugganan MD, Dogimont C, Bendahmane A: **A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons.** Science 2008, 321(5890):836-838, doi:10.1126/science.1159023.

Brown RN, Myers JR: **A genetic map of squash (*Cucurbita* ssp.) with randomly amplified polymorphic DNA markers and morphological markers.** J Am Soc Hortic Sci 2002, 127:568-575.

Clepet C, Joobeur T, Zheng Y, Jublot D, Huang MY, Truniger V, Boualem A, Hernandez-Gonzalez ME, Dolcet-Sanjuan R, Portnoy V, Mascarell-Creus A, Cano-Delgado AI, et

al.: **Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon.** BMC Genomics 2011, 12:252, doi: 10.1186/1471-2164-12-252.

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R: **Development and implementation of high-throughput SNP genotyping in barley.** BMC Genomics 2009, 10:582.

Cooper JL, Till BJ, LaportRG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N, Bilyeu KD, Meksem K, Comai L, Henikoff S: **TILLING to detect induced mutations in soybean**. BMC Plant Biol 2008, 8:9.

Cucurbigene [http://cucurbigene.net/]

Dai N, Cohen S, Portnoy V, Tzuri G, Harel-Beja R, Pompan-Lotan M, Carmi N, Zhang GF, Diber A, Pollock S, Karchi H, et al.: **Metabolism of soluble sugars in developing melon fruit: a global transcriptional view of the metabolic transition to sucrose accumulation.** Plant Mol Biol 2011, 76(1-2):1-18, doi:10.1007/s11103-011-9757-1.

de Boer JM, Borm TJA, Jesse T , Brugmans B , Tang XM, Bryan GJ, Bakker J, van Eck HJ, Visser RGF: **A hybrid BAC physical map of potato: a framework for sequencing a heterozygous genome.** BMC Genomics 2011, 12:594, doi: 10.1186/1471-2164-12-594.

Deleu W, Esteras C, Roig C, Gonzalez-To M, Fernandez-Silva I, Gonzalez- Ibeas D, Blanca J, Aranda MA, Arus P, Nuez F, Monforte AJ, Picó MB, Garcia- Mas J: **A set of EST-SNPs for map saturation and cultivar identification in melon.** BMC Plant Biol 2009, 9:90.

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, Aubert G: **Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea.** Genomics 2010, 11:468.

Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub JE, Zalapa JE, Cuevas HE, Dace G, Oliver M, Boissot N, Dogimont C, Pitrat M, Hofstede R, Koert P, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Schaffer A, Katzir N, Xu Y, Zhang H, Fukino N, Matsumoto S, Garcia-Mas J, Monforte AJ: **A consensus linkage map for molecular markers and Quantitative Trait *Loci* associated with economically important traits in melon (*Cucumis melo* L.).** BMC Plant Biol 2011, 11:111.

Eduardo I, Arus P, Monforte AJ: **Development of a genomiclibrary of near isogenic lines (NILs) in melon (*Cucumis melo*L.) from the exotic accession PI161375.** Theor Appl Genet 2005, 112(1):139-148.

Elias R, Till BJ, Mba C, Al-Safadi B: **Optimizing TILLING and Ecotilling techniques for potato (*Solanum tuberosum* L).** BMC Research Notes 2009, 2:141 doi:10.1186/1756-0500-2-141.

Esteras C, Gómez P, Monforte AJ, Blanca J, Vicente-Dólera N, Roig C, Nuez F, Picó B: **High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait *loci* mapping.** BMC Genomics 2012a, 13:80, doi: 10.1186/1471-2164-13-80.

Esteras C, Lunn J, Sulpice R, Blanca J, Garcia-Mas J, Pitrat M, Nuez F, Picó B: **Phenotyping a highly diverse core melon collection to be screened using Ecotilling.** 8 Plant Genomics European Meetings (Plant Gem), Lisbon (Portugal), 7–10 octubre 2009a: p214.

Esteras C, Pascual L, Saladie M, Dogimont C, Garcia-Mas J, Nuez F, Picó B: **Use of Ecotilling to identify natural allelic variants of melon candidate genes involved in fruit**

**ripening.** 8 Plant Genomics European Meetings (Plant Gem), Lisboa (Portugal), 7-10 octubre 2009b: p213.

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools.** En: Cucurbits: Genetics, Genomics and Breeding of Cucurbits. Editado por Wang Y, Behera TK y Kole C. Science Publishers Inc, Enfield, New Hampshire. 2012b: 140-198.

Esteras C, Sifres A, Nuez F, Picó B: **Variabilidad de *Cucurbita maxima* en su zona de origen: un recurso de interés para la mejora de esta hortaliza.** En Actas de Horticultura 54, VI Congreso Ibérico, XII Nacional de Ciencias Hortícolas. Resúmenes. Editado por Pardo A, Suso ML, Vázquez N. Logroño (España), 25-29 mayo 2009c:1284-1290.

Fan JB, Chee MS, Gunderson KL: **Highly parallel genomic assays**. Nat Rev Genet 2006, 7:632–644, doi:10.1038/nrg1901.

FAOSTAT [http://faostat.fao.org/site/291/default.aspx], a fecha 3 abril de 2012.

Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Picó B, Nuez F, Arus P, Garcia-Mas J, Monforte AJ: **Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.).** Theor Appl Genet 2008, 118:139-150.

Formisano G, Roig C, Esteras C, Ercolano MR, Nuez F, Monforte AJ, Picó B: **Genetic diversity of Spanish *Cucurbita pepo* landraces: an unexploited resource for summer squash breeding.** Genet Resour Crop Evol 2011, doi:10.1007/s10722-011-9753-y.

Gabriel S, Ziaugra L, Tabbaa D: **SNP genotyping using the Sequenom MassARRAY iPLEX platform**. Current Protocols in Human Genetics 2009, 60, 2.12.11–12.12.18.

Gallo M, Zorzoli R, Rodriguez GR, Pratta GR: **Genetic linkage among tomato fruit quality traits and polypeptides expressed at two ripening stages.** Rev Fac Cienc Agrar 2011, 43(2):145-156.

Ganal M, Polley A, Durstewitz G, Thompson M, Hansen M: **A 50K SNP Infinium chip for maize genome analysis.** 52nd Annual Maize Genetics Conference. Riva del Garda (Trento, Italia), 18-21 marzo 2010: p172.

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Picó B, Gabaldón T, Roma G, Guigó R, Casacuberta JM, Arús P, Puigdomènech P: **The genome of melon (*Cucumis melo* L.). Genome amplification in the absence of recent duplication in an old widely cultivated species.** (en fase de publicación)

Gong L, Stift G, Kofler R, Pachner M, Lelley T: **Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L**. Theor Appl Genet 2008, 117:37-48.

Gonzalez-Ibeas D, Blanca J, Donaire L, Saladie M, Mascarell-Creus A, Cano-Delgado A, Garcia-Mas J, Llave C, Aranda MA: **Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing.** BMC Genomics 2011, 12:393, doi: 10.1186/1471-2164-12-393.

Gonzalez-Ibeas D, Blanca J, Roig C, González-To M, Picó B, Truniger V, Gómez P, Deleu W, Caño-Delgado A, Arús P, Nuez F, Garcia-Mas J, Puigdomènech P, Aranda MA: **MELOGEN: an EST database for melon functional genomics.** BMC Genomics 2007, 8:306, doi:10.1186/1471-2164-8-306.

Gonzalez-Ibeas D, Cañizares J, Aranda MA: **Microarray analysis shows that recessive resistance to watermelon mosaic virus in melon is associated with the induction of**

**defense response genes.** Mol Plant Microbe Int 2012, 25(1):107-118, doi:10.1094/MPMI-07-11-0193.

González M, Xu M, Esteras C, Roig C, Monforte AJ, Troadec C, Pujol M, Nuez F, Bendahmane A, Garcia-Mas J, Picó B: **Towards a TILLING platform for functional genomics in Piel de Sapo melons.** BMC Research Notes 2011, 4:289.

Gonzalez VM, Garcia-Mas J, Arus P, Puigdomenech P: **Generation of a BAC-based physical map of the melon genome.** BMC Genomics 2010a, 11:339, doi:10.1186/1471-2164-11-339.

Grewal D, Manito C, Bartolome V: **Doubled haploids generated through anther culture from crosses of elite Indica and Japonica cultivars and/or lines of rice: large-scale production, agronomic performance, and molecular characterization.** Crop Science 2011, 51(6):2544-2553, doi: 10.2135/cropsci2011.04.0236.

Gu JF, Yin XY, Struik PC, Stomph TJ, Wang HQ: **Using chromosome introgression lines to map quantitative trait *loci* for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well-watered field conditions.** J Exp Bot 2012, 63(1):455-469, doi: 10.1093/jxb/err292.

Guo SG, Xu Y, Zhang HY, Gong GY, Huang SW, et al.: **Latest advances in watermelon genomics.** En Acta Horticulturae 871, IV International Symposium on Cucurbits. 2010:599–606.

Guo S, Tian RX, Feng ZW, Zhao CF: **Cloning and Expression Analysis of the Smooth-Edge Seed Gene GSE-1 from *Cucurbita pepo* L**. Plant Mol Biol Rep 2011, 29(2):498-504, doi:10.1007/s11105-010-0253-6.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, et al.: **The genome of the cucumber, *Cucumis sativus*. L.** Nat Genet 2009, 41:1275-1281.

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB: **High-throughput genotypingwith the GoldenGate assay in the complex genome of soybean.** Theor Appl Genet 2008, 116(7):945-52.

IbizaVP, Cañizares J, Nuez F: **EcoTILLING in *Capsicum* species: searching for new virus resistances.** BMC Genomics 2010, 11:631.

ICuGI (International Cucurbit Genomics Initiative) [http://www.icugi.org/].

Jeffrey C: **A new system of *Cucurbitaceae*.** Bot Zhurn 2005, 90: 332–335.

Kim MY, Lee S, Van K, Kim T-H, Jeong S-C, Choi I-Y, Kim D-S, Lee Y-S, Park D, Ma J, Kim W-Y, Kim BC, Park S, Lee KA, Kim D-H, Kim KH, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee YH, Kim K-H, Moon J-K, Schmutz J, Jackson SA,Bhak J, Lee SH: **Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome.** PNAS 2010, 107(51): 22032–22037.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R: **A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*.** PLoS Genetics 2009, 5(7):e1000551.

Lee YH, Jeon HJ, Hong KH, Kim BD: **Use of random amplified polymorphic DNA for linkage group analysis in an interspecific cross hybrid F2 generation of *Cucurbita*.** J Kor Soc Hortic Sci 1995, 36:323-330.

Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K: **RNA-Seq improves annotation of protein-coding genes in the cucumber genome.** BMC Genomics 2011, 12:540

Ma J, Yan GJ, Liu CJ: **Development of near-isogenic lines for a major QTL on 3BL conferring Fusarium crown rot resistance in hexaploid wheat.** Euphytica 2012, 183(2):147-152, doi: 10.1007/s10681-011-0414-1.

Mammadov J, Chen W, Mingus J, Thompson S, Kumpatla S: **Development of versatile gene-based SNP assays in maize (*Zea mays* L.).** Mol Breeding 2012, 29:779–790, doi:10.1007/s11032-011-9589-3.

Manzano S, Martínez C, Gómez P, Garrido D, Jamilena M: **Cloning and characterisation of two CTR1-like genes in *Cucurbita pepo*: regulation of their expression during male and female flower development.** Sex Plant Reprod 2010, 23(4):301-13.

MAPA [http://www.magrama.gob.es/es/], a fecha de 4 abril de 2012

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A: **A transposon-induced epigenetic change leads to sex determination in melon.** Nature 2009, 461(7267):1135-U237, doi: 10.1038/nature08498.

Martínez C, Manzano S, Megías Z, Rosales, R, Valenzuela JL, Garrido D, Jamilena M: **Ethylene production and expression of CpACS1 and CpACO1 genes in zucchini fruits during their postharvest storage at different temperatures.** Postharvest Biol Tec 2011 (en prensa).

Mascarell-Creus A, Cañizares J, Vilarrasa J, Mora-García S, Blanca J, Gonzalez-Ibeas D, Saladié M, Roig C, Deleu W, Picó B, et al.: **An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon.** BMC Genomics 10:467, doi:10.1186/1471-2164-10-467.

McCord PH; Sosinski BR, Haynes KG, Clough ME, Yencho GC: **Linkage mapping and QTL analysis of agronomic traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*).** Crop Science 2011, 51(2):771-785, doi:10.2135/cropsci2010. 02.0108.

MELOGEN [http://www.melogen.upv.es]

MELONOMICS [http://www.melonomics.upv.es]

Mueller LA, Lankhorst RK, Tanksley SD, et al.: **A snapshot of the emerging tomato genome sequence**. Plant Genome 2009, 2:78–92.

NCBI, NCBI-dbEST database [http://www.ncbi.nlm.nih.gov/dbEST].

Nelson JC, Wang S, Wu Y, Li X, Antony G, White FF, Yu J: **Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum.** BMC Genomics 2011, 12:352.

Obrero A, Die JV, Roman B, Gomez P, Nadal S, Gonzalez-Verdejo CI: **Selection of reference genes for gene expression studies in Zucchini (*Cucurbita pepo*) using qPCR.** J Agr Food Chem 2011, 59(10):5402-5411, doi: 10.1021/jf200689r.

Okabe Y, Asamizu E, Saito T, Matsukura C, Ariizumi T, Bre`s C, Rothan C,Mizoguchi T, Ezura H: **Tomato TILLING Technology: Development of a Reverse Genetics Tool for the Efficient Isolation of Mutants from Micro-Tom Mutant Libraries.** Plant Cell Physiol 2011, 52(11):1994–2005, doi:10.1093/pcp/pcr134.

Picó B, Sifres A, Esteras C, Nuez F**: *Cucumis* SSRs markers applied to the study of the genetic diversity in the *Cucurbita* genus.** Cucurbit Genet Coop Rep 2005-2006, 28-29:70-72.

Portnoy V, Diber A, Pollock S, Karchi H, Lev S, Tzuri G, Harel-Beja R, Forer R, Portnoy VH, Lewinsohn E, Tadmor Y, Burger J, Schaffer A, Katzir N: **Use of Non-Normalized, Non-Amplified cDNA for 454-Based RNA-Seq of Fleshy Melon Fruit.** The Plant Genome 2011, 4(1):36-46, doi: 10.3835/plantgenome2010.11.0026.

Qin HD, Feng SP, Chen C, Guo YF, Knapp S, Culbreath A, He GH, Wang ML, Zhang XY, Holbrook CC, Ozias-Akins P, Guo BZ: **An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations.** Theor Appl Genet 2012, 124(4):653-664, doi:10.1007/s00122-011-1737-y.

Ren Yi, Zhao Hong, Kou Qinghe, Jiang Jiao, Guo Shaogui, Zhang Haiying, Hou Wenju, Zou Xiaohua, Sun Honghe, Gong Guoyi, Levi Amnon, Xu Yong: **A high resolution genetic**

map anchoring scaffolds of the sequenced watermelon genome. PloS One 2012, 7(1):e29453.

Sahin-Cevik M, Moore GA: Quantitative trait *loci* analysis of morphological traits in *Citrus*. Plant Biotechnol Rep 2012, 6(1):47-57, doi:10.1007/s11816-011-0194-z. *Shirasawa K, Isobe S, Hirakawa H, Asamizu E, Fukuoka H, Just D, Rothan C, Sasamoto S, Fujishiro T, Kishida Y, Kohara M, Tsuruoka H, Wada T, Nakamura Y, Sato S, Tabata S: SNP Discovery and linkage map construction in cultivated tomato. DNA Res 2010, 17:381–391, doi:10.1093/dnares/dsq024.

Steermers FJ, Gunderson KL: Whole genome genotyping technologies on the Bead Array TM platform. Biotechnol J 2007, 2:41–49.

Sugiyama A, Omura M, Matsumoto H, Shimada T, Fujii H, Endo T, Shimizu T, Nesumi H, Ikoma Y: Quantitative Trait *Loci* (QTL) analysis of carotenoid content in *Citrus* fruit. J Jpn Soc Hortic Sci 2011, 80(2):136-144.

The Potato Genome Sequencing Consortium: Genome sequence and analysis of the tuber crop potato. Nature 2011, 475:189–195, doi:10.1038/nature10158.

Uauy C, Paraiso F , Colasuonno P, Tran RK, Tsai H , Berardi S, Comai L, Dubcovsky J: A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol 2009, 9:115, doi: 10.1186/1471-2229-9-115.

van Heusden S, Vosman B, Bovy A: Infinium Beadarray technology in tomato: marker density and recombination. Plant & Animal Genomes XIX Conference. Town & Country Convention Center, San Diego, CA. 15-19 enero 2011.

Wang J, Kong L, Zhao S, Zhang H, Tang L, Li Z, Gu X, Luo J, Gao G: Rice-Map: a new-generation rice genome browser. BMC Genomics 2011a, 12:165.

Wang Z, Wang F, Zhou R, Wang J, Zhang H: **Identification of quantitative trait *loci* for cold tolerance during the germination and seedling stages in rice (*Oryza sativa* L.).** Euphytica 2011b, 181(3):405-413, doi: 10.1007/s10681-011-0469-z.

Xu Y, Badea C, Tran F, Frick M, Schneiderman D, Robert L, Harris L, Thomas D, Tinker N, Gaudet D, Laroche A: **Next-Gen sequencing of the transcriptome of triticale.** Plant Genetic Resources-Characterization and Utilization 2011, 9(2):181-184, doi:10.1017/S1479262111000281.

Zhang WW, Pan JS, He HL, Zhang C, Li Z, Zhao JL, Yuan XJ, Zhu LH, Huang SW, Cai R: **Construction of a high density integrated genetic map for cucumber (*Cucumis sativus* L.).** Theor Appl Genet 2012, 124(2):249-259, doi: 10.1007/s00122-011-1701-x.

Zraidi A, Stift G, Pachner M, Shojaeiyan A, Gong L, Lelley T: **A consensus map for *Cucurbita pepo*.** Mol Breed 2007, 20:375-388.

## Abreviaturas:

**AFLP**: Amplified Fragment Length Polymorphisms

**ASO**: allele-specific oligo

**BC**: backcross

**CAPS**: Cleaved Amplified Polymorphic Sequence

**cDNA**: complementary DNA

**Chr**: chromosome

**CL60**: filtro "el SNV se encuentra a menos de 60 pb del extremo de la secuencia

**cpDNA**: chloroplast DNA

**COMAV**: Instituto de Conservación y Mejora de la Agrodiversidad Valenciana

**CS60**: filtro "el SNV se encuentra a menos de 60 pb de otro SNP o INDEL"

**DHL**: Double Haploid Line

**dCAPS**: derived Cleaved Amplified Polymorphic Sequence

**EST**: Expressed Sequence Tag

**GO**: Gene Ontology

**GS**: Genome Sequencer

**gSSR**: genomic SSR

**HVR4**: filtro "SNV localizado en una región con más de 4 SNVs por 100 pb"

**I59**: filtro "un intrón se localiza a menos de 59 pb del SNV"

**ICuGI**: International Cucurbit Genomics Initiative

**IL**: Introgression Line

**INDEL**: insertion-deletion

**ISSR**: Inter Simple Sequence Repeats

**ITS**: Internal Transcribed Spacer

**LG**: linkage group

**LSO**: *locus*-specific oligo

**MAGIC**: Multiparent Advanced Generation Inter-Cross populations

**MAS**: Marker-Assisted Selection

**mRNA**: messenger RNA

**mtDNA**: mitochondrial DNA

**NCBI**: National Center for Biotechnology Information

**NGS**: Next Generation Sequencing

**NJ**: Neighbour-Joining

**NPGS**: United States National Plant Germplasm System

**NVCant**: filtro "variable dentro del grupo *cantalupensis* (Vedrantais, Charentais y Dulce)"

**NVCon**: filtro "variable dentro del grupo de *conomon* (Pat81 y PI161375)"

**NVDul**: filtro "variable dentro de Dulce"

**NVmelo**: filtro "variable dentro del grupo de entradas pertenecientes a la ssp. *melo*"

**NVPat81**: filtro "variable dentro de Pat81"

**NVPI161**: filtro "variable dentro de PI161375"

**NVPiñ**: filtro "variable dentro de Piñonet"

**NVPs**: filtro "variable dentro del grupo de Piel de sapo (Piñonet y T111)"

**NVT111**: filtro "variable dentro de T111"

**NVVed**: filtro "variable dentro de Vedrantais"

**ORF**: Open Reading Frame

**PCA**: Principal Component Analysis

**PCoA**: Principal Coordinate Analysis

**PCR**: Polymerase Chain Reaction

**PIC**: Polymorphism Information Content

**PS**: Piel de sapo

**pSCH**: SNP putativos, identificados *in silico*, en los que para al menos un alelo sólo existe una lectura

**pSNP**: SNP putativos, identificados *in silico*, con al menos 2 lecturas por alelo

**qPCR**:  quantitative PCR

**QTL**: Quantitative Trait *Loci*

**RAPD**: Randomly Amplified Polymorphic DNA

**RFLP**: Restriction Fragment Length Polymorphism

**RIL**: Recombinant Inbred Line

**S**: Scallop

**SBAP**: Sequence Based Amplified Polymorphism

**SC**: Songwhan Charmi (PI161375)

**SCAR**: Sequence Characterized Amplified Region

**SDS-PAGE**: sodium dodecyl sulfate polyacrylamide gel electrophoresis

**SFP**: Single Feature Polymorphisms

**SGS**: Second Generation Sequencing

**SNP**: Single Nucleotide Polymorphism

**SNV**: Single Nucleotide Variant

**SRAP**: Sequence Related Amplified Polymorphism

**SSR**: Simple Sequence Repeats, microsatellite

**TILLING**: Targeting Induced Local Lesions in Genomes

**UCR**: filtro "SNV situado en una región no única o no contigua"

**UTR**: Untranslated Region

**VKS**: filtro "no es un SNP"

**VSCant-Con**: filtro "no variable en el grupo de *cantalupensis* y *conomon*"

**VSPs-Con**: filtro "no variable en el grupo de Piel de sapo y *conomon*"

**Z**: Zucchini

# I. INTRODUCCIÓN GENERAL

# I. INTRODUCCIÓN GENERAL

Revisión bibliográfica:

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools.** En: Cucurbits: Genetics, Genomics and Breeding of Cucurbits. Editado por Wang Y, Behera TK y Kole C. New Hampshire: Science Publishers Inc, Enfield; 2012:140-198.

**Abstract**

*Cucurbitaceae* are among the largest and most diverse plant families. These include several economically important cucurbits such as watermelon (*Citrullus lanatus*), cucumber (*Cucumis sativus*), melon (*Cucumis melo*) and squashes (*Cucurbita* spp.), but also minor crops and wild species distributed worldwide. It is assumed to be of Asian origin, but cucurbits have diversified around the world. In this chapter, we describe studies that aim to know the extant genetic diversity among cucurbits. Main tribes, comprising the most important genera are first described, ***Benincaseae*** (*Cucumis, Citrullus, Benincasa and Lagenaria*) and ***Cucurbiteae*** (*Cucurbita*). Most of the studies deal with variability within the main crops of each genus, melon and cucumber (*Cucumis*), watermelon (*Citrullus*), and pumpkins, *C. pepo, C. moschata* and *C. maxima* (*Cucurbita*), but also studies dealing with minor crops are included, African horned cucumber (*Cucumis metuliferus*), West Indian gherkin (*Cucumis anguria*), white gourd (*Benincasa hispida*), bottlegourd (*Lagenaria siceraria*) and malabar melon (*Cucurbita ficifolia*). In addition to the main tribes, we also review ***Joliffieae***, ***Luffeae*** and **Sicyeae**, including minor crops such as bitter melon (*Momordica charantia*), chayote (*Sechium edule*) and sponge gourd (*Luffa cylindrica*), respectively.

For each genus, a general view is provided on the taxonomy, distribution and intra-genus relationships. For each species, we summarize the information on taxonomy, morphological studies and molecular assays using different marker systems. Reported works describe the variability of wild types, landraces or commercial cultivars in the centers of origin, and primary and secondary diversification centers. The evolution of the molecular analyses is shown. From those based on the protein polymorphisms, through the use of chloroplastic and mitochondrial DNA, to different types of DNA-based markers, dominant (ISSRs, RAPDs and AFLPs) and codominant and sequence-

specific (SRAPs, SCARs and SSRs), including the most recently used SNPs, coming from genomic or ESTs sequencing projects. Knowing the amount and distribution of genetic variability will be essential to conserve and classify all this variation and to exploit it for cucurbits breeding.

## I.1 Introduction

The family *Cucurbitaceae* comprises some of the most economically important crops. Cucurbits have been used by humans as food, containers, musical instruments and as a source of medicine, for more than 12,000 years (Whitaker and Davis, 1962; Brothwell and Brothwell, 1969; Lira-Saade, 1995).

Jeffrey (1990) and Robinson and Decker-Walters (1997) described 118 genera and about 825 species within the family. However, according to the most recent classifications (Jeffrey, 2005; Jeffrey and De Wilde, 2006) *Cucurbitaceae* comprises 130 genera, including about 800 species. The family is divided into two subfamilies: Nhandiroboideae, also called Zanonioideae (Jeffrey 1990), with 19 genera and about 60 species without economic value; and Cucurbitoideae, with 111 genera and 740 species. The subfamily Cucurbitoideae comprises some of the most important tribes and genera within cucurbits: Benincaseae (*Benincasa*, *Citrullus*, *Coccinia*, *Lagenaria*, *Cucumeropsis*, *Cucumis*), Luffeae (*Luffa*), Cucurbiteae (*Cucurbita*), Sicyeae (*Cyclanthera*, *Sechium*), Joliffieae (*Momordica*, *Telfairia*), Schizopeponeae (*Schizopepon*), Trichosantheae (*Hodgsonia*, *Trichosanthes*), Bryonieae (*Ecballium*), Herpetospermae and Coniandreae (Jeffrey, 1990; Rubatzky and Yamaguchi, 1997; Jeffrey, 2005). There are some recent phylogenetic studies of the family *Cucurbitaceae* using chloroplast DNA sequences (cpDNA) (Kocyan et al., 2007). Molecular data weakly support the traditional subfamilies Cucurbitoideae and Nhandiroboideae, and

recover most of the 11 tribes, but almost none of the subtribes. Within the subfamily Cucurbitoideae, a clade comprising all the tribes but Joliffieae (polyphyletic) was reported as a "fused stamen" clade due to their fused filaments or connectives. In this clade several tribes clustered together in four groups: Herpetospermae-Schizopeponeae group, Bryonieae group, Coniandreae-Benincaseae-Cucurbiteae group and Luffeae-Sicyeae-Trichosantheae group (Kocyan et al., 2007). The most important diagnostic characters for the genera and tribes of *Cucurbitaceae* come from androecium and gynoecium morphology (number of free styles, fusion of filaments and/or anthers), tendril type, pollen structure and seed coat. Subfamily Nhandiroboideae is characterized by free styles, small pollen grains with a striate exine and branched tendrils with a sensitive basal part. Subfamily Cucurbitoideae is characterized by having styles united in a single column, tectate to semitectate pollen with a reticulate or echinate exine and simple, bifid or multifid tendrils with non-spiralling basal part. These traits correlate quite well with the phylogeny obtained with molecular data.

The genera *Cucumis*, *Cucurbita* and *Citrullus* include species (cucumber, melon, watermelon and squash) that are among the most widely cultivated crops worldwide. Apart from them, there are other notable cucurbits of local or regional economic importance, such as *Lagenaria, Momordica, Benincasa*, *Luffa* and *Sechium* (Lira Saade and Montes-Hernández, 1994; Bates et al., 1995; Lira-Saade, 1995).

Cucurbits have diversified around the world. A large amount of genetic resources adapted to many different environmental and growing conditions can be found in different areas. Knowing the extant genetic diversity among cucurbits is important in order to optimize collection and conservation programs and to facilitate the ongoing efforts by plant breeders worldwide to improve melon, cucumber, watermelon and squash with traits from wild relatives.

**I.2 Tribe *Benincaseae***

**I.2.1 Genus *Cucumis***

This genus comprises 33 species, including the most recent to be discovered, *C. canoxyi* Thulin & A.N. Al-Gifri (Thulin and Al-Gifri, 1994). Apart from *C. melo* L. and *C. sativus* L., two species are also consumed: *C. anguria* L. (West Indian gherkin) and *C. metuliferus* E. Meyer ex Naudin (African horned cucumber or jelly melon). Others, such as *C. dipsaceus* Ehrenberg ex Spach and *C. myriocarpus* Naudin, are used as ornamentals. The genus has been traditionally conceived of as being an African genus divided into two subgenera: *cucumis* and *melo* (Jeffrey, 1980; Kirkbride, 1993). The subgenus *cucumis* (two species, *C. sativus* and *C. hystrix* Chakr.) is confined to Asia and has chromosome numbers $n = 7$ (*C. sativus*) and $n = 12$ (*C. hystrix*) (Chen et al., 1997a). Figure I-1 shows the diversity centers of the main species within the genus. The subgenus *melo* (30 species, including *C. melo*), with most of its species in Africa, has a chromosome number $n = 12$ (Kirkbride, 1993), although there exists several tetraploid and hexaploid forms. Ashurmetov (1995) proposed separating *Cucumis* into two genera, *Cucumis* and *Melo*, but it was not broadly accepted.

Some molecular phylogenetic studies of *Cucumis* have been conducted using seed storage proteins (Singh and Matta, 2008), chloroplast restriction sites (Perl-Treves and Galun, 1985), nuclear isozymes (Perl-Treves et al., 1985), nuclear ribosomal DNA from the Internal Transcribed Spacer (ITS) region along with nuclear Simple Sequence Repeat (SSR) markers (Garcia-Mas et al., 2004) and chloroplast SSRs and sequence analysis (Chung et al., 2003, 2006). One of the most recent studies, conducted by Renner et al. (2007), tested the genetic relationships among *Cucumis,* combining cpDNA sequences and nuclear ITS. This study extended the analysis performed in previous studies considerably, using representatives of other potentially related genera

(*Cucumella, Dicaelospermum, Mukia, Muellerargia, Myrmecosicyos* and *Oreosyce*) included in the most recent morphology-based classification of *Cucurbitaceae* (Jeffrey, 2005). The *Cucumis* species relationships found by these authors differ from those found in earlier studies. The deepest divergence lies between the common ancestor of *C. hirsutus* and *C. humifructus* and the stem lineage of the remainder of the genus. The authors also concluded that the closest relative of *Cucumis* is *Muellerargia*, and showed that the genera *Cucumella, Dicaelospermum, Mukia, Myrmecosicyos* and *Oreosyce* are grouped among species of *Cucumis*. Similar results were obtained by Ghebretinsae et al. (2007) and Schaefer (2007), who proposed the expansion of the genus *Cucumis* to include these related genera. Renner et al. (2007) extend the range of the genus throughout the Malesian regionand into Australia. According to this study, *Cucumis* comprises an old Australian/Asian component. *Cucumis sativus* would have evolved within this Australian/Asian clade. *Cucumis melo* is sister to this Australian/Asian clade, rather than being close to African species as previously thought. In fact, a more recent research, carried out by Sebastian et al. (2010), reports a species from Australia (*Cucumis picrocarpus*) as the melon's closest relative. According to these studies (Renner et al., 2007; Sebastian et al., 2010), *C. melo* might have originated in Asia and then arrived in Africa.

### I.2.1.1 *Cucumis melo*

### I.2.1.1.1 Origin and taxonomy

*C. melo* is one of the most important fruit vegetables cultivated in tropical and temperate regions, as it is highly valued for the quality of its fruits. Traditionally, *C. melo* is considered to be divided into two subspecies according to ovary hairiness, subspecies *melo* (long hairs, distributed from India to Europe and in America) and

subspecies *agrestis* (short hairs, in Eastern Asia from India to Japan and in Africa) (Jeffrey, 1980). Today, the species comprises wild, feral and cultivated varieties, including sweet melons used for dessert and nonsweet ones consumed raw, pickled or cooked.

The origin of *C. melo* remains unclear (Robinson and Decker Walters, 1997; Yashiro et al., 2005; Lebeda et al., 2006). African origins have been traditionally assumed, but, as previously mentioned, recent studies suggest that *C. melo* could have originated somewhere in Asia and then reached Africa from there (Renner et al., 2007). In fact, wild melons are very frequent in East and West Africa, but also from Central Asia to India (Whitaker and Bemis, 1976; Staub et al., 1987; McCreight et al., 1993; McCreight and Staub, 1993; Rubatzky and Yamaguchi, 1997). It is also possible that in Africa and Asia different domestication events took place (Bates and Robinson, 1995), the most extensive of which may have occurred in Asia (more edible types). Independent domestication has also been proposed for the two subspecies.

Melon has suffered an intense process of diversification, and today shows great variation in morphological and physiological characters. Primary and secondary centers of genetic diversity are located from eastern Asia to the Mediterranean Sea (Afghanistan, Iran, Iraq, Saudi Arabia, Turkey, China, Russia and India) (Robinson and Decker-Walters, 1997; Akashi et al., 2002). Seeds of wild *C. melo* ssp. *agrestis* and possibly selected forms more closely resembling currently cultivated *C. melo* could have been introduced from Africa into the Middle East (Turkey, Iraq and Iran) and Asia (India, China and Japan) through land and sea commerce routes (Kajale, 1979; Walters, 1989; Fujishita, 1992). It seems that melon was cultivated in Iran and China 3000 BC, in India 2000 BC, in Egypt 1500 BC and western Japan as early as 100 BC (Walters, 1989; Fujishita, 1992; Decker-Walters, 1999; Stepansky et al., 1999; Karchi, 2000;

Luan et al., 2008). Melon spread through Asia and subsequently to Europe (Roman-Greek periods) (Szabó et al., 2005). Three independent introductions to Europe (from the east: Russia, Bulgaria, Hungary; from the south-east: Greece, Albania, Romania; from the south: Italy) were hypothesized by Pitrat et al. (1999). Later, Columbus introduced this crop to America, where it became popular and dispersed quickly, producing a wide range of new cultivars.

*C. melo* is considered to be the most variable species in the genus *Cucumis* (Kirkbride, 1993; Bates and Robinson, 1995). Great diversity in fruit shape, size, color and taste exists among melons. This variability was first classified as different species (*C. melo*, *C. flexuosus*, *C. dudaim*, *C. callosus*, *C. chate*, *C. conomon* and *C. momordica*) that now are considered varieties or morphotypes within the species. Naudin (1859) defined 9 "tribes" of cultivated melons and one wild form. Many scientists have since added or merged types (reviewed in Pitrat et al., 2000). Munger and Robinson (1991) proposed a simplified version of Naudin´s taxonomy, which is still used in many studies, with seven cultivar groups: *agrestis* Naud. (wild melon), *cantalupensis* Naud. (cantaloupe or muskmelon; Middle East), *inodorus* Naud (winter melons, honeydew, Cassaba; Middle East, southern Europe), *conomon* Mak. (pickling melon, Chinese white cucumber; Asia), *chito-dudaim* Naud (mango melon-queen's pocket melon; Asia), *flexuosus* Naud. (snake melon; Middle East), and *momordica* (Phoot or snap melon; Asia) (Robinson and Decker-Walters, 1997). More recently, Pitrat et al. (2000) described 16 botanical varieties: var. *conomon*, var. *makuwa*, var. *chinensis*, var. *momordica*, var. *acidulus* (included in ssp. *agrestis*) and var. *cantalupensis*, var. *reticulatus*, var. *adana*, var. *chandalak*, var. *ameri*, var. *inodorus*, var. *flexuosus*, var. *chate*, var. *tibish*, var. *dudaim*, var. *chito* (included in ssp. *melo*). In later revisions, Pitrat (2008) merged some groups. The description of the main botanical groups is included in Table I-1. Many of these

botanical groups include different cultivar-groups that are highly popular in different parts of the world. *Cantalupensis* and *inodorus* are of commercial interest in the United States, as well as in many European, Mediterranean and Asian countries (McCreight et al., 1993), and include cultivars belonging to different market classes. For example, while the popular market classes Charentais, Shipper, Ogen and Galia are in the *cantalupensis* group, the *inodorus* group houses an array of Cassaba market class types (e.g., Rochet, Piel de Sapo, Tendral, Crensahw, Honeydew, Kirkagac and Yellow Canari).

### I.2.1.1.2 Molecular markers used in genetic diversity studies

Different types of molecular markers have been used to assess genetic diversity in melon and to study genetic relationships among commercial types and different botanical groups. Isozymes were first utilized (Esquinas-Alcázar, 1981; Perl-Treves et al., 1985; Staub et al., 1997; Akashi et al., 2002), but soon DNA markers were applied: Restriction Fragment Lengh Polymorphisms (RFLPs) (Neuhausen, 1992; Silberstein et al., 1999), Random Amplified Polymorphic DNA (RAPDs) (García et al., 1998; Stepansky et al., 1999; Silberstein et al., 1999; Staub et al., 2000; Zhuang et al., 2004), Amplified Fragment Lengh Polymorphisms (AFLPs) (Garcia-Mas et al., 2000), microsatellites or Simple Sequence Repeats (SSRs) (Katzir et al., 1996; Staub et al., 2000; Danin-Poleg et al., 2001; Mliki et al., 2001; Monforte et al., 2003; Zhuang et al., 2004), Internal Simple Sequence Repeats (ISSRs) (Stepansky et al., 1999) and the sequence and structural analysis of the Internal Transcribed Spacer (ITS) (Fantaccione et al., 2008). In general, similar clustering is obtained with the different types of markers. An array of selected RAPD markers and reference accessions has been frequently employed to assess the genetic diversity of melon landraces and cultivars

from Europe, the USA and different centers of diversity (Staub et al., 2000). Also, due to the reproducibility and the discriminating capacity of AFLPs, and to the reliable, co-dominant and multi-allelic nature of SSRs, these marker systems have been broadly used to define genetic relationships among botanical groups and commercial market classes. Some recent studies have focused on the development of SSR collections using genomic microsatellite enriched libraries (Chiba et al., 2003; Ritschel et al., 2004), and more recently using available collections of melon Expressed Sequence Tags (ESTs) (Kong et al., 2007; Fernández-Silva et al., 2008).

Melon has become a model within Cucurbits for genetic and genomic studies. Different initiatives are generating sequence collections (González et al., 2010, González-Ibeas et al., 2007, http://www.melogen.upv.es; www.icugi.org) which are facilitating Single Nucleotide Polymorphism (SNP) detection. Morales et al. (2004) first detected SNPs in melon using a small set of available ESTs. These authors indicated an average frequency of 1 SNP per 441 bp between two *inodorus* genotypes. In a more complete study, screening 30,000 ESTs sequences from four genotypes, 356 high-quality SNPs were found (González-Ibeas et al., 2007). Some of them proved to study diversity in a set of melon accessions, giving genetic relationships similar to that found with SSRs (Deleu et al., 2009). In addition to genetic diversity studies, these markers have been used to construct several melon maps (Baudracco-Arnas and Pitrat, 1996; Wang et al., 1997; Liou et al., 1998; Brotman et al., 2000; Oliver et al., 2001; Danin-Poleg et al., 2002; Périn et al., 2002; Gonzalo et al., 2005; Deleu et al., 2009) that are being merged within the International Cucurbit Genomics Initiative (ICuGI) (http://www.icugi.org). The availability of increasing amounts of melon sequences is allowing the application of new methods to study variability in candidate genes of breeding interest (Nieto et al., 2007). In a recent study reported by Esteras et al. (2009b) EcoTILLING techniques are

being applied to study polymorphisms in genes involved in quality and ripening processes using a highly variable core collection of melons.

Molecular markers have been used to characterize elite melon germplasm (commercial cultivars, hybrids and breeding lines, mainly from the USA and European markets) mostly belonging to *cantalupensis* (Charentais, Shipper, European, western and eastern USA types, Galia and Ogen) and *inodorus* (Honey dew type and Cassaba Rochet, Piel de Sapo and Yellow Canari types) types. Despite the fact that molecular analysis discriminates between these market cultivars, the groupings were somewhat ambiguous, most likely due to intogressions during plant breeding. They also found a limited genetic diversity in some groups (Garcia et al., 1998; Staub et al., 2000).

Much higher genetic diversity is reported when exotic germplasm (wild, feral, landraces) from these and other botanic groups is also included. Most of the molecular studies support quite well the division into two major groups (ssp. *melo* and ssp. *agrestis*) (Staub et al., 1997; Silberstein et al., 1999; Danin-Poleg et al., 2001; Monforte et al., 2003; Nakata et al., 2005; Deleu et al., 2009; Esteras et al., 2009a). In general, higher molecular variability (number of alleles and polymorphic *loci*) is reported in Central Africa and India than in the extremes of melon distribution (the Mediterranean area and China Sea). One of the most complete studies covering representatives of most of the botanical groups was conducted by Stepansky et al. (1999).

They combined both phenotypic and molecular data (RAPDs and ISSRs), analyzing accessions from 23 countries, including wild, feral and cultivated forms representing the primary and secondary centers of diversity (Africa, southern and western Asia and the Far East). Phenotyping was based on a set of traits usually employed to define different botanical groups (seed size, stem thickness, pubescence, sex type, ovary shape, ovary pubescence, fruit shape and size, skin color, texture and design, splitting, abscission,

external aroma, flesh color, taste, sucrose, glucose and fructose and pH). According to these traits, the subdivision into most of the varietal groups persisted, indicating that the traditional classification is mainly based on consistent and highly informative characters. Some traits with taxonomic value show high variability. For example, wild types defined by small fruit size, thought to be *agrestis*, can show ovaries with long or short hairs. Also within the *flexuosus* types there exist both kinds of ovaries. The molecular results did not substantially contradict the phenotype-based dendrogram. The sweet-fruited *cantalupensis* and *inodorus* clustered together, in spite of their ripening differences, and the nonsweet varieties *agrestis*, *conomon* and *momordica* grouped together. There exist discrepancies in the classification of some botanical groups. For example, *dudaim* and *chito* cultivars often are grouped with *agrestis* types, although they have been reported to belong to ssp. *melo*. The *flexuosus* types clustered closer to the nonsweet genotypes in the phenotypic tree, and dispersed with the "dessert" ones in the molecular tree. The non existence of reproductive barriers within the species makes the crosses among the different cultivar groups possible, giving rise to a continuous distribution-pattern of variation. Esteras et al. (2009a) found similar results in a recent analysis with AFLPs of a melon core collection of 212 accessions, representing all the genetic diversity of the species, with *dudaim* and *chito* types being intermediate between ssp. *melo* and *agrestis* and *momordica* and *flexuosus* types which, being highly variable, interspersed among both groups.

## I.2.1.1.3 Molecular diversity of melon landraces in the centers of origin and diversity

Most of the molecular studies are focused on studying the genetic diversity of landraces and cultivars in all the significant primary and secondary centers of diversity.

Africa is a putative center of origin for melon, and so the study of African germplasm is essential to know the extent of genetic diversity within the species. Few studies have focused on African germplasm (Mliki et al., 2001; Akashi et al., 2006). Mliki et al. (2001), using RAPDs, clustered African accessions into two groups according to their origin: northern (mostly from Egypt, Tunisia, Libya, Morocco, Algeria, Ethiopia, Niger and Sierra Leone) *vs*. central-southern Africa (Zimbabwe, Zambia, Mali, Senegal and South Africa). Some accessions from Kenya, Senegal and Ghana grouped apart. The relationship of African germplasm to accessions from secondary centers of diversity suggests that major Indian introduction events originated from southern Africa, whereas the Middle East was the beneficiary of North African introductions (Staub et al., 2004; Sensoy et al., 2007). However, other studies support that large- and small-seeded types found in India (Akashi et al., 2002, 2006; Yashiro et al., 2005) were derived from northern and southern Africa, respectively (Tanaka et al. 2007). Molecular results support the polymorphic nature of the African germplasm.

The main center of diversity for melon and perhaps the origin of some of the principal commercial types, such as *cantalupensis* or *inodorus,* is located in the Near East and Central Asia (Jeffrey, 1980). Groups like *conomon*, *makuwa*, *momordica* and *flexuosus* also have an Asian origin. Many of the studies aimed at elucidating the genetic structure of Asian melons focus on these botanical groups, as they include most of the accessions used in melon breeding as sources of resistance genes to different pests and diseases (Akashi et al., 2002; Nakata et al., 2005; Yashiro et al., 2005; Tanaka et al., 2007). Yashiro et al. (2005) selected a representative collection of Asian accessions (from India, Myanmar, China, Korea and Japan). Using AFLPs they found that East Asian melons (*makuwa* and *conomon*) grouped apart and showed less variability than South Asian melons (especially those from India), which supported previous results (Akashi et

al., 2002). Tanaka et al. (2007) confirmed these results. Using RAPDs these authors found a high genetic diversity in India and a low variability in *makuwa* and *conomon* types.

The large genetic variation in India could be explained by the diverse climatic conditions. India is divided into 21 agro-ecological regions and 131 agroclimatic subregions. This is the area of origin of var. *momordica*. This botanic group presents most of the genes of resistance to diseases and pests, as well as to abiotic stress used for breeding cultivated melons. Dhillon et al. (2007, 2009) conducted the most complete morphological and molecular characterization of snapmelon landraces using accessions collected in many of the Indian agroclimatic subregions. A high level of variability was found for certain morphological traits (fruit cracking, abscission, flesh texture, acidity, sugar content, resistance to fungus and virus). The *momordica* accessions analyzed showed high levels of genetic diversity and were not closely related to melon accessions from other parts of the world, then supporting an independent origin of this botanic group (Staub et al., 2004; Dhillon et al., 2007, 2009). Moreover, regional differentiation among Indian accessions was reported using a set of 16 SSRs (Dhillon et al., 2009). Most of the accessions grouped according to agroclimatic subregions. Eastern Indian snapmelon has unique traits, so it is important that more germplasm from this region be sampled and preserved. These studies also included wild, small-fruited types from India. These *agrestis* accessions showed a narrow genetic diversity and cluster together with the *momordica* group. These data confirm that India is a primary center of melon diversity.

There exists a clear divergence between *momordica* and East Asian melons (*conomon-makuwa*). Oriental pickling melons are considered to be the most ancient forms of domesticated melons in China. Different hypotheses exist about where *makuwa* and

*conomon* types were domesticated. Kitamura (1950) proposed India from which point *makuwa* later established itself in northern China and *conomon* in South China (Jeffrey, 1980). Akashi et al. (2002) suggested that these types derive from a small-seed type, which adapted in East India. Recent molecular assessments of genetic diversity of Chinese melons from diverse geographical origins and belonging to different market classes [thick-skinned melon (netted and non-netted), non-netted thin-skinned and vegetable types] also support this proposal. Luan et al. (2008), using a previously defined standard RAPD marker array (Staub et al., 2000), group Chinese cultivars in two groups. Results suggest that certain melon types were introduced to western China *via* the Silk Road (from the Middle East, Iran or Iraq), whereas oriental Asian melon types (*conomon* and *makuwa* types) may have been introduced into China from India. This study also reports stark molecular differences between Indian and Chinese accessions and lends support to the occurrence of bottlenecks and/or geographic or political isolation. Chinese accessions are therefore a rich source of genetic diversity for plant improvement. Other studies relate *conomon* types with African accessions (Nakata et al., 2005). These authors, using RAPDs and SSRs, studied the variability of Japanese *conomon*. Cluster analysis separated *conomon* accessions into two groups, one more closely related to South African melons, and the other close to Japanese cultivars belonging to *cantalupensis* and *inodorus* groups. The *conomon*-African clade was separate from the rest of the examined genotypes, which might imply an Asian origin of the African *conomon*-like accessions or an independent domestication from similar ancestors. Some of the analyzed Japanese market classes were relatively rich in genetic variation, more than the USA or European accessions. Other studies, like that of Mosuk et al. (1999) on Korean diversity, also show the importance of the genetic diversity on the Asian continent.

A recent work using RAPDs and morphological traits studies the genetic diversity of melons in Myanmar (Yi et al., 2009). These authors indicated that the genetic diversity of Indian melon is conserved in Myanmar. They found that *conomon* and *agrestis* types from India group together and are apart from sweet melons. They also found a high diversity for the *conomon* group, equivalent to that of Indian melon populations, and higher than that of *conomon* from East Asia. Genetic introgression among melon groups through spontaneous hybridization is also indicated and considered important for maintaining or increasing the genetic diversity in Myanmar.

Asia is not only rich in nonsweet melons but also in melons belonging to the two most important commercial types. The Middle East (Iran, Iraq and Turkey) exhibits one of the highest variabilities in *cantalupensis*, *inodorus* and *flexuosus* types. This area has been reported as the origin of *inodorus* types and is very important for the diffusion of these types to Europe. It is thought that *cantalupensis* melons spread to Europe from the eastern part of Turkey (Zhukovsky, 1951; Günay, 1993). Sensoy et al. (2007) used RAPDs to analyze a collection of local Turkish melon genotypes. Most nonsweet melon types differed from sweet types. However, *flexuosus* types and some *momordica* grouped with sweet genotypes. The diversity found was very high, even higher than the examined African landraces. Distincion between *cantalupensis* and *inodorus* cultivars was not possible. Intermediate forms might have been formed between the *inodorus* and *cantalupensis* group due to the ancient farming practices employed by some local small-scale melon producers for centuries. Several melon genotypes grow together in many regions of Turkey and introgression of genotypes occurs naturally. Some types collected from the southeastern part of Turkey were related to some *conomon* and *momordica* accessions. Turkey has some unique genotypes, among them can be found several *dudaim* types or the *inodorus* kirkagac. Iran also has a wide diversity of the *inodorus*

61

group that has begun to be analyzed (Kohpayegani, 2004; Kohpayegani and Behbahani, 2008). Kohpayegani and Behbahani (2008) reported high variability in the Iranian melon, comparable to that of Turkish melons and much higher than landraces from Europe. A significant differentiation from reference *inodorus* has also been reported, suggesting the singularity of this germplasm. The variability in this area, including other countries of western Asia and eastern Europe (Iraq, Russia, etc.) needs to be studied further, including the variability of types belonging to other sweet and highly variable botanical groups such as *ameri*, *adana* or *chandalack*. A recent study reports the distinction between the three classical morphotypes of *adana* known in the Ukraine (Nimmakayala et al., 2009). The analyzed genotypes represent a major non-US and non-western Europe source of melon germplasm. In fact, the *adana* melons are considered to have been introduced into Europe from Asia in the 15th century and to be the ancestors of the *cantalupensis* group.

The Iberian Peninsula is considered to be a secondary diversification center for melon and is a major world producer of both *cantalupensis* and *inodorus* cultivars. The assessment of the genetic diversity of Spanishcultivars, not only as reference types (Staub et al., 2000), has become of a great interest. Typical Spanish types are Piel de Sapo, Tendral, Rochet, Amarillo and Blanco melons (var. *inodorus*) as well as several landraces belonging to var. *flexuosus*, which differ from types produced in other European countries (Staub et al., 2000). Morphological variation surveys of Spanish landraces and their market types have been carried out by several authors such as Nuez et al. (1986, 1988), Gómez-Guillamón et al. (1985, 1995, 1998) and Costa et al. (1989). López-Sesé et al. (2002, 2003), using the RAPD array reported by Staub et al. (2000) and SSR *loci* (Katzir et al., 1996; Danin-Poleg et al., 2001) evaluated a complete set of Spanish accessions. They found a high polymorphism between, but low within

accessions. While cluster analysis using fruit characteristics grouped accessions into cultivars, RAPD-based genetic distance estimates did not provide consistent accession groupings either by cultivar or geographic origin. The highest level of polymorphism was detected among melons originating from the central region of Spain, and in the Rochet cultivar, while accessions from the Andalusia region and Green cultivars were comparatively less diverse. The distinctive morphological characteristics among Spanish melon cultivars (texture and specialized taste) have prevented the introgression of genes from other germplasm of diverse origin, despite the lack of geographic isolation, as has been confirmed by molecular analysis. Other studies have evaluated traditional cultivars from different melon-producing regions in order to fingerprint those traditional landraces (Escribano et al., 2008).

Genetic variability of landraces from other European countries has also been reported. Fanourakis et al. (2000) and Staub et al. (2004) analyzed Greek landraces, finding significant differences from Spanish melons. The *flexuosus* types were the most variable and possessed affinities with *conomom* types from western Asia. It is possible that some Spanish and Greek *flexuosus* accessions may originate from or have ancestral relationships with melons from western Asia. Lotti et al. (2005) analyzed Italian landraces, including the typical Carosello (cucumber melon) belonging to the botanical group *chate,* which is closer to *flexuosus* than to *inodorus*. This variety has been suggested as the first cultivated melon in Africa, as it appeared on Egyptian mural paintings from 2000–1500 BC. Italian accessions were widely genetically different, and their clustering was related to their geographical origin. The variability within Hungarian landraces has also been studied, as Hungary is one of the first European countries where melon cultivation was reported (Szabó et al., 2008).

Most of the reported studies include accessions from Africa, Asia, Europe and the USA

as a set of reference commercial cultivars. However, there are also North American wild populations of melons. These have also been investigated in order to elucidate their origin (Decker-Walters et al., 2002a). Whereas these populations were assumed to be escaped forms of var. *chito* or var. *dudaim*, morphological-physiological and molecular (RAPDs and SSRs) data reveal that enough distinctiveness exists for them to be classified as ssp. *agrestis* var. *texanus*. This variety shows more similarity to var. *chito* and to cultivars from eastern Asia. The possible introduction to America from Asia is discussed, dating to pre-Columbian and post-Columbian times. Diversity of melons in Australia has been studied less, but this continent is also a center of complex morphological variation of *C. melo* (Telford, 1982).

### I.2.1.2 *Cucumis sativus*

### I.2.1.2.1 Origin and taxonomy

The species *C. sativus* and *C. hystrix* are found within *Cucumis* genus, subgenus *Cucumis*. Fertile amphidiploids have been synthesized between these two species, obtaining a species called *C. hystivus* (Chen et al., 1997a,b; Chung et al., 2006). This fact along with chloroplast polymorphisms observed among different *Cucumis* species, support the hypothesis that *C. hystrix* is the progenitor of *C. sativus*. Both species may share a common ancestral lineage as well (Chung et al., 2006). *C. sativus* houses several botanical varieties, including the cultivated cucumber (var. *sativus*) and the wild var. *hardwickii* (R.) Alef., cross-compatible with *C. sativus* and thought to be a progenitor or a feral form of *sativus* (Horst and Lower, 1978; Knerr et al., 1989). It grows in the Himalayan foothills and has medicinal properties. Var. *hardwickii* has been used for yield improvement in cultivated cucumber due to its multiple fruiting and branching (Staub and Bacher, 1997).

Cucumber, *C. sativus* var. *sativus*, is one of the largest crops in terms of worldwide production value, mainly in East Asia. It is thought that cucumber originated and was domesticated in Asia, likely on the Indian subcontinent by 3000 BC, and was disseminated thanks to the Silk Road and oceanic routes. China is considered a secondary center of diversity (Leppik, 1966; Meglic et al., 1996; Staub et al., 1999, 2008). Today this species is less variable than melons, and cultivars are grouped basically into two cultivar groups: those eaten fresh and those consumed as a processed product (Staub et al., 2008).

**I.2.1.2.2 Molecular markers used in genetic diversity studies**

Various marker systems have been applied to discriminate between cucumber cultivars. The first isozyme-based studies analyzed US and European cucumber germplasm (Staub et al., 1985; Knerr et al., 1989; Staub and Meglic, 1993), but the discrimination ability of these markers was limited. Later, RFLPs were also assayed (Dijkhuizen et al., 1996), detecting a low degree of variation, similar to that observed in melons by Neuhausen (1992). More recent RAPD-based studies resulted in grouping patterns consistent with accession origins, accepted dispersal routes and discriminating morphological characters (i.e., sex expression and fruit length to diameter ratio) (Horejsi and Staub, 1999). RAPDs have also been successful in discriminating elite accessions, but have detected limited genetic diversity (Bernet et al., 2003; Duca et al., 2008; Onto et al., 2008). Similar to what occurred in melons, several authors have made a big effort to develop cucumber SSRs. Katzir et al. (1996) reported seven highly polymorphic genomic SSRs in cucumber and melon. Later, Danin-Poleg et al. (2001), Fazio et al. (2002) and Kong et al. (2006) increased the number of microsatellites. SSR-enriched genomic libraries have also been developed to increase SSR availability (Fukino et al.,

2008; Watcharawongpaiboon and Chunwongse, 2008). Transferability to melon, bitter gourd, watermelon and pumpkin has been assessed. These markers have been tested in diversity studies using a different set of cucumber cultivars and have also been used for constructing genetic maps (Park et al., 2000; Bradeen et al., 2001; Fazio et al., 2003). EST-derived SSRs have also been used (Kong et al., 2006). The draft genome sequence of *Cucumis sativus* var. *sativus* was recently published. The availability of this sequence will facilitate the high-throughput discovery of new markers, such as SNPs (Huang et al., 2009).

In general, molecular studies report a low degree of genetic diversity within *C. sativus* var. *sativus* compared to other cross-fertilized species of the genus, such as melons (Dane, 1976, 1983; Esquinas-Alcazar, 1977; Knerr et al., 1989; Dijkhuizen et al., 1996; Horejsi et al., 1999), but they also describe higher levels of polymorphisms in var. *hardwickii* and consistently separate both varieties (Dijkhuizen et al., 1996; Meglic et al., 1996; Horejsi et al., 1999). Indeed, Staub et al. (2005) tried to identify a useful reference marker array (from a set of 155 markers, SSRs and Sequence Characterized Amplified Region (SCARs)) in order to distinguish very closely related varieties and elite breeding lines, which is essential for variety protection. They found difficulties in discriminating this genetic material suggesting that other markers, such as SNPs are needed to better define these cultivars.

### I.2.1.2.3 Molecular diversity of cucumber landraces in the centers of origin and diversity

Several general studies characterized the genetic diversity of the cucumber collection (about 1,000 accessions) maintained at the United States National Plant Germplasm System (NPGS) using isozymes and RAPDs (Meglic et al., 1996; Staub and Ivandic,

2000). Meglic et al. (1996) found that discrimination between the two varieties and within var. *sativus* was possible. Accessions were grouped by continent or sub-continent, and a high level of heterogeneity was detected in accessions belonging to var. *hardwickii*. As expected, Staub and Ivandic (2000) found that NPGS accessions were genetically more diverse than commercial varieties and presented a different genetic structure. Accessions grown commercially exhibited a remarkably narrow genetic base and thus could benefit from the introgression of exotic genes present in the NPGS accessions.

The Asian gene pool is the most variable compared to the other pools reported (the USA, Europe and Africa). Staub et al. (1997a, 1999) and Horejsi and Staub (1999) described genetic diversity in the primary center of origin (India) and secondary center of diversity (China). In these two works the variability of cucumber collections from India and China collected in independent expeditions was assayed. Differences in genetic variability between the different expeditions reinforced the necessity of collecting germplasm from the origin and diversity centers before they disappear. Cases of genetic erosion were detected, as that of landraces of interest from northern Rajasthan (India), which are tolerant to diverse environmental stresses (Staub et al., 1997a). Chinese and Indian accessions were different from each other and from all other groupings, including *hardwickii*. This differentiation was also found in previous studies (Dijkhuizen et al., 1996) and correlates with morphological observations. Differences between these two countries are likely the result of the geographical and political isolation of China. Molecular analysis weakly supports morphological differentiation of cucumber cultigens from northern China (longer and larger fruits, thinner skinned, warted, white-spined, slightly netted, yellow-to-brown epidermis and more resistant to abiotic stresses) and southern China (black-spined, netted and brown skinned). Chinese

cultivars from northern and southern regions are thought to have different origins. Southern germplasm has been much more isolated due to the Himalayas and the social structure, whereas the genetic diversity of northern China has taken advantage of the Silk Road, which introduced types from India and eastern Europe. These studies concluded that China and India represent the most diverse genetic variation in this species.

Despite the fact that Africa is a continent where cucumber is not a major crop, many cucumber landraces are traditionally cultivated on small-farms. Thus, there exists a differentiated African genetic pool that is of interest to analyze and maintain. Knerr et al. (1989) and Meglic et al. (1996) described and analyzed African diversity in NPGS germplasm with isozymes. They found differences among African cultigens and between them and germplasm from other continents. A more recent study with African accessions using RAPDs (Mliki et al., 2003) reports the existence of three groups (one comprising genotypes from Egypt-Ethiopia-Libya, separated from Kenyan and Algerian genotypes and the third with several accessions from Egypt). The first group clearly differentiated from the other, which was close to Chinese accessions, and therefore constitutes a different source of variation. This could be of great interest for future breeding programs, especially considering that Egyptian genotypes are known for their resistance to several pathogens.

Although distinct genepools of cucumber have been reported in the USA, Europe, Africa, and Asia, the genetic diversity within var. *sativus* is limited. The wild relative var. *hardwikii* is an underexploited resource. Until the collections performed by Bisht et al. in 2004 var. *hardwickii* was poorly represented in genebanks. These authors reported the collection and characterization of *C. sativus* var. *hardwickii* from different regions of India: north-western Himalayas (widely and abundantly distributed), the Western

Ghats (fairly distributed), Eastern Ghats, Chhota Nagpur plateau and the central plateau region at elevations from 800 to 1,700 m.s.a.l (sporadic distribution). Some of the accessions collected were morphologically and molecularly characterized, using RAPDs. A high level of diversity was found, but grouping did not correlate to the geographical origin. This study reported the existence of segregating populations and hybrids between var. *sativus* and var. *hardwickii* in natural habitats, showing intermediate traits (plant vigor, fruit skin color, fruit weight and fruit number) that are useful for the improvement of commercial cucumbers.

### I.2.1.3 Other *Cucumis* spp.

### I.2.1.3.1 *C. metuliferus*

*C. metuliferus* E. Meyer ex Naudin, known as the African horned cucumber, is mainly used for human consumption in Africa, where it originated. However, most studies carried out on this species only have evaluated it for breeding cucumber and melon because of its resistance to many important pests such as root-knot nematode, powdery mildew and downy mildew. Different crosses have been attempted to transfer resistances to *C. sativus* (Nikolova et al., 2002; Walters and Wehner, 2002) and *C. melo* (Beharav and Cohen, 1995), even using biotechnological tools like *in vitro* culture (Tabei, 1997).

Evaluation and characterization of some *C. metuliferus* accessions has been carried out to select for higher yield, disease resistance and plant vigor (Marsh, 1993) and also for fruit characteristics (size, color, firmness, moisture content and chemical composition of protein, lipids, sugars, fibre, organic acids and mineral elements) (Romero-Rodriguez et al.,1992; Krauze-Baranowska and Cisowski, 2001).

Helm and Hemleben (1997) studied relationships among some cucurbits analyzing

satellite DNAs. A new satellite from *C. metuliferus* was compared to satellites from other species of the genus, having found the most similarity with some satellite types from cucumber and melon. Results suggested that an ancestral satellite type existed in a progenitor of all *Cucumis* spp. being modified during evolution. The pattern of satellite distribution was in agreement with the taxonomy.

### I.2.1.3.2 *C. anguria*

*C. anguria* L., commonly called West Indian gherkin, has been considered to have originated in America, but now it is supposed to come from tropical Africa, having been introduced into the New World by African slaves (reviewed by Baird and Thieret, 1988). This species has been screened for resistances to several diseases, like powdery and downy mildew (Nikolova et al., 2002) and increased yield (Oliveira et al., 2009). Crosses between *C. anguria* and *C. anguria* var. *longaculeatus* have been attempted to obtain elite lines (Modolo and Costa, 2003). Some phylogenetic studies on *Cucurbitaceae* have also included this species, as well as *C. myriocarpus* Naudin, to better represent genus *Cucumis*, but few accessions have been assayed (Renner et al., 2007; Singh and Matta, 2008). *C. myriocarpus*, which has toxic fruits, has also been investigated for resistances, having been reported as resistant to CVYV (Marco et al., 2003), and adequate mineral ratios (Flyman and Afolayan, 2007).

### I.2.2 Genus *Citrullus*

The genus comprises four diploid ($2n = 2x = 22$) species with variable degrees of cross-compatibility. *C. lanatus* ((Thunb.) Matsum & Nakai), divided into var. *citroides* (citron) and var. *lanatus* (watermelon crop), and *C. rehmii* De Winter, newly discovered by De Winter (1990), are annuals. The remaining two are perennial: *C. colocynthis* L.

Schrad. (colocynth or bitter apple) and *C. ecirrhosus* Cogn. (Robinson and Decker-Walters, 1997; Wehner, 2008). The geographical distribution of these species is shown in Figure I-2.

Several works have established the phylogenetic relationships among *Citrullus* spp. and related species. Results obtained with seed proteins and allozymes were consistent with the characteristics of seed coat, grouped separately from those previously classified as *Citrullus* spp. *Acanthosicyos naudinianus*, from southern Africa, and *Praecitrullus fistulosus*, from India and Pakistan (Navot and Zamir, 1987). Other studies, using ISSRs and RAPDs, suggested that *P. fistulosus* is a distant relative of *Cucumis* and *Citrullus*, which is in agreement with the crossability barriers reported between *C. lanatus* and *C. colocynthis* with *P. fistulosus*, as no viable seeds have been achieved from the attempted crosses, thus making the transference of genes of interest from these species to cultivated watermelon difficult (Levi et al., 2005).

Within *Citrullus* spp., isozyme studies indicate that *C. ecirrhosus* is more closely related to *C. lanatus* than to *C. colocynthis*. Similar results were reported using nuclear ITS, also placing *C. rehmii* closer to cultivated watermelon than *C. colocynthis* and *C. ecirrhosus* (Jarret and Newman, 2000). Crossability and morphology traits, like foliage shape, also support these findings. Genetic similarity between annual (*C. lanatus* and *C. rehmii*) and between the perennial species (*C. ecirrhosus* and *C. colocynthis*) supports the observation of Jobst et al. (1998) regarding the derivation of annual forms from perennial ones. Dane (2002) and Dane et al. (2004) and Levi and Thomas (2005), using PCR-RFLPs, studied the haplotypes of chloroplast DNA in *Citrullus* species and close relatives. Most of the cpDNA regions studied were invariant because the substitution rate found in this genus is low compared to other species (Parducci and Szmidt, 1999; Mohanty et al., 2001; Davis et al., 2002; Yang et al., 2002). In fact, similar analyses

with mithocondrial DNA were abandoned due to lack of variant sites (Demesure et al., 1995). Seven haplotypes within *Citrullus* were identified. *C. lanatus* var. *lanatus* displayed only one haplotype, which suggests the occurrence of genetic bottlenecks during domestication and low levels of outcrossing. Higher variability was found in var. *citroides*. The presence of a similar mutation pattern in citron and in the *C. rehmii* haplotype suggests that this species may be the progenitor of citron melon.

*C. ecirrhosus* had a haplotype quite similar to *C. lanatus*, and might be considered the ancestral species of watermelon (Dane and Liu, 2007). *C. colocynthis* was the most variable species according to Dane and Lang (2004), showing five haplotypes associated with different geographic origins (I: Mediterranean area, Morocco and Cyprus; II: Chad; III: Pakistan; IV: Afghanistan; and V: Ethiopia), supporting previous reports which described different races in this species (Yanev et al., 1999). Similar results were obtained sequencing coding and non-coding cpDNA regions (Dane and Lang, 2004).

*C. lanatus* var. *lanatus*, *C. ecirrhosus* and *C. rehmii* lacked molecular variability and within the genus *Citrullus* two major clades separated *C. colocynthis* from the other three species. In the latter clade, *C. ecirrhosus* and *C. lanatus* appeared closer. The lack of strong crossing barriers between *C. colocynthis* and watermelon (despite the wide genetic distance) make the gene flow possible in order to enhance genetic diversity in this crop, marking this species as the main source of genes of interest together with citron types. Maggs-Kolling et al. (2000) analyzed chloroplast haplotypes in wild and landraces of *Citrullus* spp. of Namibia. One accession from Swaziland and another from South Africa are supposed to possess the ancestral chlorotype, whereas accessions from South Africa, Botswana and Namibia present the most recent chlorotype. This supports the Kalahari Desert as the area of origin of var. *citroides*. The divergence found between

var. *lanatus* and var. *citroides* makes one suppose that they evolved independently from a common ancestor, possibly *C. ecirrhosus*.

### I.2.2.1.*Citrullus lanatus*

### I.2.2.1.1 Origin and taxonomy

*C. lanatus* ((Thunb.) Matsum & Nakai) includes wild, cultivated and feral forms. Fursa (1972) described three subspecies: ssp. *vulgaris* (divided into var. *vulgaris* and var. *cordophanus*, including red sweet fruited cultivated forms), ssp. *lanatus* (including tsamma types from the Kalahari Desert (var. *caffer*), and the citron) and ssp. *mucosospermus* (including the egusi types from West Africa). Recently, the species has been reclassified and divided into two botanical varieties: var. *lanatus* (the cultivated forms, including egusi types), distributed in tropical and subtropical regions worldwide, and var. *citroides* (Bailey) Mansf. (including citron and tsamma types), which grows in southern Africa.

Despite the numerous studies, the origin, distribution and domestication of this species still remains unclear. Jeffrey (1967, 2001) and Zeven and Zhukovsky (1975) proposed var. *caffer* as the ancestor of the species while Navot and Zamir (1987) proposed the var. *citroides*. Other theories propose *C. colocynthis* or, as previously noted, *C. ecirrhosus* (Dane and Liu, 2007). Primitive watermelons are supposed to have had nonsweet, bitter, white-fleshed fruits, similar to those of citron or colocynth. Today it is generally accepted that watermelon originated in Africa where it reaches maximum diversity (DeCandolle, 1883). The two putative ancestors, var. *citroides* and *C. colocynthis,* can be found growing wild in Africa.

Different regions in Africa have been postulated as centers of origin of this species: southern Africa, principally around the Kalahari Desert (Meeuse, 1962; Esquinas-

Alcázar and Gulick, 1983), Central Africa (Mallick and Masui, 1986) and northern Africa (Keay and Hepper, 1985). In fact, watermelon seeds (about 5,000 years old) recently discovered at an archaeological site in southwest Libya (Wasylikowa and van der Veen, 2004) support northern Africa as the most probable domestication center. There is evidence of watermelon cultivation in the Nile Valley by 2000 BC, when southwest Africans did not yet practice farming (Zohary and Hopf, 2000). Colocynth seeds have been found at early sites in Egypt, Libya and the Near East, indicating that they could have been used first. Some authors assume that cultivation of watermelon began in ancient Egypt and India, from where it spread to the Mediterranean area, Near East and Asia. The Romans introduced this crop to Europe, and later the Muslims increased the number of varieties on the continent. However, watermelon did not become as popular there as it did in China where it arrived in the 10th–12th centuries. Subsequently, watermelon reached America (17th century) (Rubatzky, 2001; Wehner, 2008). Today, southern Africa and, to some extent, western Africa, are considered primary centers of diversity. China constitutes a secondary center of diversity, whereas a great variety of landraces and wild accessions can also be found in other regions; India, where *C. colocynthis* grows wild, the Middle East and countries in the Mediterrranean basin.

**I.2.2.1 Molecular markers used in genetic diversity studies**

Cultivated watermelon (var. *sativus*) is morphologically highly variable, mainly for fruit traits, size, shapes, flesh and rind colors and patterns (Ellul et al., 2007; Wehner, 2008). A very characteristic type is the well-known African Egusi, commonly known in Nigeria and the Congo as wild watermelon, whose fruits present bitterness and firm white flesh. Its seeds are coated by an adherent layer of tissues. This type is used for its

nutritional seeds and for cattle consumption.

Despite this morphological variability, molecular variation is limited in commercial cultivars (Jarret et al., 1997; Maggs-Kolling et al., 2000; Levi et al., 2004; Levi and Thomas, 2005). This poor variability can be increased using germplasm from the diversification areas. Germplasm of *Citrullus* from China and South Africa is represented in genebank collections, but India, south, southwest and tropical Africa and the southern areas of the former URSS and Iran are still priorities for collection (Wehner, 2008).

Molecular markers have been used to estimate genetic relatedness of watermelon cultivars, and can be used to evaluate inbred lines for purity. Studies with isozymes (Zamir et al., 1984; Navot and Zamir, 1987; Biles et al., 1989; Walters et al., 1991) reflect a low degree of genetic diversity, and only a few informative isozyme markers are available. Diversity within *C. lanatus* (20 cultivated, 70 citron watermelon and reference types) has been recently assessed by Dane and Liu (2007) by using PCR-RFLPs and sequencing of cpDNA of several non-coding regions. RAPD markers were more efficient at detecting genetic variation (Hashizume et al., 1993; Zhang et al., 1994). Lee et al. (1996) used RAPDs with a representative collection of watermelon cultivars, obtaining four groups. Molecular clusters correlated with fruit quality traits such as sugar content. Moreover, Levi et al. (2001b) used this marker system to characterize a collection of *C. lanatus* and *C. colocynthis* exhibiting several disease resistances. They found three clusters, one with watermelon cultivars, one with the *C. lanatus* var. *citroides* accessions, and the third with *C. colocynthis* accessions. Levi and Thomas (2005) performed their study using citoplasmatic markers, with which they also obtained a clear differentiation of the accessions belonging to var. *lanatus* (the five cultivars assayed), var. *citroides* and *C. colocynthis* as previously reported. They found

a closer relationship between *citroides* and colocynth vs *lanatus* and colocynth. Levi et al. (2001a) found higher levels of variability in *C. colocynthis* and var. *citroides* than in cultivated watermelon and differentiated some watermelon accessions with introgressions from var. *citroides*. Similar studies have been performed with ISSRs and AFLPs (Levi et al., 2004). This study showed that AFLPs and ISSRs are highly informative and much more efficient at differentiating between American heirloom cultivars with a narrow genetic base. AFLPs were also successfully used to detect variability among watermelon cultivars (Ke-peng et al., 2003). The polymorphism rate detected with this kind of markers proved higher than with other previously tested kinds, such as isozymes, and three groups were obtained among the 30 genotypes surveyed. Although low genetic diversity was found, classical American ecotypes, breeding and selected lines and cultivars originating from Japanese and Chinese pedigrees could be differentiated.

As in other cucurbits, both genomic SSRs and EST-SSRs have been identified and used in genetic diversity studies (Jarret et al., 1997; Guerra-Sanz, 2002; Joobeur et al., 2006; Verma and Arya, 2008). Jarret et al. (1997) differentiated *C. lanatus* var. *lanatus* including egusi types, from wild and cultivated citron (var. *citroides*) and *C. colocynthis*. The fact that most of the alleles in var. *lanatus* are common in var. *citroides* supports the hypothesis of citron as the wild progenitor. Joobeur et al. (2006) developed genomic SSRs from a BAC library constructed for watermelon finding that 95% were polymorphic, whereas Verma and Arya (2008) tested EST-SSRs in a set of seven Indian genotypes showing 22% polymorphism. They found high transferability to other genera of the family.

**I.2.2.2 Molecular diversity of watermelon in secondary centers of diversity**

The germplasm of secondary centers like northeast Brazil is also of interest. Romao (2000) proposed that watermelon was introduced to this country by African slaves about 300 years ago. A high level of diversity has been reported in this area. The locally denominated "melancia de cavalo" had been classified as *C. colocynthis*, but this author proposed adscribing it to *C. lanatus* var. *citroides*. Similar morphological studies with commercial types and landraces in Kenya have been carried out and have found low diversity within the commercial ones (Gichimu et al., 2009) in comparison with wild material and landraces. Namibia wild types and local landraces have also been morphologically compared to commercial cultivars (Maggs-Kölling et al., 2000). Indigenous classification into seed, cooking and fresh-eating types was coherent with the clustering obtained, while commercial types grouped separately. Diversity studies within this species in countries like Turkey and Korea also demonstrated that high diversity exists there for morphological traits despite their not being centers of origin (Huh et al., 2008). Korean and Turkish accessions were easily separated. Korean landraces were divided into two groups, whereas a continuous variation was found in Turkish germplasm.

**I.2.3 Genus *Benincasa***

*Benincasa hispida* (Thunb.) Cogn., known as ash gourd, white gourd, wax gourd or white pumpkin, is the most important species of the genus. In addition to their immature fruits, their young leaves and shoots are consumed. It is considered to be one of the most polymorhic crops with regard to fruit and certain phenotypic traits (Parkash et al., 2000; Singh, 2002), but few studies have been carried out in spite of its high nutritional value and traditional medicinal properties. Its center of diversity is the Indo-China

region (Rubatazky and Yamaguchi, 1999), where the crop is very popular.

RAPD and ISSR markers have been used to identify cultivars and hybrids (Meng et al., 1996), to analyze genetic diversity (Sureja et al., 2006) and to predict and develop hybrids with the highest heterosis value (Verma et al., 2007). Morphological markers were used before to assess relationships among *B. hispida* genotypes. Pandey et al. (2008), in a recent study using RAPDs, concluded that accessions from northeastern India are quite different from the other regions of the country.

### I.2.4 Genus *Lagenaria*

The most important species is *Lagenaria siceraria* (Mol.) Stand., better known as bottle gourd or calabash. It is an African crop grown for its fruit, which can either be harvested young and used as a vegetable or harvested mature, dried, and used as a bottle, utensil or pipe. Japan, China and India are great consumers of this fruit. Despite the lack of any early remains in Africa, this continent is believed to be the origin of the species (Whitaker, 1971). Wild *Lagenaria* species are distributed in northern Africa. Morphological analyses and archaeological data suggest that this species dispersed across the ocean from Africa to Asia and America, where different domestication events took place. Analyses using RAPD markers have been performed to clarify its evolutionary history (Decker-Walters et al., 2001). Landraces, cultivar accessions and a wild relative *L. sphaerica* were examined, revealing that southern African germplasm is a divergent lineage from which some cultivars have derived, and that American germplasm is distinct, but with an African origin. According to their results, landraces from New Guinea were not related to American germplasm as previously supposed, and commercial cultivars have very different origins and genetic backgrounds. Subsequent studies of diversity in this species have been carried out, especially in countries where

this crop is more important. For instance, Kenyan landraces of *L. siceraria* and wild relatives *L. sphaerica*, *L. abyssinica* and *L. breviflora* were morphologically characterized by Morimoto et al. (2005), and it was found that more diversity exists within *L. siceraria* than within the wild relatives supposedly due to human selection. In order to have material resistant to diseases and pests, such as root-knot nematodes, whiteflies, ZYMV and powdery mildew, recent researchers have examined *L. siceraria* (Levi et al., 2009) and its relationship to other cucurbits. Moreover, variability was detected within species with the accessions clustering into two major groups. One cluster included accessions collected mostly in India and a few collected in the Mediterranean region and in Northeast Africa. The second cluster included accessions from southern Africa, North, Central and South America, China, Indonesia and Cyprus.

**I.3 Tribe *Cucurbiteae***

**I.3.1 Genus *Cucurbita***

The genus *Cucurbita* comprises 20 species ($2n = 2x = 40$) according to the most recent classification of the family (Jeffrey, 2005). The most economically important are *C. pepo* L., *C. moschata* Duchesne and *C. maxima* Duchesne. *C.argyrosperma* C. Huber and *C. ficifolia* Bouché are also grown, but they have a narrower distribution. Based on ecological adaptation, *Cucurbita* species can be divided into two groups: the mesophytic annuals or short-lived perennials with fibrous root systems, which includes the five cultivated species mentioned above, and the xerophytic long-lived perennials with fleshy storage roots. Among the latter group, *C. foetidissima* Kunth (buffalo gourd) shows characteristics of interest to be domesticated (De Veaux and Schultz, 1985). Geographical distribution of the main *Cucurbita* spp. is shown in Figure I-3.

The genus *Cucurbita* is native to America, from where the domesticated species spread

worldwide. Each cultivated species is thought to have been domesticated independently from the others in distinct regions of the continent, but all in the pre-Columbian era. The centers of origin for these species are: *C. maxima* in southern South America, *C. moschata* in the lowlands of southern Central America or northern South America, *C. pepo* in northern Mexico or south-central USA, *C. ficifolia* probably in the northern or Central South American highlands and *C. argyrosperma* in southern Mexico.

A numerical study of taxonomic relationships in *Cucurbita* was carried out by Bemis et al. (1970) using a set of 160 characters. They found five groups of *Cucurbita* spp., one comprised species indigenous to arid regions of the southwestern USA and northwestern Mexico (*C. digitata, C. palmata, C. californica, C. cylindrata* and *C. cordata*) and the second comprised tropical species (*C. okeechobeensis, C. martinezii* and *C. ludelliana*); the third grouped mesophytic species (*C. sororia, C. gracilior, C. palmeri*) and the cultivated *C. argyrosperma*; the fourth was constituted by *C. maxima* and *C. andreana*; and the fifth by *C. pepo* and *C. texana*. *C. moschata, C. ficifolia, C. pedatifolia, C. foetidissima* and *C. ecuadorensis* did not join other species in clusters.

Molecular systematic work on the genus *Cucurbita* has also been carried out by various investigators by using both nuclear and cytoplasmic markers. Most of these studies mainly focus on the cultivated species and their wild relatives, many suggesting that the crop plants are not derived from a common ancestor. Decker-Walters et al. (1990), using isozymes, support a common ancestor for *C. moschata* and *C. pepo* that is not shared by *C. maxima*. Wilson et al. (1992), using cpRFLPs, separated *C. ficifolia* from the other cultivated species and considered it basal to the mesophytic species in the genus. Jobst et al. (1998) also indicate a polyphyletic origin of *Cucurbita* on the basis of nuclear ITS sequences. They found extensive allele sharing among these species, which led to an inconclusive phylogenetic analysis, suggesting a high frequency of

introgression during domestication or polyploidization events in the genus. Analysis using ISSRs and microsatellite DNAs has also been reported (King et al., 1995; Katzir et al., 2000a, b). One of the most relevant studies has been that by Sanjur et al. (2002), using mtDNA, which suggests six independent domestication events from distinct wild ancestors.

### I.3.1.1 *Cucurbita pepo*

### I.3.1.1.1 Origin and taxonomy

*C. pepo* is native to North America (Trumbull, 1876; Erwin, 1931; Whitaker, 1947). It grows wild in northeastern Mexico and the southern, southeastern and Central USA (Nee, 1990). Its wild range most likely extends to the northeastern USA (Petersen and Sidell, 1996). In northern Mexico and the eastern USA there is evidence of the presence of this primitive crop from over 4,000 years ago. Moreover, recent studies based on re-examination of early domesticate assemblages from caves in Mexico indicate that this species is the earliest domesticate (~8000 BC) in Meso-America (Smith, 1997, 2005). Taxonomically this species is divided into three subspecies according to Decker-Walters et al. (2002b): ssp. *pepo*, ssp. *ovifera* (L.) D.S. Decker, which comprises var. *ovifera* (L.) D.S. Decker, var. *orkazana* D.S. Decker-Walters and var. *texana* (Scheele) D.S. Decker, and ssp. *fraterna* (L.H. Bailey) Andres. The division of the cultivars of *C. pepo* into two distinct major lineages, ssp. *pepo* and ssp. *ovifera,* is supported by many studies based on allozymatic, cpDNA, mtDNA and nuclear markers (Wilson et al., 1992; Decker-walters et al., 1993; Katzir et al., 2000a, b; Ferriol and Picó, 2008; Paris, 2008). Most authors have also supported the theory of two independent domestication events for these two subspecies (Decker, 1985; Kirkpatrick and Wilson, 1988; Wilson et al., 1992; Decker-Walters et al., 1993, 2002b; Sanjur et al., 2002). Different wild

types have been proposed as the ancestor of cultivated *C. pepo*. Var. *texana* might be the ancestor of ssp. *ovifera* domesticated types. Both species are cross-compatible and share some morphological similarity. However, allozymic patterns differed in several studies and the possibility of it being a feral form instead of the ancestor has been also suggested (Kirkpatrick et al., 1985; Decker, 1988; Kirkpatrick and Wilson, 1988; Nee, 1990). The precursor might be an individual similar to var. *texana* that has been wiped out.

The wild ssp. *fraterna* has also been suggested as the ancestor of ssp. *ovifera*. Accessions of the ssp. *fraterna* and var. *texana* were shown by Wilson et al. (1992) to be closely aligned with cultivars of ssp. *ovifera*, but no wild taxon studied was considered a likely progenitor for ssp. *pepo*. In addition, in Katzir et al. (2000a), with ISSRs and SSRs, var. *fraterna* clustered with the specimens representing ssp. *ovifera*. Var. *orkazana* is another wild form related to the ssp. *ovifera* cultivars and has also been reported to be the ancestor of var. *ovifera* after RAPD analysis (Decker-Walters et al., 2002b). These three extant wild forms in this species are thought to have differentiated from each other before domestication (Decker-Walters et al., 1993), although they grouped together based on mtDNA comparisons (Sanjur et al., 2002). According to their results, ssp. *fraterna* is the most probable predecessor due to its higher genetic affinity.

Regarding the ancestor of ssp. *pepo*, some researchers think that it is extinct while others maintain that it is unknown (Wilson et al., 1992; Decker-Walters et al., 2002b). Sanjur et al. (2002) suggested that var. *fraterna* was distributed in the past in small and nearly isolated populations, genetically diverging. This situation explains that one of these populations, to date uncollected, was the ancestor of ssp. *pepo*. Paris et al. (2003) found that the accession Miniature Ball, previously placed within the orange gourd group of ssp. *pepo*, possessed wild-type characteristics and yet shows genetic affinity to

a wide range of domesticated *C. pepo*, suggesting that it may represent the wild ancestor preserved in cultivation. Teppner (2004) reported a new subspecies, *C. pepo* ssp. *gumala* Teppner, which is cultivated in Guatemala and Mexico. This subspecies might be the ancestor of ssp. *pepo* according to this author.

### I.3.1.1.2 Phenotypic diversity studies

*C. pepo* is likely the most polymorphic species with regard to fruit traits (Naudin, 1856). This species, along with other *Cucurbita* spp., has suffered great diversification in America, Europe and Asia after Columbus arrived to the Americas (Decker, 1988). Edible cultivated types of this species have been traditionally grouped in eight morphotypes according to fruit shape (Paris, 1986, 1989, 2001b, 2008): Pumpkin, Vegetable marrow, Cocozelle and Zucchini which belong to ssp. *pepo* and Scallop, Acorn, Crookneck and Straightneck, which belong to ssp. o*vifera* (Table I-2). Most of them are usually employed as summer squashes, except for Pumpkin and Acorn. The subspecies *pepo* includes ornamental types with orange, round and smooth or warty fruits and ssp. *ovifera* var. *ovifera* also includes ornamental types with oviform and pyriform fruits. Many studies report the morphological variation of the species mainly for traits of agronomic interest (quality traits of fruits and seeds, resistance to pests and diseases). A greater variation in ssp. *pepo* than in ssp. *ovifera* is often reported (Paris, 1998, 2001a; Paris and Nerson, 1998; Lebeda et al., 1999; Danilchenko et al., 2000; Kristkova and Lebeda, 2000; Nerson et al., 2000; Younis et al., 2000).

### I.3.1.1.3 Molecular markers used in genetic diversity studies

Apart from phenotypic variation, molecular variation within the species has been assessed with different markers. Isozymes, RFLPs, RAPDs, AFLPs, SCARs, Sequence-

Related Amplified Polymorphisms (SRAPs), ISSRs and SSRs have been used to date to evaluate genetic diversity within *C. pepo* (Ignart and Weeden, 1984; Decker, 1985; Kirkpatrick et al., 1985; Torres Ruiz and Hemleben, 1991; Lebeda et al., 1999; Stachel et al., 1998; Baranek et al., 2000; Decker-Walters et al., 2002b; Heikal et al., 2008). Most of the studies report a high variability within the species. When different species are compared, allelic diversity is greatest in *C. pepo* and *C. moschata*. Katzir et al. (2000a), using SSRs and ISSRs grouped the cultigens of *C. pepo* ssp. *ovifera* more-or-less according to fruit shape, whereas in *C. pepo* ssp. *pepo*, subclustering differentiated the cocozelle group from the Zucchini group. Some of the SSR *loci* used for *Cucurbita* analysis were transferred from C*ucumis,* as only a few microsatellites were available for *Cucurbita*. Paris et al. (2003) attempted to apply cucumber and melon SSRs described by Katzir et al. (1996) and Danin-Poleg et al. (2001) attempted to assess diversity in *C. pepo*. A set of 102 *Cucumis*-SSR primers developed by Fazio et al. (2002) were also proved in this species for mapping, but none turned out to be polymorphic. Sixty *Cucumis* SSRs (gSSRs and EST-SSRs) were tested in *Cucurbita* spp. accessions, but over 63% did not amplify in any of them (Picó et al., 2005-2006). Transferability from *Cucumis* EST-SSRs has also been assayed (Fernández-Silva et al., 2008), but the fact that these markers are located in expressed regions of the genome did not increase the transferability rate and only 5.4% were polymorphic. Therefore, SSRs are not easily transferable between genera within *Cucurbitaceae,* and because of this, a wide collection of SSRs has been developed recently using SSR-enriched partial genomic libraries from *C. pepo* ssp. *pepo* and *C. moschata*. These markers show high inter-species transferability and have been used to construct the first published *C. pepo* map (Zraidi et al., 2007; Gong et al., 2008a, b). Stift et al. (2004) had already reported a better transferability rate from *C. pepo*-SSRs to *C. moschata*, *C. maxima* and *C.*

*ecuadorensis* and Gong et al. (2008a) found with their new set of 500 SSRs (from SSR-enriched partial genomic libraries) a higher percentage of *C. pepo* markers transferred to *C. ecuadorensis* in comparison to *C. moschata*, which implies a closer relatedness between *C. pepo* and *C. ecuadorensis* than between *C. moschata* and *C. ecuadorensis*. Two of the most complete studies performed to date in *C. pepo* are those by Paris et al. (2003) and Ferriol et al. (2003a). Paris et al. (2003) assayed a set of *C. pepo* accessions belonging to the three subspecies with three different marker systems: AFLPs, ISSRs and SSRs, finding a high correlation between them. Whereas previous studies describe high levels of variation in wild genotypes (Decker-Walters et al., 2002b), Paris et al. (2003) found higher variation among domesticated than among wild populations. In general, their results were coherent with botanical and horticultural classification and with other studies with allozymes (Ignart and Weeden, 1984) and DNA markers (Torres Ruiz and Hemleben, 1991; Katzir et al., 2000a). Clustering agreed with the division into ssp. *pepo*, ssp. *ovifera* (syn. *texana*) and ssp. *fraterna*. Subspecies *fraterna* and *ovifera* appeared more closely related to each other than to ssp. *pepo*. Results show that the cultivar-groups are genetically quite distinct. In fact subclusters within ssp. *ovifera* were in accordance with morphotypes, supporting six groups: Acorn, Crookneck, Scallop, Straightneck, ovifera gourds and wild forms. Scallop group, thought to be among the oldest, was closely related to the *texana* gourds, supporting the theory of its having been developed directly from inedible fruit ancestors by American natives. Straightneck, the most recent group among ssp. *ovifera* (Paris, 2000), was consistently the most dissimilar to *texana* gourds, and little phenotypic variation within the group was observed. Within ssp. *pepo*, Zucchini is the most recent (Paris, 2000) and the most different from the other morphotypes of its subspecies, showing limited variability. Relationships within Pumpkin accessions and between this group and Cocozelle and Vegetable marrow were

not as clearly defined, perhaps due to the large amount of diversity found in these three groups.

Ferriol et al. (2003a), using SRAPs and AFLPs, focused on the diversity of Spanish landraces, with representation from the different morphotypes described in the species as well as several unclassified types. Spain was an important country for *Cucurbita* diffusions and diversification in the Old World as it acted as a bridge between Europe and America. Results were similar to those reported by Paris et al. (2003). Both subspecies were also clearly differentiated, *pepo* being more polymorphic. A more defined sub-clustering according to morphotype within ssp. *ovifera* than wihin ssp. *pepo* was observed with SRAP markers. In the ssp. *ovifera* cluster, Crookneck and Straightneck grouped separately from scallop, whereas Acorn was more dispersed.

### I.3.1.2 *Cucurbita moschata*

### I.3.1.2.1 Origin and taxonomy

Mexico was initially proposed as the domestication center of this species (Cutler and Whitaker, 1967), however, subsequent studies at some archeological sites also placed the oldest *C. moschata* remains in Ecuador (5,170 to 3,780 BC) and on the Central Pacific coast of Panama (5000 BC), while the oldest remains in the southwestern USA dated back 2,300 years (Decker-Walters and Walters, 2000; Piperno et al., 2000). Nowadays, South America is considered to be the domestication center or a secondary center of diversity because some landraces from Colombia, Panama and Bolivia display primitive traits, such as dark seeds, small fruits, bitter flesh, lignified and warty rind and indeterminate growing habits (Nee, 1990; Wessel-Beaver, 2000; Sanjur et al., 2002). Two independent domestication events in Mexico and northern South America have also been proposed (Lira-Saade, 1995; Robinson and Decker-Walters, 1997; Decker-

Walters and Walters, 2000). After its domestication, *C. moschata* spread to the Caribbean islands and subsequently to the rest of the world´s continents where it adapted to different conditions (Piperno et al., 2000).

The wild ancestor of *C. moschata* still remains unknown. *C. lundelliana*, located on the plains of the Yucatan Peninsula, was considered to be the ancestor, but morphological, isozymatic and crossing studies do not support it (Merrick, 1990). According to ecological, morphological and molecular studies using mtDNA, cpDNA and nuclear markers, *C. moschata* and *C. argyrosperma* are closely related (Wilson et al., 1992; Jobst et al., 1998; Sanjur et al., 2002). In fact, they were considered to be a single species until Pangalo (1930) proposed *C. argyrosperma* (syn. *C. mixta*) as a different species. That is the reason why a wild taxon of *C. argyrosperma*, the ssp. *sororia*, has also been suggested as the ancestor of *C. moschata*, although some crossability barriers and different isozymatic patterns exist between the species. Gene flow between *C. argyrosperma* ssp. *sororia* growing wild near milpas, where *C. argyrosperma* ssp. *argyrosperma* and *C. moschata* are cultivated, were reported in Mexico by Montes-Hernández and Eguiarte (2002), which contributes to increasing polymorphism levels. Wild gourds discovered in Bolivia, could be key in the elucidation of the origin of this species (Decker-Walters and Walters, 2000).

### I.3.1.2.2 Phenotypic diversity studies

*C. moschata* is the species of the genus *Cucurbita* for which the most secondary centers of diversity have been described, with a large amount of landraces with variable characteristics developed in different regions. Filov (1966) classified over 20 varieties of *C. moschata* in several geographic subspecies, which revealed the existence of important diversity centers in Colombia, Mexico, Central America, the western USA,

Florida, India and Asia Minor and Japan. Despite the great variability of *C. moschata* landraces, only a few types have been introduced in the commercial circuit, which initially comprised three groups of cultivars: "Cheese" (variable in shape with leather-colored rind), "Crookneck" (round end with a long straight or curved neck) and "Bell" (bell-shaped to nearly cylindrical) (Castetter, 1925; Whitaker and Davis, 1962; Robinson and Decker-Walters, 1997). "Butternut" was the first commercial cultivar, obtained from a Crookneck type in the 1930s (Mutschleer and Pearson, 1987).

Landraces from different centers of diversity have been morphologically and agronomically characterized. Chung et al. (1998) assessed accessions from Korea, where flattened and round fruited types predominated. Wessel-Beaver (2000) observed a high frequency of primitive traits in Colombian landraces and Ríos et al. (1997) characterized Cuban landraces according to their morphology and yield without pesticides and fertilizers. A great morphological diversity in Spanish landraces has also been reported (Esteras et al., 2008). Cultivars exhibiting bushy growth habit have been studied too (Carle et al., 2000; Wu et al., 2007). This interesting agronomic trait (bushy-growth) already known in modern cultivars of *C. pepo* and *C. maxima*, has recently been evaluated in *C. moschata*, leading to the discovery of the gene responsible for the inhibition of cell elongation. Some studies also describe variation in quality traits such as the great variability in β-carotene content observed in landraces from Zambia (Gwanama et al., 2002). Loche types, a highly valued landrace from Peru, are also being characterized due to their increasing interest in typical gastronomy. Variability for herbicide and pest and disease resistance has also been assessed (Poe et al., 1988; Wessel-Beaver, 1993; Maynard, 2001).

**I.3.1.2.3 Molecular markers used in genetic diversity studies**

Some of the molecular studies performed with *C. moschata* tried to establish relationships to other *Cucurbita* spp. Isozyme analysis (Decker-Walters et al., 1990; Puchalski and Robinson, 1990) and comparisons of chloroplastic, mithocondrial and nuclear ribosomal DNA (Wilson et al., 1992; Jobst et al., 1998; Sanjur et al., 2002) established the closer relatedness of this species with *C. argyrosperma*. Great diversity within *C. moschata*, higher than in *C. maxima* and similar to or higher than in *C. pepo*, has been reported using allozymes, RAPDs and chloroplastic and nuclear ribosomal DNA comparisons (Decker-Walters et al., 1990; Wilson et al., 1992, Jeon et al., 1994; Jobst et al., 1998; Baranek et al., 2000).

Only a few studies have assessed infraspecific variation in this species, most with dominant markers that do not need previous knowledge of sequences. Youn and Chung (1998) and Gwanama et al. (2000) analyzed genetic diversity in landraces from Africa and South Korea, revealing groups consistent with agroclimatic origin and not with morphological traits. One of the most remarkable works on *C. moschata* is that of Ferriol et al. (2004a) using AFLPs and SRAPs, which studied both morphological and genetic diversity within the species in a germplasm collection of Spanish landraces and some American accessions. The variability found was comparable to that reported for some secondary centers of diversity like Korea, indicating that Spain maintains a great amount of diversity. In fact, many landraces could not be classified in any of the horticultural groups of commercial importance. A clear grouping according to geographical origin was observed. A clear separation between the Spanish, Central American and South American accessions was detected, the latter showing some primitive traits. *C. moschata* landraces from the Canary Islands differed molecularly from those of the peninsula, suggesting different germplasm introductions from

America or earlier/better adaptation of some types to the more tropical climate of the islands (Ferriol et al., 2005).

AFLPs have also been used to study the genetic diversity of *C. moschata* from Brazil for the establishment of a core collection (Ramos, 2007). Until recently few SSR markers have been developed for *C. moschata*. In this genus, most molecular tools had been generated mainly for *C. pepo* or transferred from more important species in the family such as *Cucumis melo*. The inter-genus transferability was quite low, both for genomic and EST-SSRs (Picó et al., 2005-2006; Watcharawongpaiboon and Chunwongse, 2007; Fernández-Silva et al., 2008). Recently, Gong et al. (2008a) developed a set of 500 SSRs from SSR-enriched partial genomic libraries of *C. pepo* and *C. moschata*, reporting a high interspecific transferability between both species. These newly developed SSRs have been used to construct the first SSR-based map for *C. moschata* (Gong et al., 2008b). A high level of macrosynteny was found comparing both *C. pepo* and *C. moschata* maps, revealing that the transferability of markers may be more easily accomplished.

### I.3.1.3 *Cucurbita maxima*

### I.3.1.3.1 Origin and taxonomy

The oldest archaeological remains that demonstrate *C. maxima* Duchesne domestication dated back to 2,000 BC and were found in coastal Peru. In pre-Columbian times, primitive forms of *C. maxima* were already cultivated in northeastern Argentina and Paraguay by the Guarani Indians as well as in the Andean valleys (Ferriol, 2003). Oliszeweski (2005) reported northern Argentinian remains dating to between 200 and 500 AD. Nowadays, it is generally accepted that *C. maxima* ssp. *andreana* is the wild ancestor of the cultivated forms of the species. Subspecies *andreana* is endemic to

South America (Argentina, Uruguay, Bolivia and possibly Paraguay), growing wild in temperate regions of Argentina and the plains of Bolivia, so these areas can be included in the domestication center (Sanjur et al., 2002). Initially, *C. ecuadorensis*, whose habitat is located in the coastal region of Ecuador, was also thought to be the ancestor (Lira-Saade, 1995).

*C. maxima* is divided into two subspecies: *C. maxima* Duch. ssp. *maxima*, which includes the cultivated and ornamental types, and *C. maxima* Duch. ssp. *andreana* (Naud) Filov, which only includes wild forms. Hybridization has been reported between cultivated types and ssp. *andreana*, which has contributed to increasing the genetic variation (Decker-Walters and Walters, 2000).

Castetter (1925) classified the cultivated *C. maxima* types in six horticultural groups (banana, delicious, hubbard, marrow, show, turban), which have persisted until the present day (Whitaker and Davis, 1962; Decker-Walters and Walters, 2000). However, many local cultivars and landraces present different characteristics and can not be included in this classification. A few studies have focused on these traditional cultivars, adapted to very different conditions since this species spread worldwide from America. Spain acted as a bridge between America and Europe after the discovery of the continent, but other types first arrived in Australia, Africa and Asia where they diversified (secondary centers) and were later exported to Europe. However, it seems that commercial types were selected in the USA from materials collected in South America.

### I.3.1.3.2 Phenotypic diversity studies and genetic diversity studies with molecular markers

Most of the studies on genetic diversity in *C. maxima* aimed at establishing genetic

relationships with other *Cucurbita* species. Bemis et al. (1970) reported the closer relationship between this species and the previously denominated *C. andreana*, now ssp. *andreana*, employing biological, geographical and ecological data. Similar results were obtained by means of nucleic acid hybridization, isozyme and mitochondrial (Sanjur et al., 2002) and chloroplast DNA (Wilson et al., 1992) comparison. Moreover, the genetic similarity between semi domesticated *C. ecuadorensis* and *C. maxima* has been established (Goldberg et al., 1972; Puchalski and Robinson, 1990; Wilson et al., 1992; Sanjur et al., 2002).

Other studies carried out in *C. maxima* include, like in the other *Cucurbita* spp., seed characterization (Joshi et al., 1993) and disease and herbicide resistance assessment (Poe et al., 1988; Keinath and DuBose, 2000; Kristkova and Lebeda, 2000).

Some molecular studies have been performed with enzymes, finding no correspondence between morphological and molecular data (Decker-Walters et al., 1990; Júnior, 1999). Ferriol et al. (2003b, 2004b) employed AFLPs, RAPDs and SBAPs to assess genetic diversity of Spanish landraces. SBAPs markers grouped the accessions according to the type of use: human consumption, cattle consumption and ornamental, which are related to morphological and agronomical traits, while AFLPs grouped based on geographical origin, indicating much more variability among the included American landraces than among the Spanish ones, and supporting the idea of a genetic bottleneck during the introduction into Europe. A recent study about diversity of *C. maxima* in its area of origin (Esteras et al., 2009c) placed the representative accessions of Spanish variability among the less variable accessions from Peru and Ecuador. In this preliminary study performed with AFLPs, the ssp. *andreana* accessions separated clearly from those of ssp. *maxima*. Some Argentinian genotypes were intermediate between ssp. *andreana* and the remainder of accessions from Bolivia, Ecuador, Paraguay, Peru and Argentina,

which supports Argentina as the center of origin. Apart from Argentina, Ecuador also showed high diversity, indicating the necessity of collecting germplasm in this area; in Bolivia, however, a low level of variation was observed, which might be due to a genetic erosion process.

### I.3.1.4 Other *Cucurbita* spp.

### I.3.1.4.1 *C. ficifolia*

*C. ficifolia* Bouché, also called Malabar melon or Angora squash, is supposed to have been originated and domesticated in the Andean region (Nee, 1990; Sanjur et al., 2002), although a Mesoamerican origin is not ruled out due to existence of several indigenous names. This species is poorly diffused outside the tropics due to its ecological requirements, and because of that this crop presents little diversity. Fruits are very uniform, only varying in color and size.

It is cultivated mainly for self-consumption. This is the reason why only a few studies on this species have been carried out. Due to its cold resistance *C. ficifolia* has been surveyed to improve yield in other crops like cucumber by means of using as rootstock (Zhou et al., 2009). Disease resistances have been described as well. Ivancic et al. (2004) evaluated hybrids *C. ficifolia* x *C. maxima* for morpho-agronomic traits. Moreover, this species partially crosses with other less important cucurbits like *C. lundelliana*, *C. foetidissima* and *C. pedatifolia*. *C. ficifolia* has been included in several phylogenetic studies to elucidate relationships among genera in the family (Kocyan et al., 2007) or among species in the genus. The last one surveyed cultivars from four *Cucurbita* species with AFLPs concluding that *C. ficifolia* and *C. pepo* had a close relationship.

**I.3.1.4.2** *C. argyrosperma*

*C. argyrosperma* C. Huber (syn. *C. mixta* Pang.) was first considered a different species from *C. moschata* by Pangalo (1930). It is believed that its domestication took place in southwestern and Central Mexico, since the oldest archeological remains found in different caves of the region dated back between 3085 and 115 BC (Merrick, 1990; Smith, 2005). The wild ancestor of cultivated forms (ssp. *argyrosperma*) is supposed to be *C. argyrosperma* ssp. *sororia* (L.H. Bailey) Merrick and Bates (Sanjur et al., 2002), which is distributed from Mexico to Central America.

Decker-Walters and Walters (2000) described a low variability within the species with few commercial cultivars existing. Due to the poor quality of the flesh, many cultivars are only grown for their seeds and for cattle consumption. Within ssp. *argyrosperma* (cultivated forms) three varieties may be distinguished: var. *argyrosperma* (the more primitive one, commonly striped with bright color)*,* var. *callicarpa* (the most recent and variable one, comprising most of the commercial cultivars and landraces) and *stenosperma* (striped fruits, mainly used for its edible seeds).

Several studies have dealt with the traditional Mexican agroecosystem called "milpa". Montes-Hernández and Eguiarte (2002) analyzed genetic structure in squashes grown in western Mexico, where *C. moschata* and *C. argyrosperma* ssp. *argyrosperma* are cultivated within maize fields next to wild populations of *C. argyrosperma* ssp. *sororia*. Results with allozymic markers suggested a high exchange of genes among populations even if they were several kilometers apart. This extent of high gene flow has increased the squash genetic variability in this area. However, this diversity is considered to be threatened in the near future due to the lack of young farmers who want to maintain this agrosystem and its landraces (Montes-Hernández et al., 2005). Morphological diversity in this kind of system has also been studied in central-eastern Yucatan (Mexico). Canul

et al. (2005) characterized 36 squash landraces (*C. moschata* and *C. argyrosperma*) using plant, fruit and seed traits. Most of the quantitative variation was explained by seed length, width and weight; days to female flowering; fleshy thickness and fruit length and width; while qualitative variation was mainly explained by leaf shape, fruit longitudinal shape and color intensity of leaf spots. Both species separated clearly in the Principal Component Analysis (PCA), *C. moschata* being more variable.

### I.4 Tribe *Joliffieae*

The most important species in this genus is *Momordica charantia* L., known also as bitter melon or bitter gourd, which is widely cultivated in India, China, Malaysia, Africa and South America. This crop possesses comparatively high concentrations of ascorbic acid and iron and is used as a traditional medicine as well. In fact, studies like the one by Krawinkel and Keding (2006) report a high content in phenolics, polyphenolic compounds and natural oxidants and antioxidants in some Indian varieties.

This country possesses great morphological diversity in sex expression, growth habit, shape and size of the fruit, etc. (Robinson and Decker-Walters, 1999; Behera et al., 2006b) that needs to be evaluated for possible uses in breeding.

Marr et al. (2004) surveyed wild and cultivated populations of this species using isozymes and morphological traits. Comparable phenotypic variation was found in both kinds of accessions although isozyme polymorphisms were higher in the wild ones. However, relatively few polymorphic markers have been identified in bitter melon (Dey et al., 2006). To date, the most complete study is that of by Behera et al. (2008), who assessed the genetic diversity among 38 diverse bitter gourd accessions from different Indian states using RAPD and ISSR markers. The percentage of polymorphism detected was higher for ISSRs and a great difference was observed beween var. *charantia*

(domesticated genotypes) and var. *muricata* (wild, free-living types). A more recent work (Kole et al., 2010) with AFLPs, fruit traits and three healthy compounds (cucurbitacin-C, charantin and plant insulin), reported a high genetic diversity among 22 genotypes from six countries. The knowledge of the phytomedicinal compounds variation in bitter gourd germplasm and the pathways and genes involved, makes this species a model for the new phytomedomics era.

With regard to within-genus studies, the most recent one is a DNA sequence phylogenetic research that includes 58 *Momordica* species (Schaefer and Renner, 2010). They conclude this genus is monophyletic and consists of 11 well-supported clades.

## I.5 Tribe *Luffeae*

The genus *Luffa* is a very popular vegetable in the tropics. Some types are also used to make sponges or to extract fiber. Within this genus two species are cultivated: *L. aegyptiaca* Mill. (syn. *L. cylindrical* M. Roem), known as smooth gourd or sponge gourd, and *L. acutangula* Roxb, known as ridged gourd or fluted loofah.

Three varieties are recognized within *L. acutangula*: var. *acutangula* (the domesticated) and two wild varieties, var. *amara* (located in India) and var. *forskalii* (Harms) Heiser & Schilling (located in Yemen) (Heiser and Schilling, 1990). While DeCandolle (1959) suggested Asia as the domestication center, Zeven and Zhukovsky (1975) and Heiser and Schilling (1990) have supported India due to the fact that var. *amara*, which is more likely to be the ancestor, is endemic there. One of the most recent studies on this topic included landraces from China, Laos and Nepal (Marr et al., 2005). High variability was described for fruits and seeds, although only one allozyme *locus* proved to be polymorphic in this species out of the 29 *loci* assayed. Nine of them showed different and fixed alleles in both *L. ocutangula* and *L. aegyptiaca,* indicating that they

are likely to be reproductively isolated, although interspecific crosses allowing gene flow have been reported between the two species.

The latest studies on *Luffa acutangula* related to its antioxidants, and also its medicine and pesticide properties, and have increased breeders´ interest in this crop, especially in finding a higher yielding variety. Because of this, the search for phenotypic and genetic diversity has become more important.

Genetic diversity and relationships among *Luffa* spp. was studied by Tolentino et al. (1997) analyzing total seed protein profile using SDS-PAGE in a collection of 215 accessions. Great similarity was detected between both species, the smooth luffa being less variable possibly because of its major domestication. Recently, Hsieh et al. (2007) have evaluated diversity among lines and cultivars of *Luffa* using RAPDs and morphological traits, finding a clear separation between both two species. As previously reported, *L. acutangula* displays more variability. Hoque and Rabbani (2009) studied 28 ridge gourd landraces collected from different parts of Bangladesh, where many wild relatives are grown, using RAPDs and detected five groups within the germplasm assessed. A significant level of variability was found. In fact, the percentage of polymorphic *loci* was higher than that found in other cucurbits like *M. charantia* (Dey et al., 2006), *C. maxima* (Ferriol et al., 2003b) and *C. melo* (Garcia-Mas et al., 2000). Regarding phylogenetic studies, several surveys have tried to elucidate the evolutionary history of this species. *L. acutangula* together with *L. aegyptiaca* were placed in a single clade apart from the remainder of the species of the genus according to phenotypic variation (Heiser and Schilling, 1990), although chloroplast DNA markers data do not agree completely with this (Chung et al., 2003).

On the other hand, several studies on qualitative and quantitative composition of triacylglycerols from seed oils of different cucurbits have been carried out including

both cultivated species of *Luffa* (Grondin et al., 2002).

### I.6 Tribe *Sicyeae*

In the genus *Sechium*, the most important species is *Sechium edule* (Jacq.) Sw., (chayote). It is, like other minor cucurbits, a vegetable crop that is important in low-income agriculture. However, commercial production of chayote is important in several countries, such as Costa Rica, Mexico, Brazil and Puerto Rico (Hord et al., 1997).

It is native to Central America, Mexico being the main center of diversity (Lira-Saade, 1996). According to morphological and biochemical data, the wild ancestor of the cultivated forms is *Sechium edule* (Jacq.) Sw. ssp. *sylvestre*, which is endemic to Mexico (Cross, 2003; Cross et al., 2006). In addition, significant genetic variation in fruit traits is present in Central America (Engels, 1983).

Some of the most recent studies include morphological and anatomical characterization to better variety classification (Cadena-Iñiguez et al., 2008) and utilization of isozyme markers (Abdelnour and Rocha, 2008). Abdelnour and Rocha (2008) reported a high degree of genetic diversity in a collection of 42 accessions from Costa Rica and highlighted the usefulness of this kind of markers in species without DNA quality markers and genomic information. They reported the necessity of maintaining the threatened diversity in this crop and the necessity of germplasm collections. It is expected that more molecular markers for a better characterization will be developed in the future.

Other countries also are making an effort to study and maintain this diversity. In India, Sanwal et al. (2008) evaluated some indigenous populations of *S. edule*, called chow-chow there, for morphological traits such as number of fruits/plant, fruit yield/plant, total soluble solid content, acidity and ascorbic acid content.

**Acknowledgements**
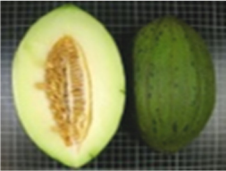
**Table I-1. Description of the main botanical groups in to which the species *C. melo* is divided, according to Pitrat (2008).**

| Variety | Distribution | Sex type | Fruit (shape/color) | Flesh color | Rind | Sweetness | Aroma | Climateric | |
|---|---|---|---|---|---|---|---|---|---|
| *conomon* | Eastern Asia | Andromonoecious | Elongated | White | Smooth, thin | No | No | No |  |
| *makuwa* | Eastern Asia | Andromonoecious | Flat-round-oval White-yellow-light green | White | Smooth, thin, with or without sutures | Yes | Little | Yes |  |
| *chinensis* | China, Korea | Andromonoecious (some hermaphrodite) | Pear-shaped Green with spots | Green-orange | Smooth | Medium | Little or no aroma | Both types |  |
| *momordica* | India | Monoecious | Flat-elongated | White | Smooth, slightly ribbed, thin | Low | Little | Yes |  |

| Variety | Distribution | Sex type | Fruit (shape/color) | Flesh color | Rind | Sweetness | Aroma | Climateric |
|---------|--------------|----------|---------------------|-------------|------|-----------|-------|------------|
| *acidulus* | India | Monoecious | Oval-elliptic Green-orange with or without spots | White | Smooth | No | No | No |
| *tibish* | Sudan | Andromonoecious | Small-oval Dark green with light green-yellow stripes | White | | No | No | No |
| *chate* | Mediterranean area and Western Asia | Monoecious (some andromonoecious genotypes) | Round-oval Green | White-light orange | Ribbed | No | No | Yes |
| *flexuosus* (snakemelon) | Northern Africa to Turkey to Iraq to India | Monoecious | Long or very long Green | White-light orange | Ribbed or wrinkled | No | No | Yes |

| Variety | Distribution | Sex type | Fruit (shape/color) | Flesh color | Rind | Sweetness | Aroma | Climateric | |
|---------|--------------|----------|---------------------|-------------|------|-----------|-------|------------|---|
| *cantalupensis* (muskmelon) | Europe, Western Asia, North and South America | Andromonoecious (usually) | Flat-oval | Orange (sometimes green) | Ribbed, smooth (sometimes with warts) | Yes | Yes | Yes |  |
| *reticulatus* | Europe, Asia, North and South America | Andromonoecious | Round-oval | Orange | Netted (sometimes ribbed) | Yes | Yes | Yes |  |
| *ameri* | Western and Central Asia | Andromonoecious | Elongated-oval Yellow-light green | White-light orange | Slightly netted | Yes | Little | Yes |  |
| *inodorus* | Central Asia, Mediterranean area, North and South America | Andromonoecious | Round-elliptic White-yellow-dark green (sometimes spotted) | White | Often wrinkled, ribbed or not | Yes | No | No |  |
| *dudaim* | Central Asia | Andromonoecious | Small-round Yellow with ochre stripes | White | Velvety | No | Strong | Yes |  |

**Table I-2. Description of the edible *C. pepo* morphotypes according to Paris (2001b, 2008).**

| Morphotype | Distribution | Fruit (shape/color) | Rind | Consumption | |
|---|---|---|---|---|---|
| Pumpkin | USA-Canada, Europe-Asia Minor, Mexico-Guatemala | Spherical-ovalated fruits with or without ribs<br><br>Orange or yellow-green pattern | Lignified/tender | Immature-mature<br><br>Seeds |  |
| Vegetable marrow | Middle-East, northern Africa | Short, tapered, cylindrical fruits | Lignified | Immature |  |
| Cocozelle | Europe (Italy), Far East, Turkey, Yugoslavia | Long, bulbous fruits<br><br>Striped, light green non-striped, ribbed or non ribbed | Smooth | Immature<br><br>Female flowers |  |
| Zucchini | Worldwide | Uniformly cylindrical fruits<br><br>Yellow and green types | Tender | Immature |  |

| Morphotype | Distribution | Fruit (shape/color) | Rind | Consumption | |
|---|---|---|---|---|---|
| Scallop | Australia, North America and Europe | Flat, scalloped fruits Yellow, white-green | Lignified | Immature |  |
| Acorn | USA-Canada | Furrowed, turbinate fruits Green | Non-lignified | Mature |  |
| Crookneck | Southeastern USA | Narrow necked fruits, usually curved and warted          Yellow | Thick, lignified | Immature |  |
| Straightneck | USA, Europe | Short-necked or constricted fruits, usually warted          Yellow | Lignified | Immature |  |
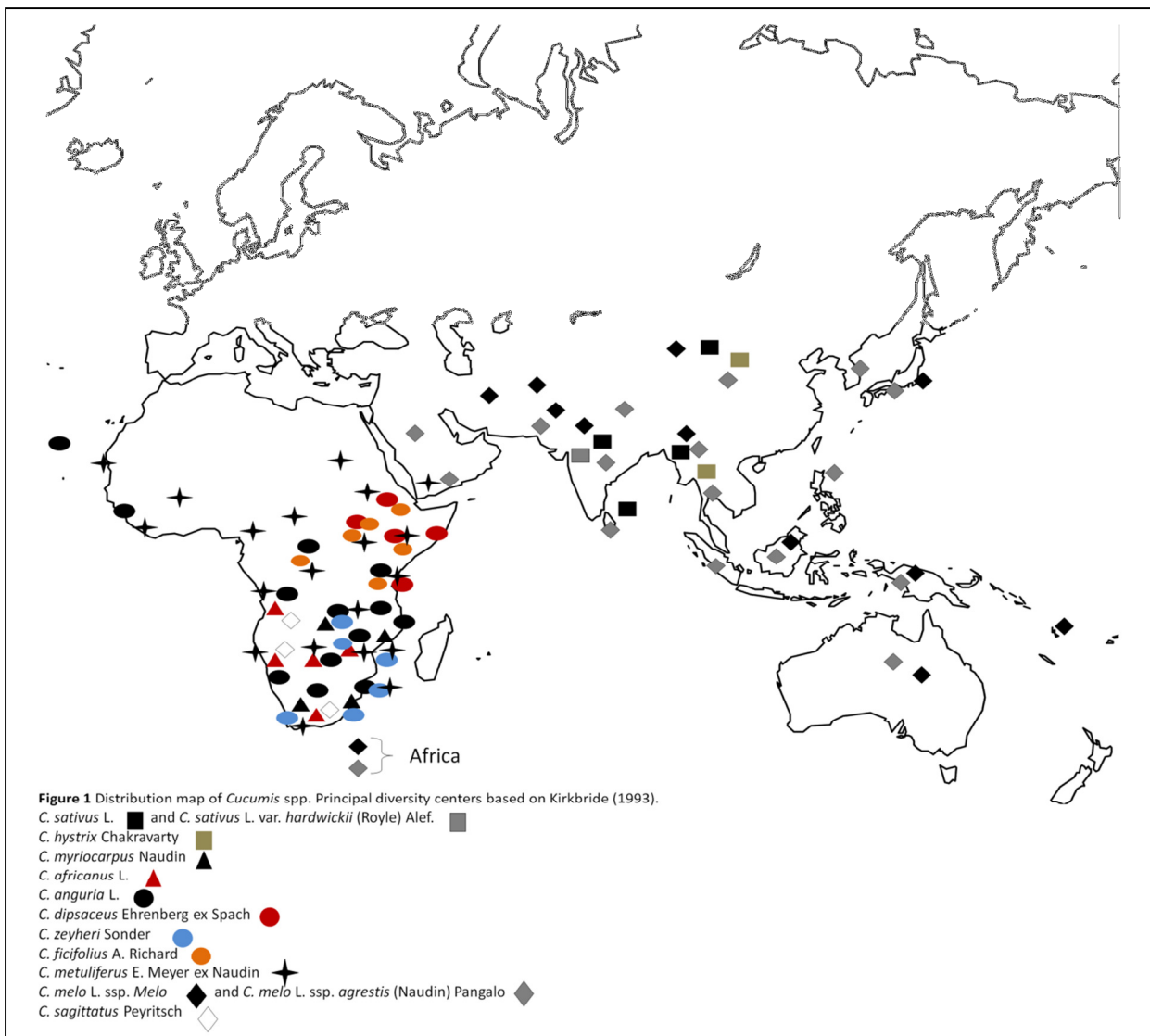
**Figure I-1. Distribution map of *Cucumis* spp. Principal diversity centers based on Kirkbride (1993).**

Figure 2. Distribution map of *Citrullus* spp. Principal diversity centers based on Jeffrey (2001).

C. colocynthis (L.) Schrad.
C. ecirrhosus Cogn.
C. lanatus ((Thunb.) Matsum & Nakai) var. *lanatus* and var. *citroides*
C. rehmii De Winter
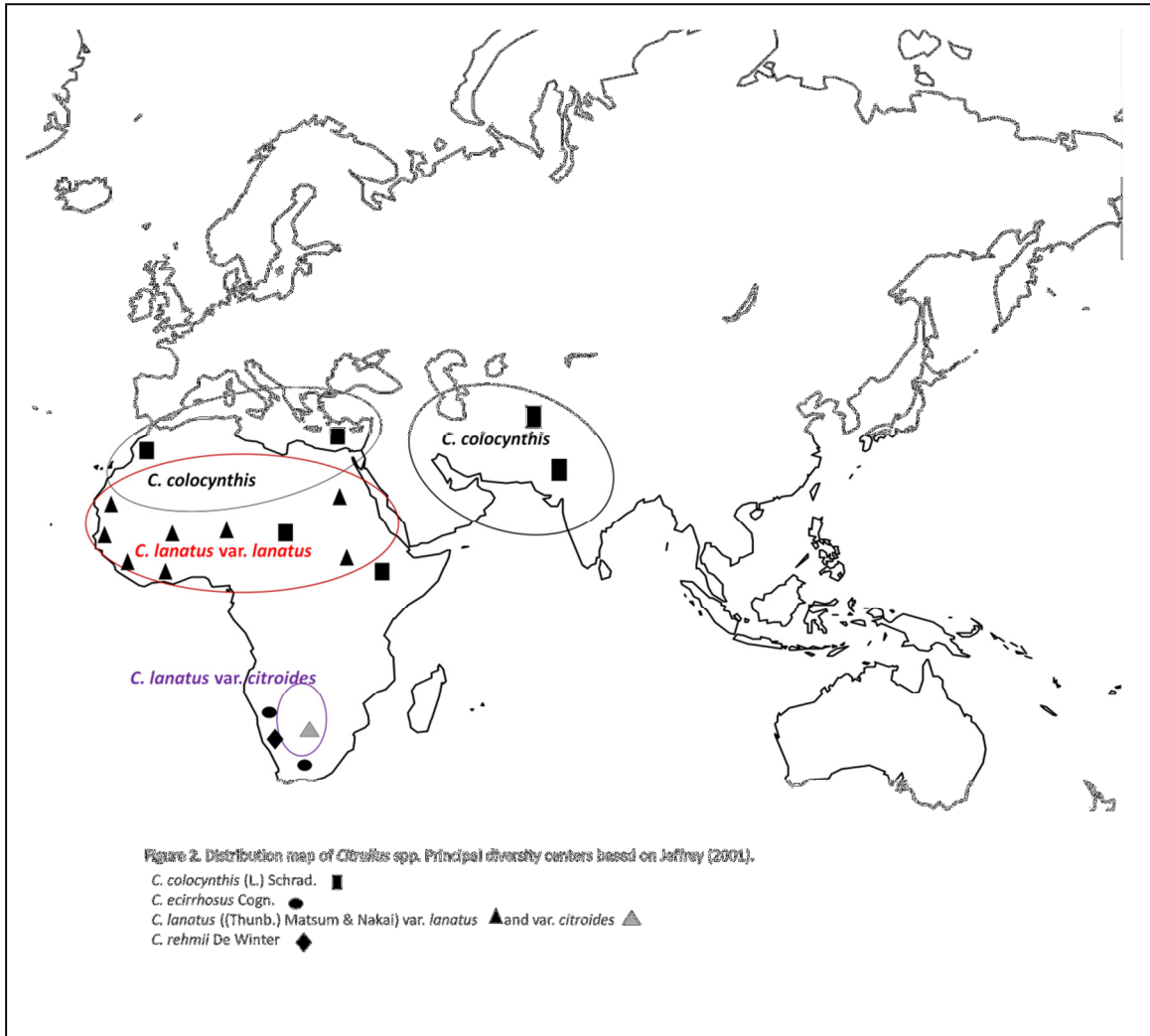
**Figure I-2. Distribution map of *Citrullus* spp. Principal diversity centers based on Jeffrey (2001).**

Figure 3. Distribution map of *Cucurbita* spp. Principal diversity centers based on Lira-Saade (1995) and Sanjur *et al.* (2002).
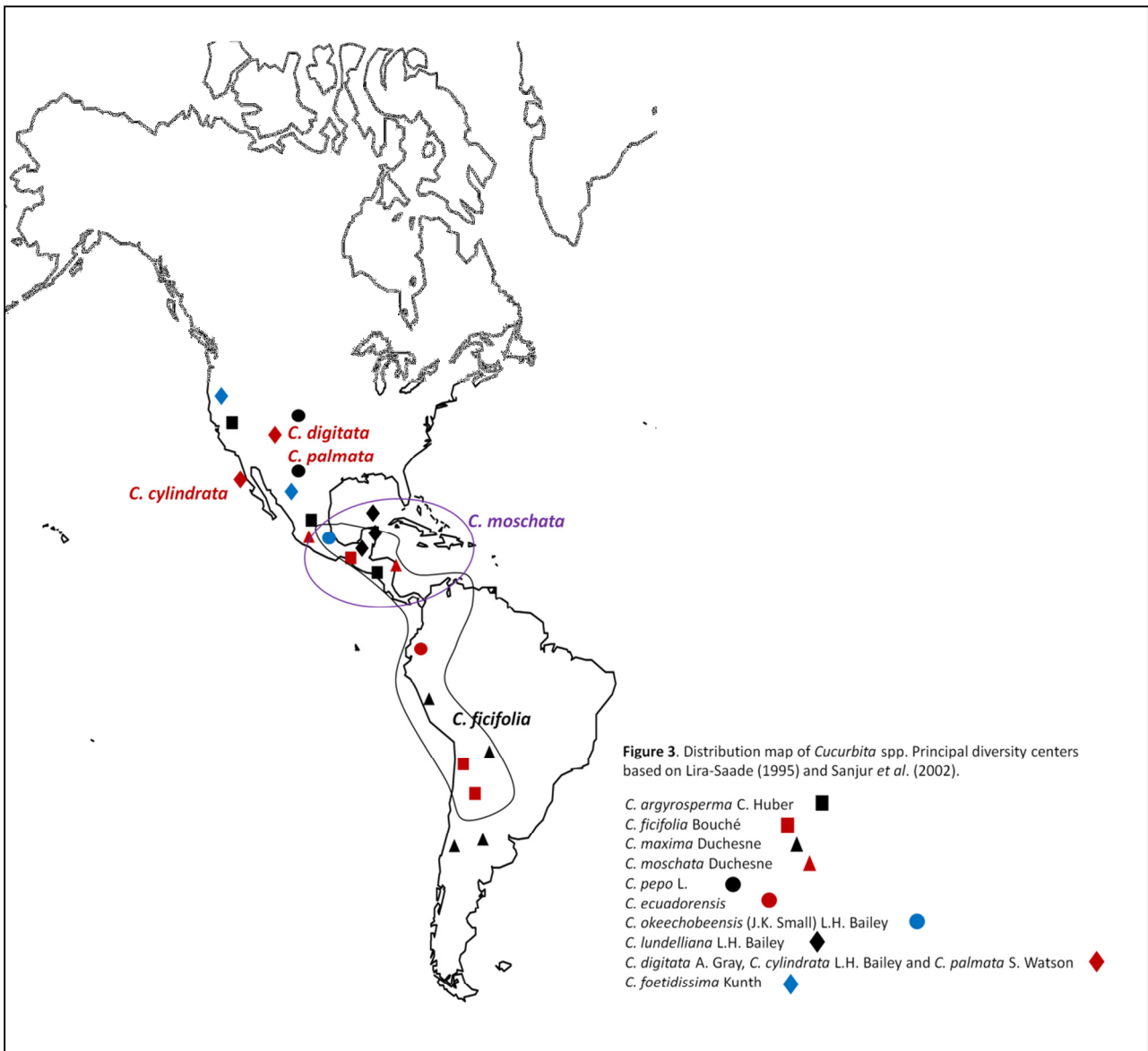
**Figure I-3. Distribution map of *Cucurbita* spp. Principal diversity centers based on Lira-Saade (1995) and Sanjur et al. (2002).**

**References**

Abdelnour A, Rocha OJ: **Genetic characterization of a collection of chayote, *Sechium edule* (Jacq.) Swartz, in Costa Rica by using isozyme markers.** Genet Resour Crop Evol 2008, 55:163–170.

Akashi Y, Fukunda N, Wako T, Masuda M, Kato K: **Genetic variation and phylogeneticrelationships in East and South Asian melons, *Cucumis melo* L., based on the analysis offive isozymes.** Euphytica 2002, 125:385–396.

Akashi Y, Tanaka K, Nishida H, Kato K, Khaning MT, Yi SS, Chou TT: **Genetic diversity and phylogenetic relationship among melon accessions from Africa and Asia revealed by RAPD analysis.** In Proc of Cucurbitaceae 2006. Edited by Holmes GJ. Asheville, North Carolina, USA; 2006:317–325.

Ashurmetov OA: **On morphology and taxonomy of the genera *Cucumis* L. and *Melo* Mill.** FEDDES Rep 1995, 106:155–159.

Baird JR, Thieret JW: **The Bur Gherkin (*Cucumis anguria* var. *anguria*, *Cucurbitaceae*).** Econ Bot 1988, 42(3):447–451.

Baranek M, Stift G, Vollmann J, Lelley T: **Genetic diversity within and between thespecies *Cucurbita pepo*, *C. moschata* and *C. maxima* as revealed by RAPD markers.** Cucurbit Genet Coop Rep 2000, 23:73–77.

Bates DM, Robinson RW: **Cucumbers, melons and watermelons.** In Evolution of crop plants, 2nd edn. Edited by Smartt J, Simmonds NW. Longman Scientific, Harlow, Essex, UK; 1995:89–96.

Bates DM, Merrick LC, Robinson RW: **Minor cucurbits-*Benincasa*, *Lagenaria*, *Luffa*, *Sechium*, and other genera (*Cucurbitaceae*).** In Evolution of crop plants, 2nd edn. Edited by Smartt J, Simmonds NW. Longman Scientific, Harlow, Essex, UK, 1995:105–111.

Baudracco-Arnas S, Pitrat M: **A genetic map of melon (*Cucumis melo* L.) with RFLP, RAPD, isozyme, disease resistance and morphological markers.** Theor Appl Genet 1996, 93:57–64.

Beharav A, Cohen Y: **Attempts to overcome the barrier of interspecific hybridization between *Cucumis melo* and *C. metuliferus*.** Isr J Plant Sci 1995, 43(2):113–123.

Behera TK, Dey SS, Sirohi PS: **'DBGy-201' and 'DBGy-202': two gynoecious lines inbitter gourd (*Momordica charantia* L.) isolated from indigenous source.** Indian J Genet 2006b, 66:61–62.

Behera TK, Singh AK, Staub JE: **Comparative analysis of genetic diversity in Indian bitter gourd (*Momordica charantia* L.) using RAPD and ISSR markers for developing crop improvement strategies.** Sci Hort 2008, 115:209–217.

Bemis WP, Rhodes AM, Whitaker TW, Carmer SG: **Numerical taxonomy applied to *Cucurbita* relationships.** Am J Bot 1970, 57(4):404–412.

Bernet GP, Bramardi S, Calvache D, Carbonell EA, Asins MJ: **Applicability of molecular markers in the context of protection of new varieties of cucumber.** Plant Breed 2003, 122:146–152.

Biles CL, Martyn RD, Wilson HD: **Isozymes and general proteins from various watermelon cultivars and tissue types.** HortScience 1989, 24:810–812.

Bisht IS, Bhat KV, Tanwar SPS, Bhandari DC, Joshi K, Sharma AK: **Distribution andgenetic diversity of *Cucumis sativus* var. *hardwickii* (Royle) Alef in India.** J Hort Sci Biotechnol 2004, 79(5):783–791.

Bradeen JM, Staub JE, Wye C, Antonise R, Peleman J: **Towards an expanded and integrated linkage map of cucumber (*Cucumis sativus* L.).** Genome 2001, 44:111–119.

Brothwell D, Brothwell P: **Food in antiquity: a survey of the diet of early peoples.** Thames and Hudson (eds) Fredrick A. Praeger Publ, New York, USA; 1969.

Brotman Y, Silberstein L, Kovalski J, K1ingler J, Thompson G, Katzir N, Perl-Treves R:**Linkage groups of *Cucumis melo*, including resistance gene homologues and known genes.** In Proc of Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:441-444.

Cadena-Iñiguez J, Avendaño-Arrazate CH, Soto-Hernández M, Ruiz-Posadas LM, Aguirre-Medina JF, Arévalo-Galarza L: **Infraspecific variation of *Sechium edule* (Jacq.) Sw. in the state of Veracruz, Mexico.** Genet Resour Crop Evol 2008, 55:835–847.

Canul J, Ramírez P, Castillo F, Chávez JL: **Diversidad morfológica de calabaza cultivada en el centro-oriente de Yucatán, México.** Sociedad Mexicana de Citogenética. Revista Fitotecnia Mexicana 2005, 28(004):339–349.

Carle RB, Maynard DN, Wessel-Beaver L:**Tropical pumpkin hybrid development:landraces to hybrid cultivars.** In Proc of Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:95–100.

Castetter EF: **Horticultural groups of cucurbits.** Proc Am Soc Hort Sci 1925, 22:338–340.

Chen JF, Staub JE, Tashiro Y, Isshiki S, Miyazaki S: **Successful interspecific hybridization between *Cucumis sativus* L. and *Cucumis hystrix* Chakr.** Euphytica 1997a, 96:413–419.

Chen JF, Isshiki S, Tashiro Y, Miyazaki S: **Biochemical affinities between *Cucumis hystrix* Chakr and two cultivated *Cucumis* species (*C. sativus* L. and *C. melo* L.) based on isozyme analysis.** Euphytica 1997b, 97:139–141.

Chiba N, Suwabe K, Numone T, Hirai M: **Development of microsatellite markers in melon (*C. melo* L.) and their application to major cucurbit crops.** Breed Sci 2003, 53:21–27.

Chung HD, Youn SJ, Choi YJ: **Ecological and morphological characteristics of the Korean native squash (*Cucurbita moschata*).** J Kor Soc Hort Sci 1998, 39(4): 377–384.

Chung S, Decker-Walters DS, Staub JE: **Genetic relationships within the *Cucurbitaceae* asassessed by consensus chloroplast simple sequence repeats (ccSSR) marker and sequence analyses.** Can J Bot 2003, 81:1–19.

Chung SM, Staub JE, Chen JF: **Molecular phylogeny of *Cucumis* species as revealed by consensus chloroplast SSR marker length and sequence variation.** Genome 2006, 49:219–229.

Costa J, Catalá MS, Cortés C, Nuez F, Abadía J, Cuartero J: **Evaluación de la variabilidad en los principales tipos de melón cultivados en España.** Invest Agri Prod Prot Veg 1989, 4:43–57 (in Spanish).

Cross HB: **Evolution, systematics, and domestication in *Sechium* and genera related (*Sicyeae, Cucurbitaceae*).** PhD Thesis, Columbia Univ, NY, USA, 182 p. 2003.

Cross HB, Lira R, Motley TJ: **Origins and diversification of Chayote.** In Darwin's harvest: new approaches to the origins, evolution,and conservation of crops. Edited by Motley TJ, Zerega NJC, Cross HB. Columbia Univ Press, New York, USA; 2006:171–194.

Cutler H, Whitaker T: **Cucurbits from the Tehuacan caves.** In The Prehistory of the Tehuacan valley, environment and subsistence, vol 1. Edited by Byers DS. Univ of Texas Press, Austin, TX, USA; 1967:212–219.

Dane F: **Evolutionary studies in the genus *Cucumis*.** PhD Dissert. Colorado State Univ,Fort Collins, TX, USA. 1976.

Dane F: **Cucurbits**. In Isozymes in plant Genetics and Breeding, part B. Edited by Tanksley SD, Orten TJ. Elsevier, Amsterdam, The Netherlands; 1983:369–390.

Dane F: **Chloroplast DNA investigations in *Citrullus* using PCR-RFLP analysis.** In: DN Maynard (ed) Cucurbitaceae 2002. ASHS Press, Naples, FL, USA, pp 100–108.

Dane F, Lang P: **Sequence variation at cpDNA regions of watermelon and related wild species: implications for the evolution of *Citrullus* haplotypes.** Am J Bot 2004, 91(11):1922–1929.

Dane F, Liu J: **Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*).** Genet Resour Crop Evol 2007, 54:1255–1265.

Dane F, Lang P, Bakhtiyarova R: **Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species.** Theor Appl Genet 2004, 108:958–966.

Danilchenko H, Paulauskiene A, Dris R, Niskanen R: **Biochemical composition and processability of pumpkin cultivars**. In Proc of Cucurbitaceae 2000. Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:493–497.

Danin-Poleg Y, Reis N, Tzuri G, Katzir N: **Development and characterisation of microsatellite markers in *Cucumis*.** Theor Appl Genet 2001, 102: 61–72.

Danin-Poleg Y, Tadmor Y, Tzuri G, Resi N, Hirschberg J, Katzir N: **Construction of agenetic map of melon with molecular markers and horticultural traits, and location of genes associated with ZYMV resistance.** Euphytica 2002, 125(3):373–384.

Davis CC, Fritsch PW, Li J, Donoghue MJ: **Phylogeny and biogeography of *Cercis* (Fabaceae): evidence from nuclear ribosomal ITS and chloroplast *ndhF* sequence data.** Syst Bot 2002, 27:289–302.

DeCandolle A: **Origine des Plantes Cultivées**. Baillière, Paris, France. 1883

DeCandolle A: **Origin of Cultivated Plants.** (Reprint of the 2nd edn, 1886).Hafner Publ, New York, USA 478 p. 1959

Decker DS: **Numerical analysis of allozyme variation in *Cucurbita pepo*.** Econ Bot 1985, 39(3):300–309.

Decker DS: **Origin(s), evolution, and systematics of *Cucurbita pepo* (*Cucurbitaceae*).** Econ Bot 1988, 42(1):4–15.

Decker-Walters DS: **Cucurbits, Sanskrit, and the Indo-Aryas**. Econ Bot 1999, 53: 98–112.

Decker-Walters DS, Walters TW: **Squash**. In The Cambridge world History of food. Edited by Kipple KF, Ornelas KC. Cambridge Univ Press, New York, USA; 2000:335–351.

Decker-Walters DS, Walters TW, Posluszny U, Kevan PG: **Genealogy and gene flow among annual domesticated species of *Cucurbita*.** Can J Bot 1990, 68:782–789.

Decker-Walters DS, Walters TW, Cowan CW, Smith BD: **Isozymic characterization of wild populations of *Cucurbita pepo*.** J Ethnobiol 1993, 13:55–72.

Decker-Walters D, Staub J, López-Sesé A, Nakata E: **Diversity in landraces and cultivars of bottle gourd (*Lagenaria siceraria*; *Cucurbitaceae*) as assessed by random amplified polymorphic DNA.** Genet Resour Crop Evol 2001, 48:369–380.

Decker-Walters DS, Chung SM, Staub JE, Quemada HD, López-Sessé AI: **The originand genetic affinities of wild populations of melon (*Cucumis melo*, *Cucurbitaceae*) in North America.** Plant Syst Evol 2002a, 233:183–197.

Decker-Walters DS, Staub JE, Chung SM, Nakata E, Quemada HD: **Diversity in free living populations of *Cucurbita pepo* (*Cucurbitaceae*) as assessed by random amplified polymorphic DNA.** Syst Bot 2002b, 27(1):19–28.

Deleu W, Esteras C, Roig C, González-To M, Fernández-Silva I, Gonzalez-Ibeas D, Blanca J,Aranda MA, Arús P, Nuez F, Monforte AJ, Picó B, Garcia-Mas J: **A set of EST-SNPs for map saturation and cultivar identification in melon.** BMC Plant Biol 2009, 9: 90.

Demesure B, Sozi N, Petit RJ: **A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants.** Mol Ecol 1995, 4:129–131.

De Veaux JS, Schultz EB Jr: **Development of buffalo gourd (*Cucurbita foetidissima*) as a semiarid land starch and oil crop.** Econ Bot 1985, 39:454–472.

De Winter B: **A new species of *Citrullus* (*Benincaseae*) from the Namib Desert, Namibia.** Bothalia 1990, 20:209–211.

Dey SS, Singh AK, Chandel D, Behera TK: **Genetic diversity of bitter gourd (*Momordica charantia* L.) genotypes revealed by RAPD markers and agronomic traits.** Sci Hort 2006, 109:21–28.

Dhillon NPS, Ranjana R, Singh K, Eduardo I, Monforte AJ, Pitrat M, Dhillon NK, Singh PP: **Diversity among landraces of Indian snapmelon (*Cucumis melo* var. *momordica*).** Genet Resour Crop Evol 2007, 54:1267–1283.

Dhillon NPS, Singh J, Fergany M, Monforte AJ, Sureja AK: **Phenotypic and moleculardiversity among landraces of snapmelon (*Cucumis melo* var. *momordica*) adapted to thehot and humid tropics of eastern India.** Plant Genetic Resources: Characterization andUtilization 2009, 7(3):291–300, doi: 10.1017/S1479262109990050.

Dijkhuizen A, Kennard WC, Havey MJ, Staub JE: **RFLP variation and genetic relationships in cultivated cucumber.** Euphytica 1996, 90:79–87.

Duca M, Port A, Leivitchi A: **Characteristics of RAPD markers in breeding of *Cucumis sativus* L.** Roum Biotechnol Lett 2008, 13(4):3843–3850.

Ellul P, Lelivelt C, Naval MM, Noguera FJ, Sanchez S, Atarés A, Moreno V, Corella P, Dirks R: **Watermelon**. In Biotechnology in agriculture and forestry, vol 60. Edited by Nagata T, Lörz H, Widholm JM. Transgenic Crops V. Edietd by Pua EC, Davey MR. Springer-Verlag Berlin, Heidelberg, Germany; 2007:129–165.

Engels JMM: **Variation in *Sechium edule* in Central America**. J Am Soc Hort Sci 1983, 108:706–707.

Erwin AT: **Nativity of the cucurbits.** Bot Gaz 1931, 91:105–108.

Escribano S, Lázaro A, Staub JE: **Genetic diversity of Spanish melons (*Cucumis melo*) of the Madrid provenance.** In Cucurbitaceae 2008, Proc IX EUCARPIA Meeting on

Genetics and Breeding of Cucurbitaceae. Edited by Pitrat M. INRA, Avignon, France, 21–24 May 2008:301–305.

Esquinas-Alcázar JT: **Alloenzyme variation and relationships in the genus *Cucumis*.** PhD Diss.Univ of California, Davis, USA. 1977.

Esquinas-Alcázar JT: **Allozyme variation and relationships among Spanish landraces of *Cucumis melo* L.** Kulturpfl anze 1981, 22:337–352.

Esquinas-Alcázar JT, Gulick PJ: **Genetic Resources of *Cucurbitaceae* - A Global Report.** International Board for Plant Genetic Resources, Rome, Italy. 1983.

Esteras C, Díez MJ, Picó B, Sifres A, Valcarcel JV, Nuez F: **Diversity of Spanish landraces of *Cucumis sativus* and *Cucurbita* ssp.** In Cucurbitaceae 2008, Proc IX EUCARPIA Meeting on Genetics and Breeding of Cucurbitaceae. Edited by Pitrat M. INRA, Avignon, France, 21–24 May 2008: 67–76.

Esteras C, Lunn J, Sulpice R, Blanca J, Garcia-Mas J, Pitrat M, Nuez F, Picó B: **Phenotyping a highly diverse core melon collection to be screened using Ecotilling.** 8 Plant Genomics European Meetings (Plant Gem), Lisbon (Portugal), 7–10 October 2009a: p214.

Esteras C, Pascual L, Saladie M, Dogimont C, Garcia-Mas J, Nuez F, Picó B: **Use of Ecotilling to identify natural allelic variants of melon candidate genes involved in fruit ripening.** 8 Plant Genomics European Meetings (Plant Gem), Lisbon (Portugal), 7–10 October 2009b: p213.

Esteras C, Sifres A, Nuez F, Picó B: **Variabilidad de *Cucurbita maxima* en su zona de origen: un recurso de interés para la mejora de esta hortaliza.** In Resúmenes. Actas de Horticultura 54. Edited by Pardo A, Suso ML, Vázquez N. VI Congreso Ibérico, XII Nacionalde Ciencias Hortícolas, Logroño, Spain, 25–29 May 2009c: 1284–1290 (in Spanish).

Fanourakis N, Tsekoura Z, Nanou E: **Morphological characteristics and powdery mildew resistance of *Cucumis melo* landraces in Greece.** In Proc Cucurbitaceae 2000, Acta Hort 510. Edited byKatzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:241–245.

Fantaccione S, Woodrow P, Pontecorvo G: **Molecular authentication of three Italian melon accessions by ARMS-PCR and ITS1 (internal transcribed spacer 1) secondary structure prediction.** Bioinformation 2008, 2(7):311–315.

Fazio G, Chung SM, Staub JE: **Development and characterization of PCR markers in cucumber (*Cucumis sativus* L.).** J Am Soc Hort Sci 2002, 127:545–557.

Fazio G, Staub JE, Stevens MR: **Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines.** Theor Appl Genet 2003, 107: 864–874.

Fernández-Silva I, Eduardo I, Blanca J, Esteras C, Picó B, Nuez F, Arús P, Garcia-Mas J, Monforte AJ: **Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.).** Theor Appl Genet 2008, 118:139–150.

Ferriol M: **Análisis de la variabilidad morfológica y molecular en el género *Cucurbita*.** Tesis doctoral. Univ Politécnica de Valencia, Spain (in Spanish). 2003.

Ferriol M, Picó B: **Pumpkin and winter squash.** In Handbook of plant breeding, vol I: Vegetables I. Edited by Prohens J, Nuez F. Springer, New York, USA; 2008:317–349.

Ferriol M, Picó B, Nuez F: **Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers.** Theor Appl Genet 2003a, 107:271–282.

Ferriol M, Picó B, Nuez F: **Genetic diversity of some accessions of *Cucurbita maxima* from Spain using RAPD and SBAP markers.** Genet Resour Crop Evol 2003b, 50:227–238.

Ferriol M, Picó B, Fernández de Córdova P, Nuez F: **Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) determined by SRAP and AFLP markers.** Crop Sci 2004a, 44:653–664.

Ferriol M, Picó B, Nuez F: **Morphological and molecular diversity of a collection of *Cucurbita maxima* landraces.** J Am Soc Hort Sci 2004b, 129(1):60–69.

Ferriol M, Picó B, Nuez F: **Genetic diversity of *Cucurbita* spp. in the Canary Islands: a bridge between America and Europe.** XVII EUCARPIA Genetic Resources Section Meeting, Castelsardo (Italy), 30 March–2 April 2005:25–32.

Filov AI: **Ekologija I klassifi katzija tykuy**. Bjulleten´ Glavnogo botaniceskogo sada 1966, 63:33–41 (in Russian).

Flyman MV, Afolayan AJ: **The implication of the mineral ratios of *Cucumis myriocarpus* Naud. And *Pergularia daemia* (Forsk.) Chiov. in human diets.** J Med Food 2007, 10(3):548–551.

Fujishita N: **Melons in the ancient Japan, revealed by excavated melon seeds.** Archaeol J 1992, 354:7–13.

Fukino N, Yoshioka Y, Kubo N, Hirai M, Sugiyama M, Sakata Y, Matsumoto S: **Development of 101 novel SSR markers and construction of an SSR-based genetic linkage map in cucumber (*Cucumis sativus* L.).** Breed Sci 2008, 58: 475–483.

Fursa TB: **K Sistematike roda *Citrullus* Schrad.** Bot Z 1972, 57:31–41.

García E, Jamilena M, Álvarez JI, Arnedo T, Oliver JL, Lozano R: **Genetic relationships among melon breeding lines revealed by DNA markers and agronomic traits.** Theor Appl Genet 1998, 96:878–885.

Garcia-Mas J, Oliver M, Gómez-Paniagua H, De Vicente MC: **Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon.** Theor Appl Genet 2000, 101:860–864.

Garcia-Mas J, Monforte AJ, Arús P: Phylogenetic **relationships among *Cucumis* species based on the ribosomal internal transcribed spacer sequence and microsatellite markers**. Plant Syst Evol 2004, 248:191–203.

Ghebretinsae AG, Thulin M, Barber JC: **Relationships of cucumbers and melons unraveled: molecular phylogenetics of *Cucumis* and related genera (Benincaseae, *Cucurbitaceae*).** Am J Bot 2007, 94(7):1256–1266.

Gichimu BM, Owuor BO, Mwai GN, Dida MM: **Morphological characterization of some wild and cultivated watermelon (*Citrullus* sp.) accessions in Kenya.** ARPN J Agri Biol Sci 2009, 4:2.

Goldberg RB, Bemis WP, Siegel A: **Nucleic acid hybridization studies within the genus *Cucurbita*.** Genetics 1972, 72:253–266.

Gómez-Guillamón ML, Abadía J, Cuartero J, Corté C, Nuez F: **Characterization of melón cultivars.** Cucurbit Genet Coop Rep 1985, 8:39–40.

Gómez-Guillamón ML, Torés JA, Soria C, Sesé AIL: **Screening for resistance to *Sphaerotheca fuliginea* and two yellowing diseases in *Cucumis melo* and related *Cucumis* species.** In Proc Cucurbitaceae 1994. Edited by Lester G, Dunlap J. Evaluation and enhancement of cucurbitgermplasm.South Padre Island, Texas, USA; 1995:205–208.

Gómez-Guillamón ML, Sánchez F, Fernández-Muñoz R: **Caracterización de cultivares de melón (Spanish).** Actas de las Jornadas de Selección y Mejora de Plantas Hortícolas. Córdoba, Spain (in Spanish). 1998.

Gong L, Stift G, Kofl er R, Pachner M, Lelley T: **Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L**. Theor Appl Genet 2008a, 117: 37–48.

Gong L, Pachner M, Kalai K, Lelley T: **SSR-based genetic linkage map of *Cucurbita moschata* and its synteny with *Cucurbita pepo*.** Genome 2008b, 51:878–887.

González-Ibeas D, Blanca J, Roig C, González-To M, Picó B, Truniger V, Gómez P, Deleu W, Caño-Delgado A, Arús P, Nuez F, Garcia-Mas J, Puigdomènech P, Aranda MA: **MELOGEN: an EST database for melon functional genomics.** BMC Genomics 2007, 8: 306. doi:10.1186/1471–2164-8-306.

González VM, Rodríguez-Moreno L, Centeno E, Benjak A, Garcia-Mas J, Puigdomènech P, Aranda MA: **Genome-wide BAC-end sequencing of *Cucumis* melo using two BAC libraries**. BMC Genomics 2010, 11: 618.

Gonzalo MJ, Oliver M, Garcia-Mas J, Monfort A, Dolcet-Sanjuan R, Katzir N, Arús P, Monforte AJ: **Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.).** Theor Appl Genet 2005, 110(5): 802–811.

Grondin I, Smadja J, Armougom R: **The qualitative and quantitative composition of triacylglycerols from four *Cucurbitaceae* seed oils of the *Lagenaria* and *Luffa* species.** OCL-Oleagineux, Corps Gras, Lipides 2002, 9(2/3):169–173.

Guerra-Sanz JM: ***Citrullus* simple sequence repeats markers from sequence databases.** Mol Ecol Notes 2002, 2:223–225.

Günay A: **Vegetable production**. V.A.Ü. Ziraat Fak. Ankara. 117s, 1993. (in Turkish)

Gwanama C, Labuschagne MT, Botha AM: **Analysis of genetic variation of *Cucurbita moschata* by random amplified polymorphic DNA (RAPD) markers.** Euphytica 2000, 113:19–24.

Gwanama C, Nichterlein K, Lungu D, Simabwachi W: **Variation of fruit b-carotene content of tropical pumpkin [*Cucurbita moschata* (Duchesne) Poirot] landraces in Zambia.** Plant Genet Resour Newsl 2002, 129:44–46.

Hashizume T, Sato T, Hirai M: **Determination of genetic purity of hybrid seed in watermelon (*Citrullus lanatus*) and tomato (*Lycopersicon esculentum*) using random amplified polymorphic DNA (RAPD).** Jpn J Breed 1993, 43:367–375.

Heikal AH, Abdel-Razzak HS, Hafez EE: **Assessment of genetic relationships among and within *Cucurbita* species using RAPD and ISSR markers.** J Appl Sci Res 2008, 4(5):515–525.

Heiser CB, Schilling EE: **The genus *Luffa*: A problem in phytogeography**. In Biology and utilization of the *Cucurbitaceae*. Edited by Bates DM, Robinson RW, Jeffrey C. Cornell UnivPress, Ithaca, New York, USA; 1990:120–133.

Helm MA, Hemleben V: **Characterization of a new prominent satellite DNA of *Cucumis metuliferus* and differential distribution of satellite DNA in cultivated and wild species of *Cucumis* and in related genera of *Cucurbitaceae***. Euphytica 1997, 94(2): 219–226.

Hoque S, Rabbani MG: **Assessment of genetic relationship among landraces of Bangladeshi ridge gourd (*Luffa acutangula* Roxb.) using RAPD markers.** J Sci Res 2009, 1(3):615–623.

Hord M, Villalobos W, Macaya-Lizano AV, Rivera C: **Chayote mosaic, a new disease in *Sechium edule* caused by a tymovirus**. Plant Dis 1997, 81:374–378.

Horejsi T, Staub JE: **Genetic variation in cucumber as assessed by random amplified polymorphic DNAs.** Genet Resour Crop Evol 1999, 46:337–350.

Horejsi T, Box JM, Staub JE: **Efficiency of randomly amplified polymorphic DNA to sequence characterized amplified region marker conversion and their comparative polymerase chain reaction sensitivity in cucumber.** J Am Soc Hort Sci 1999, 124:128–135.

Horst EK, Lower RL: *Cucumis hardwickii*: **A source of germplasm for the cucumber breeder.** Cucurbit Genet Coop Rep 1978, 1:5.

Hsieh L, Liu M, Hsiao C, Lin T, Yang W: **Genetic diversity among the lines and cultivars of** *Luffa* **species.** J Tai Soc Hort Sci 2007, 53(1):99–110.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, LiJ, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EAG, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H,Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S: **The genome of the cucumber,** *Cucumis sativus* **L**. Nat Genet 2009 41: 1275–1281.

Huh YC, Solmaz I, Sari N: **Morphological characterization of Korean and Turkish watermelon germplasm.** In Cucurbitaceae 2008, Proc IXth EUCARPIA. Meeting on Genetics and Breeding of Cucurbitaceae. Edited by Pitrat M. INRA, Avignon, France, 21–24 May 2008:327–333.

Ignart F, Weeden NF: **Allozyme variation in cultivars of** *Cucurbita pepo* **L.** Euphytica 1984, 33:779–785.

Ivancic A, Sisko M, Bohanec B, Siftar S: **Morpho-agronomic characteristics of the interspecific hybrid** *Cucurbita ficifolia* **x** *C. maxima*. Agricultura (Slovenia) 2004, 3(1):1–5.

Jarret RL, Newman M: **Phylogenetic relationships among species of *Citrullus* and the placement of *C. rehmii* De Winter as determined by Internal Transcribed Spacer (ITS) sequence heterogeneity.** Genet Resour Crop Evol 2000, 47:215–222.

Jarret RL, Merrick LC, Holms T, Evans J, Aradhya MK: **Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum.& Nakai).** Genome 1997, 40:433–441.

Jeffrey C: *Cucurbitaceae*. In Flora of East TropicalAfrica. Edited by Milne-Redhead E, Polhill RM. Crown Agents for Oversea Governments and Administrations, London, UK; 1967, 17:1–156.

Jeffrey C: **A review of the *Cucurbitaceae*.** Bot J Linn Soc 1980, 81: 233–247.

Jeffrey C: **Appendix: An outline classification of the *Cucurbitaceae*.** In Biology and utilization of the *Cucurbitaceae*. Edited by Bates DM, Robinson RW, Jeffrey C. Univ Press, Ithaca, New York, USA; 1990:449–463.

Jeffrey C: *Cucurbitaceae*. In Mansfeld's Encyclopedia of agricultural and horticultural crops. Edited by Hanelt P. Institute of Plant Genetics and Crop Plant Research. Springer, Berlin, Germany; 2001, 3:1510–1557.

Jeffrey C: **A new system of *Cucurbitaceae*.** Bot Zhurn 2005, 90: 332–335.

Jeffrey C, De Wilde WJJO: **A review of the subtribe *Thladianthinae* (*Cucurbitaceae*).** Bot Zhurn 2006, 91:766–776.

Jeon HJ, Been CG, Hong KH, Om YH, Kim BD: **Identification of *Cucurbitaceae* cultivars by using RAPD markers.** J Kor Soc Hort Sci 1994, 35(5):449–456.

Jobst J, King K, Hemleben V: **Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of the family *Cucurbitaceae*.** Mol Phylogenet Evol 1998, 9(2):204–219.

Joobeur T, Gusmini G, Zhang X, Levi A, Xu Y, Wehner TC, Oliver M, Dean RA: **Construction of a watermelon BAC library and identifi cation of SSRs anchored to melon or *Arabidopsis* genomes.** Theor Appl Genet 2006, 112:1553–1562.

Joshi DC, Das SK, Mukherjee RK: **Physical properties of pumpkin seeds**. J Agri Engg Res 1993, 54:219–229.

Júnior ATA: **Divergencia genética entre acessos de moranga do banco de germoplasma de hortaliças de Universidade Federal de Viçosa.** Hort Bras 1999, 17:3–6 (in Portuguese).

Kajale MD: **On the occurrence of ancient agricultural patterns during the Chalcolithic period (ca. 1,600–1,000 BC) at Apegaon, district Aurangabad in the central Godavarivalley, Maharashtra.** In Apegaon excavation. Edited by Deo SB, Dhavalikar MK, Ansari ZD.Pune Deccan College Postgraduate and Research Institute, Pune, India; 1979:50–56.

Karchi Z: **Development of melon culture and breeding in Israel**. In Proc Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:13–17.

Katzir N, Danin-Poleg Y, Tzuri G, Karchi Z, Lavi U, Cregan PB: **Length polymorphism and homologies of microsatellites in several cucurbitacean species.** Theor Appl Genet 1996, 93: 1282–1290.

Katzir N, Tadmor Y, Tzuri G, Leshzeshen E, Mozes-Daube N, Danin-Poleg Y, Paris H S: **Further ISSR and preliminary SSR analysis of relationships among accessions of *Cucurbita pepo*.** In Proc Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel,19–23 March 2000a:433–439.

Katzir N, Mozes-Daube N, Danin-Poleg Y, Paris HS: **Potential usefulness of SSR markers for studying infraspecific variability in *Cucurbita pepo*.** Cucurbit Genet Coop Rep 2000b, 23:71–72.

Keay WJ, Hepper FN: *Cucurbitaceae*. In The flora of west tropical Africa. 2nd edn. Edited by Keay WJ, Hepper FN. Royal Botanic Gardens, London, UK; 1985:570–574.

Keinath AP, DuBose VB: **Evaluation of pumpkin cultivars for powdery and downy midew resistance, virus tolerance, and yield.** HortScience 2000, 35(2):281–285.

Kepeng C, Liang C, Wang Y, Jin D, Wang B, Xu Y, Kang G, Zhang H: **Genetic assessment of watermelon germplasm using the AFLP technique.** HortScience 2003, 38: 81–84.

King K, Jobst J, Hemleben V: **Differential homogenization and amplification of two satellite DNAs in the genus *Cucurbita* (*Cucurbitaceae*).** J Mol Evol 1995, 41:996–1005.

Kirkbride JH Jr: **Biosystematic monograph of the genus *Cucumis* (*Cucurbitaceae*).** Parkway Publ, Boone. North Carolina, USA; 159 p. 1993.

Kirkpatrick KJ, Wilson HD: **Interspecific gene flow in *Cucurbita*: *C. texana* vs *C. pepo*.** Am J Bot 1988, 75: 517–525.

Kirkpatrick KJ, Decker DS, Wilson HD: **Allozyme differentiation in the *Cucurbita pepo* complex: *C. pepo* var. *medullosa* vs *C. texana*.** Econ Bot 1985, 39:289–299.

Kitamura S: **Notes on *Cucumis* of Far East**. Acta Phytotaxon Geobot 1950, 14:41–44.

Knerr LD, Staub JE, Holder DJ, May BP: **Genetic diversity in *Cucumis sativus* L. assessed by variation at 18 allozyme coding *loci*.** Theor Appl Genet 1989, 78: 119–128.

Kocyan A, Zhang LB, Schaeffer H, Renner SS: **A multi-*locus* chloroplast phylogeny for the *Cucurbitaceae* and its implications for character evolution and classification**. Mol Phylogenet Evol 2007, 44:553–577.

Kohpayegani JA: **Study of genetic diversity of some Iranian melon and effects of seed production on genetic erosion**. PhD Thesis, Univ Tehran, Iran, 124 p. 2004.

Kohpayegani JA, Behbahani M: **Genetic diversity of some populations of Iranian melon using SSR markers.** Biotechnology 2008, 7(1):19–26.

Kole C, Olukolu B, Kole P, Abbott AG: **Towards phytomedomics with bitter melon (*Momordica charantia* L.) as a model.** Abstract. Plant &Animal Genomes XVIII Conference, San Diego CA, 9–13 January 2010.

Kong Q, Xiang C, Yu Z: **Development of EST-SSRs in *Cucumis sativus* from sequence database.** Mol Ecol Notes 2006, 6:1234–1236.

Kong Q, Xiang C, Yu Z, Zhang C, Liu F, Peng C, Peng X: **Mining and charactering microsatellites in *Cucumis melo* expressed sequence tags from sequence database.** Mol Ecol Notes 2007, 7:281–283.

Krauze-Baranowska M, Cisowski W: **Flavonoids from some species of the genus *Cucumis*.** Biochem Syst Ecol 2001, 29(3):321–324.

Krawinkel MB, Keding GB: **Bitter gourd (*Momordica charantia*): a dietary approach to hyperglycemia.** Nutr Rev 2006, 64:331–337.

Kristkova E, Lebeda A: **Resistance in *C. pepo* and *C. maxima* germplasm to watermelon mosaic potyvirus-2.** Plant Genet Resour Newsl 2000, 121:47–52.

Lebeda A, Kristkova E, Dolezal K: **Peroxidase isozyme polymorphism in *Cucurbita pepo* cultivars with various morphotypes and different level of field resistance to powdery mildew.** Sci Hort 1999, 81:103–112.

Lebeda A, Widrlechner MP, Staub JE, Ezura H, Zalapa JE, Krístková H: **Cucurbits (*Cucurbitaceae*; *Cucumis* spp., *Cucurbita* spp., *Citrullus* spp.).** In Genetic resources, chromosome engineering, and crop improvement. Edited by Singh RJ. Taylor & Francis Group, Boca Raton, FL, USA; 2006:271–376.

Lee SJ, Shin JS, Park KW, Hong YP: Detection **of genetic diversity using RAPD-PCR and sugar analysis in watermelon [*Citrullus lanatus* (Thunb.) Mansf.] germplasm**. Theor Appl Genet 1996, 92: 719–725.

Leppik EE: **Searching gene centers of the genus *Cucumis* through host-parasite relationship.** Euphytica 1966, 15:323–328.

Levi A, Thomas CE: **Polymorphisms among chloroplast and mitochondrial genomes of *Citrullus* species and subspecies.** Genet Resour Crop Evol 2005, 52: 609–617.

Levi A, Thomas CE, Wehner TC, Zhang XP: **Low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon.** HortScience 2001a, 36:1096–1101.

Levi A, Thomas CE, Keinath AP, Wehner TC: **Genetic diversity among watermelon (*Citrullus lanatus* and *Citrullus colocynthis*) accessions.** Genet Resour Crop Evol 2001b, 48:559–566.

Levi A, Thomas CE, Newman M, Reddy OUK, Zhang X, Xu Y: **ISSR and AFLP markers differ among American watermelon cultivars with limited genetic diversity.** J Am Soc Hort Sci 2004, 129(4):553–558.

Levi A, Thomas CE, Simmons AM, Thies JA: **Analysis based on RAPD and ISSR markers reveals closer similarities among *Citrullus* and *Cucumis* species than with *Praecitrullus fistulosus* (Stocks) Pangalo.** Genet Resour Crop Evol 2005, 52: 465–472.

Levi A, Thies J, Ling K, Simmons AM, Kousik C, Hassell R: **Genetic diversity among *Lagenaria siceraria* accessions containing resistance to root-knot nematodes, whiteflies, ZYMV or powdery mildew.** Plant Genet Resour 2009, 7: 216–226.

Liou PC, Chang YM, Hsu WS, Cheng YH, Chang HR, Hsiao CH: **Construction of a linkage map in *Cucumis melo* L. using random amplified polymorphic DNA markers.** In Int Symp on Biotechnology of Tropical and Subtropical Species, part 2. Acta

Horticulturae 461. Edited by Drew RA. International Society for Horticultural Science, Brisbane, Australia; 1998:123–131.

Lira-Saade R: *Cucurbita* **L. Estudios taxonómicos y ecogeográficos de las** *Cucurbitaceae* **latinoamericanas de importancia económica.** Systematics and Ecogeographic Studies on Crop Genepools. International Plant Genetic Resources Institute, Rome, Italy; 1995, 9: 281.

Lira-Saade R: **Chayote.** *Sechium edule* **(Jacq.) Sw. Promoting the conservation and use of underutilized and neglected crops.** Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany/International Plant Genetic Resources Institute. Rome, Italy; 1996, 8:58 p.

Lira-Saade R, Montes-Hernández S: **Cucurbits (***Cucurbita* **spp.).** In Neglected crops: 1492 from a different perspective. Edited by Hernández JE, León J. Plant Production and Protection Series. FAO, Rome, Italy; 1994, 26:63–77.

López-Sesé AI, Staub J, Katzir N, Gómez-Guillamón ML: **Estimation of between and within accession variation in selected Spanish melon germplasm using RAPD and SSR markers to assess strategies for large collection evaluation.** Euphytica 2002, 127: 41–51.

López-Sesé AI, Staub JE, Gómez-Guillamón ML: **Genetic analysis of Spanish melon (***Cucumis melo* **L.) germplasm using a standardized molecular-marker array and geographically diverse reference accessions.** Theor Appl Genet 2003, 108(1): 41–52.

Lotti C, Albo M, Ricciardi L, Conversa G, Elia A: **Genetic diversity in 'Carosello' and 'Barattiere' ecotypes (***Cucumis melo* **L.).** Colture Protette 2005 N5 (Suppl): 44–46.

Luan F, Delannay I, Staub JE: **Chinese melon (***Cucumis melo* **L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns.** Euphytica 2008, 164: 445–461.

Maggs-Kolling G, Madsen S, Christiansen JL: **A phenetic analysis of morphological variation in *Citrullus lanatus* in Namibia.** Genet Resour Crop Evol 2000, 47:385–393.

Mallick MFR, Masui M: **Origin, distribution and taxonomy of melons**. Sci Hort 1986, 28: 251–261.

Marco CF, Aranda MA, Montoro T, Gomez-Guillamon ML: **Evaluation of several accessions and wild relatives of *Cucumis melo* against cucumber vein yellowing virus (CVYV).** Cucurbit Genet Coop Rep 2003, 26: 7–8.

Marr KL, Mei XY, Bhattarai NK: **Allozyme, morphological and nutritional analysis bearing on the domestication of *Momordica charantia* L. (*Cucurbitaceae*).** Econ Bot 2004, 58: 435–455.

Marr KL, Bhattarai NK, Xia YM: **Allozymic, morphological, and phonological diversity in cultivated *acutangula* (*Cucurbitaceae*) from China, Laos, and Nepal, and allozyme divergence between *L. acutangula* and *L. aegyptiaca.*** Econ Bot 2005, 59(2): 154–165.

Marsh DB: **Evaluation of *Cucumis metuliferus* as a speciality crop for Missouri.** New crops. Proc 2nd Natl Symp on New Crops, Exploration, Research and Commercialization, Indianapolis, IN, USA; 6–9 October 1991–1993: 558–559.

Maynard DN: **Variation among tropical pumpkin (*Cucurbita moschata*) cultivars in susceptibility to silver leaf.** Cucurbit Genet Coop Rep 2001, 24:71–72.

McCreight JD, Staub JE: **Indo–US Cucumis germplasm expedition**. HortScience 1993, 28:467.

McCreight JD, Nerson H, Grumet R: **Melon, *Cucumis melo* L.** In Genetic improvement of vegetable crops. Edited by Kallos G, Bergh BO. Pergamon Press, Oxford (GB); 1993: 267–294.

Meeuse AD: **The Cucurbitaceae of Southern Africa.** Bothalia 1962, 8:1–111.

Meglic V, Serquen F, Staub JE: **Genetic diversity in cucumber (*Cucumis sativus* L.): I. A reevaluation of the US. germplasm collection.** Genet Resour Crop Evol 1996, 43: 533–546.

Meng XD, Wei YY, Ma H, Zhang WH, Li JR: **Identification of Chinese wax gourd and chiehqua cultivars using RAPD markers.** Acta Agri Shanghai 1996, 12: 45–49.

Merrick LC: **Systematics and evolution of a domesticated squash, *Cucurbita argyrosperma*, and its wild and weedy relatives.** In Biology and utilization of the *Cucurbitaceae*. Edited by Bates DM, Robinson RW, Jeffrey C. Cornell Univ Press, Ithaca, NY, USA; 1990:77–95.

Mliki A, Staub JE, Zhangyong S, Ghorbel A: **Genetic diversity in melon (*Cucumis melo* L.): An evaluation of African germplasm.** Genet Resour Crop Evol 2001, 48: 587–597.

Mliki A, Staub JE, Zhangyong S, Ghorbel A: **Genetic diversity in African cucumber (*Cucumis sativus* L.) provides potential for germplasm enhancement.** Genet Resour Crop Evol 2003, 50: 461–468.

Modolo VA, da Costa CP: **Avaliação de linhagens de maxixe paulista cultivadas em canteiros com cobertura de polietileno.** Hort Bras 2003, 21(3): 534–538 (in Portuguese).

Mohanty A, Martin JP, Aguinagalde I: **A population genetic analysis of chloroplast DNA in wild populations of *Prunus avium* L. in Europe.** Heredity 2001, 87: 421–427.

Monforte AJ, Garcia-Mas J, Arús P: **Genetic variability in melon based on microsatellite variation.** Plant Breed 2003, 122: 153–157.

Montes-Hernández S, Eguiarte LE: **Genetic structure and indirect estimates of gene flow in three taxa of *Cucurbita* (*Cucurbitaceae*) in western Mexico**. Am J Bot 2002, 89(7):1156–1163.

Montes-Hernández S, Merrick LC, Eguiarte LE: **Maintenance of squash (*Cucurbita* spp.) landrace diversity by farmers' activities in Mexico.** Genet Resour Crop Evol 2005, 52(6):697–707.

Morales M, Roig E, Monforte A J, Arús P, Garcia-Mas J: **Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.).** Genome 2004, 47: 352–360.

Morimoto Y, Maundu P, Fujimaki H, Morishima H: **Diversity of landraces of the white-fl owered gourd (*Lagenaria siceraria*) and its wild relatives in Kenya: fruit and seed morphology.** Genet Resour Crop Evol 2005, 52(6): 737–747.

Mo-Suk Y, Im-Sung H, Go-Gawn D, Ann-Chong M, Kim-Doo H, Mo-Suk Y, Im SH: **RAPD analsysis of genetic diversity of melon species.** Kor J Hort Sci Technol 1999, 16:21–24.

Munger HM, Robinson RW: **Nomenclature of *Cucumis melo* L.** Cucurbit Genet Coop Rep 1991, 14:43–44.

Mutschleer MA, Pearson OH: **The origin, inheritance, and instability of Butternut squash (*Cucurbita moschata* Duchesne).** HortScience 1987, 22(4):535–539.

Nakata E, Staub JE, López-Sesé AI, Katzir N: **Genetic diversity of Japanese melon cultivars (*Cucumis melo* L.) as assessed by random amplified polymorphic DNA and simple sequence repeat markers.** Genet Resour Crop Evol 2005, 52:405–419.

Naudin C: **Nouvelles recherches sur les caractères spécifi ques et les variétes des plantes du genre *Cucurbita*.** Ann Sci Nat IV 1856, 6:5–73 (in French).

Naudin C: **Essais dune monographiedes espèces et des varieties du genre *Cucumis*.** Ann Sci Nat 1859, 11:5–87 (in French).

Navot N, Zamir D: **Isozyme and seed protein phylogeny of the genus *Citrullus* (*Cucurbitaceae*).** Plant Syst Evol 1987, 156: 61–67.

Nee M: **The domestication of *Cucurbita* (*Cucurbitaceae*).** Econ Bot 1990, 44(3):56–68.

Nerson H, Paris HS, Paris MP: **Fruit shape, size and seed yield in *Cucurbita pepo*.** In Proc of Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:227–230.

Neuhausen SL: **Evaluation of restriction fragment length polymorphisms in *Cucumis melo*.** Theor Appl Genet 1992, 83:379–384.

Nieto C, Piron F, Dalmais M, Marco CF, Moriones E, Gómez-Guillamón ML, Truniger V, Gómez P, Garcia-Mas J, Aranda MA, Bendahmane A: **EcoTILLING for the identification of allelic variants of melon *eIF4E*, a factor that controls virus susceptibility.** BMC Plant Biol 2007, 7:34, doi: 10.1186/1471-2229-7-34.

Nikolova V, Alexandrova M, Stoeva V: **Possibilities for the use of remote hybridization in the genus *Cucumis* for the development of genetic diversity**. Acta Hort 2002, 579: 39–43.

Nimmakayala P, Tomason YR, Jeong J, Vajja G, Levi A, Gibson P, Reddy UK: **Molecular diversity in the Ukrainian melon collection as revealed by AFLPs and microsatellites.** Plant Genet Resour 2009, 7:127–134.

Nuez F, Anastasio G, Cortés C, Cuartero J, Gómez-Guillamón ML, Costa J: **Germplasm resources of *Cucumis melo* L from Spain.** Cucurbit Genet Coop Rep 1986, 9: 60–63.

Nuez F, Ferrando C, Díez MJ, Costa J, Catalá MS, Cuartero J, Gómez-Guillamón ML: **Collecting *Cucumis melo* L. in Spain**. Cucurbit Genet Coop Rep 1988, 11:54–56.

Oliszeweski N: **Archaeobotany of mound structures in Campo del Pucará, Catamarca, Argentina (1750–1450 b.p.): ceremonial use or rubbish dumps?** Veg Hist Archaeobot 2005, 14(4):465–471.

Oliveira ANP, Oliveira AP, Leonardo FAP, Cruz IS, Silva DF: **Yield of gherkin in response to doses of bovine manure.** Hort Brasil 2009, 27(1):100–102.

Oliver M, Garcia-Mas J, Cardus M, Pueyo N, López-Sesé AI, Arroyo M, Gómez-Paniagua H, Arús P, De Vicente MC: **Construction of a reference linkage map for melon.** Genome 2001, 44:836–845.

Onto S, Laosat N, Suksawat W, Popluechai S, Eungwanichayapant PD, Chukeatirote E: **Phylogenetic analysis of *Cucumis sativus* using RAPD molecular markers**. J Plant Sci 2008, 3(1):105–110.

Pandey S, Kumar S, Mishra U, Rai A, Singh M, Rai M: **Genetic diversity in Indian ash gourd (*Benincasa hispida*) accessions as revealed by quantitative traits and RAPD markers.** Sci Hort 2008, 118:80–86.

Pangalo KI: **A new species of cultivated pumpkin**. Bull Appl Bot Genet Plant Breed 1930, 23:253–265.

Parducci L, Szmidt AE: **PCR-RFLP of cpDNA in the genus *Abies***. Theor Appl Genet 1999, 98:802–808.

Paris HS: **A proposed subspecific classification for *Cucurbita pepo*.** Phytologia 1986, 61:133–138.

Paris HS: **Historical records, origins, and development of the edible cultivar groups of *Cucurbita pepo* (*Cucurbitaceae*).** Econ Bot 1989, 473(4):423–443.

Paris HS: **Some observations concerning diversity in the subspecies and horticultural groups of *Cucurbita pepo*.** Cucurbit Genet Coop Rep 1998, 21:51–53.

Paris HS: **History of the cultivar-groups of *Cucurbita pepo***. In Horticulture Review. Edited by Janick J. John Wiley, New York, USA; 2000, 25:71–170.

Paris HS: **Characterization of the *Cucurbita pepo* collection at the Newe Yaár Research Center, Israel.** Plant Genet Resour Newsl 2001a, 126:41–45.

Paris HS: **History of the cultivar-groups of *Cucurbita pepo*.** Hort Rev 2001b, 25: 71–170.

Paris HS: **Summer squash**. In Handbook of plant breeding, volI: Vegetables I. Edited by Prohens J, Nuez F.Springer, New York, USA; 2008:351–379.

Paris HS, Nerson H: **Association of seed size and dimensions with fruit shape in *Cucurbita pepo*.** In *Cucurbitaceae* 1998, Evaluation and enhancement of Cucurbit germplasm. Edited by McCreight JD. Alexandria, Va, USA; 1998:230–234.

Paris HS, Yonash N, Portnoy V, Mozes-Daube N, Tzuri G, Katzir N: **Assessment of genetic relationships in *Cucurbita pepo* (*Cucurbitaceae*) using DNA markers.** Theor Appl Genet 2003, 106:971–978.

Park YH, Sensoy S, Wye C, Antonise R, Peleman J, Havey MJ: **A genetic map of cucumber composed of RAPDs, RFLPs, AFLPs, and *loci* coditioning resistance to papaya ringspot and zucchini yellow mosaic viruses.** Genome 2000, 43:1003–1010.

Parkash C, Singh KP, Kalloo G: **Variability analysis and cause and effect relationship in ash gourd [*Benincasa hispida* (Thunb.) Cogn.].** Indian J Plant Genet Res 2000, 13:298–301.

Périn C, Hagen S, De Conto V, Katzir N, Danin-Poleg Y, Portnoy V, Baudracco-Arnas S, Chadoeuf J, Dogimont C, Pitrat M: **A reference map of *Cucumis melo* based on two recombinant inbred line populations.** Theor Appl Genet 2002, 104:1017–1034.

Perl-Treves R, Galun E: **The *Cucumis* plastome: physical map, intrageneric variation and phylogenetic relationships.** Theor Appl Genet 1985, 71:417–429.

Perl-Treves R, Zamir D, Navot N, Galun E: **Phylogeny of *Cucumis* based on isozyme variability and its comparison with plastome phylogeny**. Theor Appl Genet 1985, 71:430–436.

Petersen JB, Sidell NA: **Mid-Holocene evidence of *Cucurbita* sp. from central Maine.** Am Antiq 1996, 61:685–698.

Picó B, Sifres A, Esteras C, Nuez F: *Cucumis* **SSRs markers applied to the study of the genetic diversity in the *Cucurbita* genus.** Cucurbit Genet Coop Rep 2005-2006, 28–29: 70–72.

Piperno DR, Andres TC, Stothert KE: **Phytolits in *Cucurbita* and other neotropical *Cucurbitaceae* and their occurrence in early archaeological sites from the lowland American tropics.** J Archeol Sci 2000, 27:193–208.

Pitrat M: **Melon (*Cucumis melo* L.).** In Handbook of crop breeding, vol I: Vegetables. Edied by Prohens J, Nuez F. Springer, New York, USA; 2008:283–315.

Pitrat M, Chauvet M, Foury C: **Diversity, history, and production of cultivated cucurbits.** Acta Hort 1999, 492:21–28.

Pitrat M, Hanelt P, Hammer K: **Some comments on infraspecific classification of cultivar of melon**. In Proc of Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel, 19–23 March 2000:29–36.

Poe RP, Coyne DP, Swisher BA, Clegg MD: **Differential *Cucurbita* spp. Tolerance to the herbicide trifluralin.** J Am Soc Hort Sci 1988, 113(1):35–40.

Puchalski JT, Robinson RW: **Electrophoretic analysis of isozymes in *Cucurbita* and *Cucumis* and its application for phylogenetic studies.** In Biology and utilization of the *Cucurbitaceae*. Edited by Bates DM, Robinson RW, Jeffrey C. Cornell Univ Press, Ithaca, New York, USA; 1990:60–76.

Ramos SRR: **Genetic diversity based on AFLP molecular markers and indicators for the establishment of a core collection for pumpkin (*Cucurbita moschata*) for north-east Brazil.** Plant Genetic Resources Newsletter 2007, Thesis Abstr 149(43)145: 66–67.

Renner SS, Schaefer H, Kocyan A: **Phylogenetics of *Cucumis* (*Cucurbitaceae*): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*).** BMC Evol Biol 2007, 7:58, doi: 10.1186/1471-2148-7-58.

Ríos H, Fernández A, Batista O: **Cuban pumpkin genetic variability under low input conditions.** Cucurbit Genet Coop Rep 1997, 20:48–49.

Ritschel PS, Lins TCL, Tristan RL, Buso GSC, Buso JA, Ferreira ME: **Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.).** BMC Plant Biol 2004, 4:9.

Robinson RW, Decker-Walters DS: **Cucurbits**. Crop Production Science in Horticulture. CAB Int, New York, USA; 1997, 6: 226 p.

Robinson RW, Decker-Walters DS: **Cucurbits**. CABI Publ, Wallingford, Oxford, UK; 1999:226 p.

Romao RL: **Northeast Brazil: A secondary center of diversity for watermelon (*Citrullus lanatus*).** Genet Resour Crop Evol 2000, 47:207–213.

Romero-Rodriguez MA, Vazquez-Oderiz ML, Lopez-Hernandez J, Simal-Lozano J: **Physical and analytical characteristics of the kiwano**. J Food Compos Anal 1992, 5(4):319–322.

Rubatzky VE: **Origin, distribution and uses.** In Watermelons. Characteristics, production and marketing. Edited by Maynard DN. ASHS Press, Alexandria, Va, USA; 2001:21–26.

Rubatzky VE, Yamaguchi M: **World vegetable principles, production and nutritive values.** 2nd edn. Chapman and Hall, International Thompson Publ, New York, USA; 1997:853 p.

Rubatazky VE, Yamaguchi M: **World Vegetables**. Chapman and Hall, New York, USA; 1999.

Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L: **Phylogenetic relationships among domesticated and wild species of *Cucurbita* (*Cucurbitaceae*) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin.** Proc Nat Acad Sci 2002, 99:535–540.

Sanwal SK, Yadav RK, Singh PK, Rai N: **Variability and genetic diversity studies in indigenous chow-chow genotypes of northeast India.** Indian J Hort 2008, 65(2):167–170.

Schaefer H: *Cucumis* (*Cucurbitaceae*) **must include** *Cucumella, Dicoelospermum, Mukia, Myrmecosicyos,* **and** *Oreosyce***: a recircumscription based on nuclear and plastid DNA data.** Blumea 2007, 52:165–177.

Schaefer H, Renner SS: **A three-genome phylogeny of** *Momordica* (*Cucurbitaceae*) **suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia.** Mol Phylogenet Evol 2010, 54(2):553–560.

Sebastian P, Schaefer H, Telford IR, Renner SS: **Cucumber (***Cucumis sativus***) and melon (***C. melo***) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia.** Proc Natl Acad Sci 2010, 107(32):14269–14273.

Sensoy S, Buyukalaca S, Abak K: **Evaluation of genetic diversity in Turkish melons (***Cucumis melo* **L.) based on phenotypic characters and RAPD markers**. Genet Resour Crop Evol 2007, 54:1351–1365.

Silberstein L, Kovalski I, Huang R, Anagnostou K, Jahn MK, Perl-Treves R: **Molecular variation in melon (***Cucumis melo* **L.) as revealed by RFLP and RAPD markers.** Sci Hort 1999, 79:101–111.

Singh DK: **Genetic analysis of yield and its components in ash gourd [***Benincasa hispida* **(Thunb.) Cogn.].** PhD Thesis. UP College, Varanasi, India. 2002.

Singh NP, Matta NK: **Variation studies on seed storage proteins and phylogenetics of the genus** *Cucumis***.** Plant Syst Evol 2008, 275:209–218.

Smith BD: **The initial domestication of** *Cucurbita pepo* **in the Americas 10.000 years ago.** Science 1997, 276:932–934.

Smith BD: **Reassessing Coxcatlan Cave and the early history of domesticated plants in Mesoamerica.** Proc Natl Acad Sci USA 2005, 102(27):9438–9445.

Stachel M, Csanádi G, Vollmann J, Lellet T: **Genetic diversity in pumpkins (*Cucurbita pepo* L.) as revealed in inbred lines using RAPD markers.** Cucurbit Genet Coop Rep 1998, 21:48–50.

Staub JE, Meglic V: **Molecular genetic markers and their relevance for cultivar discrimination: A case study in cucumber**. HortTechnology 1993, 3:291–299.

Staub JE, Bacher J: **Cucumber as a processed vegetable.** In Processing vegetables: Science and Technology IV. Edited by Smith DS, Cash JN, Nip W, Hui YH. Technomic Publishing Co, Inc Lancaster, PA, USA; 1997:129–193.

Staub JE, Ivandic V: **Genetic assessment of the United States National cucumber collection.** In Proc of Cucurbitaceae 2000, Acta Hort 510. Edited by Katzir N, Paris HS. Ma'aleh Hahamisha, Israel; 19–23 March 2000:113–122.

Staub JE, Kupper RS, Schuman D, Wehner TC, May B: **Electrophoretic variation and enzyme storage stability in cucumber.** J Am Soc Hort Sci 1985, 110: 426–431.

Staub JE, Fredrick L, Marty T: **Electrophoretic variation in cross-compatible wild diploid species of *Cucumis*.** Can J Bot 1987, 65:792–798.

Staub JE, Box J, Meglic V, Horejsi TF, McCreight JD: **Comparison of isozyme and random amplified polymorphic DNA data for determining intraspecific variation in *Cucumis*.** Genet Resour Crop Evol 1997, 44:257–269.

Staub JE, Serquen FC, Horejsi T, Chen J: **Genetic diversity in cucumber (*Cucumis sativus* L.): IV. An evaluation of Chinese germplasm.** Genet Resour Crop Evol 1999, 46:297–310.

Staub JE, Danin-Poleg Y, Fazio G, Horejsi T, Reis N, Katzir N: **Comparative analysis of cultivated melon groups (*Cucumis melo* L.) using random amplified polymorphic DNA and simple sequence repeat markers.** Euphytica 2000, 115:225–241.

Staub JE, López-Sesé I, Fanourakis N: **Diversity among melón landraces (*Cucumis melo* L.) from Greece and their genetic relationships with other melon germplasm of diverse origins.** Euphytica 2004, 136:151–166.

Staub JE, Chung S-M, Fazio G: **Conformity and genetic relatedness estimation in crop species having a narrow genetic base: the case of cucumber (*Cucumis sativus* L.).** Plant Breed 2005, 124:44–53.

Staub JE, Robbins MD, Wehner TC: **Cucumber**. In Handbook of plant breeding, vol I: Vegetables I. Edited by Prohens J, Nuez F. Springer, New York, USA; 2008:241–282.

Stepansky A, Kovalski I, Perl-Treves R: **Intraspecific classification of melons (*Cucumis melo* L.) in view of their phenotypic and molecular variation.** Plant Syst Evol 1999, 217:313–332.

Stift G, Zraidi A, Lelley T: **Development and characterization of microsatellite markers (SSR) in *Cucurbita* species.** Cucurbit Genet Coop Rep 2004, 27:61–65.

Sureja AK, Sirohi PS, Behera TK, Mohapatra T: **Molecular diversity and its relationship with hybrid performance and heterosis in ash gourd [*Benincasa hispida* (Thunb.) Cogn.].** J Hort Sci Biotechnol 2006, 81(1):33–38.

Szabó Z, Gyulai G, Humphreys M, Horvath L, Bittsansky A, Lagler R, Heszky L: **Genetic variation in melon (*C. melo*) compared to an extinct landrace from the Middle Ages (Hungary**). I. RDNA, SSR, and SNP analysis of 47 cultivars. Euphytica 2005, 146:87–94.

Szabó Z, Gyulai G, Tóth Z, Heszky L: **Morphological and molecular diversity of 47 melon (*Cucumis melo*) cultivars compared to an extinct landrace excavated from the 15th**

**century.** In *Cucurbitaceae* 2008, Proc IXth EUCARPIA Meeting on Genetics and Breeding of Cucurbitaceae. Edited by Pitrat M. INRA, Avignon, France; 21–24 May 2008:313–321.

Tabei Y: **Study on breeding of *Cucurbitaceae* using biotechnology**. Bull Natl Inst Agrobiol Resour 1997, 11:1–107.

Tanaka K, Nishitani A, Akashi Y, Sakata Y, Nishida H, Yoshino H, Kato K: **Molecular characterization of South and East Asian melon, *Cucumis melo* L., and the origin of Group Conomon var. makuwa and var. conomon revealed by RAPD analysis.** Euphytica 2007, 153:233–247.

Telford IR: *Cucurbitaceae*. Flora Aust 1982, 205:158–198.

Teppner H: **Notes on *Lagenaria* and *Cucurbita* (*Cucurbitaceae*)-review and new contributions.** Phyton 2004, 44(2):245–308.

Thulin M, Al-Gifri AN: *Cucumis canoxyi* (*Cucurbitaceae*): a new species from Yemen.** Nord J Bot 1994, 14:315–317.

Tolentino MIS, Laude RP, dela Viña AC: **Genetic diversity analysis of *Luffa* species based on seed protein profile using SDS-PAGE.** Phil J Crop Sci 1997, 22(3): 141–146.

Torres Ruiz RA, Hemleben V: **Use of ribosomal DNA spacer probes to distinguish cultivars of *Cucurbita pepo* L. and other *Cucurbitaceae*.** Euphytica 1991, 53: 11–17.

Trumbull JH: **Vegetables cultivated by the American Indians**. Bull Torrey Bot Club 1876, 6:69–71.

Verma M, Arya L: **Development of EST-SSRs in watermelon (*Citrullus lanatus* var. *lanatus*) and their transferability to *Cucumis* spp.** J Hort Sci Biotechnol 2008, 83(6):732–736.

Verma VK, Behera TK, Munshi AD, Parida SK, Mohapatra T: **Genetic diversity of ash gourd [*Benincasa hispida* (Thunb.) Cogn.] inbred lines based on RAPD and ISSR markers and their hybrid performance.** Sci Hort 2007, 113: 231–237.

Walters SA, Wehner TC: **Incompatibility in diploid and tetraploid crosses of *Cucumis sativus* and *Cucumis metuliferus*.** Euphytica 2002, 128(3):371–374.

Walters TW: **Historical overview on domesticated plants in China with special emphasis on the *Cucurbitaceae*.** Econ Bot 1989, 43:297–313.

Walters TW, Decker-Walters DS, Posluszny U, Kevan PG: **Determination and interpretation of comigrating allozymes among genera of the *Benincaseae* (*Cucurbitaceae*).** Syst Bot 1991, 16:30–40.

Wang YH, Thomas CE, Dean RA: **A genetic map of melon (*Cucumis melo* L.) based on amplified fragment length polymorphism (AFLP) markers.** Theor Appl Genet 1997, 95:791–798.

Wasylikowa K, van der Veen M: **An archaeobotanical contribution to the history of watermelon, *Citrullus lanatus* (Thunb.) Matsum. & Nakai (syn. *C. vulgaris* Schrad.).** Veg Hist Archaeobot 2004, 13:213–217.

Watcharawongpaiboon N, Chunwongse J: **Development of microsatellite markers from an enriched genomic library of pumpkin (*Cucurbita moschata* L.).** Songklanakarin J Sci Technol 2007, 29(5):1217–1223.

Watcharawongpaiboon N, Chunwongse J: **Development and characterization of microsatellite markers from an enriched genomic library of cucumber (*Cucumis sativus*).** Plant Breed 2008, 127: 74–81.

Wehner TC: **Watermelon**. In Handbook of plant breeding, volI: Vegetables I. Edited by Prohens J, Nuez F. Springer, New York, USA; 2008:381–418.

Wessel-Beaver L**: Powdery and downy mildew resistance in *Cucurbita moschata* accessions.** Cucurbit Genet Coop Rep 1993, 16:73–74.

Wessel-Beaver L: **Evidence for the cener of diversity of *Cucurbita moschata* in Colombia.** Cucurbit Genet Coop Rep 2000, 23:54–55.

Whitaker TW: **American origin of the cultivated cucurbits.** Ann MO Bot Gard 1947, 34:101–111.

Whitaker TW: **Endemism and pre-Columbian migration of the bottle gourd, *Lagenaria siceraria* (Mol.) Standl.** In Man across the sea. Edited by Kelley JC, Pennington CW, Rands RL. Univ of Texas Press, Austin, TX, USA; 1971:320–327.

Whitaker TW, Davis GN: **Cucurbit, botany, cultivation and utilization.** Interscience Publ, New York, USA; 1962: 249 p.

Whitaker TW, Bemis WP: **Cucurbits, *Cucumis*, *Citrullus*, *Cucurbita*, *Lagenaria* (*Cucurbitaceae*).** In Evolution of crop plants. Edited by Simmonds NW. Longrams, NewYork, USA; 1976:64–69.

Wilson HD, Doebley J, Duvall M: **Chloroplast DNA diversity among wild and cultivated members of *Cucurbita* (*Cucurbitaceae*).** Theor Appl Genet 1992, 84:859–865.

Wu T, Zhou J, Zhang Y, Cao J: **Characterization and inheritance of a bush-type in tropical pumpkin (*Cucurbita moschata* Duchesne).** Sci Hort 2007, 114:1–4.

Yanev Z, Shabelsky E, Schafferman D: **Colocynth: potential arid land oil seed from an ancient cucurbit.** In Perspectives on new crops and new uses. Edited by Janick J. ASHS Press, Alexandria, Va, USA; 1999:257–261.

Yang Y-W, Tai P-Y, Chen Y, Li W-H: **A study of the phylogeny of *Brassica rapa*, *B. nigra*, *Raphanus sativus* and their related genera using noncoding regions of chloroplast DNA.** Mol Phylogenet Evol 2002, 23:268–275.

Yashiro K, Iwata H, Akashi Y, Tomita K, Kuzuya M, Tsumura Y, Kato K: **Genetic relationship among East and South Asian melon (*Cucumis melo* L.) revealed by AFLP analysis.** Breed Sci 2005, 55:197–206.

Yi SS, Akashi Y, Tanaka K, Cho TT, Khaing MT, Yoshino H, Nishida H, Yamamoto T, Win K, Kato K: **Molecular analysis of genetic diversity in melon landraces (*Cucumis melo* L.) from Myanmar and their relationship with melon germplasm from East and South Asia**. Genet Resour Crop Evol 2009, 56:1149–1161, doi 10.1007/s10722-009-9438-y.

Youn SJ, Chung HD: **Genetic relationship among the local varieties of the Korean native squashes (*Cucurbita moschata*) using RAPD technique**. J Kor Soc Hort Sci 1998, 39(5):517–521.

Younis YMH, Ghirmay S, Al-Shihry SS: **African *Cucurbita pepo* L.: properties of seed and variability in fatty acid composition of seed oil.** Phytochemistry 2000, 54: 71–75.

Zamir D, Navot N, Rudich J: **Enzyme polymorphism in *Citrullus lanatus* and *C. colocynthis* in Israel and Sinai.** Plant Syst Evol 1984, 146: 163–137.

Zeven AC, Zhukovsky PM: **Dictionary of cultivated plants and their centres of diversity.** Centre for Agriculture Publishing and Documentation, Wageningen, The Netherlands; 1975:219–244.

Zhang XP, Rhodes BB, Skorupska HS: **RAPD molecular markers in watermelon.** Cucurbit Genet Coop Rep 1994, 17:116–119.

Zhou Y, Zhou J, Huang L, Ding X, Shi K, Yu Ji: **Grafting of *Cucumis sativus* onto *Cucurbita ficifolia* leads to improved plant growth, increased light utilization and reduced accumulation of reactive oxygen species in chilled plants.** J Plant Res 2009, 122(5):529–540.

Zhuang FY, Chen JF, Staub JE, Qian CT: **Assessment of genetic relationships among** *Cucumis* **ssp. by SSR and RAPD marker analysis**. Plant Breed 2004, 123:167–172.

Zhukovsky P: **Agricultural structure of Turkey (Anatolia).** Türkiye Seker Fab.AS.Yay.20. 1951(in Turkish).

Zohary D, Hopf M: **Domestication of plants in the Old World**. Oxford Univ Press, Oxford, UK; 2000.

Zraidi A, Stift G, Pachner M, Shojaeiyan A, Gong L, Lelley T: **A consensus map for** *Cucurbita pepo*. Mol Breed 2007, 20:375–388.

# II. OBJETIVOS

## II. OBJETIVOS

El principal objetivo del trabajo realizado durante estos años ha sido el desarrollo de herramientas moleculares para dos especies cucurbitáceas de gran importancia económica, *Cucumis melo* y *Cucurbita pepo*, con el fin de impulsar la mejora eficiente de estos cultivos para características de calidad y resistencia a enfermedades. Para ello se han abordado los siguientes objetivos específicos:

-Desarrollo inicial de marcadores tipo SNP en melón a partir de los primeros ESTs disponibles para la especie (desarrollados con la metodología de Sanger) y su empleo para la saturación del mapa genético de melón.

-Obtención y caracterización del transcriptoma de melón e identificación de grandes colecciones de marcadores de alta calidad de tipo SSR y SNP a partir de estas secuencias.

-Aplicación de marcadores tipo SSR, identificados *in silico* en el primer transcriptoma de calabaza, en una colección de cultivares europeos, que incluyen tipos comerciales y tradicionales españoles.

-Aplicación de marcadores tipo SNP en calabaza mediante plataformas de genotipado masivo y construcción del primer mapa genético para el género basado en SNPs.

# III. DESARROLLO Y APLICACIÓN DE HERRAMIENTAS GENÓMICAS EN MELÓN

# III-1. A set of EST-SNPs for map saturation and cultivar identification in melon

Deleu W*, Esteras C*, Roig C, González-To M, Fernández-Silva I, Gonzalez-Ibeas D, Blanca J, Aranda MA, Arús P, Nuez F, Monforte AJ, Picó B, Garcia-Mas J: **A set of EST-SNPs for map saturation and cultivar identification in melon.** BMC Plant Biology 2009, 9:90, doi:10.1186/1471-2229-9-90.

*misma contribución

**Abstract**

Background: There are few genomic tools available in melon (*Cucumis melo* L.), a member of the *Cucurbitaceae*, despite its importance as a crop. Among these tools, genetic maps have been constructed mainly using marker types such as Simple Sequence Repeats (SSRs), Restriction Fragment Length Polymorphisms (RFLPs) and Amplified Fragment Length Polymorphisms (AFLPs) in different mapping populations. There is a growing need for saturating the genetic map with Single Nucleotide Polymorphisms (SNPs), more amenable for high throughput analysis, especially if these markers are located in gene coding regions, to provide functional markers. Expressed Sequence Tags (ESTs) from melon are available in public databases, and resequencing ESTs or validating SNPs detected *in silico* are excellent ways to discover SNPs.

Results: EST-based SNPs were discovered after resequencing ESTs between the parental lines of the PI 161375 (SC) × 'Piel de sapo' (PS) genetic map or using *in silico* SNP information from EST databases. In total 200 EST-based SNPs were mapped in the melon genetic map using a bin-mapping strategy, increasing the map density to 2.35 cM/marker. A subset of 45 SNPs was used to study variation in a panel of 48 melon accessions covering a wide range of the genetic diversity of the species. SNP analysis correctly reflected the genetic relationships compared with other marker systems, being able to distinguish all the accessions and cultivars.

Conclusion: This is the first example of a genetic map in a cucurbit species that includes a major set of SNP markers discovered using ESTs. The PI 161375 × 'Piel de sapo' melon genetic map has around 700 markers, of which more than 500 are gene-based markers (SNPs, RFLPs and SSRs). This genetic map will be a central tool for the construction of the melon physical map, the step prior to sequencing the

complete genome. Using the set of SNP markers, it was possible to define the genetic relationships within a collection of forty-eight melon accessions as efficiently as with SSR markers, and these markers may also be useful for cultivar identification in Occidental melon varieties.

**Background**

Single Nucleotide Polymorphisms (SNPs) are the most frequent type of variation found in DNA (Brookes, 1999) and are valuable markers for high-throughput genetic mapping, genetic variation studies and association mapping in crop plants. Several methods have been described for SNP discovery (Ganal et al., 2009): SNP mining from Expressed Sequence Tag (EST) databases (Batley et al., 2003); based on array hybridization (Borevitz et al., 2003) or amplicon resequencing (Choi et al., 2007); from the complete sequence of a genome (Velasco et al., 2007) and more recently, using high-throughput sequencing technologies (Barbazuk et al., 2007). The discovery of SNP markers based on transcribed regions has become a common application in plants because of the large number of ESTs available in databases, and EST-SNPs have been successfully mined from EST databases in non-model species such as Atlantic salmon (Hayes et al., 2007), catfish (Wang et al., 2008), tomato (Yamamoto et al., 2005) and white spruce (Pavy et al., 2006).

Melon (*Cucumis melo* L.) is an important crop worldwide. It belongs to the *Cucurbitaceae* family, which also includes cucumber, watermelon, pumpkin and squash. The melon genome has an estimated size of 450 Mb (Arumuganathan and Earle, 1991) and is a diploid with a basic chromosome number of x = 12. In recent years research has been carried out to increase the genetic and genomic resources for this species, such as the sequencing of ESTs (Gonzalez-Ibeas et al., 2007), the construction

of a BAC library (van Leeuwen et al., 2003), the development of an oligo-based microarray (Mascarell-Creus et al., 2009) and the development of a collection of near isogenic lines (NILs) (Eduardo et al., 2005). Genetic maps have also been reported for melon, but they have been constructed with different types of molecular markers and genetic backgrounds (Wang et al., 1997; Danin-Poleg et al., 2000; Perin et al., 2002; Gonzalo et al., 2005; Fernandez-Silva et al., 2008), making it difficult to transfer markers from one map to another. The aim of the International Cucurbit Genomics Initiative (ICuGI) (http://www.icugi.org), currently in progress, is to obtain a consensus genetic map by merging genetic maps available using a common set of SSRs as anchor markers.

A Double Haploid Line (DHL) population from the cross between the Korean accession PI 161375 (SC) and the *inodorus* type 'Piel de sapo' T111 (PS) was the basis for the construction of a genetic map with 221 codominant, transferable RFLP and SSR markers (Fernandez-Silva et al., 2008). New EST-derived SSR markers, added to this map using a bin-mapping strategy with only 14 mapping individuals, gave a new map with 296 markers distributed in 122 bins and a density of 4.2 cM/marker (Fernandez-Silva et al., 2008). There is a need for saturating the SC × PS genetic map with more markers that are amenable for large-scale genotyping, as are SNPs. In a preliminary experiment with melon, amplicon resequencing of 34 ESTs in SC and PS was used for SNP discovery, obtaining a frequency of one SNP every 441 bp and one INDEL every 1,666 bp (Morales et al., 2004). The availability of more than 34,000 melon ESTs from normalized cDNA libraries from different melon genotypes and tissues (Gonzalez-Ibeas et al., 2007) is a valuable resource for the identification of SNPs to be added to the current genetic map.

Genetic markers can also be used for variability analysis studies. In melon, there have

been several attempts to elucidate intraspecific relationships among melon germplasm, using isozyme (Staub et al., 1999), RFLP (Neuhausen, 1992), RAPD (Stepansky et al., 1999), AFLP (Garcia-Mas et al., 2000) and SSR (Monforte et al., 2003) markers, with SSRs the preferred marker for fingerprinting and genetic variability analysis in melon (Monforte et al., 2003). Due to the absence of a known set of SNPs in the species, this marker has not been compared with other types for variability analysis. It would be of special interest to have a set of these markers for a high-throughput system to identify the germplasm used in breeding programs, mainly from *inodorus* and the *cantalupensis* melon types.

The objectives of this work were to increase the marker resolution in the melon genetic map, discovering EST-SNPs in a melon EST database, and to study the performance of a subset of EST-SNPs for variability analysis in a collection of melon accessions.

**Results and discussion**

SNP discovery

Two strategies were used to discover SNPs in melon. The first was based on producing amplicons from randomly selected melon ESTs and resequencing the parental lines of the melon genetic map PI 161375 (SC) × 'Piel de sapo' T111 (PS). Primers were designed from 223 melon ESTs (Table III.1-1). After discarding primers that did not amplify a PCR product, amplicons that did not produce high quality sequences and monomorphic amplicons, 93 ESTs (56.3%) showed at least one polymorphism between SC and PS.

The second strategy was the validation of *in silico* SNPs from the ICuGI database (http://www.icugi.org). Three hundred and sixty-six *in silico* SNPs found in the database were selected, belonging to two types of SNPs: pSNP and pSCH (Table III.1-

1; see methods). Primers were designed from 269 ESTs containing pSNPs and 97 containing pSCHs. Putative *in silico* SNPs were validated in 51.8% and 21.3% of the amplicons for pSNPs and pSCHs, respectively. In some instances additional SNPs were detected in the sequenced regions, giving a slightly higher percentage of polymorphic amplicons (69.7% and 31.3% for pSNP and pSCH amplicons, respectively). From the ESTs reported by Gonzalez-Ibeas et al. (2007), 47.3% were obtained from two accessions of the 'Piel de sapo' cultivar type (Piñonet and PS), and the remainder from two genotypes, the C-35 cantaloupe accession (29.3%) and the pat81 *agrestis* accession (23.4%). The pSNPs and pSCHs were deduced from this set of EST sequences, with a high proportion found between pat81 and 'Piel de sapo', and SNPs experimentally validated after resequencing amplicons from PS and SC. SC belongs to the ssp. *agrestis* var. *conomon* as the accession pat81 but has a different origin, so, as expected not all the SNPs were conserved between SC and PS, giving a pSNP validation of 51.8%. On the other hand, only 21.3% of the pSCHs were validated, indicating that many may represent sequencing errors or mutations introduced during the cDNA synthesis procedure. The SNPs in a subset of amplicons containing *in silico* SNPs between 'Piel de Sapo' and pat81 were validated using different genotyping methods (see below) rather than resequencing in PS and SC.

A total of 368 amplicons (random and containing *in silico* SNPs) were resequenced in PS and SC and produced 177.5 kb of melon DNA, with 431 SNPs and 59 short INDELs, at an average of one SNP every 412 bp and one INDEL every 3.0 kb, (Table III.1-2). This is in agreement with the values obtained in a previous small-scale experiment using the same two melon accessions, which gave one SNP every 441 bp and one INDEL every 1.6 kb (Morales et al., 2004). SC and PS belong to the *agrestis* (*C. melo* ssp. *agrestis*) and *inodorus* (*C. melo* ssp. *melo*) melon groups, respectively,

which are two of the more distant groups in the species (Monforte et al., 2003). This may explain the relatively high frequency of SNPs between the cultivars.

SNP detection

Various detection methods were used for genotyping the SNPs in each EST. A restriction site around the SNP position, different in the parental sequences, was used to develop a CAPS marker for 103 EST-SNPs. When more than one SNP was discovered in one amplicon, we selected the most suitable SNP for detection using CAPS. When no restriction enzyme was available to produce a CAPS marker, we used the SNaPshot SNP detection system. Seventy-seven EST-SNPs were genotyped with SNaPshot. For 14 ESTs, PS and SC gave a different amplicon size, so they could be genotyped as SCAR markers. Four EST-SNPs were genotyped using DNA sequencing and two were converted into dCAPS. The SNP detection method used for each mapped EST-SNP is shown in Additional file III.1-1.

SNP variability

Forty-five SNPs (see Additional file III.1-2) were randomly chosen to study their variability in a set of melon accessions of worldwide cultivar and botanical types (see Additional File III.1-3). The *inodorus* cultivars were overrepresented in order to assess whether SNPs between distant melon accessions (SC and PS) were also variable among more closely related genotypes.

All SNPs were polymorphic and the mean major allele frequency was 0.69 (Table III.1-3). Only one SNP (AI_24-H05) had a rare allele (frequency = 0.08), whereas the frequencies of the two alleles were similar in 28 SNPs (major allele frequency < 0.65). Average gene diversity (He) was 0.4 (ranging from 0.14 to 0.5). Forty-three SNPs

yielded He > 0.20, demonstrating that most of the chosen SNPs were highly informative, as found for SNPs in rye (Varshney et al., 2007) but contrasting with crops such as soybean (Yoon et al., 2007) and wheat (Ravel et al., 2006) where SNPs yielding rare alleles are more frequent.

The mean gene diversity index for SNPs was considerably lower than the values reported for SSRs in melon (e.g. PIC = 0.58 (Fernandez-Silva et al., 2008), He = 0.66 (Monforte et al., 2003)). To ensure the difference was not due to sampling, gene diversity indexes were estimated using a subset of genotypes that had been included in a previous study with SSRs (Monforte et al., 2003) (see Additional file III.1-3). The differences in gene diversity were confirmed, demonstrating that they were intrinsic to the different marker type. SNPs are biallelic, implying that the He value can not exceed 0.5, whereas SSRs are multiallelic and so it can be higher. Haplotypes may yield higher gene diversity values than individual SNPs and provide more efficient application of SNP markers (Varshney et al., 2007).

All *inodorus* genotypes could be distinguished with the set of SNPs, although polymorphism was notably reduced (Table III.1-3). Fourteen SNPs were monomorphic and 18 were informative (minor allele frequency > 0.1). As most of the SNPs were discovered between the *agrestis* and *inodorus* cultivar and not within *inodorus*, we expected the SNP polymorphism within *inodorus* to be lower. Nevertheless, these results demonstrate that SNPs discovered using a germplasm sample can be successfully transferred to different germplasm samples in melon.

The genetic relationships among accessions based on SNP polymorphism were investigated by cluster analysis. The NJ dendrogram (Figure III.1-1) fits very well with previous classifications using different markers (Stepansky et al., 1999; Monforte et al., 2003; Monforte et al., 2005). Comparing the common genotype set in Monforte et al.

(2003), the average pair-wise distances based on SNPs and SSRs were 0.47 and 0.64, respectively. The correlation between the two distance matrices was 0.73 (P < 0.00001) according to Mantel's test, confirming that the current SNP set is as effective as SSRs in establishing genetic relationships among melon accessions, as shown in species such as rye (Varshney et al., 2007) and soybean (Yoon et al., 2007).

The population structure was estimated using the STRUCTURE software (Pritchard et al., 2000a). The *a posteriori* probability of the data increased rapidly from K = 1 to 4 and begun to reach a plateau for K = 5, inferring that our collection can be divided in five populations. Genetic variability among melon germplasm seems to be highly structured. The subdivision of the accessions in 5 populations agrees with the botanical classification and the cluster analysis (Figure III.1-1): group 1 included all the *inodorus* cultivars from Spain; group 2, a diverse group of traditional *inodorus* landraces and similar ones from the Near-East region such as elongated (*chate* and *flexuosus*) and Asiatic *ananas* and *chandalak* types; group 3, modern *cantalupensis* cultivars; group 4, mainly traditional varieties and wild melons from India and Africa and group 5 included *conomon* accessions from the Far East. The population structure should be taken into account when establishing a collection of genotypes for association mapping studies in melon and models including population structure should be used (Pritchard et al., 2000b). Alternatively, melon collections without structure, as we found with the *inodorus* melon accessions included in our studies, could be used.

These results demonstrate that SNPs discovered using a small germplasm sample can be transferred to different cultivar groups, being useful for depicting genetic relationships as well as for cultivar identification.

SNP mapping using a bin-mapping strategy

Two hundred and seventy-eight SNP-containing ESTs (Table III.1-1) plus twelve additional SNP-containing ESTs previously discovered between the two parental lines (ICuGI) were used for mapping in the SC × PS genetic map using 14 DHLs of the melon bin-mapping population (Fernandez-Silva et al., 2008). In total, 199 EST-derived SNPs were mapped, yielding 200 new markers (Figure III.1-2). F112 produced two SCAR markers (F112a and F112b) that mapped to groups I and V, respectively. Our previous melon bin-map contained 296 markers distributed in 122 bins, with a density of 4.2 cM/marker and 2.4 markers per bin (Fernandez-Silva et al., 2008). With the addition of 35 candidate genes previously reported for resistance to virus and fruit ripening (Morales et al., 2004; Moreno et al., 2008; Essafi et al., 2009)  and the SNPs now described, the new bin-map contains 528 markers, distributed in 145 bins, with an increased density of 2.35 cM/marker and 3.64 markers per bin. The SNP-based markers defined 23 new bins with an average bin length of 8.55 cM. Some of the new bins were located in regions with poor marker density in the previous SC × PS melon map (Fernandez-Silva et al., 2008), such as HS_30-B08 in group XI, AI_12-B08 in group VII, A_38-F04 in group VI or P06.05 in group III.

Essentially the new version of the melon bin-map is a gene-based map, with 412 markers (78%) obtained from gene sequences. Additionally, 114 RFLPs derived from ESTs were previously mapped in an F2 population from the cross SC × PS (Oliver et al., 2001), and their approximate position can also be plotted in the corresponding bin-map. As a large proportion of the markers are codominant and based on gene sequences, this makes this map a very useful tool for melon breeding and comparative analysis in cucurbit species.

With the advent of next generation sequencing technologies, SNP discovery has become

more feasible in non-model crop species, allowing the discovery of thousands of SNPs in a single experiment (Barbazuk et al., 2007). In *Eucalyptus grandis* more than 23,000 SNPs were discovered using 454 sequencing technology, with a validation rate of 83% (Novaes et al., 2008). In melon, a preliminary analysis of 100,000 reads obtained after 454 sequencing of leaf cDNAs from SC and PS produced more than 1,000 SNPs (Garcia-Mas, unpublished). This indicates that the use of next generation sequencing technologies is the next step towards saturation of the melon genetic map.

**Conclusion**

The set of 200 SNP markers discovered and mapped have increased the marker resolution of the melon genetic map by defining new bins. The genetic map contains more than 500 gene-based codominant markers (SNPs, RFLPs and SSRs), which can be used as anchor points with other genetic maps in this species. This genetic map is also a useful resource for comparative mapping in the *Cucurbitaceae*, for the construction of the melon physical map and for sequencing the melon genome. Additionally, the set of SNPs has proven to be as useful as microsatellites for studying genetic relationships in melon and for varietal identification.

**Methods**

Plant material and DNA extraction

The parent lines of the melon double haploid line (DHL) mapping population, PI 161375 'Songwan Charmi' (SC) and 'Piel de sapo' line T111 (PS), were used for SNP discovery (Gonzalo et al., 2005). Fourteen DHLs from the SC × PS segregating population were used to bin-map the SNP set (Fernandez-Silva et al., 2008). The 48 melon genotypes selected for analysis with a subset of SNPs (see Additional file III.1-3)

were obtained from the germplasm collection maintained at COMAV (Valencia, Spain) and from a previous study of germplasm variability using SSRs (Monforte et al., 2003). DNA from all genotypes was extracted using a modified CTAB method (Garcia-Mas et al., 2000). DNA of the forty-eight melon accessions was extracted from leaves of five individuals per accession to take into account the genetic variability within heterogeneous accessions.

SNP discovery and detection

SNPs were discovered using two different strategies. Firstly, random ESTs were selected from the International Cucurbit Genomics Initiative (ICuGI) webpage (http://www.icugi.org). Primer pairs were designed from each EST using the Primer3 software (Rozen and Skaletsky, 2000) with an average length of 20 nucleotides, a melting temperature around 60°C and an expected PCR product of 500-700 bp. Genomic DNA from the parental lines of the melon mapping population was amplified with each primer pair as previously described (Morales et al., 2004). Amplified fragments were purified with Sepharose columns and sequenced using the ABI Prism BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) in an ABI Prism 3130 sequencer (Applied Biosystems, Foster City, CA, USA). Sequences were aligned and screened for polymorphism with the Bioedit software (Hall, 1999). Putative SNP positions were visually verified on the sequence chromatogram, and the genomic sequences compared with the original EST sequence to identify any introns. In the second strategy, *in silico* SNPs previously identified (Gonzalez-Ibeas et al., 2007) using EST2uni (Forment et al., 2008) were classified as i) pSNPs, corresponding to SNPs presenting at least two EST sequences from the same genotype in a given contig and with the same base change and ii) pSCHs, corresponding

to single nucleotide variations insequence that did not follow the above criteria for pSNPs.

Selected pSNPs and pSCHs were verified in most cases after resequencing the parental lines of the melon mapping population. For a small subset, the SNP was verified with an appropriate SNP detection method. Bioedit software was used to generate restriction maps from sequences obtained from SC and PS. SNPs (or INDELs) showing differential restriction maps were used to develop Cleaved Amplified Polymorphic Sequence (CAPS) markers. When no differential restriction maps were available, the ABI Prism SNaPshot ddNTP Primer Extension Kit (Applied Biosystems) was used for SNP genotyping (Morales et al., 2004).

Markers F112, 46d_11-A08, FR12J11, 15d_17-G01, P01.45, PSI_26-B12, F012, PS_18-F05, PS_16-C09, F088, A_02-H11, AI_13-G03 and FR15D10 produced amplicons of different sizes in the parental lines, which were not sequenced and were genotyped as Sequence Characterized Amplified Region markers (SCARs) after electrophoresis in agarose gels or using a LICOR IR2 sequencer (Li-CorInc, Lincoln, Nebraska, USA). Markers PSI_12-D08 and PSI_35-F11 were converted into dCAPS markers (Neff et al., 1998). Markers F028, F149, F080 and PSI_25-B05 were genotyped using direct sequencing.


SNP mapping

SNPs and INDELs were mapped by selective genotyping using the bin-mapping strategy (Howad et al., 2005), adapted for the melon mapping population (Fernandez-Silva et al., 2008). Fourteen out of 72 DHLs from the melon mapping population were selected to obtain the maximum resolution with a minimum number of genotypes. SNPs

and INDELs were placed in the bin-map by visual inspection of the genotypes predicted by the markers and genotypes in the bin set.

Genetic variability analysis

Forty-five SNPs from 44 amplicons (two SNPs were selected from F241) were chosen for genetic variability analysis. SNPs were genotyped as CAPS or by pyrosequencing as shown in Additional file III.1-2. Thirty SNPs, described in Additional file III.1-1, were used. Twelve SNPs that were not polymorphic between SC and PS were also included in the variability analysis, and the primers for each amplicon are provided in Additional file III.1-2. The SNPs CmERF1, CmPm3 and CmXTH5 have been previously described (Moreno et al., 2008).

Eight SNPs were genotyped by minisequencing the region surrounding the polymorphism (two SNPs were detected for F241 in the same reaction). Pyrosequencing was performed using a PSQ™ HS 96 system (Pyrosequencing AB, Uppsala, Sweden) following the manufacturers' instructions. Primers were designed with the Pyrosequencing™ Assay Design Software (Biotage AB, Uppsala, Sweden). One of the amplifying primers was 5' end labeled with biotin, allowing the immobilization of the fragment onto M-280 streptavidin coated Sepharose™ dynabeads (DynalAS, Oslo, Norway). The genotyping primer was hence designed to anneal several nucleotides upstream of the SNP. After denaturation of the streptavidin-captured PCR fragments, the single stranded DNA fragments were released into the wells of the PSQ HS 96 plate. Pyrosequencing was performed using the PSQ HS SNP Reagent kit (Pyrosequencing AB, Uppsala, Sweden), and bioluminometric quantification of pyrophosphate (Ppi) released as a result of nucleotide incorporation during DNA synthesis was measured with the PSQ™ HS 96 system.

Allele frequencies, major allele frequency, gene diversity (measured as expected heterozygosity, He (Nei et al., 1983)), genetic distances and neighbor-joining (NJ) tree were calculated using Powermarker 3.25 (Liu and Muse, 2005). The NJ tree was plotted with MEGA 3.0 (Tamura et al., 2007). Distance matrices were compared by the Mantel test (Mantel, 1967).

The number of populations in our collection was deduced with the STRUCTURE software (Pritchard et al., 2000a). This package uses a Bayesian clustering approach to identify subpopulations and to assign individuals to these populations on the basis of their genotypes. Given a sample of individuals, K populations are assumed (where K may be unknown) and individuals are assigned to these populations. *A posteriori* probability for each K (Pr(K)) can be calculated, which is very small for K values lower than the appropriate value. Usually, the researcher fixes a minimum K (for example K= 1), recording Pr(K) after the analysis, and tests increasing Ks, plotting K against Pr(K). The final K is defined when Pr(K) reaches a plateau for higher K values. Consequently, in the current report, several number of populations (from K = 1 to 8) were tested with the software and the total number of populations was set when the probability reached a plateau for higher K.

**Authors' contributions**

WD discovered and mapped the SNPs and performed the genotyping for the variability analysis. CE and MGT discovered and mapped SNPs. CR discovered SNPs. IFS mapped SNPs. DGI identified and selected *in silico* SNPs. JB carried out the bioinformatics analyses for *in silico* SNPs. AJM performed the variability analysis, coordinated the SNP mapping and participated in the drafting of the manuscript. MBP prepared DNAs for the melon accessions and participated in the genotyping for the

variability analysis and in the drafting of the manuscript. JGM, PA, FN, MBP and MAA were involved in the conception of the study. JGM is the principal researcher of this work, supervised it and wrote the manuscript. All authors read and approved the final manuscript.

**Additional material**

**Additional file III.1-1. SNPs markers mapped in the SC × PS genetic map**.

Shown here, for each SNP marker: the EST and accession number from where it was obtained; best BlastX hit and E-value for each EST; amplicon primer sequences; SNP/INDEL position; SNP detection method; linkage group and BIN where the marker maps to: [a]sequence available in http://www.icugi.org or http://www.melogen.upv.es without accession number, [b]SNPs published by Morales et al (2004), [c]SNP position is provided when located in exons and referred to EST in first column, [d]third primer was used for SNaPshot genotyping.

Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-9-90-S1.xls].

**Additional file III.1-2. SNP markers used for genotyping the melon accessions**.

The EST from where the SNPs were discovered, the genotyping method (CAPS or pyrosequencing), linkage group where the marker maps to, and source of the marker are given. For unmapped SNP markers, the amplicon primer sequences are given. For SNP markers genotyped using pyrosequencing, forward, reverse and internal primers were used for genotyping. 5'bio: Forward or reverse primer was 5' labeled with biotine. ST1: Additional file III.1-1.

Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-9-90-S2.xls].


**Additional file III.1-3. Forty-eight melon accessions that were examined in this study**.

Plant assignation (or common name), code used in the current study, accession number from the respective gene banks, cultivar group, origin and seedbank donor (COMAV, Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Valencia, Spain; USDA/ARS/NCRPIS, North Central Regional Plant Introduction Station, Ames, IA, USA; IPK, Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany; INRA, Institute Nationale de la Recherche Agronomique, Montfavet, Avignon, France; Semillas Fitó SA, Barcelona, Spain; ARO, Agricultural Research Organization, Ramat Yishay, Israel) are specified for each genotype. Accessions marked with (*) were previously used by Monforte et al. (2003) for an SSR study.

Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2229-9-90-S3.xls].

**Table III.1-1. Amplicons designed from ESTs for SNP discovery.**

| | Amplicons | Failed | Monomorphic | Polymorphic | Polymorphic amplicons* | *In silico* SNP validation |
|---|---|---|---|---|---|---|
| Random ESTs | 223 | 58 | 72 | 93 | 56.3% | |
| *in silico* pSNPs | 269 | 41 | 69 | 159 | 69.7% | 51.8% |
| *in silico* pSCHs | 97 | 14 | 57 | 26 | 31.3% | 21.3% |
| TOTAL | 589 | 113 | 198 | 278 | 58.4% | |

ESTs were selected at random or chosen because they contained pSNPs or pSCHs in the MELOGEN database. Columns show the number of amplicons that failed to amplify or gave bad quality sequences, and monomorphic and polymorphic amplicons between SC and PS. The percentages of polymorphic amplicons and *in silico* SNPs that were validated are shown in the last two columns. (*) Polymorphic amplicons rate was calculated without considering failed amplicons.

**Table III.1-2. Frequency of SNPs and INDELs found after resequencing EST-derived amplicons.**

| Amplicons sequenced in SC and PS | Length sequenced (bp) | SNPs | bp per SNP | INDELs | bp per INDEL | Reference |
|---|---|---|---|---|---|---|
| 368 | 177,518 | 431 | 411.9 | 59 | 3,008.8 | this report |
| 34 | 15,000 | 34 | 441.2 | 9 | 1,666.6 | Morales et al. (2004) |

Data from a previous report using the same two melon parental lines is shown as a comparison.

**Table III.1-3. Gene diversity indexes for SNP and SSR alleles using all,** *inodorus* **or genotypes described in a previous study (Monforte et al., 2003).**

| Genotypes | Marker type | Major allele frequency | Ho | He | He range |
|---|---|---|---|---|---|
| all | SNP | 0.69 | 0.10 | 0.40 | 0.14–0.50 |
| *inodorus* | SNP | 0.85* | 0.07 | 0.15 | 0–0.50 |
| group used in Monforte et al. (2003) | SNP | 0.63 | 0.09 | 0.47 | 0.16–0.50 |
| group used in Monforte et al. (2003) | SSR | 0.47 | 0.14 | 0.64 | 0.51–0.83 |

Ho, observed heterozygosity; He expected heterozygosity. * Major allele frequency was only calculated for polymorphic SNPs.
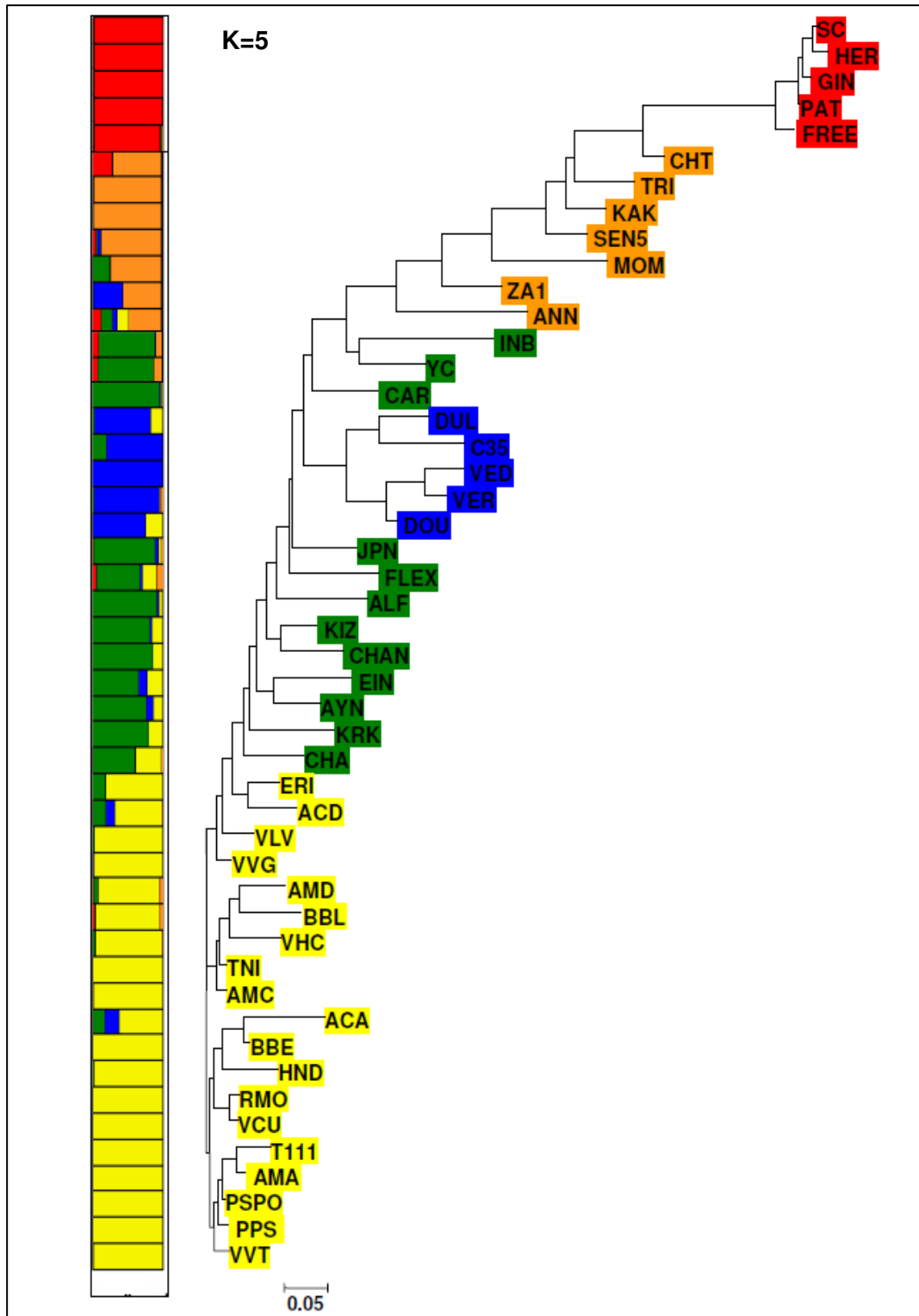
**Figure III.1-1. Dendogram and population structure based on the variability of 45 SNPs in 48 melon accessions**.

The neighbor-joining (NJ) tree based on Nei genetic distances (Nei et al., 1983) for the selected melon accessions is shown on the right. The subdivision based on

STRUCTURE is shown on the left; each accession on the NJ is colored according to its group assignation defined from STRUCTURE analysis. Group 1 (yellow), group 2 (green), group 3 (blue), group 4 (orange) and group 5 (red).

Although information about accessions is presented in Additional file III.1-3, code, plant assignation, accession name and type is detailed below to better understand the dendrogram.

Legend:

| Code | Plant assignation | Accession | Type |
|------|-------------------|-----------|------|
| PSP | Piñonet, Piel de sapo | PS | *inodorus* |
| T111 | T111, Piel de sapo | T111 | *inodorus* |
| PSPO | Pipa de oro, Piel de sapo | C-64 | *inodorus* |
| RMO | Mochuelo, Rochet | CM-C-50 | *inodorus* |
| TNI | Negro de invierno, Tendral | AN-C-8 | *inodorus* |
| VLV | Largo de Villaconejos, Green | CM-C-53 | *inodorus* |
| VCU | Cuarenteno, Green | C-C-39 | *inodorus* |
| VHC | Hilo carrete, Green | E-C-3 | *inodorus* |
| VVT | Verde tardío, Green | MU-C-44 | *inodorus* |
| VVG | Verde gordo, Green | AN-C-7 | *inodorus* |
| ACD | Caña dulce, Yellow | C-48 | *inodorus* |
| ACA | Amarillo Canario, Yellow | MU-C-35 | *inodorus* |
| AMC | Melon de calamonte, Yellow | E-C-1 | *inodorus* |
| AMD | Madura amarilla, Yellow | C-58 | *inodorus* |
| AMA | Amarillo, Yellow | PI 271755 | *inodorus* |
| BBL | Blanco liso, White | C-199 | *inodorus* |
| BBE | Blanco escriturado, White | AN-C-70 | *inodorus* |
| HND | Honeydew Green flesh | A-049 | *inodorus* |
| ERI | Eriço | C-75 | *inodorus* |
| KIZ | Kizal Assal | PI 288236 | *inodorus* |
| KRK | Kirkagac | PI 169305 | *inodorus* |
| VER | Verdejo | CL-C-1 | *cantalupensis* |
| CAR | Carraqueño | CA-C-6 | *cantalupensis* |
| DOU | Doublon | C-100 | *cantalupensis* |
| JPN | Pearl | PI 266947 | *cantalupensis* |

| | | | |
|---|---|---|---|
| AYN | Ananas Yokneam | PI 280079 | *cantalupensis, ananas* |
| DUL | Dulce | Dulce | *cantalupensis* |
| VED | Vedrantais | Vedrantais | *cantalupensis* |
| C35 | PMR 45 | C-35 | *cantalupensis* |
| EIN | Ein dor | PI 385966 | *cantalupensis, ananas* |
| CHAN | Chandalak | PI 276660 | *chandalak* |
| ALF | Alficoz | V-C-131 | *flexuosus* |
| FLEX | Snake cucumber | PI 435288 | *flexuosus* |
| CHA | Carosello | Cum 363 | *chate* |
| INB | PI 124112 | PI 124112 | *momordica* |
| MOM | PI 414723 | PI 414723 | *momordica* |
| YC | PI 124111 | PI 124111 | *momordica* |
| ANN | Queen Anne´s pocket melon | PI 273438 | *dudaim* |
| SC | Sonwhang Charmi | PI 161375 | *conomon* |
| FREE | Freeman´s cucumber | PI 420149 | *conomon* |
| GIN | Ginsen makuwa | PI 420193 | *conomon* |
| HER | Hermafrodite B | PI 420150 | *conomon* |
| PAT | Pat-81 | C-32 | *conomon* |
| CHT | Velleri | PI 164320 | *chito* |
| KAK | Kakru | PI 164493 | *agrestis* |
| ZA1 | ZM/A 5317 | PI 505599 | *agrestis* |
| SEN5 | G 22841 | PI 436532 | *agrestis* |
| TRI | *Cucumis trigonus* | Ames 24297 | *agrestis* |

**VII**

5.0 ECM50 F072
GCM181 MC373
ECM79 AI_05-F11
8.0 CmeIF(iso)4G-2
A_06-A03
AI_12-B08 AI_27-F07
PSI_26-B12

MC311
MC44
CMAGN75 ECM182
ECM227
AI_25-C11 F271
P06.69

39.0 MC253 CMTCN30
PSI_33-F04 HS_04-F11
CSAT425B ECM84
47.0 TJ38 ECM77 CMCAN90
48.0 CMBR53 CMBR27 P06.02
AI_03-G06 CmACS2
51.0 CMAGN21 CMBR84 GCM521
55.0 F149 F012
CSWCT12T CI_08-C08
60.0 CM139 AI_08-H11

ECM204

CMGA15 ECM172 CmERF2
PSI_37-G01 CI_37-H11
76.0

MC125 PS_19-E06
P05.15 AI_16-D09

99.0

**VIII**

15d_01-B03 A_30-G06
CSGA057 CMBR24 CMBR7
ECM88 P4.35 PS_28-E01
F125 PSI_29-D11
3.0 F080 PS_18-F05
6.0 TJ3 MC301
ECM217
ECM128 PSI_23-A11
HS_25-A10 CmAco3
16.0 MC68

TJ2 CmACS3
CI_33-B09

29.0 ECM221 AI_02-A08

CMTC13 AI_21-G05
41.0 AI_21-D08 P01.8

MC356 MC11 MC78
ECM200

A_04-B10 FR13O21
54.0
CMACC146 GCM241
MC208 CMAG59 F013
A_32-B01 CI_58-C10
68.0
CMAT141
PSI_25-H03
77.0

F129

ECM55
MC138

94.0 HS_39-A03
97.0 CMATN56 CmEXP1
102.0

**IX**

0.0 MRGH21

A_08-H06

MC92 CMTC47
ECM186 P05.64
13.0

ECM150 ECM177
MRGH7 CM98
ECM66
PSI_12-C05 AI_17-B03
AI_39-A12 PSI_23-G11
FR18J20 A_17-A08
35.0 F036 AI_04-D08

ECM56

61.0 PSI_21-D01

69.0 CM91

76.0 MC325

CMTCN1
CMCTN7
AI_21-E10
CmERF1 CmPME3

97.0 AI_08-F01 A_20-H12
AI_35-E03 P01.17

MC237
CMATN22
110.0

**X**

HS_23-E06 ECM86 ECM82
PS_15-H02 PS_40-E11
0.0 ECM175 MC149 ECM78
PS_16-C09 PSI_35-F11
PS_33-E12 AI_36-F12
CMCTN71 CMCTN19
10.0
MC17 CSWCT01
CSWCT22A
17.0

CMTCN67

31.0 CmXTH5
35.0 CMGA172 ECM228
MC112 MC67 CM40
CMTA134B CMCTN65
CMGA165 ECM232
GCM153 ECM49
ECM116 ECM220
ECM101 ECM344
AI_38-B09 F088
54.0 46d_21-E02 CmEXP3
57.0 CM101B
CMTCN8
62.0

**XI**

0.0 TJ33
2.0 CMTCN62
MC146 ECM210 P05.50
CMTC160 ECM63
8.0 ECM183 P4.39 A_02-H01

HS_30-B08

MC264 CMGAN12
MC375 CMAGN45
CSWCT18B CmeIF4A-1
31.0 HS_35-E11 AI_22-A08
34.0 A_05-A02 FR12O13 CmACO5
MC231 ECM164
HS_02-E07 PS_24-E03
PSI_41-B07

CMATN89 MC349
CMGA104 P06.79
PSI_35-H10

66.0
MC93
71.0

MC265
ECM147
ECM145
A_08-D10 15d_27-B02
AI_13-G03
93.0 P02.75

**XII**

MC132 GCM206
PSI_22-B02 PSI_12-D12
0.0
5A6U nsv MC320
ECM67 ECM105
CmeIF4E AI_35-A08
AI_09-G07 15d_01-B03
FR12P24

17.0
TJ29
MC330
HS_23-D06 FR15D10

29.0

CSAT425A
MC286
CI_57-E03

49.0
53.0 CMTCN14

CMAGN33 CMAGN32
CMAGN80
ECM123
ECM218
P02.03 FR14F22

81.0

**Figure III.1-2. EST-SNP bin map of *Cucumis melo* obtained by selective genotyping of fourteen DHLs**.

Linkage groups are represented by vertical bars, divided in bins defined by the joint genotype of the selected DHLs. The mapped SNPs in this report are shown in bold. Underlined markers are candidate genes previously reported (Morales et al., 2004; Moreno et al., 2008; Essafi et al., 2009). The other markers have been described in Fernandez-Silva et al. (2008). Genetic distances are shown on the left, indicating the position of the last marker included in the bin according to the framework map in Fernandez-Silva et al. (2008). Markers defining new bins are shown in italics. The hypothetical position of the last marker of these bins is indicated by a dashed horizontal line within the linkage group bar, without the genetic distance.

**References**

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

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS: **SNP discovery via 454 transcriptome sequencing.** Plant J 2007, 51(5):910-918.

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D: **Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data.** Plant Physiol 2003, 132(1):84-91.

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, BerryCC, Winzeler E, Chory J: **Large-scale identification of single-featurepolymorphisms in complex genomes.** Genome Res 2003, 13(3):513-523.

Brookes AJ: **The essence of SNPs.** Gene1999, 234(2):177-186.

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV,Chase K, Lark KG, Reiter RS, Yoon MS, et al.: **A soybean transcriptmap: gene distribution, haplotype and single-nucleotide polymorphism analysis.** Genetics 2007, 176(1):685-696.

Danin-Poleg Y, Reis N, Baudracco-Arnas S, Pitrat M, Staub JE, OliverM, Arus P, deVicente CM, Katzir N: **Simple sequence repeats in *Cucumis* mapping and map merging.** Genome 2000, 43(6):963-974.

Eduardo I, Arus P, Monforte AJ: **Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI161375.** Theor Appl Genet 2005, 112(1):139-148.

Essafi A, Diaz-Pendon JA, Moriones E, Monforte AJ, Garcia-Mas J, Martin-Hernandez AM: **Dissection of the oligogenic resistance to *Cucumber mosaic virus* in the melon accession PI 161375.** Theor Appl Genet 2009, 118(2):275-284.

Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Pico B, Nuez F, ArusP, Garcia-Mas J, Monforte AJ: **Bin mapping of genomic and EST derived SSRs in melon (*Cucumis melo* L.).** Theor Appl Genet 2008, 118(1):139-150.

Forment J, Gilabert F, Robles A, Conejero V, Nuez F, Blanca JM: **EST2uni: an open, parallel tool for automated EST analysis and database creation, with a data mining web interface and microarray expression data integration.** BMC Bioinformatics 2008, 9:5.

Ganal MW, Altmann T, Roder MS: **SNP identification in crop plants.** Curr Opin Plant Biol 2009, 12(2):211-217.

Garcia-Mas J, Oliver M, Gómez H, de Vicente MC: **Comparing AFLP, RAPD and RFLP markers to measure genetic diversity in melon.** Theor Appl Genet 2000, 101:860-864.

Gonzalez-Ibeas D, Blanca J, Roig C, Gonzalez-To M, Pico B, TrunigerV, Gomez P, Deleu W, Cano-Delgado A, Arus P, et al.: **MELOGEN: an EST database for melon functional genomics.** BMC Genomics 2007, 8:306.

Gonzalo MJ, Oliver M, Garcia-Mas J, Monfort A, Dolcet-Sanjuan R, Katzir N, Arus P, Monforte AJ: **Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.).** Theor Appl Genet 2005, 110(5):802-811.

Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** Nucleic Acids Symp Ser 1999, 41:95-98.

Hayes B, Lærdahl JK, Lien S, Moen T, Berg P, Hindar K, Davidson WS, Koop BF, Adzhubei A, Høyheim B: **An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences.** Aquaculture 2007, 265:82-90.

Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, CiprianiG, Monforte AJ, Georgi L, Abbott AG, Arus P: **Mapping witha few plants: using selective mapping for microsatellite saturationof the *Prunus* reference map.** Genetics2005, 171(3):1305-1309.

Liu K, Muse SV: **PowerMarker: an integrated analysis environment for genetic marker analysis.** Bioinformatics 2005, 21(9):2128-2129.

Mantel N: **The detection of disease clustering and a generalized regression approach.** Cancer Res 1967, 27(2):209-220.

Mascarell-Creus A, Cañizares J, Vilarrasa J, Mora-García S, Blanca J,Gonzalez-Ibeas D, Saladié M, Roig C, Deleu W, Picó B, et al.: **An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon.** BMC Genomics 2009, 10:467, doi:10.1186/1471-2164-10-467.

Monforte AJ, Eduardo I, Abad S, Arus P: **Inheritance mode of fruit traits in melon-heterosis for fruit shape and its correlation with genetic distance.** Euphytica 2005, 144:31-38.

Monforte AJ, Garcia-Mas J, Arús P: **Genetic variability in melon based on microsatellite variation.** Plant Breeding 2003, 122:153-157.

Morales M, Roig E, Monforte AJ, Arus P, Garcia-Mas J: **Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo L.*).** Genome 2004, 47(2):352-360.

Moreno E, Obando JM, Dos-Santos N, Fernandez-Trujillo JP, MonforteAJ, Garcia-Mas J: **Candidate genes and QTLs for fruit ripening and softening in melon.** Theor Appl Genet 2008, 116(4):589-602.

Neff MM, Neff JD, Chory J, Pepper AE: **dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in Arabidopsis thaliana genetics.** Plant J 1998, 14(3):387-392.

Nei M, Tajima F, Tateno Y: **Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data.** J Mol Evol 1983, 19(2):153-170.

Neuhausen SL: **Evaluation of restriction fragment length polymorphismsin *Cucumis melo*.** Theor Appl Genet 1992, 83:379-384.

Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D,Sederoff RR, Kirst M: **High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome.** BMC Genomics 2008, 9:312.

Oliver M, Garcia-Mas J, Cardus M, Pueyo N, Lopez-Sese AL, ArroyoM, Gomez-Paniagua H, Arus P, de Vicente MC: **Construction of a reference linkage map for melon.** Genome 2001, 44(5):836-845.

Pavy N, Parsons LS, Paule C, MacKay J, Bousquet J: **Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs.** BMC Genomics 2006, 7:174.

Perin C, Hagen S, De Conto V, Katzir N, Danin-Poleg Y, Portnoy V, Baudracco-Arnas S, Chadoeuf J, Dogimont C, Pitrat M: **A reference map of *Cucumis melo* based on two recombinant inbred line populations.** Theor Appl Genet 2002, 104(6-7):1017-1034.

Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** Genetics 2000a, 155(2):945-959.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** Am J Hum Genet 2000b, 67(1):170-181.

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, BalfourierF, Dufour P, Chalhoub B, Brunel D, et al.: **Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.).** Genome 2006, 49(9):1131-1139.

Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** Methods Mol Biol 2000, 132:365-386.

Staub JE, Box J, Meglic V, Horejsi TF, Mccreight JD: **Comparison of isozyme and random amplified polymorphic DNA data for determining intraspecific variation in *Cucumis*.** Genet Res Crop Evol 1999, 44:257-269.

Stepansky A, Kovalski I, Perl-Treves R: **Intraspecific classification of melons (*Cucumis melo* L.) in view of their phenotypic and molecular variation.** Plant Syst Evol 1999, 217:313-332.

Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** Mol Biol Evol 2007, 24(8):1596-1599.

The International Cucurbit Genomics Initiative (ICuGI) [http://www.icugi.org]

van Leeuwen H, Monfort A, Zhang HB, Puigdomenech P: **Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*.** Plant Mol Biol 2003, 51(5):703-718.

Varshney RK, Beier U, Khlestkina EK, Kota R, Korzun V, Graner A, Borner A: **Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery, frequency, and applications for genome mapping and diversity studies.** Theor Appl Genet 2007, 114(6):1105-1116.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, PrussD, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, et al.: **A high quality draft consensus sequence of the genome of a heterozygous grapevine variety.** PLoS ONE 2007, 2(12):e1326.

Wang S, Sha Z, Sonstegard TS, Liu H, Xu P, Somridhivej B, PeatmanE, Kucuktas H, Liu Z: **Quality assessment parameters for EST derived SNPs from catfish.** BMC Genomics 2008, 9:450.

Wang YH, Thomas CE, Dean RA: **A genetic map of melon (*Cucumis melo* L.) based on amplified fragment length polymorphism (AFLP) markers.** Theor Appl Genet 1997, 95:791-798.

Yamamoto N, Tsugane T, Watanabe M, Yano K, Maeda F, Kuwata C,Torki M, Ban Y, Nishimura S, Shibata D: **Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar MicroTom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars.** Gene 2005, 356:127-134.

Yoon MS, Song QJ, Choi IY, Specht JE, Hyten DL, Cregan PB: **BARCSoySNP23: a panel of 23 selected SNPs for soybean cultivar identification.** Theor Appl Genet 2007, 114(5):885-899.

# III-2. Melon transcriptome characterization: Simple Sequence Repeats and Single Nucleotide Polymorphisms discovery for high throughput genotyping across the species.

Blanca J, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó B: **Melon transcriptome characterization: Simple Sequence Repeats and Single Nucleotide Polymorphisms discovery or high throughput genotyping across the species.** The Plant Genome 2011, 4(2):118-131.

184

## Abstract

Melon (*Cucumis melo* L.) ranks among the highest-valued fruit crops worldwide. Some genomic tools are available for this crop, including a Sanger transcriptome. We report the generation of 689,054 *C. melo* high-quality Expressed Sequence Tags (ESTs) from two 454 sequencing runs, using normalized and nonnormalized complementary DNA (cDNA) libraries prepared from four genotypes belonging to the two *C. melo* subspecies and the main commercial types. 454 ESTs were combined with the Sanger available ESTs and *de novo* assembled into 53,252 unigenes. Over 63% of the unigenes were functionally annotated with Gene Ontology (GO) terms and 21% had known orthologs of *Arabidopsis thaliana* (L.) Heynh. Annotation distribution followed similar tendencies than that reported for *Arabidopsis thaliana*, suggesting that the dataset represents a fairly complete melon transcriptome. Furthermore, we identified a set of 3,298 unigenes with microsatellite motifs and 14,417 sequences with Single Nucleotide Variants of which 11,655 Single Nucleotide Polymorphism met criteria for use with high-throughput genotyping platforms, and 453 could be detected as Cleaved Amplified Polymorphic Sequence (CAPS). A set of markers were validated, 90% of them being polymorphic in a number of variable *C. melo* accessions. This transcriptome provides an invaluable new tool for biological research, more so when it includes transcripts not described previously. It is being used for genome annotation and has provided a large collection of markers that will allow speeding up the process of breeding new melon varieties.

## Background

Melon (*Cucumis melo* L.) is an important vegetable crop that is grown worldwide, mainly in temperate, subtropical, and tropical climates. It ranks as the ninth

horticultural crop in terms of total world production. This species belongs to the botanical family *Cucurbitaceae*, commonly known as cucurbits, which includes several economically and nutritionally important vegetable crops, such as cucumber (*Cucumis sativus* L.), watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai], and pumpkins, gourds, and squashes (*Cucurbita* spp.) (Schaefer et al., 2009).

Melon is a diploid species ($2n = 2x = 24$) with an estimated genome size of 450 Mbp, which is similar to that of rice (*Oryza sativa* L.) or cucumber (Huang et al., 2009) and approximately three times the size of the model species *Arabidopsis thaliana* (L.) Heynh. (Arabidopsis Genome Initiative, 2000). The species displays a rich diversity of many traits and has become a primary model for sex expression and fruit ripening analysis (Boualem et al., 2008). *Cucumis melo* is considered to be divided into two subspecies, each one with several botanical varieties (Pitrat, 2008): ssp. *melo* and ssp. *agrestis* (Naudin) Pangalo, including the main commercial types and the most important sources of resistances and quality traits, respectively.

The availability of genetic and genomic resources for this crop has increased significantly in recent years due to several national and international projects, such as the Spanish initiative Melogen (González-Ibeas et al., 2007; Melogen, 2003) and the International Cucurbit Genomics Initiative (ICuGI) (ICuGI, 2007). A broad range of genomic tools are available today (Ezura and Fukino, 2009), such as mapping populations of double haploids, recombinant inbreeds and near-isogenic lines (Eduardo et al., 2007; Harel-Beja et al., 2010), genetic maps (Deleu et al., 2009), a detailed physical map (González et al., 2010a), and an oligo-based microarray that is providing the first expression studies (Mascarell-Creus et al., 2009). These tools include a collection of Expressed Sequence Tags (ESTs), obtained using Sanger sequencing technologies, from a number of genotypes belonging to both subspecies. These ESTs

have been assembled to generate a melon transcriptome. The last version of the assembly (version 4.0, released in May 2010), available at the Cucurbit Genomics Database (ICuGI, 2007), includes 24,444 unigenes. Also, an effort is in progress through a Spanish initiative to obtain the whole genome sequence of this crop (González et al., 2009, 2010b).

A complete transcriptome is a basic resource for gene discovery, large-scale expression analysis, and genome annotation. The cost and the limited parallelization of the conventional Sanger method has prevented the large scale generation of EST that is necessary to generate complete transcriptomes. Recent advances in next-generation sequencing technologies allow us a very deep EST sequencing, efficiently and cost-effectively (Shendure and Ji, 2008). There are increasing studies in which 454 sequencing technology, combined or not with Solexa/Illumina, are used to characterize transcriptomes in cereals, legumes, and other crops (Cheung et al., 2006; Emrich et al., 2007; Vega Arreguin et al., 2009; Folta et al., 2010; Guo et al., 2010). Even in model species, such as *Arabidopsis thaliana*, this deep sequencing is allowing us to identify new transcripts not present in previous EST collections (Weber et al., 2007). Also, specific transcriptomes are being generated in species for which previous genomic resources are lacking (Alagna et al., 2009; Barakat et al., 2009; Wang et al., 2009; Li et al., 2010; Sun et al., 2010). When this massive amount of data is produced from a number of different genotypes, it offers a means to identify and characterize the genetic polymorphisms underlying the phenotypic variation. Single Nucleotide Polymorphisms (SNPs) are the most abundant variations in genomes and, therefore, constitute a powerful tool for mapping and marker-assisted breeding. The new transcripts produced with next-generation sequencing are being used for high-throughput SNP identification. Single nucleotide polymorphism detection is performed by aligning raw reads from

different genotypes to a reference genome or transcriptome previously available, as in maize (*Zea mays* L.), cucumber, and even in polyploid species such as oilseed rape (*Brassica napus* L.) (Barbazuk et al., 2006; Trick et al., 2009; Guo et al., 2010). *De novo* assembly of raw sequences coming from a set of genotypes, followed by pairwise comparison of the overlapping assembled reads, has also successfully been used in species lacking any significant genomic or transcriptomic resources (Novaes et al., 2008; Blanca et al., 2011).

In this study, we describe the generation of 689,054 new *C. melo* high-quality ESTs from two 454 sequencing runs, one Genome Sequencer (GS) FLX and one GS FLX Titanium, using four libraries prepared from a number of tissues and four genotypes belonging to the main commercial types, with contrasting fruit phenotypes, and to the two subspecies of *C. melo*. By combining these new ESTs with 125,908 Sanger ESTs previously available, a melon transcriptome has been generated with 53,252 unigenes. These unigenes have been accurately annotated, screened for Simple Sequence Repeat (SSR) motifs, and used to identify a large SNP collection suited for high-throughput mapping purposes. All these tools will allow accelerating genetics and breeding of this crop.

**Material and Methods**

Plant material

Two nonnormalized complementary DNA (cDNA) libraries were constructed using leaf material from the T111 Piel de sapo line (Semillas Fitó, Barcelona, Spain), belonging to *C. melo* ssp. *melo* var. *inodorus* H. Jacq., and the accession PI161375, belonging to *C. melo* ssp. *agrestis* var. *conomon* (Thunb.) Makino. These two geno-types are the parents of the melon genetic map (Deleu et al., 2009) and belong to the

two subspecies of *C. melo*. Two normalized cDNA libraries were constructed using total RNA extracted from different tissues (leaf, root, male and female flowers at different stages, seedlings, and dark-grown and ethylene treated seedlings) of two genotypes, the Piñonet cultivar (a Piel de sapo type, belonging to *C. melo* ssp. *melo* var. *inodorus*) and the Vedrantais cultivar (belonging to *C. melo* ssp. *melo* var. *cantalupensis* Naudin), belonging to the main melon commercial types, having contrasting phenotypes for fruit ripening interesting in melon breeding (both genotypes maintained at the germplasm collection of the Cucurbits breeding group of the Institute for the Conservation and Breeding of Agricultural Biodiversity (COMAV, Valencia, Spain). All tissues were collected and immediately frozen in liquid nitrogen and stored at –80°C till use.

Complementary DNA preparation and sequencing

Total RNA from T111 and PI161375 lines was extracted from young leaves with TRI Reagent (Sigma-Aldrich, Saint Louis, MO). Double-stranded cDNA was synthesized using SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, Carlsbad, CA) and an oligo d(T)-primer. Fragments below 300 bp were removed with Agencourt AMPure beads (Beckman Coulter Genomics, Danvers, MA). Final double-stranded cDNA size distribution was checked using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Three micrograms of size-selected double-stranded cDNA was used for each GS FLX library and two quarters of a sequencing plate were used for each sample. Sequencing was performed at the CRAG Sequencing Service (Barcelona, Spain).

Total RNA from Piñonet and Vedrantais was extracted from each tissue using the TRI Reagent. Equivalent amounts of RNA from each tissue were combined into one pool per

cultivar. Messenger RNA (mRNA) was purified from the total RNA using the illustra mRNA Purification Kit (GE Healthcare, Amersham Bioscience, Buckinghamshire, UK). Double-stranded cDNA was then synthesized from the RNA pools with the SMART cDNA Library Construction Kit (Clontech, Palo Alto, CA). A normalization step was performed with the TRIMMER cDNA normalization Kit (Evrogen, Moscow, Russia) to prevent over-representation of the most common transcripts. The polymerase chain reaction (PCR) products of cDNA were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Normalization quality of cDNAs libraries was checked by quantitative PCR. The cDNA length and normalization are critical factors to have a good transcriptome representation, to have SNPs along the whole gene sequence, and to have a high quality SNP prediction. Approximately 1 μg of double-stranded cDNA from each of the two normalized cDNA pools were used for sequencing on a GS FLX Titanium Sequencing platform. A half sequencing plate was performed for each sample at Creative Genomics (New York, NY).

Complementary DNA sequence processing and assembly

The whole sequence analysis was performed by using the ngs_backbone pipeline developed at COMAV (Bioinformatics at COMAV, 2010). The tools and analysis mentioned in the following sections were all performed by ngs_backbone, but here the third party tools, databases, and parameters used by ngs_backbone are described.

For assembly, 454 sequences were combined with Sanger sequences, previously available at ICuGI database (ICuGI, 2007). Sanger sequences were obtained from different libraries and genotypes; the same four genotypes sequenced with 454 (the *inodorus* Piel de sapo T111 and Piñonet, the *cantalupensis* Vedrantais, and the *conomon* PI161375) and six additional genotypes, the cultivar TamDew (*C. melo* ssp. *melo* var.

*inodorus),* the cultivars Dulce, Noy Israel, and Charentais, the accession C35 of the germplasm collection of Estación Experimental La Mayora - Consejo Superior de Investigaciones Científicas (EELM-CSIC, Málaga, Spain) (*C. melo* ssp. *melo* var. *cantalupensis*), and the exotic accession Pat81 of the COMAV collection (*C. melo* ssp. *agrestis* var. *conomon*).

Sequences were processed before the assembly. To remove the adaptors, an alignment with the primers and adaptors used during the sequencing process was done by Exonerate (Exonerate, 2005). The low quality regions from the reads were trimmed by using Lucy (Chou and Holmes, 2001). Sequences shorter than 100 bp were discarded and not used for the assembly. The processed sequences were assembled with Mira (Chevreux et al., 2004). Default ngs_backbone options were used to assemble the unigenes.

Gene annotation

Structural and functional annotation was performed by sequence comparison with public databases. All unique assembled sequences (unigenes) were sequentially compared using blast (cutoff e -value of $10^{-20}$) with the sequences in three major public protein databases, prioritizing nonmachine curate databases. The used database order was Swiss-Prot (uniprot_sprot_ release of 2010 04 23) (UniProt Consortium, 2010a, b), *Arabidopsis* proteins (Tair_9_pep_ release 2009 06 19) (Arabidopsis Information Resource, 2009; Swarbreck et al., 2008), and UniRef90 (uniref90_release 2010 04 23) (European Bioinformatics Institute, 2010; UniProt Consortium, 2010b). Once a sequence had a good enough blast hit in one of the databases, a description was build from the description of that hit. Also, a bidirectional blast search comparison was performed to obtain a set of putative orthologs with *Arabidopsis thaliana* and

equivalents with melon, using the melon unigenes contained in the ICuGI database (ICuGI, 2007).

Additionally, we performed a functional classification of the unigenes following the Gene Ontology (GO) scheme. Blast2GO (Conesa and Götz, 2008) was used for this purpose. Blast2GO used the results of a blast search against the National Center for Biotechnology Information (NCBI) nonredundant protein database (release of 2010 - 03-27) (NCBI, 2010a) (cutoff e-value of $10^{-20}$) to infer the relevant GO terms for every sequence. Open Reading Frames (ORFs) were predicted in the unigenes with the aid of the ESTScan software (Iseli et al, 1999). We used the *Arabidopsis* codon usage table to perform the ORF searching. Introns were assigned by aligning the unigenes with the melon genomic sequence (a draft is being produced within the MELONOMICS consortium and is available for the partners of the funding project [González et al., 2010b]) using the Emboss: est2 genome (European Bioinformatics Institute, 2001).

Identification of Simple Sequence Repeats and Single Nucleotide Polymorphisms

Simple sequence repeats were annotated using the Sputnik software (Abajian, 1994). Sequences containing greater than or equal to four di-, tri-, or tetra-nucleotide repeats were selected. A set of SSRs were validated using the genotypes sequenced with 454 (T111, Piñonet, Vedrantais, and PI161375), some of the genotypes sequenced with Sanger (TamDew, Dulce, Noy Israel, and Pat 81) described previously, and two additional genotypes belonging to the *C. melo* L. ssp. *melo* var. *momordica* (Roxb.) Duthie & J. B. Fuller group (PI124111 and PI414743, provided by National Plant Germplasm System [NPGS]-USDA, and maintained at the COMAV collection) used as parentals of other melon genetic maps. Primer pairs flanking each SSR *locus* were designed using the Primer3 program (Rozen and Skaletsky, 2000).

Polymerase chain reactions were performed in a final volume of 15 µL with 1x PCR buffer (100 mM Tris-HCl, 15 mM MgCl$_2$, 500 mM KCl, and pH 8.3), 200 µM deoxyribonucleotide triphosphates (dNTPs), 0.15 µM each primer, and 1 µL of template (approximately 10 ng µL$^{-1}$) 0.5 U Taq polymerase. The cycling conditions were as follows: denaturation at 94°C for 2 min, followed by seven cycles of 45 s at 94°C, 45 s at 68°C (with each cycle the annealing temperature decreasing 2°C), and of 1 min at 72°C. Products were subsequently amplified for 30 cycles at 94°C for 45 s, 54°C for 45 s, and 72°C for 1 min, with a final extension at 72°C for 7 min.

The forward primer was designed adding an M13 tail to its 5′ end. Polymerase chain reaction products were separated using 6% polyacrylamide gels, 1x Tris-borate-ethylenediaminetetraacetic acid (EDTA) (TBE) buffer in a LI-COR 4300 (Li-COR BioScience, Lincoln, NE). IRD700 and IRD800-labeled amplicons were visualized by adding to PCR mixture 0.2 µM of fluorescent label (IR700 or IR800) M13 tail. Number of alleles, frequency of the most common allele and the Polymorphism Information Content (PIC) were calculated for each *locus* for the 10 melon genotypes using the PowerMarker software (Liu and Muse, 2005). Cluster analysis was also performed with the Nei's genetic distance (Nei et al., 1983). The support values for the degree of confidence at the nodes of the dendrogram were analyzed by bootstrap resampling 1,000 times.

This *C. melo* EST collection has been produced using several genotypes belonging to different subspecies and morphotypes of *C. melo* appropriate for SNP discovery. ngs_backbone was also used to detect Single Nucleotide Variants (SNVs) (SNPs and INDELs) by mapping the 454 and Sanger processed reads against the unigene assembly using BWA (Burrows-Wheeler Aligner) (Li and Durbin, 2010). We kept only SNVs meeting stringent quality criteria: (i) minimum allele quality (accumulated sequence

quality for every allele) and (ii) minimum mapping quality. The default threshold set by ngs_ backbone was set for both parameters.

Despite satisfying the quality criteria, not all the SNVs seemed equally reliable. Several filters were applied to maximize a successful validation and/or implementation in high-throughput genotyping platforms (Fan et al., 2006; Gupta et al., 2008). Single nucleotide variants include both SNPs and INDELs and the VKS filter (VKS [it is not an SNP]) classifies them in those categories. Some filters dismiss SNVs in redundant or highly variable regions (UCR [region is not unique or noncontiguous] and HVR4 [the region has more than four SNVs per 100 bp']). Other filters used were CS60 (SNV is closer than 60 bp to another SNP or INDEL), I59 (an intron is located closer than 59 bp), and CL60 (SNV is closer than 60 bp to the sequence edge) that facilitate the use of variants in GoldenGate genotyping platforms. Those SNVs with only one allele (sequenced one or more times) within each genotype can be selected by filtering out those that are variable with NVPiñ, NVT111, NVVed, NVDul, NVPI161, and NVPat81 (it is variable in Piñonet, T111, Vedrantais, Dulce, PI161375, and Pat81, respectively). Also, those with only one allele, with one or more reads, within a group of genotypes can be selected by using NVPs for the Piel de sapo group (it is variable in T111 and Piñonet), NVCant for the *cantalupensis* group (it is variable in Vedrantais, Charentais, and Dulce), NVCon for the *conomon* group (it is variable in Pat81 and PI161375), and NVmelo for the accessions of the ssp. *melo* (it is variable in T111, Piñonet, Dulce, Vedrantais and Charentais). These filters allow us to identify markers uniform within varieties or subspecies. Combined with additional filters VSCant-Con (it is not variable in Dulce, Vedrantais, Charentais, Pat81 and PI161375) and VSPs-Con (it is not variable in T111, Piñonet, Pat81 and PI161375), they allow to select those SNPs polymorphic among groups.

We also detected those SNPs that can be analyzed *via* Cleaved Amplified Polymorphic Sequence (CAPS) (searching for allele-specific restriction targets) filtering out with nCAP, and validated a subset of them using the same set of genotypes that in SSRs analysis. Polymerase chain reactions were performed in a final volume of 25 µL with 1x PCR buffer (100 mM Tris-HCl, 15 mM $MgCl_2$, 500 mM KCl, and pH 8.3), 200 µM dNTPs, 0.15 µM each primer, and 2 µL of template (approximately 10 ng $µL^{-1}$). The cycling conditions were as follows: denaturation at 95°C for 3 min, followed by 30 cycles of 30 s at 95°C, 30 s at 55°C, and of 60 s at 72°C, with a final extension at 72°C for 7 min. Polymerase chain reaction products were digested with the corresponding enzymes and detected by 2% agarose gel electrophoresis. Cluster analysis was as described for SSRs.

**Results and discussion**

<u>Expressed Sequence Tag sequencing and assembly</u>

We performed two 454 sequencing runs, one with the GS FLX reagents and one with GS FLX Titanium reagents on the GS FLX System. A half GS FLX run was performed on each of the two libraries constructed from leaves of the genotypes used as parentals of the melon genetic map (Deleu et al., 2009), belonging to the two subspecies of *C. melo* ssp. *melo* var. *inodorus* cv. Piel de sapo T111 and ssp. *agrestis* var. *conomon* PI161375. The two GS FLX Titanium half runs were performed on two normalized libraries constructed from a pool of tissues (leaves, roots, female and male flowers at different stages, seedlings, dark-grown seedlings, and ethylene-treated seedlings) using two cultivars of the ssp. *melo*, belonging to the main melon commercial classes, var. *inodorus* cv. Piel de sapo Piñonet and var. *cantalupensis* cv. Vedrantais. A total of 763,352 reads were obtained from the four libraries (Table III.2-1). The GS FLX

Titanium run provided reads ~1.7x longer than GS FLX (average length of 392 *versus* 228 bp). Similar lengths have been reported in previous studies (Gedye et al., 2010; Li et al., 2010; Sun et al., 2010; Blanca et al., 2011). Reads were processed using the ngs_backbone software (Bioinformatics at the Institute for the Conservation and Breeding of Agricultural Biodiversity, 2010) to eliminate adaptor sequences, low quality chromatograms, and sequences of less than 100 bp. After processing, we obtained 689,054 high quality ESTs, comprising 243 Mbp, with an average length of 400 and 235 bp for the two sequencing methods, respectively. The length distribution of these ESTs is shown in Figure III.2-1.

More than 89% of the ESTs fell between 200 and 550 bp in length. All reads were deposited in the NCBI database and can be accessed in the Sequence Read Archive (NCBI, 2010b) under the accession number SRP005268. The 454 ESTs generated in this study were combined for clustering and *de novo* assembly with a previously available collection of 125,908 ESTs (average length 613 bp and total length 81.3 Mbp) produced using traditional Sanger sequencing methods. Sanger ESTs were obtained mainly by two transcriptome sequencing initiatives, Melogen (Gonzalez-Ibeas et al., 2007) and ICuGI (ICuGI, 2007), from different tissues (callus, leaves, roots, flowers, and fruits at different stages) and from 10 genotypes (including the four sequenced in this paper with 454 and one additional *inodorus*, four *cantalupensis*, and one *conomon*) and are publicly available at ICuGI (ICuGI, 2007). A summary of the Sanger dataset using for this assembly, before and after processing, is detailed in Additional file III.2-1. Finally, 753,004 ESTs were assembled using the Mira assembler (Chevreux et al., 2004) yielding a total of 53,252 high-confident tentative consensus sequences (nonredundant sequences or unigenes). Approximately 55% of the unigenes (29,566) were composed only of 454 ESTs, whereas only 4% (1,873) were assembled

exclusively from Sanger ESTs and 41% (21,813) included sequences of both datasets. About 28, 32, and 96% of the unigenes included sequences from Melogen (Gonzalez-Ibeas et al., 2007), ICuGI (ICuGI, 2007), and 454 datasets, respectively. These results agree with those of previous studies combining 454 and Sanger ESTs (Vega-Arreguinet al., 2009; Guo et al., 2010; Li et al., 2010; Ueno et al., 2010) in which next-generation sequencing contributes significantly to the identification of novel unigenes.

The distribution of the number of ESTs per unigene is shown in Figure III.2-2A. The majority of unigenes were assembled from a moderate number of ESTs (from 2 to 10), with an average of 14.1 ESTs per unigene. This relatively low redundancy is probably due to the success of the normalization process, responsible for the suppression of superabundant transcripts. Also, the EST assemblers can in some cases split ESTs generated from the same transcript into several unigenes. This behavior is usually due to the difficulty of distinguishing sequence variations due to individual provenance to those due to close paralogs.

However, part of the 454 and the Sanger ESTs came from nonnormalized libraries. Consistently, we were able to identify a number of highly abundant transcripts. Around 3,103 transcripts (5.8% of all the unigenes) have more than 50 EST members. These abundant transcripts contain ~60% of the assembled EST reads.

The assembled unigenes had an average length of 774.3 bp comprising approximately 42.2 Mbp in total. The length distribution of the unigenes is shown in Figure III.2-2B. The analysis revealed that more of the 50% of unigenes were larger than 673 bp, and only 5% of the sequences were shorter than 266 bp. The number of assembled unigenes is similar to that obtained in previous transcriptome analyses performed with massive sequencing in maize, *Eucalyptus grandis* W. Hill ex Maiden, *Artemisia annua* L., chestnut (*Castanea* spp.), olive (*Olea europaea* L.), and cucumber. However, our *de*

*novo* assembly with the longer reads obtained with the GS FLX Titanium platform and the combination with Sanger sequences render unigenes that on average are almost two times longer than those reported in other studies that used Genome Sequencer 20 and Genome Sequencer FLX platforms (Novaes et al., 2008; Alagna et al., 2009; Barakat et al., 2009; Vega-Arreguin et al., 2009; Wang et al., 2009; Guo et al., 2010).

Our assembled unigenes were also larger than those reported for other plant trancriptomes obtained using the GS FLX Titanium platform combined with Sanger (Li et al., 2010; Sun et al., 2010). These differences in length might also be due to the different assemblers used.

*Cucumis melo* unigenes length is comparable to that reported for previous melon Sanger-based transcriptome (Gonzalez-Ibeas et al., 2007), but the addition of 454 reads brought to 53,252 the number of unigenes, two times the number of unigenes reported in the last ICuGI assembly (24,444 unigenes in the v.4) (ICuGI, 2007). The sequences of the unigenes in fasta format are available in the Additional file III.2-2 with unigene numbers from METC000001 to METC053252.

Structural and functional annotation

The long unigenes generated in this study present the advantage of being more accurately annotated. Most unigenes, 49,610 (93%) were predicted to have one ORF. By aligning the unigenes with the genomic sequence of melon, intron locations were predicted in 33,217 unigenes (63%). Annotation results regarding ORF and introns position are included in Additional file III.2-3. To identify *C. melo* unigenes potentially encoding proteins with known function, a BLAST analysis (Altschul et al., 1997) was performed in a sequential way using Swiss-Prot, *Arabidopsis* protein, and Uniref90 protein databases (e-value cutoff of $10^{-20}$). Over 67% of the unigenes (35,740) had at

least one significant hit. Most unigenes had significant hits in the Swissprot (59%) and in the *Arabidopsis* (33%) databases and less in Uniref (8%). Blast results are listed in Additional file III.2-4. Gene Ontology terms were further assigned to melon unigenes using Blast2GO (Conesa and Gotz, 2008) based on the automated annotation of each unigene using BLAST results against the GenBank nonredundant protein database (nr) from NCBI (NCBI, 2010a). A total of 33,575 unigenes (63%) could be annotated with one or more GO terms. The unigenes distribution regarding the number of GOs to which they were assigned is shown in Additional file III.2-5. The number of GO terms per unigene varied from 1 to 34. More than the 78% of the unigenes could be assigned to more than one GO term, the majority of the unigenes being mapped to two to seven GO terms. In total, 121,302 GO terms were retrieved. The GO annotation analysis reinforces the idea that abroad diversity of genes was sampled in our multi-tissue normalized and nonnormalized cDNA libraries. We used the GO annotations to classify each unigene into different functional categories using a set of GO slims, which are a list of high-level GO terms providing a broad overview of the ontology content. Figure III.2-3 shows the functional classification of melon unigenes into GO slims within the categories of biological process and molecular function in comparison with that obtained with the assembly of Sanger ESTs (available at ICuGI database, v4 [ICuGI, 2007]). Gene Ontology annotations showed fairly consistent sampling of functional classes. Cellular process and metabolic process were among the most highly represented groups in the category of biological process, indicating extensive metabolic activities. Genes involved in other important biological processes such as stress response, signal transduction, and ripening were also identified through GO annotations. Under the molecular function category, assignments were mainly to the binding and catalytic activities. A large number of hydrolases, kinases, and transferases were annotated. Also,

transcription and translation factors were well represented. It is worth noting that percentages of genes involved in DNA, RNA, and protein binding and transcription factors were higher in the new unigene collection than in the Sanger ESTs annotation that was richer in transferases, hydrolases and kinases. The distribution of *C. melo* unigenes also follow similar tendencies to that reported for *Arabidopsis thaliana* (Gonzalez-Ibeas et al., 2007), suggesting that the dataset represents a fairly complete melon transcriptome. All GO annotations are compiled in the Additional file III.2-6. Doing a reciprocal blast search, we have also identified 11,253 *C. melo* unigenes (21%) with an ortholog in *Arabidopsis thaliana*, and 18,405 (35%) with a melon equivalent in the ICuGI databases (ICuGI, 2007). A list of the identified orthologs is included in Additional file III.2-7. Only 10,710 (20%) of the unigenes could not be annotated (no significant blast hits, no GO terms, and no ortholog or equivalent). Short sequences are less likely to align with a significant e-value. However, the average length of these nonannotated unigenes was 440 bp, with 50% being longer than 523 bp. For homology searches against known genes, unigenes longer than 200 bp are widely accepted for the effective assignment of functional annotations (Li et al., 2010). In previous studies performed with massive sequencing techniques a similar or even higher number of unigenes did not match with any reported transcribed sequences, potentially representing newly detected transcripts (Barbazuk et al., 2006; Weber et al., 2007; Vega-Arreguin et al., 2009), although a certain degree of contamination of transcriptome libraries with genomic DNA is also possible.

Simple Sequence Repeat and Single Nucleotide Polymorphism discovery and validation
We performed a general screen on the *C. melo* unigene dataset for the presence of microsatellites, analyzing its nature and frequency. A search for di-, tri- and tetra-

nucleotide repeats yielded a total of 3,593 potential SSRs in 3,298 unigenes; that is, approximately 6% of the unigenes contained at least one of the considered SSR motifs. This percentage agrees with previous studies using EST databases that show that 3 to 7% of expressed sequences contain putative SSR motifs (Thiel et al., 2003; Guo et al., 2010).

Approximately 45% (1,631) of the SSRs were in genes with a melon equivalent in ICuGI (ICuGI, 2007), whereas 17% (624) were in nonannotated genes. The maximum and minimum lengths of the repeats were 124 and 17, and the average length was 27 nucleotides. These were mostly tri-nucleotide (67.7%) and less di- (25.2%) and tetra-nucleotide (7.1%). The most common repeat motifs are indicated in Table III.2-2. The melon SSRs annotated in the Sanger collection (available at ICuGI [ICuGI, 2007]) showed a higher percentage of di-nucleotide motifs (41%) and differences in the distribution of trinucleotide motifs. A similar bias toward AG, AAG, and AAAG and against CG repeats has been reported in EST-SSRs of many crops, including previous studies of other cucurbits like cucumber and *Cucurbita pepo* L. (Guo et al., 2010; Blanca et al., 2011). It has been proposed that this may be due to the tendency of CpG sequences to be methylated, which potentially might inhibit transcription. The complete list of SSRs and their corresponding information are provided in Additional file III.2-8. It is known that the Untranslated Regions (UTRs) are richer in SSRs than the coding regions (Morgante et al., 2002; Thiel et al., 2003). Most melon SSRs with known positions were located in the UTRs (57%), mainly in the 3′ region, and less in the ORFs (43%) (Table III.2-3). An analysis of the localization of di-, tri- and tetra-repeats showed that tri-nucleotides localized preferentially in ORFs, consistently with maintenance of the ORFs coding capacity, whereas di- and tetra-nucleotides were more frequent in UTRs. These results agree with those reported previously in melon and other

species (Gonzalez-Ibeas et al., 2007; Ueno et al., 2010). We selected a set of 43 EST-SSRs for validation, all located in ORFs, 40 (93%) amplified polymorphic fragments in a set of 10 genotypes of *Cucumis melo*, belonging to different subspecies and morphotypes of the species (three *inodorus*: T111, Piñonet, and TamDew; three *cantalupensis*: Vedrantais, Dulce, and Noy Israel; two *conomon*: PI161375 and Pat81; and two *momordica*: PI124112 and PI414723). Details of these validated SSRs are included in the Additional file III.2-9.

On average we found 3.6 alleles (from 2 to 7) per primer pair and PIC values ranged from 0.25 to 0.85 (mean = 0.54). Most SSRs (92%) were useful to detect variability between ssp. *melo* and *agrestis*, with unique alleles of one or both subspecies. The genetic relationships among accessions based on SSRs were investigated by cluster analysis (Figure III.2-4). The dendrogram fits very well previous classifications with different marker systems (Esteras et al., 2012). Simple sequence repeats successfully separated accessions of both subspecies and differentiated all the analyzed genotypes. Eighty, sixty-three, and twelve percent were polymorphic within the *momordica*, *cantalupensis*, and *conomon* groups, respectively, and 15% detected variation between the two genetically close Piel de sapo types. These SSRs can be useful for fingerprinting genetically close commercial lines and also for mapping purposes, as T111, Vedrantais, Dulce, PI161375, PI124112 and PI414723 are parentals of several melon genetic maps (Perin et al., 2002: Perchepied et al., 2005; Gonzalez-Ibeas et al., 2007; Harel Beja et al., 2010).

Massive sequencing of ESTs in a number of diverse genotypes has been previously used for developing large SNPs collections (Barbazuk et al., 2006; Novaes et al., 2008; Trick et al., 2009; Guo et al., 2010). Since the ESTs generated under the present study, using the 454 Sequencing technology, are from four different genotypes belonging to the two

subspecies, and the previously available Sanger sequences come from six additional genotypes, we expected SNPs to be frequent in our collection. The SNP calling was done with the default parameters recommended by the ngs_backbone software (Bioinformatics at the Institute forthe Conservation and Breeding of Agricultural Biodiversity, 2010). A total of 810,882 sequences were mapped to the assembled transcriptome for SNPs identification (84% 454 and 26% Sanger). The mapped reads were mostly from the *cantalupensis* Vedrantais (35.8%), the two *inodorus* Piel de sapo cultivars, Piñonet (28.7%) and T111 (17.7%), and the *conomon* PI 161375 (13.9%). The remainder genotypes were less represented: less than 1% the *cantalupensis* C35 and Dulce and the *conomon* Pat81 and less than 0.5% the *cantalupensis* Noy Israel and Charentais and the *inodorus* TamDew.

Stringent quality criteria were used for distinguishing sequence variations from sequencing errors and mutations introduced during the cDNA synthesis step. Only variations with allele and mapping quality over the established thresholds were annotated. By applying these criteria, we initially identified a total of 38,587 SNPs and 5,795 INDELs distributed in 14,417 unigenes (27.1%), averaging a total of 3.1 single variations per unigene. Of this only 3,728 (8%) were detected exclusively with Sanger sequences, whereas the vast majority were detected using only 454 or combining 454 and Sanger. Within the detected SNPs, transitions (74%) were much more common than transversions (26%) (Table III.2-4). A similar number of A/G and C/T transitions and also similar percentages of the four transversion types (A/T, A/C, G/T, and C/G) were found. A set of SNPs could be accurately located with respect to putative initiation and termination codons, being mostly located in ORFs (82%). Detailed information about the identified SNPs and INDELs is included in the Additional file III.2-10, including the nucleotidic change, the number of alleles, and the allelic frequency in each genotype.

Different filters were applied to facilitate the management of the identified variants. INDELs can be distinguished from SNPs (those with VKS filter tag [it is not an SNP] in Additional file III.2-10). Other filters were applied for the *in silico* selection of the SNPs to identify the ones more suited for different research purposes. Unigenes were mapped to the available draft genome sequence of melon. All SNPs mapping in nonunique or noncontiguous genomic regions can be filtered out (those with UCR filter tag in Additional file III.2-10) (9,179, or 24%). Also, all SNPs located in sequences with more than four SNPs or INDELs per 100 bp (with HVR4) may be discarded (986, or 3%). These requirements allow minimizing the discovery of false polymorphisms due to the alignment of paralogs, a potentially significant problem when aligning short sequence reads. Therefore, only nucleotide variants in relatively conserved or recently derived paralogs may have been incorrectly identified as SNPs. The drawback is that some true SNPs in hotspots of genetic diversity or genes under high diversifying selection may be discarded.

From the remaining 28,832 high-confidence SNPs, we implemented different filters to facilitate selection of those appropriated for detecting variability within or between different groups of genotypes:

1. Single nucleotide polymorphisms that are not sequenced or are variable (have more than one allele) within the genotypes Piñonet, T111, Vedrantais, Dulce, PI161375 and Pat81 can be filtered out using NVPin, NVT111, NVVed, NVDul, NVPI161 and NVPat81, respectively (those having the corresponding filter tag in Additional file III.2-10). The absence of these filters indicates that the SNP is sequenced and has only one allele within each genotype.

2. Single nucleotide polymorphisms that are not sequenced or are variable within groups of genotypes- all the ssp. *melo* (T111, Piñonet, Dulce, Vedrantais and Charentais), the

Piel de sapo (T111 and Piñonet), the *cantalupensis* (Vedrantais, Charentais and Dulce), and the *conomon* (PI161375 and Pat81)- can be filtered out with NVmelo, NVPs, NVCant and NVCon. The absence of these filters indicates that the SNP is sequenced and has only one allele within each group.

3. Single nucleotide polymorphisms that are not sequenced or are uniform (have only one allele) in groups of genotypes- the *cantalupensis* and *conomon* (Dulce, Vedrantais, Charentais, Pat81 and PI161375) and the Piel de sapo and *conomon* (T111, Piñonet, Pat81 and PI161375)- can be filtered out with VSCant-Con and VSPs-Con. The absence of these filters indicates that the SNP is sequenced and is variable within the group.

Combining these filters we can, for example, select those SNPs that are sequenced and are uniform within each Piel de sapo genotype (lacking NVPin and NVT111 in Additional file III.2-10), but are variable between them (with NVPs). One thousand, seven hundred eighty-eight SNPs met those criteria (288 with two or more reads per allele), being potentially useful for fingerprint genetically close varieties belonging to this main commercial market class. Only 24% of these SNPs could have been detected only from Sanger sequences, which reinforces the importance of deep sequencing for SNP detection. Among those SNPs (5,980) that were uniform within and between Piel de sapo cultivars (lacking NVPin, NVT111 and NVPs), 2,122 were polymorphic between Piel de sapo and one or both *conomon* sequenced genotypes (lacking NVCon and VSPs-Con). Another SNP set (3,235) was selected for detecting variation between *cantalupensis* and *conomon,* being uniform within each group (lacking VSCant-Con, NVCon and NVCant). These sets of markers could be more efficient for mapping purposes, using crosses between Piel de sapo × conomon or *cantalupensis* × *conomon*, using as parentals not only the sequenced accessions but also additional accessions of each group.

Another set of filters allow to select those SNPs that met different criteria for facilitating validation and for their use in a GoldenGate genotyping assay (Fan et al., 2006; Gupta et al., 2008), discarding those that were closer than 60 bp to another SNP or INDEL (CS60) and/or were predicted to be located closer than 59 bp to an intron (I59) and/or were closer than 60 bp to the unigene edge (CL60). Finally, 11,655 SNPs were selected that met all criteria. From these, 453 SNPs were identified that can be detected as CAPS (lacking nCAP filter tag) as they generate allele-specific restriction targets. We selected 45 of these putative CAPS (all with a minimum of two reads per allele) for validation. All amplified in the corresponding genotypes and were validated by sequencing. Eighty-nine percent (40) provided the expected polymorphism affecting the corresponding restriction target. All were also validated as CAPS, 98% being polymorphic between the expected genotypes after digestion with the corresponding enzyme. This percentage of validation is comparable to that reported in previous studies (Novaes et al., 2008; Blanca et al., 2011). The genetic relationships among accessions based on CAPS were also investigated by cluster analysis. Results were similar to those obtained with SSRs. Cleaved amplified polymorphic sequences efficiently differentiate all cultivars, even the genetically close Piel de Sapo, and separate accessions of both subspecies (Figure III.2-5). Information of the validated SNPs is included in Additional file III.2-11. These CAPS markers are especially useful when experience or equipment for SNPs detection using other methods is not available.

All annotation results (ORFs, introns, descriptions, GO terms, *Arabidopsis thaliana*, melon orthologs, SNVs and SSRs) have been also added in Additional file III.2-12 using the GFF3 standard file format of the Sequence Ontology Resources (Sequence Ontology Project, 2010; Eilbeck et al., 2005).

**Conclusions**

Our results demonstrate how massive sequencing can contribute to improve the melon Sanger-based transcriptome, providing new unigenes that might represent newly detected transcripts. The use of GS FLX Titanium reads and the combined assembly with longer Sanger ESTs has generated long unigenes, allowing an accurate annotation. The detailed annotation provided in this paper will facilitate the use of the collection for gene discovery. It is also being used for genome annotation. Also, the markers collection provided here will be useful for mapping purposes and genotyping studies. The fact that these markers have been detected in a wide set of genotypes of this highly variable species make them suitable to design genotyping assays with broad utility. The filtering process is also very useful to optimize their use with high throughput genotyping platforms.

**Additional material**

**Additional file III.2-1 (excel xls). Sanger ESTs statistics.**

Number of sequences, average length, quality and total length are indicated for Sanger sequences used in this study. Initial data and data after processing are indicated. Data are presented for the different libraries, including tissue and genotype information. These data were produced within the Melogen and ICUGI initiatives and are available at the ICUGI webpage.

**Additional file III.2-2 (fasta format zip comprised). *C. melo* unigenes.**

The fasta sequence of the 53,252 *C. melo* unigenes assembled from 454 and Sanger ESTs is included.

**Additional file III.2-3 (excel xls). Annotation of ORFs and introns.**

Unigene length and predicted position of ORFs and introns is indicated for the whole *C. melo* unigene collection.

**Additional file III.2-4 (excel xls). Blast hits.**

Descriptions build from the blast hit obtained by a sequential blast search of three protein databases Swissprot, *Arabidopsis* protein, and Uniref90 for the whole collection unigene.

**Additional file III.2-5 (power point ppt). Distribution of GO terms.**

The unigenes distribution regarding the number of GOs to which they were assigned is shown.

**Additional file III.2-6 (excel xls). GO terms.**

GO annotations for the whole *C. melo* unigene collection are compiled.

**Additional file III.2-7 (excel xls). *Arabidopsis* orthologs and melon equivalents.**

Orthologs found by reciprocal search with *Arabidopsis* and melon database ICuGI are indicated.

**Additional file III.2-8 (excel xls). *C. melo* SSRs.**

The table provides the list of SSRs identified from the *C. melo* unigene dataset, their length, motif sequences, position in the unigene and scores.

**Additional file III.2-9 (excel xls). Validated *C. melo* SSRs.**

The table provides the list of SSRs experimentally validated using a collection of *C. melo*. Primers used for validation, number of alleles, frequency of the most common allele, Polymorphism Information Content (PIC) across species, and polymorphism detected between species and within Piel de sapo, *cantalupensis*, *conomon*, and *momordica* groups.

**Additional file III.2-10 (excel xls). *C. melo* SNPs and INDELs.**

The table provides the list of INDELs and SNPs identified from the *C. melo* unigene dataset, their position, the nucleotide change, insertion or deletion and the quality of the polymorphic base. The different filters applied for the *in silico* selection of the SNPs and the number of reads per genotype and allele are also indicated. Filter description is included in the Methods section.

**Additional file III.2-11 (excel xls). Validated *C. melo* SNP-CAPS.**

The table provides the list of SNPs that affect restriction targets and were validated by sequencing and *via* CAPS, their position, location, primers used for validation. For those validated as CAPS also frequency of the most common allele and Polymorphism Information Content (PIC) across species is indicated.

**Additional file III.2-12 (GFF3 format, comprised with zip). Annotation results in GFF3.**

All the annotation results (ORFs, introns, descriptions, GO terms, *Arabidopsis* and melon orthologs, SNVs and SSRs) are provided also in the standard format GFF3 of The Sequence Ontology Resources.

**Authors' Contributions**

CE, GM, JC, and BP prepared the cDNA libraries for sequencing. BP and CE selected and validated the SSRs, SNPs and CAPS. JB and PZ performed the bioinformatic analysis. BP, JB, JC, JG and FN participated in the conception of the study. BP was primarily responsible for drafting and revising the manuscript with contributions from coauthors. All authors read and approved the final manuscript.

**Table III.2-1. Sequence statistics of *Cucumis melo* 454 ESTs.**

| Library | Raw reads Number/average length | Total length (bp) | Sequence quality average | Processed reads Number/average length | Total length (bp) | Sequence quality average |
|---|---|---|---|---|---|---|
| Piel de sapo T111† | 117,352 / 225 | 26,388,460 | 33 | 107,756/ 233 | 25,067,194 | 34 |
| PI161375† | 98,460 / 231 | 22,768,835 | 33 | 91,236/ 237 | 21,658,724 | 34 |
| **TOTAL** | **215,812 / 228** | **49,157,295** | **33** | **198,992/ 235** | **46,725,918** | **34** |
| Piel de sapo Piñonet‡ | 249,902 / 397 | 99,083,949 | 33 | 226,650/ 403 | 91,284,481 | 34 |
| Vedrantais‡ | 297,638 / 387 | 115,306,933 | 33 | 263,412/ 398 | 104,745,240 | 34 |
| **TOTAL** | **547,540 / 392** | **214,390,882** | **33** | **490,062/ 400** | **196,029,721** | **34** |

†Summary of the *Cucumis melo* expressed sequences generated with two half runs of 454 Genome Sequencer (GS) FLX. Statistics of reads before and after processing are indicated.

‡Summary of the *Cucumis melo* expressed sequences generated with two half runs of Titanium pyrosequencing. Statistics of reads before and after processing are indicated.

**Table III.2-2. Simple Sequence Repeats (SSRs) statistics.**

| Di-nucleotide repeat | Number of di-SSRs† | % |
|---|---|---|
| AG | 680 | 75 |
| AT | 166 | 18 |
| AC | 60 | 7 |
| Total | 906 | 100 |
| Tri-nucleotide repeat | Number of tri-SSRs | |
| AAG | 1,582 | 65 |
| AGG | 178 | 7 |
| ATC | 174 | 7 |
| AAT | 148 | 6 |
| AAC | 114 | 5 |
| Other tri-nucleotide repeats (%≤ 5 each one) | | |
| AGC,ACC,ACG,CCG,ACT | 237 | 10 |
| Total | 2,433 | 100 |
| Tetra-nucleotide repeat | Number of tetra-SSRs | % |
| AAAG | 120 | 47 |
| AAAT | 34 | 13 |
| AAAC | 19 | 8 |
| AAGG | 15 | 6 |
| Other tetranucleotide repeats (%≤ 5 each one) | | |
| ACTC,AACC,ACAG,AGGC,AACG,AATT, AGCC,AGCG,AAGC,AGGG,AGAT,AAGT, ACCG, ATCG, ATCC, AATC, ACAT, AATG | 66 | 26 |
| Total | 254 | 100 |

†The number of di-, tri-, and tetra-nucleotide repeats identified in the *Cucumis melo*

unigene dataset is shown for the complete set of putative SSRs.

**Table III.2-3. Localization of Simple Sequence Repeats (SSRs) with respect to putative initiation and termination codons in the *Cucumis melo* unigene dataset. Unigenes were checked for the presence of the start and stop codons.**

|  | di-SSRs | | tri-SSRs | | tetra-SSRs | | all-SSRs | |
|---|---|---|---|---|---|---|---|---|
|  | *N°* | *%* | *N°* | *%* | *N°* | *%* | *N°* | *%* |
| 5´-UTR† | 213 | 23.5 | 281 | 11.6 | 90 | 35.4 | 584 | 16.3 |
| ORF‡ | 132 | 14.6 | 1177 | 48.4 | 44 | 17.3 | 1353 | 37.7 |
| 3´-UTR | 436 | 48.1 | 686 | 28.2 | 88 | 34.6 | 1210 | 33.7 |
| Other§ | 125 | 13.8 | 289 | 11.9 | 32 | 12.6 | 446 | 12.4 |
| **Total** | **906** | **100** | **2433** | **100** | **254** | **100** | **3593** | **100** |

†UTR, Untranslated Region.

‡ORF, Open Reading Frame.

§Other means imprecise localization of the SSRs with respect to putative initiation or termination codons.

**Table III.2-4. Single Nucleotide Polymorphism (SNPs) statistics.**

| SNPs | Number | SNPs† | Number | Number |
|---|---|---|---|---|
| Transitions | | Transversions | | Complex |
| A<->G | 13,911 | A<->T | 3,347 | 70 |
| C<->T | 14,716 | G<->T | 2,650 | |
| | | C<->G | 1,866 | |
| | | A<->C | 2,027 | |
| Total | 28,627 (74.2%) | Total | 9,890(25.6%) | 70 (0.2%) |

†Type and number of transition and transversions are shown for putative high quality SNPs identified in the *Cucumis melo* database.
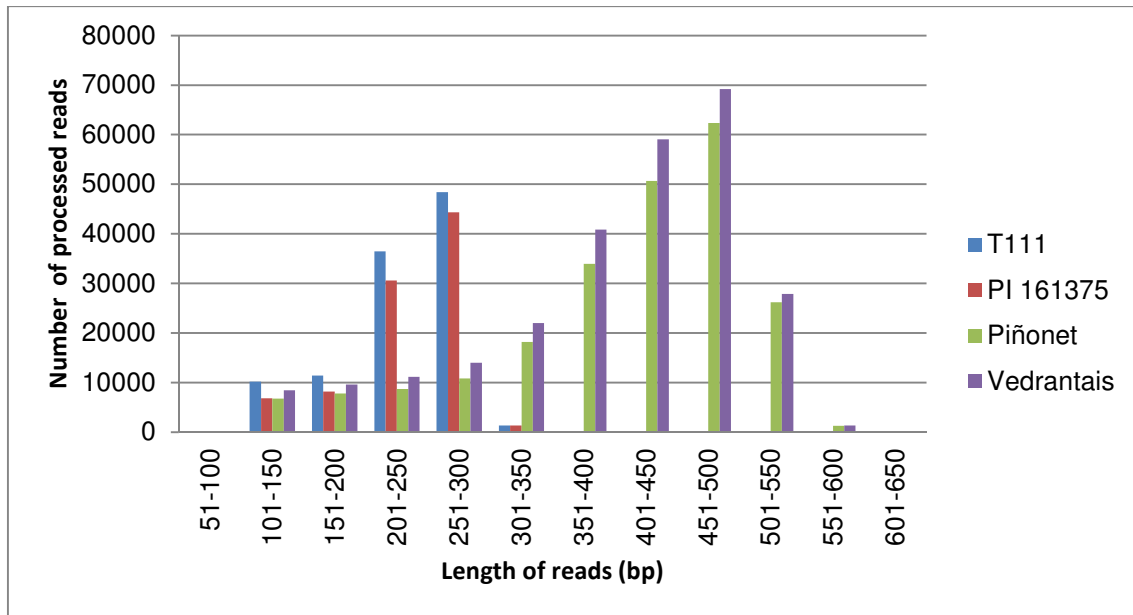
**Figure III.2-1. Length distribution of the *Cucumis melo* Expressed Sequence Tags (ESTs).**

Data obtained after sequencing with two half runs of 454 Genome Sequencer (GS) FLX sequencing each one of two *C. melo* complementary DNA (cDNA) libraries (ssp. *melo* var. *inodorus* H. Jacq. cv. Piel de sapo T111 and ssp. *agrestis* var. *conomon* PI161375) and with two half runs of GS FLX Titanium sequencing each one of two *C. melo* cDNA libraries (ssp. *melo* var. *inodorus* cv. Piel de sapo Piñonet and var. *cantalupensis* Naudin cv. Vedrantais). Data after processing reads are presented.
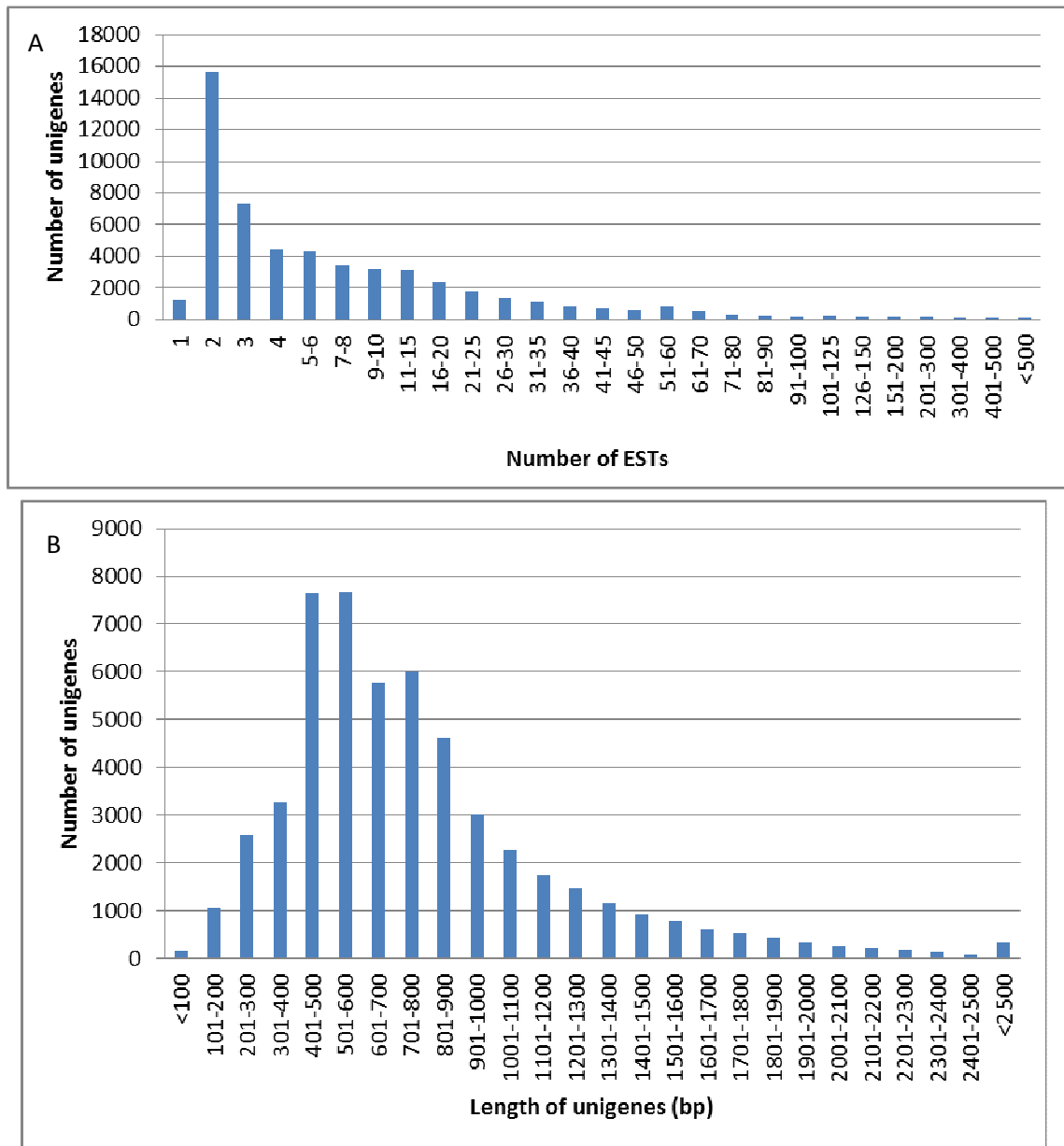
**Figure III.2-2. Distribution of number of Expressed Sequence Tags (ESTs) and length of *Cucumis melo* unigenes.**

A) Unigenes are classified according to the number of 454 and Sanger ESTs used in their assembly. B) Length distribution of *C. melo* unigenes *de novo* assembled combining 454 and Sanger ESTs.

A BIOLOGICAL PROCESS

B MOLECULAR FUNCTION
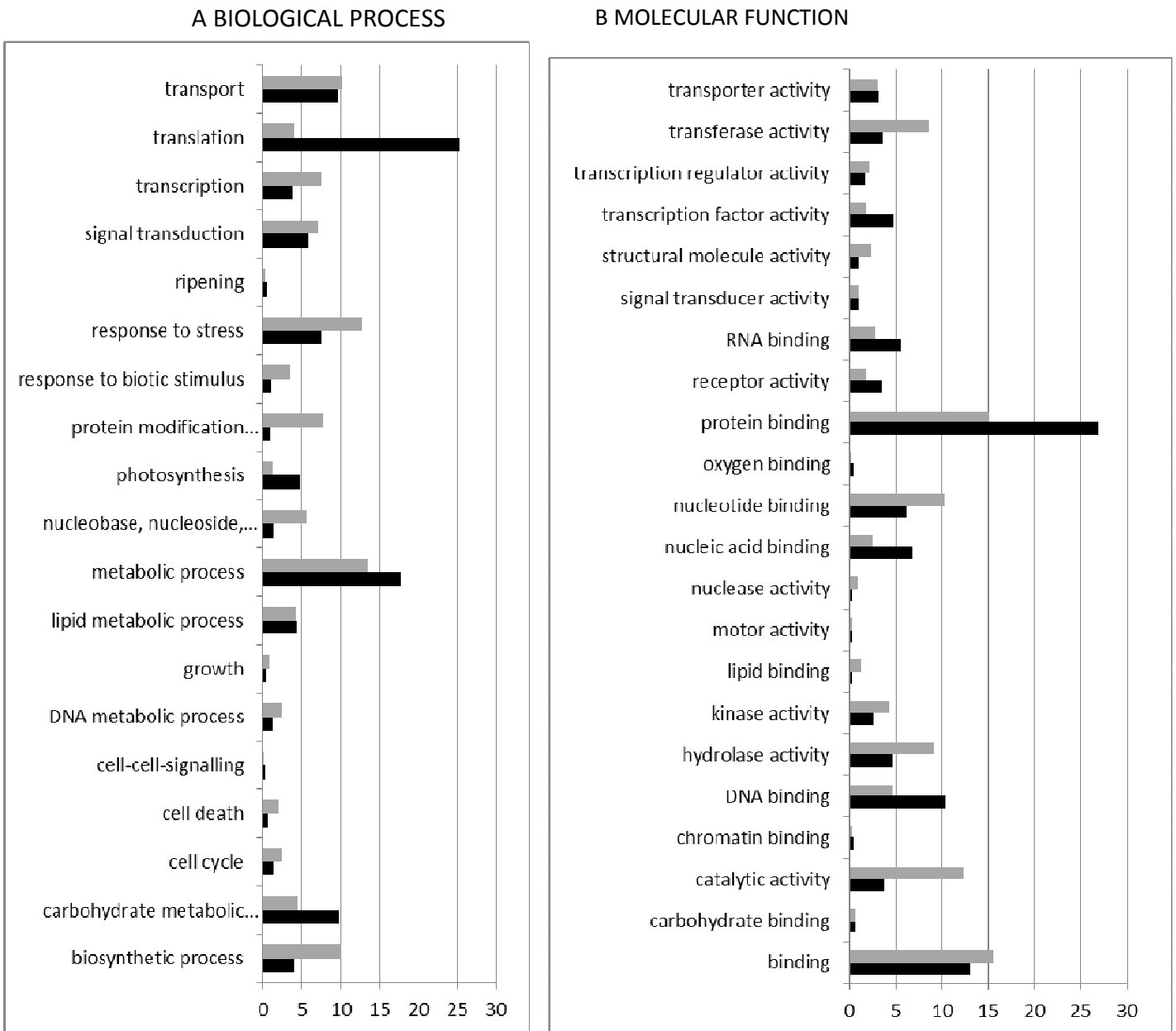
**Figure III.2-3. Percentage of *Cucumis melo* unigenes in each functional category.** Detailed legend: *C. melo* unigenes were classified into different functional groups based on a set of Gene Ontology (GO) slims in the biological process (A) and the molecular function category (B). Bars indicate the percentage of genes annotated in each GO term (black: blast2GO annotation of the unigene collection presented in this paper; gray: GO annotations of the Sanger-based transcriptome available at ICuGI database [ICuGI, 2007]).
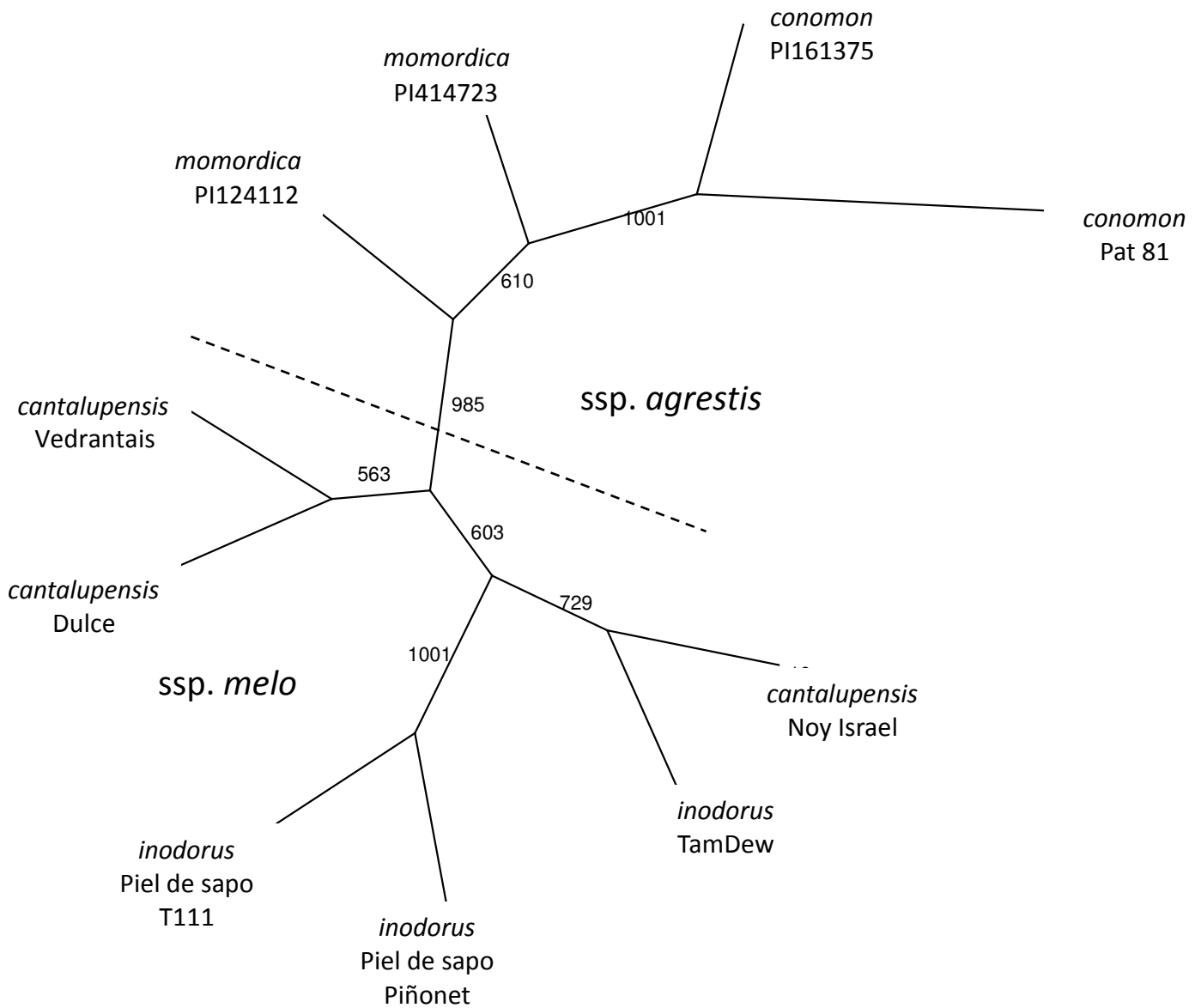
**Figure III.2-4. Genetic relationships among *Cucumis melo* genotypes analyzed with Expressed Sequence Tag (EST)-Simple Sequence Repeats (SSRs).**

Cluster analysis of the genetic relationships among a set of 10 *C. melo* genotypes belonging to two varieties of ssp. *agrestis* (*conomon* and *momordica*) and two varieties of ssp. *melo* (*cantalupensis* Naudin and *inodorus* H. Jacq.). Genetic distance of Nei et al. (1983) and bootstrap (resampling 1,000 times) were calculated.

218
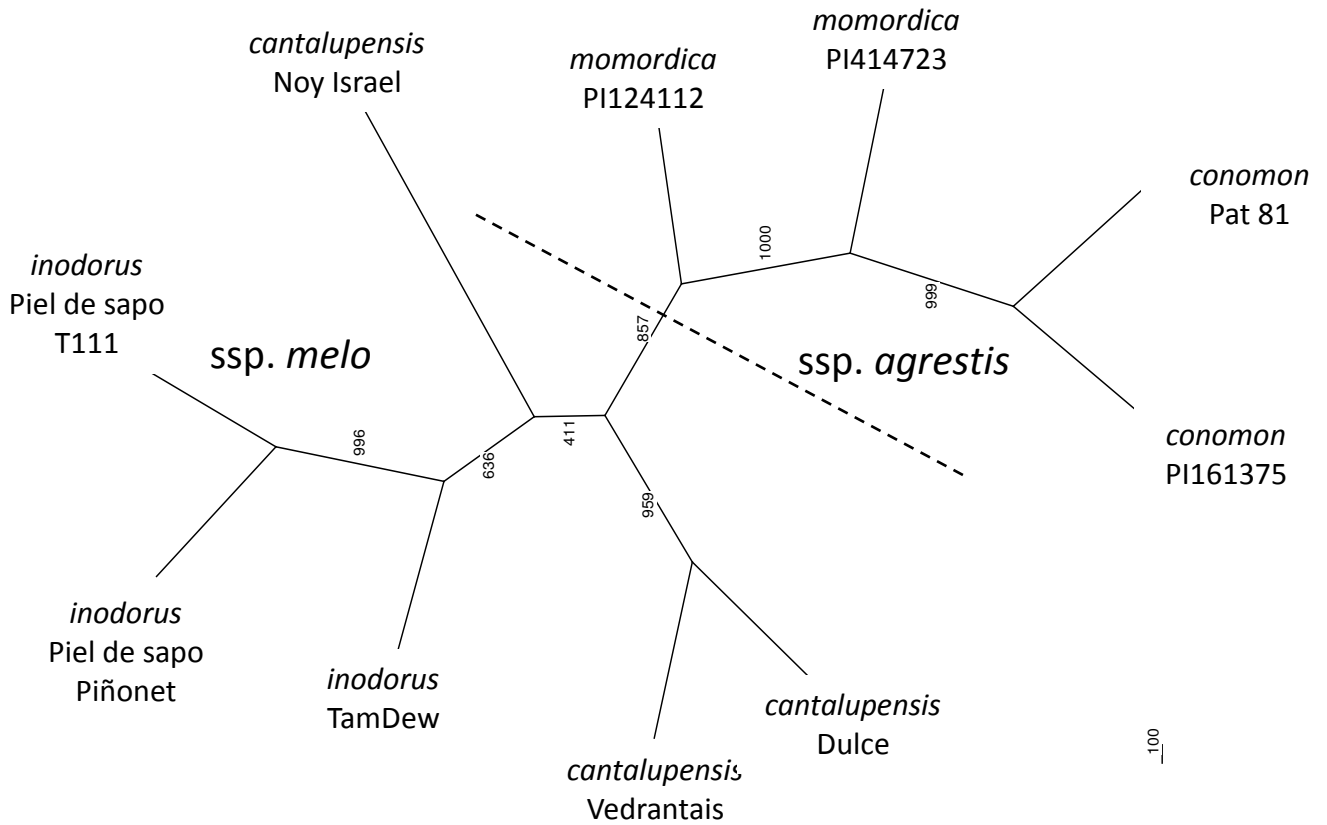
**Figure III.2-5. Genetic relationships among *Cucumis melo* genotypes analyzed with Single Nucleotide Polymorphisms (SNPs)-Cleaved Amplified Polymorphic Sequence (CAPS)** (for genotypes and cluster analysis see Figure III.2-4).

**References**

Abajian C: **Sputnik: DNAS microsatellite repeat search utility.** Chris Abajian, Seattle WA, 1994 [http://espressosoftware.com/sputnik/index.html] (verified 24 May 2010).

Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G: **Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development**. BMC Genomics 2009, 10:399, doi:10.1186/1471-2164-10-399.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: A new generation of protein database search programs.** Nucleic Acids Res 1997, 25:3389–3402, doi:10.1093/nar/25.17.3389.

Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** Nature 2000, 408:796–815, doi:10.1038/35048692.

Arabidopsis Information Resource: **The TAIR database.** Carnegie Institution, Department of Plant Biology, Stanford, CA. 2009 [http://www.arabidopsis.org/] (verified 24 May 2011).

Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE: **Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut light infection.** BMC Plant Biol. 2009, 9:51, doi:10.1186/1471-2229-9-51.

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS: **SNP discovery via 454 transcriptome sequencing**. Plant J 2006, 51:910–918, doi:10.1111/j.1365-313X.2007.03193.x.

Bioinformatics at the Institute for the Conservation and Breeding of Agricultural Biodiversity (COMAV): **Ngs_backbone.** COMAV, Universitat Politècnica de Valencia, Valencia, Spain. 2010 [http://bioinf.comav.upv.es/ngs_backbone] (verified 24 May 2011).

Blanca JM, Cañizares J, Roig C, Ziarsolo P, Nuez F, Picó B: **Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (*Cucurbitaceae*).** BMC Genomics 2011, 12:104, doi:10.1186/1471-2164-12-104.

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari MA, Collin F, Flowers JM, Pitrat M, Purugganan MD, Dogimont C, Bendahmane A: **A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons.** Science 2008, 321:836–838, doi:10.1126/science.1159023.

Cheung F, Haas BJ, Goldberg SMD, May GD, Xiao Y, Town CD: **Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology**. BMC Genomics 2006, 7:272, doi:10.1186/1471-2164-7-272.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, Suhai S: **Using the Mira EST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs.** Genome Res 2004, 14:1147–1159, doi:10.1101/gr.1917404.

Chou HH, Holmes MH: **DNA sequence quality trimming and vector removal.** Bioinformatics 2001, 17:1093–1104.

Conesa A, Gotz S: **Blast2GO: A comprehensive suite for functional analysis in plant genomics.** Int J Plant Genomics 2008, 2008:619832, doi:10.1155/2008/619832.

Deleu W, Esteras C, Roig C, Gonzalez-To M, Fernandez-Silva I, Blanca J, Aranda MA, Arus P, Nuez F, Monforte AJ, Picó B, Garcia-Mas J: **A set of EST-SNPs for map saturation and cultivar identification in melon.** BMC Plant Biol 2009, 9:90, doi:10.1186/1471-2229-9-90.

Eduardo I, Arus P, Monforte AJ, Obando J, Fernandez-Trujillo JP, Martinez JA, Alarcon AL, Alvarez JM, van der Knapp E: **Estimating the genetic architecture of fruit quality**

**traits in melon using a genomic library of near isogenic lines.** J Am Soc Hortic Sci 2007, 132:80–89.

Eilbeck K, Lewis S, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M: **The sequence ontology: A tool for the unification of genome annotations.** Genome Biol 2005, 6(5):R44.

Emrich SJ, Barbazuk WB, Li L, Schnable PS: **Gene discovery and annotation using LCM-454 transcriptome sequencing.** Genome Res 2007, 17:69–73, doi:10.1101/gr.5145806.

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools.** In: Cucurbits: Genetics, Genomics and Breeding of Cucurbits. Edited by Wang Y, Behera TK, Kole C. New Hampshire: Science Publishers Inc, Enfield; 2012:140-198.

European Bioinformatics Institute: **The Emboss: est2genome.** European Bioinformatics Institute, Hinxton, UK. 2001 [http://emboss.sourceforge.net/apps/cvs/emboss/apps/est2%20genome.html] (not verified)

European Bioinformatics Institute: **UniProt reference clusters (UniRef) databases**. European Bioinformatics Institute, Hinxton, UK. 2010 [http://www.ebi.ac.uk/uniref] (verified 24 May 2011).

Exonerate: **A generic tool for sequence alignment**. European Bioinformatics Institute, Hinxton, UK. 2005 [http://www.ebi.ac.uk/~guy/exonerate/] (verified 25 May 2011).

Ezura H, Fukino N: **Research tools for functional genomics in melon (*Cucumis melo* L.): Current status and prospects.** Plant Biotechnol 2009, 26:359–368, doi:10.5511/plantbiotechnology.26.359.

Fan JB, Chee MS, Gunderson KL: **Highly parallel genomic assays**. Nat Rev Genet 2006, 7:632–644, doi:10.1038/nrg1901.

Folta KM, Clancy MA, Chamala S, Brunings AM, Dhingra A, Gomide L, Kulathina RJ, Peres N, Davis TM, Barbazuk WB: **A transcript accounting from diverse tissues of a**

**cultivated strawberry.** Plant Gen 2010, 3:90–105, doi:10.3835/plantgenome2010. 02.0003.

Gedye K, Gonzalez-Hernandez J, Ban Y, Ge X, Thimmapuram J, Sun F, Wrigh Ch, Ali S, Boe A, Owens V: **Investigation of the transcriptome of prairie cord grass, a new cellulosic biomass crop.** Plant Gen 2010, 3:69–80, doi:10.3835/plantgenome2010. 06.0012.

Gonzalez VM, Garcia-Mas J, Arus P, Puigdomenech P: **Generation of a BAC-based physical map of the melon genome.** BMC Genomics 2010a, 11:339, doi:10.1186/1471-2164-11-339.

Gonzalez VM, Mir G, Arus P, Puigdomenech P, Garcia-Mas J: **Abstract: Towards the whole sequence of the melon genome.** In Plant Animal Genome XVII Conf., San Diego, CA. Publisher unknown. 10–14 January 2009.

Gonzalez VM, Rodriguez-Moreno L, Centeno E, Benjak A, Garcia-Mas J, Puigdomenech P, Aranda MA: **Genome-wide BAC-end sequencing of *Cucumis melo* using two BAC libraries.** BMC Genomics 2010b, 11:618, doi:10.1186/1471-2164-11-618.

Gonzalez-Ibeas D, Blanca J, Roig C, Gonzalez-To M, Pico B, Truniger V, Gomez P, Deleu W, Cano-Delgado A, Arus P, Nuez F, Garcia-Mas J, Puigdomenech P, Aranda MA: **Melogen: An EST database for melon functional genomics.** BMC Genomics 2007, 8:306, doi:10.1186/1471-2164-8-306.

Guo S, Zheng Y, Joung JG, Liu S, Zhang Z, Crasta OR, Sobral BW, Xu Y, Huang S, Fei Z: **Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types.** BMC Genomics 2010, 11:384, doi:10.1186/1471-2164-11-384.

Gupta PK, Rustgi S, Mir RR: **Array-based high throughput DNA markers for crop improvement.** Heredity 2008, 101:5–18, doi:10.1038/hdy.2008.35.

Harel-Beja R, Tzuri G, Portnoy V, Lotan-Pompan M, Lev S, Cohen S, Dai N, Yeselson L, Meir A, Libhaber S, Avisar E, Melame T, Van Koert P, Verbakel H, Hofstede R, Volpin H, Oliver M, Fougedoire A, Stalh C, Fauve J, Copes B, Fei Z, Giovannoni J, Ori N, Lewinsohn E, Sherman A, Burger J, Tadmor Y, Schaffer A, Katzir N: **A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes.** Theor Appl Genet 2010, 12:511–533, doi:10.1007/s00122-010-1327-4.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan , Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S: **The genome of the cucumber, *Cucumis sativus*.** L Nat Genet 2009, 41:1275-1281, doi:10.1038/ng.475

International Cucurbit Genomics Initiative (ICuGI): **Cucurbit genomics database**. USDA-ARS, Boyce Thompson Institute, Cornell Univ, Ithaca, NY. 2007 [http://www.icugi.org] (verified 24 May 2011).

Iseli C, Jongeneel CV, Bucher P: **ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.** Proc Int. Conf Intell Syst Mol Biol 1999, 138–148.

Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** Bioinformatics 2010, 26:589–595, doi:10.1093/bioinformatics/btp698.

Li Y, Luo HM, Sun C, Song JY, Sun YZ, Wu Q, Wang N, Yao H, Steinmetz A, Chen SL: **EST analysis reveals putative genes involved in glycyrrhizin biosynthesis.** BMC Genomics 2010, 11:268, doi:10.1186/1471-2164-11-268.

Liu K, Muse SV: **Powermarker: Integrated analysis environment for genetic marker data.** Bioinformatics 2005, 21:2128–2129, doi:10.1093/bioinformatics/bti282.

Mascarell-Creus A, Cañizares J, Vilarrasa J, Mora-Garcia S, Blanca J, Gonzalez-Ibeas D, Saladie M, Roig C, Deleu W, Pico B, Lopez-Bigas N, Aranda MA, Garcia-Mas J, Nuez F, Puigdomenech P, Cano-Delgado A: **An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon (*Cucumis melo* L.).** BMC Genomics 2009, 10:467, doi:10.1186/1471-2164-10-467.

Melogen: **Development of genomics tools in melon.** COMAV, Universitat Politècnica de Valencia, Valencia, Spain. 2003 [http://www.melogen.upv.es] (verified 25 May 2011).

Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.** Nat Genet 2002, 30:194–200. doi:10.1038/ng822.

National Center for Biotechnology Information (NCBI): **Nonredundant protein database.** NCBI, Bethesda, MD. 2010a [ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz] (verified 24 May 2011).

National Center for Biotechnology Information (NCBI): **Sequence read archive.** NCBI, Bethesda, MD. 2010b [http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi] (verified 24 May 2011).

Nei M, Tajima F, Tateno Y: **Accuracy of estimated phylogenetic trees from molecular data.** J Mol Evol 1983, 19:153–170.doi:10.1007/BF02300753.

Novaes E, Drost D, Farmerie W, Pappas G, Grattapaglia D, Sederoff RR, Kirst M: **High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome.** BMC Genomics 2008, 9:312, doi:10.1186/1471-2164-9-312.

Perin C, Gomez-Jimenez M, Hagen L, Dogimont C, Pech JC, Latche A, Pitrat M, Lelievre JM: **Molecular and genetic characterization of a non-climacteric phenotype in melon reveals two *loci* conferring altered ethylene response in fruit.** Plant Physiol 2002, 129:300–309, doi:10.1104/pp.010613.

Perchepied L, Bardin M, Dogimont C, Pitrat M: **Relationship between *loci* conferring downy mildew and powdery mildew resistance in melon assessed by quantitative trait *loci* mapping**. Phytopathology 2005, 95:556–565, doi:10.1094/PHYTO-95-0556.

Pitrat M: **Melon (*Cucumis melo* L.).** In Handbook of plant breeding.Vegetables (vol. I). Edited by Prohens J, Nuez F.Springer, New York, NY; 2008:283-315.

Rozen S, Skaletsky HJ: **Primer3 on the WWW for general users and for biologist programmers.** In Bioinformatics methods and protocols: Methods in molecular biology. Edited by Krawetz S, Misener S. Humana Press, Totowa, NJ; 2000:365-386.

Schaefer H, Heibl C, Renner SS: Gourds afloat: **Adated phylogeny reveals an Asian origin of the gourd family (*Cucurbitaceae*) and numerous oversea dispersal events**. Proc Biol Sci 2009, 276:843–851, doi:10.1098/rspb.2008.1447.

Sequence Ontology Project: **Generic feature format version 3**. Karen Eilbeck, Salt Lake City, UT. 2010 [http://www.sequenceontology.org/gff3.shtml] (verified24 May 2011).

Shendure J, Ji H: **Next-generation DNA sequencing**. Nat Biotechnol 2008, 26:1135–1144, doi:10.1038/nbt1486.

Sun Ch, Li Y, Wu Q, Luo H, Sun Y, Song J, Lui EMK, Chen S: **De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis.** BMC Genomics 2010, 11:262, doi:10.1186/1471-2164-11-262.

Swarbreck D, Wilks C, Lamesch P, Berardini TA, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, et al.: **The Arabidopsis information resource (TAIR): Gene structure and function annotation.** Nucleic Acids Res 2008, 36:1009–1014.

Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.).** Theor Appl Genet 2003, 106:411–422.

Trick M, Long Y, Meng J, Bancroft I: **Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing.** Plant Biotechnol J 2009, 7:334–346, doi:10.1111/j.1467-7652.2008.00396.x.

Ueno S, Le Provost G, Leger V, Klopp Ch, Noirot C, Frigerio JM, Salin F, Salse J, Abrouk M, Murat F, Brendel O, Derory J, Abadie P, Leger P, Cabane C, Barre A, Daruvar A, Couloux A, Wincker P, Reviron MP, Kremer A, Plomion Ch: **Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species.** Oak BMC Genomics 2010, 11:650.doi:10.1186/1471-2164-11-650.

UniProt Consortium: **The UniProtKB/Swiss-Prot database.** Uniprot Consortium, UK, Switzerland, United States. 2010a [http://www.uniprot.org/downloads] (verified 24 May2011).

UniProt Consortium: **The universal protein resource (UniProt) in 2010**. Nucleic Acids Research 2010b, 38:142–148.

Vega-Arreguin JC, Ibarra-Laclette E, Jimenez-Moraila B, Martinez O, Vielle-Calzada JP, Herrera-Estrella L, Herrera-Estrella A: **Deep sampling of the Palomero maize**

**transcriptome by a high throughput strategy of pyrosequencing.** BMC Genomics 2009, 10:299, doi:10.1186/1471-2164-10-299.

Wang W, Wang Y, Zhang Q, Qi Y, Guo D: **Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencin**g. BMC Genomics 2009, 10:465, doi:10.1186/1471-2164-10-465.

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB: **Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing**. Plant Physiol 2007, 144:32–42, doi:10.1104/pp.107.096677.

# IV. DESARROLLO Y APLICACIÓN DE HERRAMIENTAS GENÓMICAS EN CALABACÍN

## IV-1. Genetic diversity of Spanish *Cucurbita pepo* landraces: an unexploited resource for summer squash breeding

**Abstract**

*Cucurbita pepo* is a worldwide cultivated vegetable of American origin. Most of the widely grown commercial types are known as summer squashes and belong to the elongated forms of *C. pepo* ssp. *pepo* (Cocozelle, Vegetable marrow and Zucchini groups). These forms were developed in Europe after the arrival of the first American landraces through a process of selection and fixation that led to a loss of genetic diversity. Part of the genetic variability of the first American cultigens remains intact in diverse landraces that are still cultivated for self-consumption and sale in local markets. Using the first collection of genomic and EST-derived microsatellites that has just become available for the species, we compared the natural variation present in a collection of Spanish landraces with that of a set of commercial varieties and hybrids, representing current summer squash market offerings. A total of 194 alleles allowed us to distinguish all the genotypes, even those that were closely related. In general, Cocozelle and Vegetable marrow, groups with considerably long histories, were more variable than the Zucchini group, of more recent origin. We found significant genetic diversity among landraces. The variation present among landraces belonging to the Zucchini group was larger than that of the commercial cultivars. Cluster, principal coordinate and population structure results suggested that the variation of the Spanish landraces has not been extensively used in breeding. Commercial summer squashes can therefore benefit from this underexploited variability, especially from certain landraces that already display favourable commercial traits.

**Background**

*Cucurbita pepo* L. is the most economically important species of the genus *Cucurbita* L. (*Cucurbitaceae*). Cultivated *C. pepo* has traditionally been considered to comprise two

subspecies with cultivated forms (Decker, 1988; Sanjur et al., 2002; Nesom, 2011), each one encompassing several cultivar-groups: *C. pepo* ssp. *pepo* L. (including Pumpkin group, Vegetable marrow group, Cocozelle group and Zucchini group) and *C. pepo* ssp. *texana* (Scheele) Filov (syn ssp. *ovifera* (L.) Decker) (including Acorn group, Scallop group, Crookneck group and Straightneck group) (Ferriol et al., 2003; Paris, 1986; Paris et al., 2003). The primary economic value of this species is based on the culinary use of immature fruits often referred to collectively as ''summer squash''. Only cultivars of the Pumpkin and Acorn groups have a widespread use as ''winter squashes'', grown for consumption of their mature fruits (Paris, 2008).

This species originated in North America (Smith, 1997), although nowadays its distribution is worldwide, and is one of the most phenotypically variable species in the plant kingdom. Wild relatives and ancient cultigens of the species, mostly round- or nearly round-fruited, are still found in the USA, Mexico and Central America, where several variable morphotypes were selected by Native Americans in pre-Columbian times (Paris, 2000 and 2008). A number of these accessions are housed in facilities in Mexico, the USA and Costa Rica (Lira and Montes, 1994).

*C. pepo* ssp. *ovifera* morphotypes were developed in America, where today they are most popular, as they can hardly be found in Europe. Since the arrival to Europe of several American cultigens, an extraordinary variety of new phenotypes has been generated through hybridisation and recombination. Selection for long-fruitedness in the subspecies *pepo* was reportedly first conducted in Italy in the sixteenth century (Paris, 2008). New cultivars were then developed, particularly the elongated forms of this subspecies: first the Vegetable marrow group (short, tapered cylindrical fruits), followed soon afterward by the Cocozelle group (long or very long, bulbous and cylindrical fruits) and much more recently, probably in the late nineteenth century, the Zucchini

group (long, uniformly cylindrical fruits). Zucchini is of the most recent origin, and is also one of the least variable morphotypes (Paris, 2000). Nowadays, cultivars of the Zucchini group dominate both the squash market and the breeding efforts of seed companies.

Summer squashes of the Vegetable marrow, Cocozelle and Zucchini groups are popular in the Middle East, North Africa and European countries bordering the Mediterranean Sea, with cultivars meeting specific consumer preferences in each area. Diverse collections of landraces can still be found in Italy, Turkey (Paris, 2008) and Spain (Ferriol and Picó, 2008). In the last century, breeders have made great strides in summer squash breeding by performing selection and fixation from some of these landraces in order to develop new and improved cultivars and hybrids, which are now replacing old landraces.

In Spain, the Genebank of the Institute for the Conservation and Breeding of Agricultural Diversity (COMAV), located at the Polytechnic University of Valencia, is dedicated to preserving this genetic heritage. This Genebank maintains seed collections of old Spanish winter and summer squash cultigens of *C. pepo* ssp. *pepo* (Ferriol et al., 2003). These were mainly collected from the farmers that still use them for self-consumption and/or sale in local markets. One of the aims of the COMAV Institute is to characterise these landraces in order to use them in summer squash breeding.

The genetic variability within *C. pepo* has previously been assessed using allozymes and different DNA marker systems [Restriction Fragment Length Polymorphism (RFLPs), Random Amplified Polymorphic DNA (RAPDs), Amplified Fragment Length Polymorphisms (AFLPs) and Inter Simple Sequence Repeats (ISSRs)] (reviewed in Lebeda et al., 2007 and in Esteras et al., in press). Most of the studies have been focused on the assessment of the genetic and evolutionary relationships between the wild and

domesticated types, between the two subspecies and among the cultivar groups, including only a few representatives of the European landraces (Ferriol et al., 2003; Paris et al., 2003).

Most of the marker systems used to date have limitations associated with their dominant and/or unreliable nature. Microsatellite markers [Simple Sequence Repeats (SSRs)], if available, are preferred due to their being reliable and codominant and at the same time multiallelic and highly polymorphic, making them well suited for detecting variation among closely related varieties. However, these markers are difficult to obtain in species for which genetic and genomic tools are lacking.

Cucurbits are becoming primary models for the study of several biological processes (Boualem et al., 2008; Ezura and Fukino, 2009; Li et al., 2009). The availability of molecular tools, including the whole genome sequence, has significantly increased in the last years, mainly for melon and cucumber (Huang et al., 2009; Gonzalez et al., 2010), but more recently for the *Cucurbita* genus, as well.

Gong et al. (2008) developed the first collection of genomic SSRs (gSSR), about 400, using an enriched genomic library from an oil-seed pumpkin cultivar. Recent advances in next-generation sequencing technologies (Metzker, 2010) have allowed us to generate the first *Cucurbita* transcriptome (Blanca et al., 2011), with 49,610 *Cucurbita* unigenes assembled *de novo* from 512,751 high quality ESTs, sequenced using Roche GS/454. These unigenes were screened for SSR motifs, leading to the discovery of a collection of 1,882 unigenes with SSR motifs (EST-SSR), which could be considered to be the first genomic resource in this genus. EST-SSRs have the advantage over gSSRs of being in the functional fraction of the genome.

In this study, we evaluate the utility of these new genomic resources for assessing genetic variation within *C. pepo*, even among closely related cultivars. Both gSSRs and

EST-SSRs have proven to be useful for genotyping a selected set of Spanish landraces, belonging to the three summer squash morphotypes of high commercial value. These varieties have also been phenotyped for the presence of traits of agronomic interest. Most of the genetic variation that still remains in these landraces is not present in the set of commercial varieties and hybrids, which is representative of the current offering of summer squashes in the European market. Their diversity and the presence of certain unique alleles make the Spanish landraces of *C. pepo* an invaluable resource with great potential for summer squash breeding, mainly for the most widely grown and least variable Zucchini group.

**Materials and methods**

Germplasm

A set of twenty-three Spanish landraces, included in the core collection of *C. pepo* held and characterised by the Cucurbits breeding group at the COMAV, representing the three most important morphotypes of *C. pepo* ssp. *pepo* (five belonging to the Zucchini group, eleven to the Vegetable marrow group and seven to the Cocozelle group) and the variability found in all the agroecological regions in the country (Table IV.1-1; Figure IV.1-1) were selected for this study. All the selected accessions are primitive landraces, cultivated in small orchards used for self-consumption or sale in local markets. Original germplasm, maintained through sibling, was used for the molecular study.

Twelve commercial varieties (three belonging to the Zucchini group, seven to the Cocozelle group and two summer squash Pumpkins) and six commercial hybrids (four belonging to the Zucchini group, one to the Vegetable marrow group and another to the Cocozelle group) were included as representatives of the main commercial varieties currently offered in the market (Table IV.1-1). Other accessions were included as

controls: one South American accession of the Zucchini group, one North African accession of the Vegetable marrow group and two Pumpkin accessions used as parentals of the mapping populations: the oil-seed styrian pumpkin (Gong et al., 2008) and the Turkish variety PI171678, which is used to increase the levels of carotenoids in *C. pepo* (Ferriol et al., 2003) and three accessions belonging to three morphotypes of the ssp. *ovifera*, one of each of the Acorn, Scallop and Crookneck groups.

Morphological characterisation

Three plants per accession were grown in the greenhouse and characterised for different vine, flower and fruit traits of commercial interest (Table IV.1-2): plant growth habit (bushy (B), compact with short inter-nodes; viny (V), spreading growth habit, with long, thin internodes; and intermediate (I), semi-bushy), lateral branch development [no-branching (0), moderately branched (0.5) and highly branched (1)], presence of spines in stems and leaves [spineless (0), sparse and small spicules (0.5) and dense, large and sharp spicules (1)], days to male and female flowering from transplant (DMF and DFF), node in which the first male/female flower appears (NMF and NFF), number of male/female flowers 7 days after the opening of the first female flower (NºMF and NºFF) and the femaleness ratio NºFF/NºMF. Mature fruits were also characterised (Figure IV.1-1).

SSR analysis

Genomic SSRs

A set of 30 gSSRs were used for the analysis (Table IV.1-3). These were selected from those developed by Gong et al. (2008) using an SSR-enriched partial genomic library prepared from the Austrian oil-pumpkin variety Gleisdorfer Olkürbis (*C. pepo* ssp.

*pepo*). We selected at least one marker in each of the 20 linkage groups of *C. pepo* ssp. *pepo* (oil-pumpkin variety) x *C. pepo* ssp. *ovifera* (a variety belonging to the Crookneck group) map developed by these authors (Table IV.1-3).

EST-SSRs

We used a set of 27 EST-SSRs, selected from those identified *in silico* by Blanca et al. (2011), after screening the 49,610 unigenes of the first *C. pepo* transcriptome [assembled using ESTs from the cultivars ZU-MU16 (Zucchini group) and UPV-196 (Scallop group)] for SSR motifs (Table IV.1-4). EST sequences containing perfect repetitions of more than seven units were selected. Most of the SSRs are within genes with a putative known function and were mainly located in ORFs (Open Reading Frames), with only a few of them in UTRs (Untranslated Regions). Most of the EST-SSRs selected were in unigenes that showed homology to annotated protein sequences (Blanca et al., 2011). Some were involved in catabolism, transport, regulation of transcription and stress and defense response (Table IV.1-4). The primers used for each SSR *locus* are listed in Blanca et al. (2011).

PCR amplification and detection of SSR *loci*

DNA was extracted using a modified CTAB method (Doyle and Doyle, 1990) from young leaves of three plants per accession. PCR reactions were performed in a final volume of 15 µL with 15 ng of genomic DNA, 1.5 U Taq DNA polymerase (Biotools, B&M Labs S.A., Madrid, Spain), 1x PCR buffer (75 mM Tris-HCl pH 9, 20 mM $(NH_4)_2SO_4$, 50 mM KCl), 3 mM $MgCl_2$, 200 µM of each dNTP, 0.15 µM each forward and reverse primer, 0.02 µM/0.2 µM dye-labelled primer (FAM, VIC, PET or NED for ABI sequencer and IRDye®700 or 800 for LI-COR). Forward primers were designed

with an M13 tail sequence added to their 5´ end (5´ -cacgacgttgtaaaacgacc-3´). The cycling conditions were as follows: denaturation at 95ºC for 3 min, followed by 10 cycles of 30 s at 95ºC, 30 s at 65ºC (with each cycle, the annealing temperature decreasing 1ºC) and 30 s at 72ºC. Products were subsequently amplified for 20 cycles at 95ºC for 30 s, 55ºC for 30 s and 72ºC for 30 s, with a final extension at 72ºC for 5 min. SSR markers were scored either on the LI-COR® 4300 DNA Analyzer (Li-Cor Bioscience, Lincoln, Nebraska, USA) or on the ABI 3130XL capillary sequencer (Applied Biosystems, Carlsbad, California, USA).

Genetic variability analysis

Due to the bulking of samples during DNA extraction, the observation of two or more SSR alleles in a single reaction could have resulted from the presence of heterozygous plants, homozygous plants for the alternative alleles or a combination of both. All the detected alleles were assumed to have a frequency of 1/n (n = number of alleles). Microsatellite allele sizes were estimated comparing their migration with the IRDye®700 or 800 50-350 bp size standards (Li-Cor Bioscience, Lincoln, Nebraska, USA). The number of alleles and the Polymorphism Information Content (PIC) were calculated for each *locus* for the 48 genotypes as well as for groups of morphotypes using the PowerMarker software (Liu and Muse, 2005). ANOVA was used to compare the mean allele number and the PIC between EST-SSRs and gSSRs. The correlation between the number of repeated units and the PIC was also studied. All statistical analyses were performed with Statgraphics Plus v.5.0 (Statistical Graphics Corporation, Inc., Rockville, MD, USA).

Genetic similarity between accessions based on the presence or absence of SSR alleles was calculated using the Dice (Nei and Li, 1979) similarity coefficient ($S_{ij}$). Genetic

distance (GD) values among genotypes were calculated as $1 - S_{ij}$. From the genetic distance matrix, a neighbour-joining (NJ) tree was obtained using PowerMarker. Distance matrices obtained with gSSRs and EST-SSRs were compared by the Mantel test (Mantel, 1967). The reliability and robustness of the tree were tested by bootstrap analysis with 1,000 replications to assess branch support using the PHYLIP 3.6 software. The population structure of our collection was estimated using STRUCTURE version 2.3 (Pritchard et al., 2000; Falush et al., 2003, 2007; Hubisz et al., 2009). Accessions CR-NSL5227, AC-PI615111 and SC-UPV196 were excluded from the STRUCTURE analysis because they belong to *C. pepo* ssp. *ovifera*, whereas the rest of the genotypes belong to *C. pepo* ssp. *pepo*. The admixture and independent allele frequency options were chosen. A burn-in period of 500,000 Markov Chain Monte Carlo iterations and the 1,000,000 iterations after burn were performed to estimate the parameters. Twenty runs were done for each K (number of populations) from 1 to 8. The K optimum was defined according to Evanno et al. (2005). Principal coordinate analysis (PCoA) was carried out using the Dice genetic distance matrix and the DCENTER and EIGEN procedures of NTSYSpc 2.02 (Rohlf, 1998).

**Results and discussion**

<u>Phenotypic variation</u>

The phenotypic variability found among the summer squash genotypes used in the present study is provided in Table IV.1-2. The morphological characterisation of the landraces allowed us to define the morphotype of each accession and to record the presence or absence of traits of commercial interest.

The Zucchini group is by far the most widely grown summer squash, and for this reason it is the morphotype in which seed companies have primarily focused their breeding

efforts. The Zucchini commercial hybrids and varieties were, as expected, bushy, unbranched, spineless and had uniformly cylindrical fruits. The bushy growth habit, conferred by a single gene (Bu) (Paris and Brown, 2005), greatly facilitates multiple harvesting, and this gene is introgressed in all the current commercial Zucchini types. The lack of branching improves fruit-picking efficiency, and few or absent spicules is a desirable trait for any cultivar. Commercial cultivars of the Zucchini group were also early-flowering, with male and female flowers starting to appear from 30 to 36 DAP (days after planting) and from 41 to 52 DAP, respectively, and with a high female/male ratio (0.3-0.6). Selection for earliness and female tendency were two other important achievements of pre-twentieth century squash breeding (Paris, 2008).

In the Cocozelle commercial group, bushy growth habit was not very common. Viny and semi-bushy, more branched and spiny varieties were found. A more delayed flowering (male and female flowers appearing from 32 to 45 DAP, and 43 to 61 DAP) associated with a decreased female/male ratio (0.1-0.5) was also displayed. The commercial Vegetable marrow and Pumpkin groups displayed intermediate traits.

Spanish landraces are mostly primitive cultivars that have not been subjected to intensive selection. They harbour a great deal of phenotypic variation. Most of the Zucchini landraces were bushy rather than viny, and developed uniformly cylindrical dark-green fruits (Figure IV.1-1). Some landraces showed desired commercial traits, being unbranched, spineless and early-flowering, and having even more positive female/male ratios than some commercial varieties (ZU-E27 and ZU-MU16). Others, related to more ancestral *C. pepo* (ZU-E10, ZU-MU20, ZU-A13) as well as the landrace from Ecuador (ZU-ECU227), had some commercially undesirable traits. Our results are in agreement with the primitive traits reported in Italian and Turkish landraces (Paris, 2008).

Phenotypic variation was more apparent among Spanish landraces of the Vegetable marrow and Cocozelle groups, as these were indeterminate or viny, having branched and from medium to highly spiculate foliage. They displayed everything from early to late flowering (male and female flowers appearing from 30 to 63 and DAP, and 45 to 77 DAP, respectively), with some accessions having a predominant male sexuality (female/male ratio 0.05-0.2), mainly in the Cocozelle group. Only one Cocozelle variety was bushy and not branched (CO-V185). High variability in fruit colour, colour pattern, rind texture and fruit size was found in Vegetable marrow and Spanish Cocozelle landraces, ranging from yellow to cream to white to light, medium and dark green, and being striped, dotted, smooth, ribbed, wrinkled and/or warted (Figure IV.1-1).

Molecular variation

gSSRs and EST-SSRs

A summary of the results obtained with genomic- and EST-SSRs are shown in Tables IV.1-3 and IV.1-4. Amplification products were obtained for 29 (97%) gSSRs, 27 (90%) of which were polymorphic among the *Cucurbita* genotypes. A total of 109 alleles were found across the full set of accessions, ranging from 2 to 9 (average 3.8) alleles per SSR. The average allele number in our collection was similar to that reported by Gong et al. (2008) (3.7), who validated these gSSRs using a set of 8 genotypes of both subspecies of *C. pepo* (one cultivar of each group: Zucchini, Cocozelle, Vegetable marrow, oil-Pumpkin, Acorn, Scallop, Crookneck and Straightneck), but also including 3 accessions of the related species, *C. moschata* (Duchesne ex Lam.) Duchesne ex Poir, and 1 of the wild species, *C. ecuadorensis* H. C. Cutler & Whitaker. Despite our larger number of accessions, here we assessed only the intraspecific variability, so we expected the SSR polymorphism to be lower in our collection. Nevertheless, our results

demonstrate that this germplasm sample is highly variable; noteworthy 12 of these markers (40%) detected more alleles in our population than in that of Gong et al. (2008).

Twenty-six EST-SSRs (96%) amplified polymorphic fragments in the set of *Cucurbita* genotypes. The total (85) and average allele number (3.2) were similar, with no significant differences (P = 0.11) from the results obtained using gSSRs, and were also similar to those previously reported for this set of EST-SSRs, which were validated using 9 genotypes representative of the diversity within *C. pepo* (4 Zucchini, 1 Vegetable marrow, 1 Pumpkin, 1 Styrian Pumpkin, 1 Crookneck and 1 Scallop) and 1 *C. moschata* accession (Blanca et al., 2011).

The allele diversity was similar for both SSR sets, with PIC values ranging from 0.1 to 0.74 (with a mean of 0.42) in the gSSRs and from 0.1 to 0.66 (with a mean of 0.36) in the EST-SSRs, with no significant differences (P = 0.26). The correlation between the number of SSR repeats and the PIC was positive and significant both for the gSSRs (r = 0.49, P = 0.02) and EST-SSRs (r = 0.62, P < 0.001). In addition, this correlation was positive in melon, but with a considerably higher mean PIC value (with a mean of 0.58 and 0.54 for ESTs and gSSRs, respectively) (Fernandez-Silva et al., 2008; Fergany et al., 2011). The differences could be due to the fact that in the melon analysis, both subspecies are well represented, whereas here we use only a few accessions of the subspecies *ovifera* as a control.

These results confirm that both gSSRs and EST-SSRs discovered in a small germplasm sample can be transferred to different cultivar groups, and can be useful for depicting genetic relationships as well as for the identification of closely related cultivars. Also, as some of the tested SSRs are in functional regions, the study of their variability may give insight into functional variability and its possible relationship with phenotypic

variability.

Molecular variability within the cultivar groups

In general, a lower genetic variability is found within commercial hybrids compared with commercial varieties and landraces (Table IV.1-5), which is consistent with the narrower origin of the hybrids and genetic erosion due to intensive breeding (Formisano et al., 2010). On the other hand, commercial cultivars and Spanish landraces showed similar genetic diversity. When we analyse the genetic variability of the different groups, the Cocozelle Group shows the highest genetic variability among commercial cultivars. This is coherent with the more ancient characteristics found in this group, which is more similar to ancient Mexican landraces, as this group is thought to have been developed by the seventeenth century (Paris, 2000). The least variable group among the commercial cultivars was, as expected, the Zucchini. In fact, 35 alleles were found exclusively in all the other morphotypes: Pumpkin, Cocozelle and Vegetable marrow, being absent in Zucchini, both the commercial and the landraces, even in the South American control. However, no differences in genetic variability among cultivar groups were found for the Spanish landraces.

Twenty-one alleles were detected exclusively within the Spanish landraces, being absent in the commercial varieties of the Zucchini, Cocozelle and Vegetable marrow types. Spanish landraces belonging to the Zucchini group showed the highest number of exclusive alleles (27), followed by the 15 alleles exclusive to the Spanish Cocozelle landraces. The number of alleles exclusive to the commercial cultivars was considerably lower (nine in each case). Therefore, despite the fact that the Spanish Zucchini landraces did not show a remarkable morphological variability, they have a level of genetic diversity higher than that of the commercial types, so they can be considered a

reservoir of alleles useful for breeding.

<u>Relationships among varieties</u>

The genetic relationships among accessions based on SSR polymorphism were investigated by cluster analysis. No significant differences were found in the average pair-wise distances based on gSSRs and EST-SSRs. The correlation between the two distance matrices was 0.83 (P < 0.002) according to Mantel's test (Mantel, 1967), confirming that the new EST-SSR set is as effective as the previously available gSSRs in establishing genetic relationships among summer squash accessions. Similar results were previously observed in melon when the polymorphism of gSSRs and EST-SSRs was compared (Fernandez-Silva et al., 2008). Comparing the different groups, the average intragroup pair-wise distances in the Cocozelle and Vegetable marrow groups (0.27 and 0.28) were significantly higher than in the Zucchini group (0.20), with a consistently higher intragroup variability within the former. Intergroup distances were also significantly higher between the Cocozelle and the Vegetable marrow groups (0.27) than between Zucchini and these two groups (0.22).

The NJ dendrogram based on 109 gSSR and 85 EST-SSR alleles, as shown in Figure IV.1-2, fits very well with previous classifications using different markers (Esteras et al., 2012; Ferriol et al., 2003; Paris et al., 2003), with two major clusters that clearly separated accessions of both subspecies (bootstrap = 1,000). Two major patterns of genetic association were observed. Most the commercial accessions, except one group of Cocozelle, one Vegetable marrow and one Pumpkin variety, were separated from the Spanish landraces. Commercial Cocozelle cultivars were distributed in several groups with some cultivars even being similar to more ancient Pumpkin types (CO-BDT and PI171678, bootstrap = 548).

Despite the lower level of polymorphism, all Zucchini genotypes could be distinguished with the set of SSRs. Only one Zucchini landrace was molecularly similar to the group of commercial Zucchini cultivars (ZU-MU16 was very similar to ZU-GIO, bootstrap = 895), whereas the others were genetically more similar to the Vegetable marrow or Cocozelle types (ZU-A13 similar to VM-V21, bootstrap = 629; ZU-E10 similar to CO-V74, bootstrap = 690). Interestingly, the accession ZU-E27, which is phenotypically similar to commercial Zucchini cultivars, was molecularly similar to the Cocozelle and Vegetable marrow landraces. This accession has potential value for introgressing molecular variation into the commercial Zucchini group without introducing commercially unfavourable traits. Cocozelle landraces were highly variable and intermingled with Vegetable marrow landraces spread across the dendrogram.

The population structure estimated by STRUCTURE analysis (Pritchard et al., 2000) revealed that the genotypes belonging to *C. pepo* ssp. *pepo* can be separated into three populations; most of the genotypes could be clearly assigned to one of the two major populations (green or red in Figure IV.1-2), whereas only accession CO-BDT could be separated clearly from the two previous populations. The remaining five accessions could not be assigned to any of the former populations. The results are consistent with the NJ cluster analysis, showing two major populations, one of them consisting of mainly commercial cultivars and the second one of Spanish landraces. All cultivar groups are represented in both populations.

Figure IV.1-3 represents the distribution of the different accessions according to the two principal axes of variation using PCoA. The first PCoA we performed explained 31% of the variation. On the basis of the first coordinate, which accounted for 15.4% of the total variation, the accessions were clearly grouped according to subspecies. We performed a second PCoA after removing the ssp. *ovifera* accessions. The first coor-

dinate, which explains 10% of the variation, separated all the commercial cultivars of the Zucchini Group from the other commercial groups and from the Spanish landraces, supporting the differentiation of this modern group. Paris et al. (2003), along with previous studies performed with first generation molecular markers (Esteras et al., 2011), also found that the Zucchini group, the most recent of the subspecies, is the most distinct of the edible-fruited cultivar-groups of ssp. *pepo*. However, ancient European landraces are not included in most of these previous studies, and our study reveals that this distinction only occurs in commercial Zucchini, but not in Zucchini landraces that are more similar to the oldest groups. The second coordinate, which explains 8% of the variation, separated the remaining commercial cultivars, mostly Cocozelle, but also Pumpkin and Vegetable marrow from the Spanish landraces.

All three previous analyses support the genetic structure of the current summer squash sampled genotypes as being mainly due to the type of cultivar (commercial or landrace), i.e., Spanish landraces are clearly separated from commercial varieties, although within Spanish landraces there is no clear genetic structure due to cultivar type. The fact that the Spanish landraces are not closely related to the commercial cultivars of the same cultivar types seems to suggest that the gene variation displayed by the Spanish landraces has not been used extensively to develop commercial cultivars, confirming that this gene pool is a reservoir of genetic variability that has been underexploited in modern summer squash breeding.

**Conclusions**

The current report is the first study on the diversity of a wide collection of European landraces belonging to the elongated forms of *C. pepo* ssp. *pepo* using codominant markers. Our results indicate that these accessions retain traits common to ancient

cultivars of this species. These germplasm resources could be useful for the enrichment of current commercial cultivars. They still conserve phenotypic and molecular variation that has been lost during the breeding process.

The existence of landraces with favourable commercial traits, such as ZU-E27 or CO-V185, while still being molecularly variable, is particularly noteworthy, as it could provide new alleles useful for breeding without altering certain necessary commercial characteristics.

This new found genetic potential in Spanish landraces not present in the restricted gene pool of modern varieties could be used to adapt market types to regional preferences of colour, secondary design or shapes, and could also provide other interesting traits, such as flavour, vitamin and mineral contents, which have long been associated with colour variation in summer squash. Other traits that are current objectives of summer squash breeding, such as fruit glossiness, parthenocarpy and disease response could also be variable in this collection.

**Author´s contributiuons**

CE participated in the phenotyping and genotyping process and assisted in the Cluster and PCoA analysis. CE also contributed to manuscript drafting.

**Table IV.1-1. Spanish summer squash (*C. pepo* ssp. *pepo*) landraces, other landraces, commercial cultivars and hybrids used in this study.**

| Code[a] | Name | Morphotype | source | Resistances[b] |
|---|---|---|---|---|
| *C.pepo* ssp. *pepo* | | | | |
| *Commercial hybrids* | | | | |
| ZU-GIO | Giove F1 | Zucchini | Petoseed | WMV, ZYMV |
| ZU-ZU1805 | ZU 1805 F1 | Zucchini | Petoseed | PM, CMV, WMV, ZYMV |
| ZU-MIK | Mikonos F1 | Zucchini | Syngenta | PM, CMV, WMV, ZYMV |
| ZU-QUI | Quine F1 | Zucchini | Syngenta | PM, CMV, WMV, ZYMV |
| VM-TON | Tonya F1 | V.Marrow | Syngenta | ZYMV |
| CO-GS2386 | GS2386 F1 | Cocozelle | Syngenta | PM, ZYMV |

| Code | Name | Morphotype | source |
|---|---|---|---|
| *Commercial varieties* | | | |
| ZU-NVM | Nano Verde di Milano | Zucchini | La Semiorto Sementi |
| ZU-TRF | True French | Zucchini | Thompson & Morgan |
| ZU-BB | Black Beauty | Zucchini | Semillas Battle |
| CO-ROM | Romanesco | Cocozelle | Semitalia |
| CO-LUF | Lungo Fiorentino | Cocozelle | La Semiorto Sementi |
| CO-ODF | Ortolana di Faenza | Cocozelle | La Semiorto Sementi |
| CO-LBS | Lungo Bianco | Cocozelle | La Semiorto Sementi |
| CO-BDT | Bianca di Trieste | Cocozelle | La Semiorto Sementi |
| CO-SPQ | San Pasquale | Cocozelle | La Semiorto Sementi |
| CO-VAL | Alberello Sel. Valery | Cocozelle | La Semiorto Sementi |
| PU-TON | Tondo chiaro di Nizza | Pumpkin | La Semiorto Sementi |
| PU-TOP | Tondo di Piacenza | Pumpkin | La Semiorto Sementi |

| Code | Name | Morphotype | Origin |
|---|---|---|---|
| *Spanish landraces* | | | |
| ZU-A13 | Calabacín Blanco | Zucchini | Huesca |
| ZU-E10 | Calabacín | Zucchini | Cáceres |

250

| Code | Name | Morphotype | Origin |
|---|---|---|---|
| ZU-E27 | Calabacín | Zucchini | Cáceres |
| ZU-MU16 | Calabacín | Zucchini | Murcia |
| ZU-MU20 | Calabacín verde | Zucchini | Murcia |
| VM-A12 | Calabacín oscuro | V.Marrow | Huesca |
| VM-AN23 | Calabacín de mesa | V.Marrow | Málaga |
| VM-AN27 | Calabaza | V.Marrow | Cádiz |
| VM-AN113 | Calabacín de freir | V.Marrow | Almeria |
| VM-CL19 | Calabacín | V.Marrow | Segovia |
| VM-CL21 | Calabacín | V.Marrow | Valladolid |
| VM-CM32 | Calabacín | V.Marrow | Cuenca |
| VM-CM47 | Calabacín | V.Marrow | Ciudad Real |
| VM-CM949 | Calabacín blanco | V.Marrow | Guadalajara |
| VM-V10 | Calabacín | V.Marrow | Valencia |
| VM-V21 | Calabacín | V.Marrow | Valencia |
| CO-A2 | Calabacín francés | Cocozelle | Teruel |
| CO-AN75 | Rastrero blanco | Cocozelle | Córdoba |
| CO-C9 | Calabazón | Cocozelle | Barcelona |
| CO-V74 | Calabaza de freir | Cocozelle | Valencia |
| CO-V116 | Calabaza parda | Cocozelle | Castellón |
| CO-V185 | Calabacín | Cocozelle | Alicante |
| CO-VPAS | Calabacín | Cocozelle | Castellón |

*Other landraces*

| | | | |
|---|---|---|---|
| ZU-ECU227 | Calabacín Zaguin | Zucchini | Ecuador |
| VM-AFR12 | Calabacín | V.Marrow | Morroco |
| PU-PI 171678[c] | pumpkin | Pumpkin | Turkey |
| PU-Styrian | Oil-seed pumpkin | Pumpkin | Austria |

| Code | | Morphotype | Origin |
|---|---|---|---|

*C. pepo* **ssp.** *ovifera*

| | | | |
|---|---|---|---|
| CR-NSL5227[c] | | Crookneck | USA |

| | | |
|---|---|---|
| AC-PI615111[c] | Acorn | USA |
| SC-UPV196 | Scallop | Valencia |

*C. pepo* ssp. *ovifera* included as controls.

[a] The accession code indicates the morphotype/group followed by the cultivar name (commercial cultivars) or COMAV's Genebank code (Spanish landraces)

[b] Resistance to: PM powdery mildew, ZYMV zucchini yellow mosaic virus, CMV cucumber mosaic virus, WMV watermelon mosaic virus, PRSV papaya ringspot virus

[c] These accessions were kindly provided by the NPGS of the USDA

**Table IV.1-2. Morphological characterization of the summer squash accessions.**

| Code | Growth habit[a] | Branching[a] | Spines[a] | DMF[a] | NMF[a] | DFF[a] | NFF[a] | NºMF[a] | NºFF[a] | NºFF/NºMF[a] |
|---|---|---|---|---|---|---|---|---|---|---|
| *Commercial hybrids* | | | | | | | | | | |
| ZU-GIO | B | 0 | 0 | 30.7 | 2 | 41.6 | 7 | 12 | 7.3 | 0.6 |
| ZU-ZU1805 | B | 0 | 0 | 32.7 | 2.3 | 45.7 | 10 | 17.7 | 4.7 | 0.3 |
| ZU-MIK | B | 0 | 0 | 35.6 | 1.7 | 42 | 8.7 | 11.3 | 5.3 | 0.5 |
| ZU-QUI | B | 0 | 0 | 30.7 | 1.7 | 43.7 | 7.3 | 18 | 6.7 | 0.4 |
| VM-TON | B | 0.5 | 0.5 | 34.3 | 1 | 52.6 | 8.7 | 23 | 7.7 | 0.3 |
| CO-GS2386 | B | 0 | 0 | 34.3 | 1.7 | 46 | 11 | 16.7 | 7.3 | 0.4 |
| *Commercial varieties* | | | | | | | | | | |
| ZU-NVM | B | 0 | 0 | 34 | 3 | 52.6 | 12 | 20.3 | 5.3 | 0.3 |
| ZU-TRF | B | 0 | 0 | 30,6 | 4,3 | 48,3 | 11 | 18.7 | 8.7 | 0.5 |
| ZU-BB | B | 0 | 0 | 33 | 3 | 45,2 | 9 | 22.2 | 8,1 | 0.4 |
| CO-ROM | I | 0 | 1 | 32 | 2 | 43.3 | 10.3 | 14 | 8.7 | 0.5 |
| CO-LUF | B | 0 | 1 | 39,3 | 2.6 | 47,3 | 13 | 14.6 | 5.3 | 0.4 |
| CO-ODF | B | 1 | 0 | 37.7 | 3 | 56.5 | 11 | 31 | 3.5 | 0.1 |
| CO-LBS | I | 0.5 | 0.5 | 44 | 1 | 53.3 | 12.7 | 29 | 5 | 0.2 |
| CO-BDT | I | 0 | 0.5 | 44.3 | 2.3 | 49.3 | 14 | 65 | 9.3 | 0.1 |
| CO-SPQ | B | 0.5 | 0.5 | 44 | 3.7 | 61.3 | 11.7 | 37.7 | 7.7 | 0.2 |
| CO-VAL | V | 1 | 0 | 45.3 | 1.7 | 53.3 | 16 | 48.3 | 4.3 | 0.1 |
| PU-TON | I | 0.5 | 1 | 32 | 1.3 | 48.3 | 12.7 | 25 | 6 | 0.2 |
| PU-TOP | B | 0 | 0.5 | 33.7 | 1 | 55.3 | 9.7 | 38 | 8 | 0.2 |
| *Spanish landraces* | | | | | | | | | | |
| ZU-A13 | I | 0.5 | 0 | 32 | 2.5 | 52 | 16.5 | 40 | 4 | 0.1 |
| ZU-E10 | B | 1 | 0.5 | 49 | 2 | 68 | 12 | 32 | 2 | 0.1 |
| ZU-E27 | B | 0 | 0 | 35.5 | 1 | 45.5 | 8.5 | 10.5 | 6 | 0.6 |
| ZU-MU16 | B | 0 | 0 | 32.5 | 1.5 | 41 | 7 | 11 | 7 | 0.6 |
| ZU-MU20 | I | 1 | 0.5 | 37 | 2 | 57 | 15 | 55 | 4 | 0.1 |
| VM-A12 | I | 1 | 0.5 | 39 | 3 | 56.5 | 15 | 60 | 10 | 0.2 |

| Code | Growth habit[a] | Branching[a] | Spines[a] | DMF[a] | NMF[a] | DFF[a] | NFF[a] | NºMF[a] | NºFF[a] | NºFF/MF[a] |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *Spanish landraces* | | | | | |
| VM-AN23 | V | 1 | 1 | 62.5 | 11.5 | 65 | 24 | 60 | 5.5 | 0.1 |
| VM-AN27 | I | 1 | 0.5 | 47 | 1 | 58.5 | 12 | 37.5 | 2.5 | 0.1 |
| VM-AN113 | V | 1 | 1 | 39.5 | 1.5 | 50.5 | 13 | 45 | 5 | 0.1 |
| VM-CL19 | V | 1 | 0.5 | 33 | 3 | 50 | 21.5 | 60 | 3.5 | 0.1 |
| VM-CM32 | V | 1 | 1 | 39.5 | 2 | 54 | 15.5 | 45 | 4.5 | 0.1 |
| VM-CM47 | I | 1 | 1 | 32.5 | 2 | 45 | 15 | 25 | 4 | 0.2 |
| VM-CM949 | V | 1 | 1 | 47.5 | 3 | 61 | 18.5 | 60 | 5.5 | 0.1 |
| VM-V10 | I | 1 | 1 | 48.5 | 2 | 56.5 | 18.5 | 60 | 3.5 | 0.1 |
| VM-V21 | I | 1 | 1 | 46 | 2 | 58 | 17 | 55 | 4 | 0.1 |
| CO-A2 | I | 1 | 1 | 47 | 4 | 62.5 | 16 | 72.5 | 2.5 | 0.05 |
| CO-AN75 | V | 1 | 1 | 43 | 1.5 | 58 | 19.5 | 60 | 7 | 0.1 |
| CO-C9 | I | 1 | 0.5 | 34 | 2 | 54 | 22.5 | 60 | 2.5 | 0.05 |
| CO-CL21 | I | 1 | 0.5 | 41 | 6 | 77 | 21 | 60 | 7 | 0.1 |
| CO-V74 | V | 1 | 1 | 42 | 3.5 | 53 | 18 | 60 | 4.5 | 0.1 |
| CO-V116 | V | 1 | 1 | 40.5 | 3.5 | 57 | 24 | 60 | 5.5 | 0.1 |
| CO-V185 | B | 0 | 1 | 30 | 1 | 58 | 12 | 14 | 6 | 0.05 |
| CO-VPAS | I | 1 | 1 | 50.5 | 3.5 | 58 | 13 | 38.5 | 6 | 0.2 |
| | | | | | *Other* | | | | | |
| ZU-ECU227 | B | 1 | 0.5 | 49 | 2 | 60 | 6 | 36 | 8 | 0.2 |
| VM-AFR12 | I | 1 | 1 | 34.5 | 1 | 2 | 11.5 | 60 | 4.5 | 0.1 |

[a] Growth habit: B bushy, V viny, I intermediate; lateral branch development: 0 = no branching, 0.5 = moderately branched,1 = highly branched; presence of spines: 0 = spineless, 0.5 = sparse and small spicules, 1 = dense, large and sharp spicules; DMF, DFF: days to male and female flowering from transplant; NMF, NFF: node in which the first male/female flower appears; NºMF and NºFF: number of male/female flowers 7 days after the opening of the first female flower; NºFF/ NºMF: femaleness ratio.

**Table IV.1-3. Characteristics of the genomic SSRs selected for genotyping the summer squash collection.**

| SSR *locus*[a] | Linkage group[a] | Allele number reported[a] /observed | Motif[a] | No. Rep[a] | Expected size[a] | Observed size | PIC |
|---|---|---|---|---|---|---|---|
| CMTp193 | 1 | 5/2 | ga | 18 | 186 | 177-200 | 0.37 |
| CMTp98 | 2 | 3/6 | aag | 9 | 213 | 224-255 | 0.55 |
| CMTp131 | 3a | 3/3 | ccg | 7 | 117 | 109-135 | 0.38 |
| CMTp187 | 3b | 3/2 | cag+caa | 6+4 | 189 | 200-300 | 0.37 |
| CMTp63 | 4 | 3/4 | ttc | 10 | 152 | 145-175 | 0.49 |
| CMTp88 | 5 | 5/3 | tc | 12 | 167 | 178-197 | 0.55 |
| CMTp235 | 5 | 6/9 | gtt | 12 | 148 | 145-200 | 0.74 |
| CMTp256 | 6 | 2/3 | atc | 5 | 154 | 175-200 | 0.10 |
| CMTp224 | 6 | 5/3 | caa | 7 | 151 | 169-175 | 0.41 |
| CMTp248 | 7 | 4/1 | gga | 5 | 154 | 145-200 | 0 |
| CMTp142 | 8 | 7/6 | tc | 12 | 158 | 145-240 | 0.64 |
| CMTp257 | 9 | 3/5 | cgt | 11 | 138 | 100-200 | 0.60 |
| CMTp58 | 9 | 2/1 | ga+t | 6+10 | 102 | 100 | 0 |
| CMTp145 | 10a | 4/4 | cat | 7 | 100 | 117-128 | 0.42 |
| CMTp66 | 10a | 3/4 | gaa | 9 | 128 | 120-175 | 0.10 |
| CMTp260 | 11 | 4/4 | cat | 7 | 155 | 140-174 | 0.63 |
| CMTp245 | 11 | 5/3 | gcg | 9 | 134 | 100-200 | 0.43 |
| CMTp36 | 12 | 3/4 | aac | 5 | 151 | 164-185 | 0.47 |
| CMTp69 | 13 | 3/4 | tatt | 4 | 70 | 77-92 | 0.19 |
| CMTp176 | 14 | 4/4 | tc | 12 | 111 | 100-145 | 0.24 |
| CMTp33 | 14 | 4/3 | gaa | 7 | 171 | 189-210 | 0.27 |
| CMTp86 | 15 | 3/3 | cca | 9 | 133 | 151-192 | 0.23 |
| CMTp169 | 15 | 2/4 | gaa | 11 | 158 | 175-200 | 0.42 |
| CMTp231 | 16 | 9/- | ag | 38 | 170 | -- | - |
| CMTp208 | 17 | 2/4 | gtt | 5 | 117 | 133-136 | 0.25 |

| SSR *locus*[a] | Linkage group[a] | Allele number reported[a] /observed | Motif[a] | No. Rep[a] | Expected size[a] | Observed size | PIC |
|---|---|---|---|---|---|---|---|
| CMTp209 | 17 | 3/3 | gtt | 6 | 116 | 100-145 | 0.10 |
| CMTp183 | 18 | 3/5 | cat | 6 | 196 | 216-303 | 0.56 |
| CMTp188 | 18 | 2/4 | cat | 7 | 147 | 145-350 | 0.69 |
| CMTp132 | 19 | 4/5 | gat | 12 | 151 | 145-200 | 0.46 |
| CMTp47 | 20 | 3/3 | ag | 9 | 154 | 165-173 | 0.44 |

[a] *Locus* name, linkage group, motif, number of repetitions and allele number reported, according to Gong et al. (2008).

**Table IV.1-4. Characteristics of the EST-SSRs selected for genotyping the summer squash collection.**

| SSR *locus*[a] | Position within EST[a] | Allele number reported[a]/observed | Motif[a] | No. rep[a] | Expected size[a] | Observed size | PIC | Arabidopsis Ortholog[b] | Melon Ortholog[b] | Gene description[b] | Functional categories[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CUTC001906 | UTR | 2/3 | acgg | 7 | 198 | 200-255 | 0.45 | AT2G36320 | - | zinc finger (AN1-like) family protein | response to stress |
| CUTC002749 | ORF | 4/4 | tgc | 10 | 183 | 175-200 | 0.53 | AT1G60030 | MU21505 | nucleobase-ascorbate transporter 7 (NAT7) | transport |
| CUTC004158 | ORF | 2/2 | aag | 7 | 188 | 200-204 | 0.35 | AT4G19100 | MU19453 | unknown | - |
| CUTC004307 | ORF | 4/5 | aag | 8 | 100 | 120-145 | 0.54 | AT5G20190 | MU27127 | tetratricopeptide repeat (TPR)-like superfamily protein | binding function |
| CUTC004399 | ORF | 3/3 | acg | 7 | 165 | 175-200 | 0.19 | AT1G12830 | MU38445 | unknown protein | - |
| CUTC004782 | UTR | 3/2 | agg | 7 | 164 | 175-200 | 0.30 | - | - | unknown protein | - |
| CUTC004991 | UTR | 3/4 | agc | 7 | 173 | 175-200 | 0.32 | - | - | hydratase,cytoplasmatic, (cucurbita maxima) | metabolism |
| CUTC005739 | ORF | 2/2 | aag | 8 | 194 | 200-204 | 0.32 | AT5G11270 | MU24155 | overexpressor of cationic peroxidase 3. Transcription factor | transcription regulator, defense response |
| CUTC005800 | ORF | 5/5 | aag | 11 | 189 | 200-240 | 0.66 | AT2G47460 | - | domain protein 12 MYB12 | transcription regulator |
| CUTC006209 | ORF | 3/3 | aac | 9 | 200 | 200-300 | 0.42 | AT3G04930 | MU21252 | transcription regulator | transcription regulator |
| CUTC006703 | ORF | 2/2 | acc | 7 | 123 | 145,00 | 0.27 | AT2G42260, | MU23778 | UV-B-insensitive 4 UVI4 | Cell cycle |

| SSR locus[a] | Position within EST[a] | Allele number reported[a]/observed | Motif[a] | No. rep[a] | Expected size[a] | Observed size | PIC | Arabidopsis Ortholog[b] | Melon Ortholog[b] | Gene description[b] | Functional categories[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CUTC006891 | ORF | 3/3 | ag | 12 | 154 | 145-175 | 0.26 | AT2G44620 | - | mitochondrial acyl carrier protein 1 | transport |
| CUTC007942 | ORF | 3/5 | aac | 8 | 152 | 175-200 | 0.16 | - | MU37594 | BEL1-LIKE homeodomain 1, BLH1 | transcription regulator |
| CUTC008357 | ORF | 3/3 | acc | 8 | 136 | 156,00 | 0.41 | AT1G23860 | - | RS containing zinc finger protein 21 | DNA splicing |
| CUTC008409 | ORF | 3/3 | tgc | 8 | 107 | 120-145 | 0.43 | AT3G12920 | MU22243 | protein binding | defense response |
| CUTC008659 | ORF | 1/1 | aag | 7 | 221 | 241,00 | 0.00 | AT1G64770, | MU34026 | NDH-dependent cyclic electron flow 1, NDF2 | catabolism |
| CUTC046645 | ORF | 4/4 | atc | 9 | 194 | 200-240 | 0.28 | - | MU26160 | Acyl-CoA binding protein | binding function |
| CUTC009316 | ORF | 3/2 | gca | 7 | 300 | 320,00 | 0.15 | - | - | RNA binding | binding function |
| CUTC009607 | UTR | 4/4 | ag | 11 | 146 | 145-200 | 0.52 | AT3G13510 | MU35247 | unknown protein | - |
| CUTC009760 | ORF | 3/2 | aac | 7 | 159 | 179,00 | 0.26 | - | - | EIN3- like protein (Cucumis melo) | transcription regulator |
| CUTC011336* | ORF | 5/5 | agg | 9 | 149 | 145-175 | 0.53 | AT3G15070 | - | zinc finger (C3HC4 Type RING FINGER) family protein | binding function |
| CUTC012342 | ORF | 3/2 | aag | 7 | 182 | 200-204 | 0.18 | AT4G34630 | - | unknown protein | - |
| CUTC017708 | ORF | 4/4 | agc | 8 | 223 | 250-300 | 0.47 | - | - | trasporter (Arabidopsis thaliana) NAT6 | transport |
| CUTC018879 | ORF | 2/2 | aag | 7 | 164 | 175-200 | 0.37 | - | - | unknown protein | - |
| CUTC020992 | ORF | 4/3 | aag | 10 | 156 | 175,00 | 0.42 | - | MU23978 | Similar to factor bHLH147 | transcription regulator |

| SSR *locus*[a] | Position within EST[a] | Allele number reported[a]/observed | Motif[a] | No. rep[a] | Expected size[a] | Observed size | PIC | Arabidopsis Ortholog [b] | Melon Ortholog[b] | Gene description[b] | Functional categories[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CUTC022867 | ORF | 2/5 | agc | 8 | 140 | 145-175 | 0.48 | - | - | Similar to AT1G73230.1,  nascent polypeptide-associated complex (NAC) domain-containing protein | response to salt stress |
| CUTC023363 | ORF | 4/2 | agc | 7 | 244 | 264,00 | 0.02 | - | - | similar to predicted protein (Populus trichocarpa) | - |

[a] *Locus* name, location in the EST (ORF Open Reading Frame, UTR Untranslated Region), motif, number of repetitions, expected size and

allele number reported, according to Blanca et al. (2011).

[b] Annotation results reported in Blanca et al. (2011) are also included, *Arabidopsis* and melon orthologues were identified by searching in

*Arabidopsis* (http://www.arabidopsis.org) and ICuGI, International Cucurbit Genomics Initiative melon, (http://www.icugi.orgdatabase).

**Table IV.1-5. Average Polymorphism Information Content (PIC) among cultivar groups relative to the SSR origin (EST-SSR or gSSR) and taking both types together (All SSRs).**

| Variety origin | EST-SSRs | g-SSRs | All SSRs |
|---|---|---|---|
| *Cocozelle* | | | |
| Commercial | 0.32 | 0.37 | 0.36 |
| Spanish landraces | 0.29 | 0.33 | 0.31 |
| *Vegetable marrow* | | | |
| Commercial | 0.19 | 0.32 | 0.26 |
| Spanish landraces | 0.24 | 0.33 | 0.31 |
| *Zucchini* | | | |
| Commercial | 0.17 | 0.27 | 0.23 |
| Spanish landraces | 0.24 | 0.31 | 0.31 |
| *All morphotypes* | | | |
| Commercial varieties | 0.32 | 0.34 | 0.34 |
| Commercial hybrids | 0.22 | 0.32 | 0.28 |
| Spanish landraces | 0.29 | 0.34 | 0.33 |
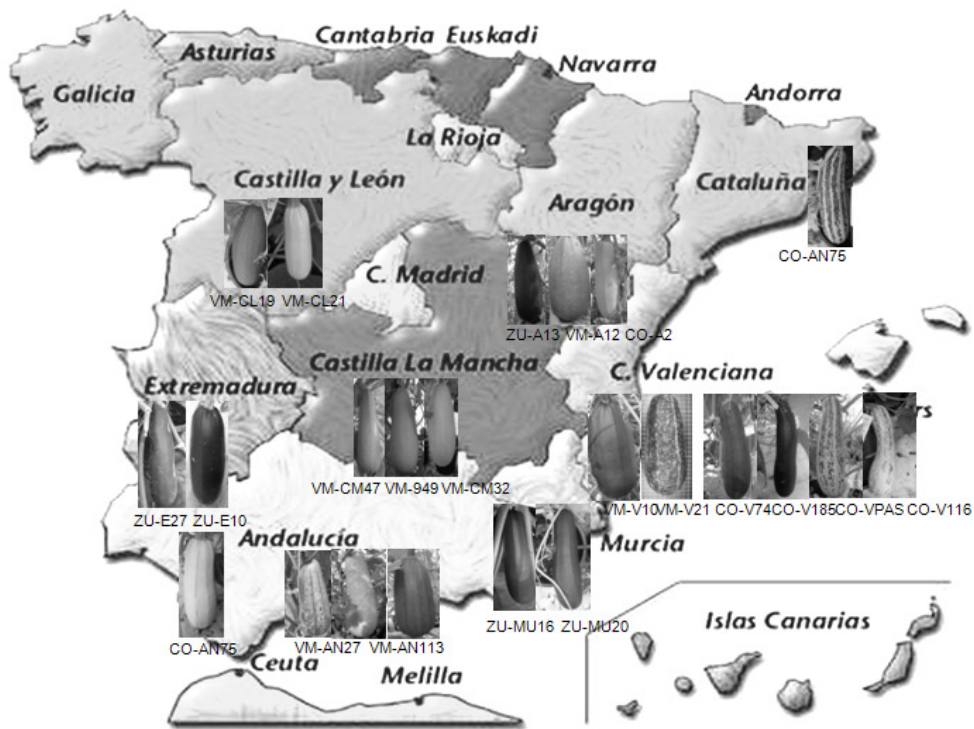
**Figure IV.1-1. Phenotypic variation of mature fruits of the Spanish summer squash landraces used in this study.**

Accessions are located on the map according with their origin.
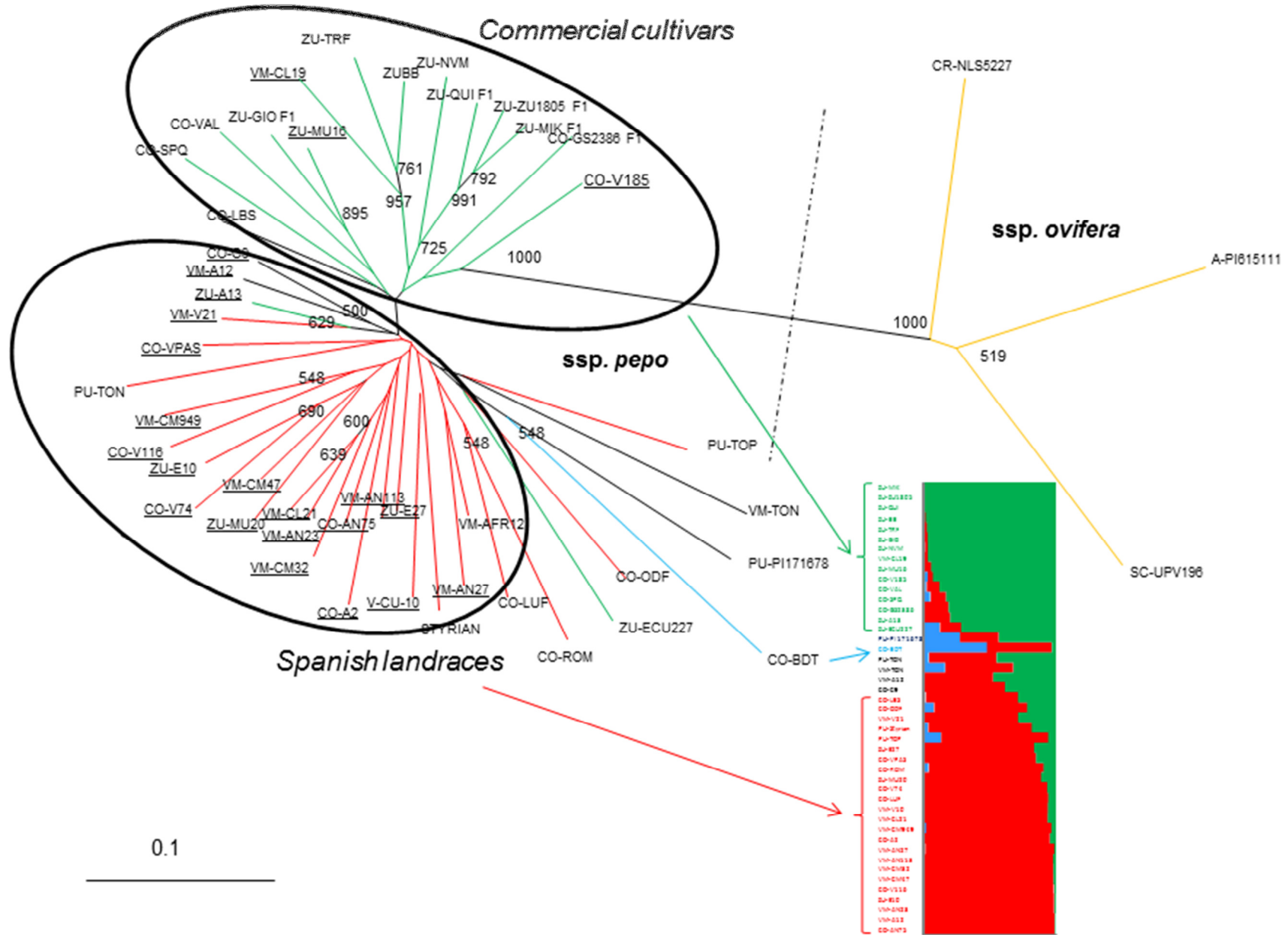
**Figure IV.1-2. NJ tree showing relationships among the 48 accession of summer squash using genomic SSR and EST–SSR markers based on DICE distance.** Bootstrap values over 500 are indicated and are based on 1,000 re-samplings of the data set. Spanish landraces are underlined. Cluster analysis clearly separates the control accessions of *C. pepo* ssp. *ovifera* (indicated with dashed lines). The population structure estimated by STRUCTURE analysis is depicted in the right lower quadrant, revealing two major populations (green or red); only accession CO-BDT could be separated clearly for the two previous populations. The populations inferred by STRUCTURE are connected with the corresponding accessions in the NJ tree by arrows.
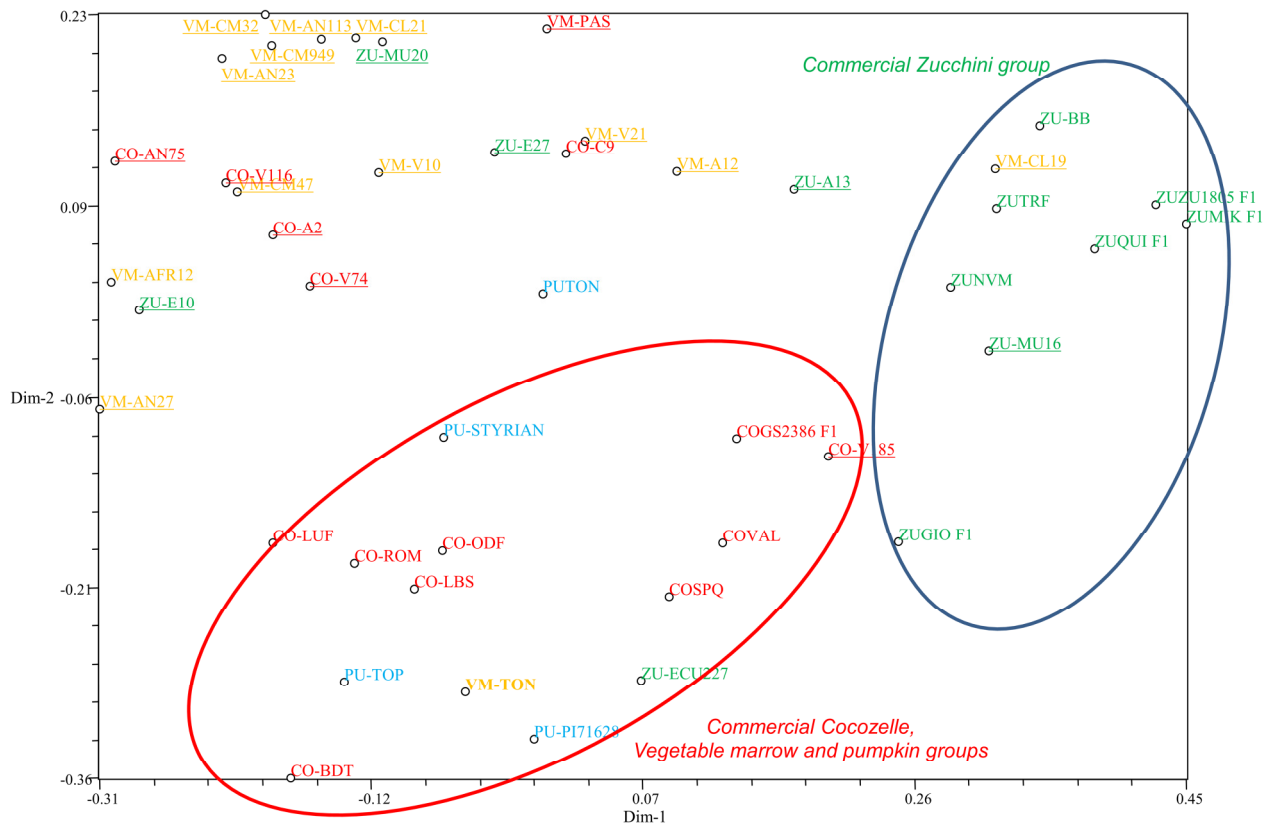
**Figure IV.1-3. Diagram showing the relationships among the 48 accessions of summer squash based on principal coordinate analysis using both genomic and EST-SSRs.**

A two-dimensional scatter plot using the first and second principal coordinates is shown. Spanish landraces are underlined. Cocozelle, Vegetable marrow, Zucchini and Pumpkin groups are indicated by letters CO, VM, ZU and PU respectively.

**References**

Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Picó B: **Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (*Cucurbitaceae*).** BMC Genomics 2011, 10(12):104.

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari MA, Collin F, Flowers JM, Pitrat M, Puru-gganan MD, Dogimont C, Bendahmane A: **A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons.** Science 2008, 321:836–838.

Decker DS: **Origin(s), evolution, and systematics of *Cucurbita pepo* (*Cucurbitaceae*).** Econ Bot 1988, 42:4–15.

Doyle JJ, Doyle JL: **Isolation of plant DNA from fresh tissue.** Focus 1990, 12:13–15.

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools**. In Genetics, Genomics and Breeding of Cucurbits. Edited by Behera TK, Wang Y, Kole C. New Hampshire: Science Publishers Inc, Enfield; 2012:140-198.

Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** Mol Ecol 2005, 14:2611–2620.

Ezura H, Fukino N: **Research tools for functional genomics in melon (*Cucumis melo* L.): current status and prospects.** Plant Biotechnol 2009, 26:359–368.

Falush D, Stephens M, Pritchard JK: **Inference of population structure using mulitlocus genotype data: linked *loci* and correlated allele frequencies.** Genetics 2003, 164:1567–1587.

Falush D, Stephens M, Pritchard JK: **Inference of population structure using mulitlocus genotype data: dominant markers and null alleles**. Mol Ecol Notes 2007, 7:574–578.

Fergany M, Balvir Kaur, Monforte AJ, Pitrat M, Rys C, Lecoq H, Dhillon NPS, Dhaliwal SS: **Variation in melon (*Cucumis melo*) landraces adapted to the humid tropics of southern**

**India.** Genet Resour Crop Evol 2011, 58:225–243.

Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Pico B, Nuez F, Arus P, Garcia-Mas J, Monforte AJ: **Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.).** Theor Appl Genet 2008, 118:139–150.

Ferriol M, Picó B: **Pumpkin and Winter Squash**. In Handbook of plant breeding, vegetables I. Edited by Prohens J, Nuez F. Springer, New York; 2008:317–349.

Ferriol M, Picó B, Nuez F: **Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers**. Theor Appl Genet 2003, 107:271–282.

Formisano G, Paris HS, Frusciante L, Ercolano MR: **Commercial *Cucurbita pepo* squash hybrids carrying disease resistance introgressed from *Cucurbita moschata* have high genetic similarity.** Plant Genet Res 2010, 8:198–203.

Gong L, Stift G, Kofler R, Pachner M, Lelley T: **Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L.** Theor Appl Genet 2008, 117:37–48.

Gonzalez VM, Rodríguez-Moreno L, Centeno E, Benjak A, Garcia-Mas J, Puigdomenech P, Aranda MA: **Genome-wide BAC-end sequencing of *Cucumis melo* using two BAC libraries**. BMC Genomics 2010, 11:618.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R,

Zhang B, Jiang S, Wang J, Du Y, Li S: **The genome of the cucumber, *Cucumis sativus* L.** Nat Genet 2009, 41:1275–1281.

Hubisz MJ, Falush D, Stephens M, Pritcard JK: **Inferring weak population structure with the assistance of sample group information.** Mol Ecol Resour 2009, 9:1322–1332.

Lebeda A, Widrlechner MP, Staub J, Ezura H, Zalapa J, Kristkova E: **Cucurbits (*Cucurbitaceae*; *Cucumis* spp., *Cucurbita* spp., *Citrullus* spp.).** In Genetic resources, chromosome engineering,  and  crop improvement, vol 3. Edited by Singh  RJ. CRC Press, Boca Raton; 2007:271–376.

Li Z, Huang S, Liu S, Pan J, Zhang Z, Tao Q, Shi Q, Jia Z, Zhang W, Chen H, Si L, Zhu L, Cai R: **Molecular isolation of the M gene suggests that a conserved-residue conversion induces the formation of bisexual flowers in cucumber plants.** Genetics 2009, 182:1381–1385.

Lira R, Montes S: **Cucurbits (*Cucurbita* spp.).** In Neglected crops, 1492 from a different perspective. Edited by Hernandez JE, Leon J. F.A.O, Rome; 1994:63–77.

Liu K, Muse SV: **PowerMarker: integrated analysis environment for genetic marker analysis.** Bioinformatics 2005, 21:2128–2129.

Mantel N: **The detection of disease clustering and a generalized regression approach**. Cancer Res 1967, 27:209–220.

Metzker ML: **Applications of next-generation sequencing. Sequencing technologies the next generation.** Nat Rev Genet 2010, 11:31–46.

Nei M, Li WH: **Mathematical model for studying genetic variation in terms of restriction endonucleases**. Proc Natl Acad Sci USA 1979, 76:5269–5273.

Nesom GL: **Toward consistency of taxonomic rank in wild/domesticated *Cucurbitaceae***. Phytoneuron 2011, 13:1–33.

Paris HS: **A proposed subspecific classification for *Cucurbita pepo***. Phytologia 1986, 61:133–138.

Paris HS: **History of the cultivar-groups of *Cucurbita pepo***. In Hort revs, vol 25. Edited by Janick J. Wiley, New York; 2000:71–170.

Paris HS: **Summer squash**. In Handbook of plant breeding, vegetables I. Edited by Prohens J, Nuez F. Springer, New York; 2008:351–379.

Paris HS, Brown RN: **The genes of pumpkin and squash.** HortScience 2005, 40:1620–1630.

Paris HS, Yonash N, Portnoy V, Mozes-Daube N, Tzuri G, Katzir N: **Assessment of genetic relationships in 4 *Cucurbita pepo* (*Cucurbitaceae*) using DNA markers.** Theor Appl Genet 2003, 106:971–978.

Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** Genetics 2000, 155:945–959.

Rohlf JF: **NTSYS: numerical taxonomy and multivariate analysis system, version 2.02.** Exeter Software, Setauket, 1998.

Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L: **Phylogenetic relationships among domesticated and wild species of *Cucurbita* (*Cucurbitaceae*) inferred from a mitochondrial gene: implications for crop plant evolution and areas of origin.** Proc Natl Acad Sci USA 2002, 99:535–540.

Smith BD: **The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago.** Science 1997, 276:932–934.

## IV-2. High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait *loci* mapping

**Abstract**

Background: *Cucurbita pepo* is a member of the *Cucurbitaceae* family, the second- most important horticultural family in terms of economic importance after *Solanaceae*. The "summer squash" types, including Zucchini and Scallop, rank among the highest-valued vegetables worldwide. There are few genomic tools available for this species.

The first *Cucurbita* transcriptome, along with a large collection of Single Nucleotide Polymorphisms (SNPs), was recently generated using massive sequencing. A set of 384 SNPs was selected to generate an Illumina GoldenGate assay in order to construct the first SNP-based genetic map of *Cucurbita* and map Quantitative Trait *Loci* (QTL).

Results: We herein present the construction of the first SNP-based genetic map of *Cucurbita pepo* using a population derived from the cross of two varieties with contrasting phenotypes, representing the main cultivar groups of the species' two subspecies: Zucchini (ssp. *pepo*) x Scallop (ssp. *ovifera*). The mapping population was genotyped with 384 SNPs, a set of selected EST-SNPs identified *in silico* after massive sequencing of the transcriptomes of both parents, using the Illumina GoldenGate platform. The global success rate of the assay was higher than 85%. In total, 304 SNPs were mapped, along with 11 SSRs from a previous map, giving a map density of 5.56 cM/marker. This map was used to infer syntenic relationships between *C. pepo* and cucumber and to successfully map QTLs that control plant, flowering and fruit traits that are of benefit to squash breeding. The QTL effects were validated in backcross populations.

Conclusion: Our results show that massive sequencing in different genotypes is an excellent tool for SNP discovery, and that the Illumina GoldenGate platform can be successfully applied to constructing genetic maps and performing QTL analysis in *Cucurbita.* This is the first SNP-based genetic map in the *Cucurbita* genus and is an invaluable new tool for

biological research, especially considering that most of these markers are located in the coding regions of genes involved in different physiological processes. The platform will also be useful for future mapping and diversity studies, and will be essential in order to accelerate the process of breeding new and better-adapted squash varieties.

**Background**

The *Cucurbita* genus, of American origin, is one of the most variable genera within the *Cucurbitaceae* family (reviewed by Esteras et al., 2012). *C. pepo* L. (2n=40), the most economically important crop of this genus (Paris, 2008), displays eight commercial morphotypes grouped into two subspecies (ssp. *pepo* L.: Pumpkin, Vegetable Marrow, Cocozelle and Zucchini; ssp. *ovifera* (L.) Decker (syn ssp. *texana* (Scheele) Filov): Scallop, Acorn, Crookneck and Straightneck). The main economic value of the species resides in the consumption of its immature fruits as vegetables, commonly known as summer squashes. Summer squashes of the Zucchini type rank among the highest-valued vegetables worldwide, whereas the "winter squash" types (fruits consumed when mature) of *C. pepo* and related *Cucurbita* spp. are food staples and rich sources of fat and vitamins in developing countries (Ferriol and Picó, 2008).

Despite its economic importance, there are few genomic tools available for *Cucurbita,* unlike other cucurbits, such as watermelon (*Citrullus lanatus* (Thunb.) Matsum & Nakai), cucumbers (*Cucumis sativus* L.) and melons (*Cucumis melo* L.), for which new mapping populations, dense genetic maps (Ren et al., 2009; Deleu et al., 2009; Diaz et al., 2011; Levi et al., 2011), microarrays (Wechter et al., 2008; Mascarell-Creus et al., 2009), reverse genetics platforms (Dahmani-Mardas et al., 2010; González et al., 2011), transcriptomes (Guo et al., 2010; Blanca et al., 2011b) and even whole genome sequences have already been generated

(Huang et al., 2009; González et al., 2010; Fei et al., 2011). Many of these resources are available at the database maintained by the International Cucurbit Genomics Initiative (ICuGI, http://www.icugi.org) and are being successfully employed by cucurbit researchers to study gene functions and their related polymorphisms.

High-throughput sequencing technologies, mainly Roche 454 and Illumina GA (Metzker, 2010), are contributing to filling this gap for non-model crops, thereby allowing the rapid generation of sequence information, even in species about which there is little prior knowledge. One of the most interesting applications of massive sequencing is the large-scale discovery of genetic variants that can be converted into genetic markers, mainly microsatellites or Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphisms (SNPs) (Deschamps and Campbell, 2010). SSRs and SNPs are now the predominant markers in plant genetic analysis. The first transcriptome of *C. pepo* was recently generated using 454 GS FLX Titanium technology. A total of 49,610 unigenes were assembled from 512,751 new ESTs (Expressed Sequence Tags) and used to generate the first large collection of EST-derived SSRs and SNPs in this species (Blanca et al., 2011a). SNPs are abundant in the genomes, and are stable, amenable to automation and increasingly cost-effective, and are therefore fast becoming the marker system of choice in modern genomics research. SSRs, however, continue to be widely used in studies with no need for automation due to their co-dominant and multiallelic nature.

A practical way of optimizing large SNP collections is that of using them with cost-effective platforms for medium- to high-density genotyping. A large number of commercial platforms for SNP genotyping are currently available (reviewed by Gupta et al. (2008)). The Illumina GoldenGate assays that genotype 384, 768 or 1,536 SNPs in parallel have been the most widely used for mid-throughput applications (Fan et al., 2003). This genotyping technique has

been used extensively in humans (International HapMap Consortium, 2003) and several animal species (McKay et al., 2008; Kijas et al., 2009; Malhi et al., 2011). SNP platforms are also available for several plant species, made up mostly of cereals, legumes and conifers (Pavy et al., 2008; Hyten et al., 2008; Akhunov et al., 2009; Close et al., 2009; Muchero et al., 2009; Yan et al., 2009; Deulvot et al., 2010; Lepoittevin et al., 2010; Massman et al., 2011). One of their main applications is the rapid development and saturation of genetic maps (Eckert et al., 2009; Yan et al., 2010).

Dense genetic maps are necessary tools for efficient molecular breeding. They are particularly useful for Quantitative Trait *Loci* (QTL) mapping and for the development of new high-quality mapping populations, such as introgression line libraries (Zamir, 2001; Collard et al., 2005). Four genetic maps have been reported in the *Cucurbita* genus to date. The first two maps were constructed using a population derived from an inter-specific cross between *C. pepo* x *C. moschata* Duchesne, a closely related species, with Random Amplified Polymorphic DNA (RAPD) markers (Lee et al., 1995; Brown and Myers, 2002). Two maps were subsequently produced from two intra-specific crosses, one using a cross between the oil-seed Pumpkin x Zucchini "True French" varieties (both of which belong to *C. pepo* ssp. *pepo*), and the other using a *C. pepo* ssp. *pepo* x *C. pepo* ssp. *ovifera* cross (oil-seed Pumpkin x Italian Crookneck, respectively). These maps consisted mainly of RAPDs and Amplified Fragment Length Polymorphisms (AFLPs) (Zraidi and Lelley, 2004; Zraidi et al., 2007). These markers are dominant and cannot be transferred readily to other populations. The first collection of SSR markers was recently produced from genomic libraries in *Cucurbita* by Gong et al. (2008b). Part of this collection, consisting of 178 SSRs, was used to increase the density of the Pumpkin x Crookneck map and also to study macrosynteny with *C. moschata*

(Gong et al., 2008a). Before the study by Blanca et al. (2011a), no SNP were available for the species, which is why these markers have not previously been used for mapping purposes.

Even though nearly one hundred major genes controlling different aspects of *Cucurbita* biology have been described (Paris and Kabelka, 2009), most have not been mapped. The available maps only include a few monogenic traits and have not yet been efficiently used for QTL mapping. There is a growing need for generating new maps with more informative and transferable markers that are amenable to large-scale genotyping. Markers linked to traits of interest are necessary for molecular breeding in these species, mainly in the Zucchini type, which by far dominates the squash market and the breeding efforts of seed companies. The current availability of a collection of 19,980 EST-SNPs, located mostly in gene-coding regions (Blanca et al., 2011a), will facilitate map development with functional markers.

In this study, we used a set of 9,043 EST-SNPs that were detected *in silico* by Blanca et al. (2011a), and which are suitable for detecting polymorphism between two main commercial types of *C. pepo* (Zucchini and Scallop) that have contrasting vine, flowering and fruit phenotypes, in order to develop an Illumina GoldenGate 384-SNP platform. This platform was employed to build the first SNP-based genetic map with an $F_2$ population (Zucchini x Scallop) and to detect QTLs for the very first time. The genotyping platform and the genetic map are invaluable new tools for molecular breeding in *Cucurbita.*

## Material and methods

<u>Plant material</u>

An $F_2$ population of 146 plants derived from the *C. pepo* ssp. *pepo* var. Zucchini MU-CU-16 x *C. pepo* ssp. *ovifera* var. Scallop UPV-196 cross was used to generate the linkage map. These are the same parental genotypes that were previously employed to generate the first *C.*

*pepo* transcriptome (Blanca et al., 2011a). Both represent the main summer squash cultivar groups of each subspecies, and have contrasting phenotypes for vine, flowering and fruit traits (Figure IV.2-1). Four $F_1$ plants and several individuals of each backcross generation to MU-CU-16 (BCZ, 30 individuals) and to UPV-196 (BCS, 30 individuals) were also included in the assay. In order to check if the selected SNPs might also be useful for genetic diversity studies and genotyping in other mapping populations, a panel of seven accessions of *C. pepo*, including representatives of the four morphotypes of the subspecies *pepo* (two Zucchini landraces from southern Spain, MU-20 and E-27; one Vegetable Marrow from Morocco, AFR-12; one Spanish Cocozelle landrace, V112; and two Pumpkin accessions, Styrian Pumpkin and the Mexican landrace, CATIE 18887) and one morphotype of the subspecies *ovifera* (the cultivar Early Summer Crookneck) were included in the genotyping assay. One accession of the related species, *C. moschata*, was also genotyped (the Spanish landrace AN-45). All these accessions belong to the *Cucurbita* core collection of the Cucurbits Breeding Group of the Institute for the Conservation and Breeding of Agricultural Biodiversity (COMAV) (Ferriol et al., 2003; Ferriol et al., 2004) except for CATIE 18887, which was kindly provided by the Genebank of the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in Costa Rica.

Total DNA was extracted from young leaves using the CTAB method (Doyle and Doyle, 1990) with minor modifications. To improve the quality of the obtained DNA, 70% ethanol containing 15 mM ammonium acetate was used in the last wash, and the DNA was treated with RNase. DNA concentrations in TE buffer were adjusted to 50 ng/μl, with the PicoGreen fluorescence being measured on an ABI7900 apparatus (Applied Biosystems). Samples were sent for genotyping to the Centro Nacional de Genotipado (CEGEN-ISCIII, CRG-Node,

Barcelona), a high-throughput genotyping service. A minimum of 200 ng of DNA were used for SNP genotyping.

SNP selection for the GoldenGate platform

Using the first *C. pepo* transcriptome (Blanca et al., 2011a) as a reference, a collection of 512,751 *C. pepo* ESTs, generated using 454 pyrosequencing, from the two genotypes used as the parentals for the mapping population (Zucchini MU-CU-16 and Scallop UPV-196), was mined for SNPs. This screening yielded a total of 19,980 putative SNPs and 1,174 INDELs, distributed in 8,147 unigenes. Using the different filters established in Blanca et al. (2011a), we selected a set of markers that, *in silico*, were monomorphic within and polymorphic between the two sequenced genotypes and suitable for genotyping with the Illumina GoldenGate system. Only SNPs were selected, as the INDELs were discarded (Blanca et al., 2011a). Sequences with more than 4 SNPs or INDELs per 100 bp were discarded (using filter HVR4) to avoid SNPs located in hypervariable regions. This selection was intended to reduce false polymorphisms provoked by the alignment of paralogs, a potentially significant problem when aligning short sequence reads. To facilitate their use in a GoldenGate genotyping assay, we also discarded those SNPs that were closer than 60 bp to another SNP or INDEL, to an intron or to the unigene edge (filtering them out with CS60, I60 and CL60, respectively). Only SNPs with two or more reads per allele were selected, since our previous experience with *in silico*-detected SNPs in melon (Blanca et al., 2011b) indicated that putative SNPs with only one read in one allele have a low percentage of validation (even when the quality of the sequenced nucleotide is high). Blanca et al. (2011a) annotated the unigene collection using the Blast2GO package (Conesa et al., 2005), which assigns Gene Ontology (GO) terms based on the BLAST definition. We used this annotation to prioritize the selection of SNPs located

in the Open Reading Frame regions (ORF) of annotated unigenes (with GO terms and significant BLAST in the Swiss-Prot, *Arabidopsis* org or Uniref90 databases (http://www.uniprot.org/downloads; http://www.arabidopsis.org/;http://www.ebi.ac.uk/uniref) and with orthologs of *Arabidopsis* and/or melon). A set of SNPs that generate allele-specific restriction targets, with the possibility of being detected *via* Cleaved Amplified Polymorphic Sequences (CAPS), was also included, even though they did not meet some of the aforementioned requirements.

The GO terms were reclassified into different functional groups based on a set of GO slims in the Molecular Function and Biological Process categories in order to provide a broad overview of the ontology content of the final platform.


SNP genotyping

The sequence of each selected *locus*, including the polymorphic nucleotide and a 60 bp flanking sequence, was submitted to the Illumina Assay Design Tool (ADT) (Illumina, San Diego, CA), and designability scores were used for final marker selection. These scores ranged from 0 to 1.0, where a score of > 0.6 means a high success rate for the conversion of an SNP into a successful GoldenGate assay. On the basis of these scores, a final set of 384 SNPs was selected, which was predicted to have a high likelihood of success. The GoldenGate genotyping assay was conducted as described elsewhere (Fan et al., 2003; Hyten et al., 2008; Deulvot et al., 2010).

To summarize, three primers were designed for each *locus*. Two were allele-specific oligos (ASOs), complementary to the sequence directly adjacent to the SNP, only differing at the 3' base complementary to each allele. The third primer was a *locus*-specific oligo (LSO), which hybridizes to the complementary sequence located downstream of the target SNP. The three

oligos had three universal primers attached at the 5' end. Each *locus*-specific oligo also had an "IllumiCode" sequence complementary to the array. The sequence of each *locus* and the 1,152 custom oligos, three at each of the 384 different SNP *loci*, are listed in Additional file IV.2-1: **"**Sequence and primers for genotyping the 384 SNPs included in the GoldenGate platform".

After DNA hybridization, an extension and ligation step was performed connecting each allele-specific oligo with the *locus*-specific oligo. A PCR step was then conducted for all 384 *loci* using common universal primers. The GoldenGate assay was deployed on the BeadXpress® platform using Veracode® technology (Illumina, San Diego, CA) (Tindall et al., 2010). The PCR products, labeled Cy3 or Cy5 depending on the allele, were hybridized to glass Veracode microbeads, each bearing a *locus*-specific barcode *via* the corresponding Illumicode sequence. Then, each SNP was identified by its IllumiCode and alleles were discriminated by their fluorescent signals on a Veracode BeadXpress Reader (Lin et al., 2009).

The automatic allele calling for each *locus* was accomplished using the GenomeStudio software (Illumina, San Diego, CA). The clusters were manually edited when necessary.

SSR selection and amplification

A set of 25 genomic SSRs (gSSR), evenly distributed in the previously published map constructed by Gong et al. (2008b) using the $F_2$ *C. pepo* ssp. *pepo* oil-Pumpkin variety "Lady Godiva" x *C. pepo* ssp. *ovifera* Crookneck variety "Bianco Friulano", were selected to be used as anchors between both maps. Information about the selected SSRs is included in Additional file IV.2-2: **"**Primers for genotyping the SSRs included in the map".

PCR reactions were carried out in a total volume of 15 μl in PCR buffer 1X (75 mM Tris-HCl pH 9, 20 mM $(NH_4)_2SO_4$, 50 mM KCl), 3 mM $MgCl_2$, 200 μM each dNTP, 0.15 μM each

primer, 0.2 µM M13 IRDye700/800 (LI-COR. Lincoln, Nebraska) tagged tail, 0.35 U Taq DNA Polymerase (Biotools B&M Labs, S.A., Madrid, Spain) and 10-15 ng DNA. Forward primers were designed with an added M13 tail sequence at their 5' end. The thermal profile was the following for all the *loci*: 3 min denaturation at 95ºC, 10 cycles of 30 s at 95ºC, 30 s at 65ºC (decreasing 1ºC every cycle) and 30 s at 72ºC, 20 cycles of 30 s at 95ºC, 30 s at 55ºC and 30 s at 72ºC with a final extension of 5 min at 72ºC. A LICOR 4300 analyzer was employed to visualize SSR-allele size differences on a denaturing polyacrylamide gel, loading a 1/10 or 1/20 dilution in formamide.

Linkage analysis and map construction

The genetic map was constructed using the genotyping results for the $F_2$ Zucchini x Scallop mapping population, obtained with the new 384-SNP GoldenGate platform and the anchor SSRs. Segregation distortion at each marker *locus* was tested against the expected ratio for $F_2$ (1:2:1) using a $\chi^2$ test. The linkage map was generated with MAPMAKER/EXP version 3.0b (Lincoln et al., 1993). Markers were associated with the "group" command with LOD> 4. Markers within groups were ordered using the "order" command. Distances in centiMorgans (cM) were calculated from the recombination frequencies using the Kosambi mapping function (Kosambi et al., 1944). The remaining markers were then located with the "try" command. The map was drawn with MapChart version 2.1 (Voorrips, 2002).

Synteny with cucumber

The colinearity of the *C. pepo* genetic map with the cucumber genome was evaluated by doing a BLAST search of the unigenes corresponding to every *C. pepo* SNP against the cucumber genome. The FASTA sequence of this genome was downloaded from the ICuGI

database (http://www.icugi.org). The hits obtained in the tBLASTx search of the *C. pepo* unigenes against the cucumber genome were considered to be significant if they had an e-value above $10^{-6}$. The locations of these significant hits were plotted in a scatter plot in which one axis represented the cucumber genome and the other the *C. pepo* map. The processing of the BLAST results was carried out with a custom Python script that is available upon request.

Phenotyping

All $F_2$, BCZ and BCS plants were cultivated in a greenhouse with a fully randomized experimental design (February to July, 2010) and extensively phenotyped. Five plants of each parental and the $F_1$ generation were also included in the assay. Fifty traits were measured for each single plant, and twelve were scored visually (Table IV.2-1). Vine traits were related to plant color, length and branching intensity, and flowering traits were related to the flowering time and male/femaleness tendency. Each plant was selfed and two fruits per plant were analyzed. One fruit per plant was analyzed when immature, 7 days after pollination, which corresponds to the commercial state of summer squashes. The second fruit per plant was analyzed at physiological maturity (ranging from 20 to 60 days after pollination). Traits measuring fruit size, shape, texture, firmness, rind and flesh color, sugar content and acidity were analyzed. More details about quantitative and qualitative traits are included in Table IV.2-1. Correlations between pairs of traits were estimated by using the Pearson correlation coefficient.

QTL analysis

QTLs for quantitative traits were analyzed by composite interval mapping with Windows QTL Cartographer 2.5 (Basten et al., 2005) using the developed genetic map and the stepwise

forward regression procedure with a walking speed of 1 cM, a window size of 15 cM and the inclusion of up to 5 maximum background marker *loci* as QTL cofactors. The LOD threshold for a Type I error $P< 0.05$ value was calculated by a permutation test (Churchill and Doerge, 1994) implemented in Windows QTL Cartographer with 1,000 permutations independently for each trait. Additive and dominant QTL effects (a and d, respectively), the degree of dominance (d/[a]) and the proportion of phenotypic variance explained by QTL ($R^2$) were estimated at the highest peaks depicted by the QTL analysis.

A positive value of additive effects (positive a) indicates that the Zucchini allele increases the trait, and, conversely, a negative value indicates the Scallop allele increases the trait. For positive a values, positive values of d indicate that the Zucchini allele is dominant, whereas negative values indicate that the Scallop allele is dominant. Conversely, for negative a values, positive and negative d values indicate dominance of the Scallop and Zucchini alleles, respectively.

In order to validate the QTL effects and the utility of the linked markers for breeding purposes (Marker-Assisted Selection, MAS), genotypic and phenotypic data of the two backcross populations, BCZ and BCS, were analyzed for the detected QTLs. ANOVA analysis conducted using the SPSS v. 16.0 software was employed to detect significant differences in the average value of homozygous backcross individuals (Zucchini, a, or Scallop, b) *versus* heterozygous individuals (h) for the markers located within or near the 1-LOD interval for the QTL. In those traits displaying QTLs confirmed in the backcrosses, broad-sense heritabilities were estimated as described by Wright (1968): $H^2 = [V_{F2} – (0.25V_Z + 0.25V_S + 0.5V_{F1})]/V_{F2}$, where V is the variance of $F_2$, Zucchini (Z), Scallop (S) and $F_1$ populations, respectively.

For the QTL analysis, the qualitative traits were coded as dummy variables, absent (0) or present (1), and analyzed with the Qgene v. 4.3.9 software (Joehanes and Nelson, 2008) using

composite interval mapping analysis and conducting 1,000 permutations to calculate the LOD threshold value for P< 0.05 using a resampling test. In order to confirm the observed linkage with the flanking markers, a contingency $\chi^2$ test was conducted in those cases in which significant LOD values were found. A null hypothesis ($H_0$) of independence of frequency between a trait (scored as 0-1) and the marker (genotyped as homozygous or heterozygous) was checked for the $F_2$, with an error type I rate of $\alpha= 0.05$ and 2 degrees of freedom (df). BCZ and BCS populations were also used for validating QTL effects in qualitative traits. We checked the frequency of the corresponding category in each group of individuals classified according to their genotype for the corresponding linked markers. The association between trait categories and linked markers was also checked using the Fisher exact probability test, as the number of individuals was too low (VassarStats: http://faculty.vassar.edu/lowry/fisher3x3.html; SISA: http://www.quantitativeskills.com /sisa/statistics/ord2.htm). P was calculated as the probability of the observed array of cell frequencies plus the sum of the probabilities of all other cell-frequency arrays that are smaller than the probability of the observed array. $H_0$ of independence was rejected when P< 0.05.

Information on those QTLs for quantitative and qualitative traits that were validated in the backcrosses was also included in the MapChart file to obtain a more complete map of the species.

**Results and Discussion**

Design of the 384-SNP GoldenGate genotyping platform

Of the 19,980 SNPs identified *in silico* (Blanca et al., 2011a), 9,043 were monomorphic within and polymorphic between the two parents of the map (Zucchini MU-CU-16 and Scallop UPV-196) (Figure IV.2-1) and were not located in highly variable regions (filtered

out with HVR4). A total of 3,538 of these high-confidence SNPs met the criteria for high-throughput genotyping platforms, i.e., being absent of any other known SNP in their vicinity and having enough sequence information up- and downstream of the SNP (filtered out with CS60, CL60 and I60). A preliminary set of 713 SNPs, located in different unigenes, was selected, prioritizing SNPs in long unigenes with well-defined functions. Designability scores were then given to each *locus* using the Illumina ADT. Only SNPs with scores of $> 0.6$ were selected. Sequences and primers of the finally selected SNP collection are included in Additional file IV.2-1. The Illumina scores and annotation details of the corresponding unigenes are also described in Additional file IV.2-3: "Annotation data and map position of the 384 *loci* included in the GoldenGate platform".

The final set of the 384 SNPs included in the GoldenGate platform had a mean designability score of 0.89. The average length of the selected unigenes was 1,057 bp (ranging from 398 to 2,336). These unigenes were previously annotated (Blanca et al., 2011a). Most SNPs (367, 95.6%) were located in the ORF of the corresponding unigene, with only 17 in the Untranslated Regions (UTR).

Blanca et al. (2011a) functionally classified the unigenes following the Gene Ontology (GO) scheme. Only 24 of the 384 selected unigenes (6.25%) with SNP could not be assigned to any GO term. We used the GO annotations to assign most unigenes (360, 93.8%) to a set of GO slims in the Biological Process and Molecular Function categories (Additional File IV.2-4: "Number of unigenes in each functional category"). The GO annotations for the unigenes showed a fairly consistent sampling of functional classes, indicating that these SNP markers represent genes with different molecular functions and that they are involved in various different biological processes. Cellular, metabolic, biosynthetic and developmental processes were among the most highly represented groups under the Biological Process category

(Additional file IV.2-4). Other abundant assignments were transcriptional regulation, translation, signal transduction, transport and oxidation-reduction functions. Stimulus, stress- and defense-responsive genes were also well represented. Genes involved in other important biological processes, such as growth, ripening and hormone-signaling processes were included. Some of these genes might play a role in the response to diseases, floral sex determination and fruit development and quality. Under the Molecular Function GO hierarchy (Additional file IV.2-4), assignments were mainly to catalytic and binding activities. A large number of hydrolases, kinases and transferases, representing genes involved in the secondary metabolite synthesis pathways, were also included. Transcription and translation factors were also well-represented.

The putative orthologs of all the unigenes were identified (Blanca et al., 2011a) by doing a reciprocal BLAST search of the *Arabidopsis* and melon databases (http://www.arabidopsis.org/; http://www.icugi.org). Most unigenes selected for the GoldenGate platform had an *Arabidopsis* ortholog (236, 61.4%) and/or a melon ortholog (228, 59.4%). Only 22.4% had no orthologs. GO terms, gene description and a list of the identified orthologs are included in Additional file IV.2-3.

Genotyping results: allele call and polymorphism

The GoldenGate genotyping assay was carried out successfully, with 90.1% of the SNPs successfully genotyped taking into account both monomorphic and polymorphic SNPs. Only 38 of the 384 SNPs included in the platform failed to give a clear genotype. Fifteen and sixteen SNPs could not be analyzed due to the absence of or low cluster separation, respectively; five displayed more than three clusters and two had low intensities, according to quality Veracode genotyping (Additional file IV.2-3). The absence of cluster separation might

be the result of a non-allele-specific match of the primers. Likewise, the existence of more than two alleles and/or the amplification of a non-unique genomic region might be the cause of the existence of more than three clusters.

The average designability score values for failed markers was significantly lower than that of the successful markers (0.86 *versus* 0.89, P< 0.05), but all scores were > 0.6, which is considered to be the optimal threshold for a GoldenGate assay. The percentage of failed markers with only 2 reads in one or both alleles (according to the sequencing results, Blanca et al. (2011a)) was higher compared to that of the successful SNPs (65.8% *versus* 43.6%).

All in all, a total of 346 SNPs were classified as successful assays. Similar success rates have been reported in soybean (Hyten et al., 2008), barley (Rostoks et al., 2006), maize (Yan et al., 2010) and pea (Deulvot et al., 2010). All of these markers amplified in nearly all the accessions of *C. pepo*. Knowing their polymorphism in diverse germplasm could help to determine their usefulness in future genetic diversity studies or mapping efforts.

One hundred and ninety-six SNPs detected variation among the morphotypes of *C. pepo* ssp. *pepo* (Zucchini, Vegetable Marrow, Cocozelle and Pumpkin), making them useful for genetic diversity studies or for mapping purposes using intra-subspecific crosses. Eighty-two detected variability among the assayed Zucchini types. Zucchini is by far the most important commercial type of summer squash and at the same time the most recently developed and the least variable. Therefore, markers detecting variability within this culti-group could be of interest for cultivar fingerprinting. Fifty-nine SNPs detected variation between the two accessions of *C. pepo* ssp. *ovifera* (Crookneck and Scallop) and seventy-eight markers yielded alleles exclusive to only one subspecies. The latter could be of interest for mapping purposes using inter-subspecific crosses. We included two accessions in our assay that belong to the morphotypes of the map parentals of Gong et al. (2008b) (Styrian Pumpkin and Early

Summer Crookneck). Two hundred and fourteen markers were polymorphic between them, and may be used to increase the density of that map, connecting both maps with common markers.

In addition, 305 SNPs (79%) were amplified in the *C. moschata* accession. This is an interesting result as most of the platform´s markers could potentially be used in introgression programs aimed at transferring traits from *C. moschata* into *C .pepo*. The only previous set of markers that proved to be transferable between *C. pepo* and *C. moschata* was a set of 76 genomic SSRs used to perform the first macrosynteny studies between the two species (Gong et al., 2008a). Our set of functional markers will be useful for further macrosynteny studies with this species and for the Marker-Assisted Selection of traits introgressed from *C. moschata* into *C. pepo*.

Details about the polymorphism detected by each SNP are included in Additional file IV.2-3. In order to facilitate the future application of this marker set, information about possible detection *via* CAPS is also provided. Sixty-two of the 384 SNPs affected restriction targets and could be easily assayed as CAPS.

The 384-SNP set was selected *in silico* for being polymorphic between Zucchini MU-CU-16 and Scallop UPV-196 (Blanca et al., 2011a). Of the 346 successfully called SNPs, 330 were polymorphic between these genotypes, yielding 3 clear clusters representing the two homozygous plus the heterozygous genotypes. Sixteen did not show polymorphism between the parentals. This could be explained by artifacts generated during the sequencing process. However, the lack of polymorphism could also be explained by the incapacity of this technique to discriminate an SNP at this *locus*, for example, because of the lack of amplification of one allele due to polymorphism in the priming site. In order to reduce false SNPs, we only selected SNPs with two or more reads per allele. Most of the monomorphic

markers have only 2 reads in one or both alleles (81.3 %). These results suggest that a higher number of reads per allele is a good criterion for selecting true SNPs from *in silico*-mined collections.

SSR results

The microsatellite transferability rate from the previous *Cucurbita* map (Gong et al., 2008b) to our mapping population proved to be low, as only 17 out of 25 SSRs (68%) amplified, and only 11 (44%) could be mapped. SSRs that displayed several amplification problems, such as nonspecific amplification (CMTp235 and CMTp245) and preferential amplification of the MU-CU-16 allele (CMTp86, CMTp188, CMTp47, CMTp256, CMTp33 and CMTp208), were discarded. The preferential amplification of the MU-CU-16 parental is consistent with the origin of these genomic SSRs, which were developed from a genomic library derived from a Pumpkin genotype (ssp. *pepo*). Four of the 17 amplified markers were monomorphic in our parentals and two resulted in a distorted $F_2$ segregation. Details of the SSR results are included in Additional file IV.2-2. SSRs are multiallelic markers, easily used by single PCR. However, SSR genotyping cannot be automated and the analysis of large populations is still time-consuming. This makes SNPs the preferred markers for different high-throughput genotyping purposes.

Genetic map of the Zucchini x Scallop population

We were able to successfully map 304 of the 330 SNPs that were polymorphic between the parentals. A set of 26 validated markers was discarded for mapping, either because the SNPs did not show the three genotypic classes in the $F_2$ or because one of the parents was heterozygous. The genetic map was constructed using a set of 315 high-quality markers (304

SNPs and 11 SSRs) using MAPMAKER at a LOD score of 4 (Figures IV.2-2, IV.2-3, IV.2-4, IV.2-5 and IV.2-6). The MU-CU-16 x UPV-196 genetic map covered 1,740.8 cM and was divided into 22 major linkage groups (LGs) and a minor group (LG23, with only 2 markers, 1.1 cM), with an average of 6.02 ± 6.65 cM between markers. The maximum gap between markers was 30.3 cM in LG13. Two SNPs, C007167 and C008395, remained unlinked.

The total number of markers included in major LGs varied from 5 in LG17 to 31 in LG2. Apart from LG23, only three groups contained less than eight markers (LG17, LG19 and LG22), with the markers being more or less evenly distributed among and within each LG group. LG length ranged from 12.2 cM in LG22 to 173.8 cM in LG2. On average, a linkage group covered 79.1 ± 34.7 cM and contained 14.1 ± 5.8 markers, resulting in an average map density of 5.56 ± 1.70 cM/marker. Less coverage was presented herein in comparison to the previous map for the species (1,936 cM and a density of 2.9 cM/marker) (Gong et al., 2008b). However, the previous map was mainly constructed with dominant, non-transferable RAPDs or AFLPs. Of the 659 *loci* mapped, only 178 correspond to co-dominant SSRs, which appeared unevenly distributed across the genome. Our results with the transferability of these markers have also been very low. The SNP-based map presented here is the first to include high-quality markers amenable to automation in the genus *Cucurbita,* many of which are putatively transferable to other populations and even to other species, and most of which are in fully annotated genes involved in diverse biological processes. In addition, distances have been reported not to be comparable between different software, as Joinmap lengths of the individual linkage groups are usually shorter than those obtained with MAPMAKER (Bradeen et al., 2001; Tani et al., 2003; Gong et al., 2008b).

Distorted segregation was observed in 30 SNPs and 2 SSRs, a larger number than in the Pumpkin x Crookneck cross (Zraidi et al., 2007; Gong et al., 2008b), but lower than that

reported in maps constructed from interspecific crosses (Brown and Myers, 2002). Grouped markers were especially observed in LG2 and LG5 (Figures IV.2-2 and IV.2-3).

Using the microsatellites as anchors to the previous *Cucurbita* map, it was possible to associate the linkage groups of both maps: LG2, LG5, LG7, LG8, LG9, LG12, LG14, LG16, LG18, LG21 and LG23 correspond to groups LGp2, 6, 9, 14, 3a, 18, 4, 10a, 8, 10a and 15 from Gong et al. (2008b), respectively. In the previous map, CMTp145 and CMTp66 mapped in the same group (10a) at LOD 3, but in this study, they appear associated with different groups (LG16 and LG21). In the future, newly developed SNPs will have to be mapped to improve the map saturation and obtain the 20 expected linkage groups, merging some of those that are less represented in the current map.

The distorted segregation found in LG2 was not reported in the corresponding LGp2 (Gong et al., 2008b), even though only three markers were mapped in this linkage group and the anchor SSR mapped in LG2 is out of this area. Scallop alleles were over-represented, suggesting that the alleles in this region may be subject to gametic or zygotic selection and/or related to preferential germination or better seedling viability. Different functions were associated with the distorted markers (Additional file IV.2-3). Some of these unigenes may be the cause of the segregation distortion, but it could also be the result of linkage to other genes.


Syteny with cucumber

Three hundred of the 304 mapped unigenes, yielded significant tBLASTx hits (threshold e-value of $10^{-6}$) and were assigned to the cucumber chromosomes. Figure IV.2-7 shows the colinearity between the genomes of the two species, *C. sativus* and *C. pepo*; details about the position of the unigenes in the cucumber genome are also included in Additional file IV.2-3.

We found syntenic blocks between most of the *C. pepo* linkage groups and *C. sativus* chromosomes.

Syntenic studies in the family *Cucurbitaceae* have been conducted with the two main cultivated species of the *Cucumis* genus: cucumber (2n = 14) and melon (2n = 24). Recent studies, using common markers and the whole genome sequence of cucumber, have shown that colinearity exists between cucumber and melon, indicating that chromosome fusions and other complex structural changes have generated cucumber chromosomes from a progenitor species with 2n= 24 (Li et al., 2011). We also found a high level of colinearity between *C. pepo* and the cucumber genome. Some *Cucurbita* linkage groups (LGs) can be considered homoeologous to cucumber chromosomes (Chr). For example, *Cucurbita* LGs 3, 5 and 18 showed syntenic blocks with cucumber Chromosome1, LG9 and 17 with Chr2, LG6 and 10 with Chr3, LG21 with Chr4, LG1, 2 and 14 with Chr5, LG7 with Chr6, and LG11 and 15 with Chr7. Some of the remaining LGs (4, 8, and 20) were syntenic to genetic blocks from two cucumber chromosomes (Chr 2-6, and Chr 4-6).

Most cucumber chromosomes contained two to three partially overlapping syntenic blocks with different LGs of *C. pepo*, which may suggest a certain level of duplication in this species. The higher chromosome number (2n=40) of *Cucurbita* suggests that this genus may be of polyploid origin. In fact, previous cytogenetic and isozyme studies indicate that this genus may be an ancient tetraploid (Weeden, 1984; Lebeda et al., 2007). Our results agree with a certain degree of duplication in this species.

A recent study on the level of macrosynteny between two species of the genus, *C. pepo* and *C. moschata*, through a comparative alignment of SSR markers, did not provide any indication of a possible ancient polyploid origin of the species (Gong et al., 2008a). In that paper, the authors studied the segregation of SSR *loci*, previously selected to be uniquely located in the

genome. However, in our study, synteny has been analyzed by blasting whole unigene sequences, which is more likely to yield significant matches in diverse genome sites than the uniquely located SSR primers. Differences in the approaches and the higher number of markers used in our study may explain the differences between the BLAST-based and SSR-based results.

QTL identification and QTL effect validation for Marker-Assisted Selection

Additional file IV.2-5: "Quantitative and qualitative traits" shows the values found for each attribute in the parents, $F_2$ and the backcross populations, clearly demonstrating phenotypic variability for most attributes. Forty-eight QTLs were detected for 31 quantitative traits and 11 QTLs were detected for 11 qualitative traits. These QTLs were distributed in 24 independent positions in 13 linkage groups. The proportion of the phenotypic variance explained by a single QTL ($R^2$) varied from 7% to 81%. Fifteen major QTLs ($R^2 > 25\%$) were detected for flowering traits (associated with late flowering and maleness tendency) and for immature and mature fruit traits (associated with fruit length and rind and flesh color). Detailed information about all these QTLs (explained variance, LOD peaks, flanking markers, additive and dominance effects and heritabilities) are shown in Additional Files IV.2-6 and IV.2-7: "QTL analysis for quantitative and qualitative traits 1 and 2".

The genetic inheritance of important agronomic traits is largely unknown in *Cucurbita*. This QTL analysis provides the first results of the genetic control of most of these plant, flowering and fruit traits. Our preliminary results should be further confirmed using additional populations and phenotypic replications. In this paper, we confirmed the utility of some of these QTLs for Marker-Assisted Selection by validating their effects on the backcross populations. Despite the limited number of plants, the effects of eleven of the 15 major QTLs

(*NoMaF_3*, *NoFeF_3*, *DMaF_3*, *DFeF_3*, *NºMaF_3*, *MaF/FeF_3*, *IFLe_6*, *MFLe_6*, *MLRCo_14*, *MaRCo_14* and *MbFCo_16*) detected in the $F_2$ were verified in one or both backcross populations (Table IV.2-2). In addition, six minor QTLs (all with $R^2 > 10\%$) (*IFLe_18, MFWi_6, MRib_11, MCaTh_6, MFFi_2* and *MLoNº_5)* and eight QTLs involved in qualitative traits (*SC_14*, *LIns_14*, *T_1*, *IFCo_20*, *MFSh_6*, *MPriRCo_14*, *MPSecRCo_14* and *MFCo_16*) were also verified in the backcrosses (Table IV.2-2). The verified QTLs segregated differently between the backcross populations, segregating only in one or in both of them. This differential segregation is in general compatible with the direction of additive effects and dominance deviation estimated in the $F_2$. Information about the QTL set validated in the backcross populations is detailed in Additional file IV.2-7. The most likely positions on the linkage map for these validated QTLs are shown in Figures IV.2-2, IV.2-3, IV.2- 4, IV.2-5 and IV.2-6. The most important QTLs displaying real effects in backcrosses related to flowering, fruit shape and color are described below in greater detail.

*Flowering*

A cluster of QTLs controlling several flowering traits (all with medium–high broad-sense heritabilities, 0.71 – 0.85) was detected in LG3, most of which had major effects ($R^2 > 25\%$) and partial or complete dominance of the Zucchini alleles (d/[a] from -0.78 to -1.05), associated with the early appearance of male and female flowers as well as an enhanced femaleness tendency of the plant (*NoMaF_3, DMaF_3, NoFeF_3, DFeF_3, NºMaF_3, MaF/FeF_3*) (Additional file IV.2-7, Figure IV.2-2). In agreement with the a and d values estimated in the $F_2$, the backcrosses show how the Scallop alleles delayed flowering and increased maleness with a recessive gene action (Table IV.2-2). Consequently, no differences between plants homozygous for the Zucchini alleles *versus* heterozygous were found in the BCZ population, whereas the mean of the plants homozygous for the Scallop alleles was

significantly higher than those of the heterozygous genotypes in the BCS population. The sex expression in *Cucurbitaceae* is known to be controlled by various genetic, environmental and hormonal factors, with ethylene being the main hormone involved in this trait. In *C. sativus* and *C. melo,* it is controlled by several major independent genes, some of which have been cloned (Boualem et al., 2008, Martin et al., 2009; Li et al, 2009). Our results also suggest the existence of a major gene controlling flowering time and the enhanced female/maleness phenotype in summer squash. Further research is necessary to determine whether the co-segregation of the flowering time traits and female/male tendency is due to pleiotropy at a single *locus* or linkage between *loci*.

*Fruit shape*

Two major QTLs ($R^2 > 25\%$) involved in fruit shape, controlling the length of immature and mature fruits (*IFLe_6* and *MFLe_6*), co-segregate in LG6, along with various minor QTLs that control mature-fruit width and cavity thickness (*MFWi_6, MCaTh_6*) and also with a QTL controlling fruit shape (MFSh_6) (Additional file IV.2-7, Figure IV.2-3). The Zucchini type contributed alleles producing elongated fruits, while the Scallop alleles produced wider fruits with wider cavities. Most of these traits presented moderate heritabilities. The two major QTLs (*IFLe_6* and *MFLe_6*), with additive gene action estimated in the $F_2$ (d/[a] 0.24 and -0.09 respectively), were verified in both the BCS and BCZ populations with the expected direction of allelic effects (Table IV.2-2). These results suggest that these QTLs can be exploited in both genetic backgrounds for hybrid or pure line development. *MFWi_6* and *MCaTh_6* were also additive in the $F_2$, but they were verified only in one of the backcross populations, which may be due to the low capacity for QTL detection in the backcross populations due to their modest sample size or to genetic background effects. An independent

QTL for fruit length was detected in LG18 (*IFLe_18*). Also, Scallop alleles of *MLoNº_5* and *MRib_11* modified fruit shape by increasing the number of locules and the ribbing intensity.

Several genes have been reported to be related to fruit shape. A dominant gene (*Di*) seems to control the discoid fruit shape of scallop squash (Paris and Kabelka, 2009). This gene was reported to be dominant over spherical or pyriform shapes. A digenic epistatic control has also been reported for summer squash fruit shape. Our results are consistent with the existence of a major gene that is, however, not dominant, and several minor modifiers.

*Fruit color*

Major QTLs for the rind color of mature fruits mapped in LG14 (*MLRCo_14* and *MaRCo_14*), with lightness (L Hunter parameter, white color) increasing with Scallop alleles and greenness increasing with Zucchini alleles (Additional file IV.2-7, Figure IV.2-5). High heritabilities were found (0.95 and 0.97) for these rind color parameters. Also, the visual scores of primary rind color and the pattern of secondary color mapped in the same region (*MPriRCo_14* and *MPSecRCo_14),* with Zucchini alleles leading to dark rind colors and Scallop alleles to the absence of a secondary color (Table IV.2-2).

The genetic control of flesh color seems to be independent. A major QTL for immature flesh color is located in LG20 (*IFCo_20*), whereas a major QTL was found for mature fruit flesh color in LG16 (*MbFCo_16*), which is consistent with the location of the qualitative trait color *MFCo_16* (Additional file IV.2-7, Figures IV.2-5 and IV.2-6).

Squash fruit color has been studied intensively, and a complex genetic control has been proposed for rind color, with major genes (one dominant, derived from Scallop *W* (weak rind coloration)) (Paris and Kabelka, 2009), complemented by modifiers, whereas less complexity is reported for flesh color. The QTLs that control rind color in mature fruits were validated in

one or both backcrosses (Table IV.2-2). Homozygous plants for the Zucchini allele are dark green or black, whereas individuals that are heterozygous or homozygous for the Scallop allele are white or cream-colored in any genetic background, consistently with the major gene *W*, which confers a white or cream color independently of genetic background (Paris and Kabelka, 2009). This gene has been reported to be complementary to the major gene, *Wf,* also from Scallop, which is dominant over colored flesh (Paris and Kabelka, 2009), as most white-rinded squashes are also white-fleshed. Accordingly, mature homozygous Zucchinis for the SNP marker C017913, which is linked to *MFCo-16,* were mostly orange/yellow-fleshed, whereas homozygous Scallops were mostly white-fleshed. However, heterozygous individuals were all white-fleshed in the Scallop background, while some yellow-fleshed fruits appeared in the Zucchini background. Consistently, a significant effect of Scallop alleles of C030754 (linked to *MbFCo_16*), reducing flesh yellowness, was only detected in the Scallop background. Therefore, it seems to be a major gene with dominance of white flesh (*Wf*), although other minor genes also seem to contribute to the control of flesh color.

Most of the QTLs reported in this paper had not been located previously. The maps that have been developed to date include six monogenic traits (precocious yellow fruit, *B*; bush growth habit, *Bu;* leaf mottling, *M*; hull-less seed coat, *n;* and mature fruit color) (Brown and Myers, 2002; Zraidi and Lelley, 2004; Gong et al., 2008b), most of which did not segregate in our population. QTLs for fruit length, width and number of fruit locules were located on a Zucchini x Crookneck map constructed using RAPD markers (Zraidi and Lelley, 2004). Other QTLs were also reportedly associated with RAPD markers for fruit shape and leaf indentation using an interspecific *C. pepo* x *C.moschata* map (Brown and Myers, 2002). However, the comparison of the results is not possible due to the lack of common markers with the current

map. These previously detected QTLs have not been used to date for MAS selection in *Cucurbita*.

**Conclusions**

Our results demonstrate the utility of the 384-SNP GoldenGate genotyping array in *Cucurbita pepo*. Next-generation sequencing, together with this cost-effective genotyping technique, have been successfully applied to constructing the first SNP-based genetic map reported in the genus. This Zucchini x Scallop map is not only an important resource for high-quality markers that are polymorphic between two highly contrasting squash types, but is also an invaluable tool for breeding purposes, since these markers are developed in coding regions involved in different physiological processes. Several preliminary QTLs related to vine, flowering and fruit traits in the mature and immature stages have been reported and mapped for the first time. QTL effects have been validated as has been the utility of various markers for Marker-Assisted Selection, which demonstrates the suitability of the current population and genetic map for dissecting genetically complex fruit traits in *Cucurbita* spp. This information will be essential for future breeding programs focused on obtaining better-adapted varieties. The SNP platform has been successfully assayed to detect variability between/within both *C. pepo* subspecies and different squash morphotypes, and has also revealed a great number of *loci* transferable to *C. moschata*. This will facilitate synteny studies with other cucurbits and subsequent diversity and mapping studies that will contribute to increasing the genomic resources for these crops.

**Authors' contributions**

BP designed and coordinated the study, developed the $F_2$ population, selected the markers for the GoldenGate assay, provided the phenotypic data and contributed to the mapping, QTL and synteny analysis. CE participated in the design of the study, and in the development of the $F_2$ population, and contributed to the phenotyping process. CE prepared the DNA of the populations, contributed to the SNP selection process, and performed the SSR analysis. AJM and CE performed the mapping and QTL analysis. JB assisted in the marker selection and in the synteny study. CR assisted in the annotation process. PG participated in the conception of the study. PG and NVD also participated in the design and development of the $F_2$ population, and contributed to the phenotyping and genotyping processes. FN is the director of the COMAV and participated in the conception of the study. BP and CE integrated all the information and drafted the manuscript, with contributions from all authors. All authors have read and approved the final manuscript.

**Additional material**

**Additional file IV.2-1. Sequence and primers for genotyping the 384 SNPs included in the GoldenGate platform.**

The sequences of each of the 384 SNPs included in the GoldenGate platform are indicated, including the polymorphic nucleotide and a 60 bp flanking sequence, along with the *locus-*specific and two allele-specific primers designed for detecting each *locus* with the GoldenGate assay, the three universal primers and the Illumicode.

**Additional file IV.2-2: Primers for genotyping the SSRs included in the map.**

The primers for genotyping the 25 SSRs selected from the previous *C. pepo* map (Gong et al., 2008b) are listed, along with the genotyping results and map position for each *locus.*

**Additional file IV.2-3: Annotation data and map position of the 384 *loci* included in the GoldenGate platform.**

Annotation data of the 384 unigenes included in the GoldenGate platform are described. Previous annotation data provided by Blanca et al. (2011a), after *in silico* detection of the unigenes, consist of unigene length, position of the SNP, in ORF or UTR, number of reads in each parental, GO terms, gene description after sequential BLAST of Swissprot, Arabidopsis org and Uniref90, orthologs detected with *Arabidopsis* and *C. melo* by reciprocal BLAST of *Arabidopsis_*pep and ICUGI databases and putative SNP-CAPS. Data generated in this paper are final CRG/CEGEN scores for GoldenGate genotyping reactions, GoldenGate genotyping results, distribution of the unigenes in the *C. sativus* genome (scaffolds or chromosomes) after BLAST against cucumber genome available at ICUGI, linkage group according to the

Zucchini x Scallop map obtained, variability of SNPs in the germplasm panel of *C. pepo* ssp. *pepo* and ssp. *ovifera* accessions and amplification in *C. moschata*.

**Additional file IV.2-4: Number of unigenes in each functional category.**

Number of unigenes, of the 384 included in the GoldenGate platform, assigned to each GO Slim in the Biological Process category (A) and the Molecular Function category (B).

**Additional file IV.2-5: Quantitative and qualitative traits.**

**A.** Scored quantitative traits are described. Mean values, ranges and standard deviation for parental, $F_2$ and backcross populations are indicated. **B.** Pearson correlations between pairs of quantitative traits. **C.** Visually scored qualitative traits are described. Relative frequency of each phenotype is shown for both parental, $F_2$ and backcross populations.

**Additional file IV.2-6: QTL analysis for quantitative and qualitative traits 1.**

QTLs whose effects have not been validated in the backcross populations are included.

**A.** Linkage group positions and flanking markers of 31 QTLs, along with their associated logarithms of odds (LOD) for 20 vine, flowering and fruit quantitative traits analyzed in the $F_2$ population derived from the Zucchini x Scallop cross. Major QTLs ($R^2 > 25\%$) are indicated in bold. **B.** Linkage group positions and flanking markers of 3 QTLs, along with their associated logarithms of odds (LOD) and contingency $\chi^2$ results for fruit qualitative traits analyzed in the $F_2$ population derived from the cross Zucchini x Scallop. Major QTLs ($R^2 > 25\%$) are indicated in bold.

**Additional file IV.2-7: QTL analysis for quantitative and qualitative traits 2.**

QTLs whose effects have been validated in the backcross populations are included. **A.** Linkage group positions and flanking markers of 17 QTLs, along with their associated logarithms of odds (LOD) for 16 flowering and fruit quantitative traits analyzed in the $F_2$ population derived from the Zucchini x Scallop cross. Major QTLs ($R^2 > 25\%$) are indicated in bold. Heritabilities have been calculated for these traits **B.** Linkage group positions and flanking markers of 8 QTLs, along with their associated logarithms of odds (LOD) and contingency $\chi^2$ results for vine and fruit qualitative traits analyzed in the $F_2$ population derived from the cross Zucchini x Scallop. Major QTLs ($R^2 > 25\%$) are indicated in bold.

**Table IV.2-1. Quantitative and Qualitative traits measured/scored in the mapping populations.**

| Trait code | Description/categories |
|---|---|
| **Quantitative traits** | |
| *Vine traits* | |
| NºBr | Number of branches 7 days after the appearance of the first female flower |
| PLe | Plant length measured at the end of the assay (cm) |
| NoNº | Number of nodes measured at the end of the assay |
| *Flowering traits* | |
| NoMaF | First node with a male flower |
| NoFeF | First node with a female flower |
| DMaF | Days from sowing to the development of the first male flower |
| DFeF | Days from sowing to the development of the first female flower |
| NºMaF | Number of male flowers measured 7 days after the opening of the first female flower |
| NºFeF | Number of female flowers measured 7 days after the opening of the first female flower |
| TNºF | Total number of flowers measured 7 days after the opening of the first female flower |
| MaF/FeF | Ratio male to female flowers |
| *Immature fruit* | |
| IPeLe | Peduncle length (mm) |
| IFLe | Fruit length (cm) |
| IFWi | Fruit width (cm) |
| IFWe | Fruit weight (g) |

| | |
|---|---|
| IRTh | Rind thickness (mm) |
| IFTh | Flesh thickness (mm) |
| **Trait code** | **Description/categories** |
| ICaTh | Cavity thickness (mm) |
| IBrix | Total soluble solids measured with refractometer (Brix degrees) |
| IRFi | Rind firmness measured with penetrometer (kg) |
| IFFi | Flesh firmness measured with penetrometer (kg) |
| IRBr | Rind brightness, scored visually as matte (0), medium (1) and bright (2) |
| ILoNº | Number of locules |
| ILRCo | Rind color measured with colorimeter, Hunter parameter L (Lightness: from white, L=100, to black, L=0) |
| IaRCo | Rind color measured with colorimeter, Hunter parameter a (from redness for positive values to greenness for negative values) |
| IbRCo | Rind color measured with colorimeter, Hunter parameter b (from yellowness for positive values to blueness for negative values) |
| ILFCo | Flesh color measured with colorimeter, Hunter parameter L |
| IaFCo | Flesh color measured with colorimeter, Hunter parameter a |
| IbFCo | Flesh color measured with colorimeter, Hunter parameter b |
| *Mature fruits* | |
| DMa | Days from pollination to maturity |
| MPeLe | Peduncle length (mm) |
| MFLe | Fruit length (cm) |
| MFWi | Fruit width (cm) |
| MFWe | Fruit weight (g) |
| MRib | Intensity of fruit ribbing, scored visually based on presence and depth of the ribs as absent (0), surface ribbing (1), intermediate ribbing (2) and strong |

| | ribbing (3) |
|---|---|
| MRTh | Rind thickness (mm) |
| MFTh | Flesh thickness (mm) |
| **Trait code** | **Description/categories** |
| MCaTh | Cavity thickness (mm) |
| MBrix | Total soluble solids measured with refractometer (Brix degrees) |
| MpH | pH measured with paper |
| MRFi | Rind firmness measured with penetrometer (kg) |
| MFFi | Flesh firmness measured with penetrometer (kg) |
| MRBr | Rind brightness, scored visually as matte (0), medium (1) and bright (2) |
| MLoNº | Number of locules |
| MLRCo | Rind color measured with colorimeter, Hunter parameter L |
| MaRCo | Rind color measured with colorimeter, Hunter parameter a |
| MbRCo | Rind color measured with colorimeter, Hunter parameter b |
| MLFCo | Flesh color measured with colorimeter, Hunter parameter L |
| MaFCo | Flesh color measured with colorimeter, Hunter parameter a |
| MbFCo | Flesh color measured with colorimeter, Hunter parameter b |

**Qualitative traits**

*Vine traits*

| SC | Stem color, scored as dark green, intermediate or light green |
|---|---|
| LIns | Green to white color change in leaf insertion scored as absent or present |
| T | Presence of tendrils scored as absent or present |

*Immature fruit*

| IFSh | Fruit shape,  scored as elongated,pear-shaped, discoid or oval |
|---|---|
| IPriRCo | Primary rind color,  scored as dark green, green, lightgreen, white-green or white |

| | |
|---|---|
| IPSecRCo | Pattern of secondary color,  scored as dotted speckled, striped or absent |
| IFCo | Flesh color,  scored as green, light green, white-green or white |

*Mature fruit*

| | |
|---|---|
| MFSh | Fruit shape,  scored as elongated, pear-shaped, discoid or oval |

| Trait code | Description/categories |
|---|---|
| MPriRCo | Primary rind color,  scored as black, green, orange, yellow, cream or white |
| MPSecRCo | Pattern of secondary color,  scored as dotted speckled, banded, striped or absent |
| MFCo | Flesh color,  scored as green, orange, yellow, cream or white |
| MRTe | Rind texture, scored as smooth or warted |

**Table IV.2-2. QTL effects validated in backcross populations.**

**A Quantitative traits**

| QTL | Marker[1] | BCZ[2] | | BCS[3] | |
|---|---|---|---|---|---|
| | | a | h | b | h |
| *Flowering* | | | | | |
| *NoMaF_3* | C001057 | 1.273 | 1.143 | **3.467** | **1.929** |
| *NoFeF_3* | C001057 | 11.546 | 9.714 | **19.400** | **14.286** |
| *DMaF_3* | C001057 | 24.818 | 23.571 | **30.200** | **26.786** |
| *DFeF_3* | C001057 | 31.364 | 27.571 | **41.000** | **36.357** |
| *NºMaF_3* | C001057 | 19.000 | 21.000 | **35.300** | **30.800** |
| *MaF/FeF_3* | C001057 | 3.974 | 2.531 | **18.900** | **11.600** |
| *Immature fruits* | | | | | |
| *IFLe_6* | C002852 | **19.633** | **15.450** | 7.890 | **11.556** |
| *IFLe_18* | C003897 | 17.717 | 17.367 | **8.300** | **10.500** |
| *Mature fruits* | | | | | |
| *MFLe_6* | C002852 | **31.806** | **26.994** | 9.728 | **13.386** |
| *MFWi_6* | C002852 | **9.361** | **10.950** | 13.469 | 13.250 |
| *MRib_11* | C004998 | 0.000 | 0.000 | **2.118** | **1.500** |
| *MCaTh_6* | C008686 | **61.847** | **75.932** | 80.080 | 85.575 |
| *MFFi_2* | C011474 | 10.614 | 10.830 | **11.648** | **8.671** |
| *MLoNº_5* | C016718 | 3.450 | 3.500 | **4.600** | **3.900** |
| *MLRCo_14* | C005730 | **50.285** | **77.989** | 82.382 | 78.662 |
| *MaRCo_14* | C005730 | **-7.902** | **0.419** | -0.4406 | 0.9755 |
| *MbFCo_16* | C030754 | 22.338 | 26.532 | **10.636** | **12.139** |

**B Qualitative traits[4]**

| QTL | Marker[1] | BCZ[2] | | BCS[3] | |
|---|---|---|---|---|---|
| | | a | h | b | h |
| *Vine traits* | | | | | |
| *SC_14* | C005730 | | | | |
| dark green | | 0.64 | 0.29 | **0.18** | **0.67** |
| intermediate | | 0.36 | 0.71 | **0.64** | **0.33** |
| light green | | 0.00 | 0.00 | **0.18** | **0.00** |
| P[5] | | 0.192 | | **0.011†** | |
| *LIns_14* | C005730 | | | | |
| absent | | **1.00** | **0.29** | 0.00 | 0.00 |
| present | | **0.00** | **0.71** | 1.00 | 1.00 |
| P | | **0.020†** | | 1.00 | |
| *T_1* | C003546 | | | | |
| present | | 0.58 | 0.33 | **0.53** | **1.00** |
| absent | | 0.42 | 0.67 | **0.47** | **0.00** |
| P | | 0.864 | | **0.011†** | |
| *Immature fruits* | | | | | |
| *IFCo_20* | C005014 | | | | |
| green | | **0.13** | **0.00** | 0.00 | 0.00 |
| light green | | **0.00** | **0.50** | 0.00 | 0.08 |
| white-green | | **0.88** | **0.50** | 0.17 | 0.46 |
| white | | **0.00** | **0.00** | 0.83 | 0.46 |
| P | | **0.051†** | | 0.404 | |
| *Mature fruits* | | | | | |

| QTL | Marker[1] | BCZ[2] | | BCS[3] | |
| --- | --- | --- | --- | --- | --- |
| *MFSh_6* | C002852 | | | | |
| elongated | | 1.00 | 0.89 | **0.00** | **0.00** |
| **QTL** | **Marker[1]** | **BCZ[2]** | | **BCS[3]** | |
| | | **a** | **h** | **b** | **h** |
| pear-shaped | | 0.00 | 0.11 | **0.06** | **0.36** |
| discoid | | 0.00 | 0.00 | **0.94** | **0.64** |
| P | | 0.500 | | **0.042†** | |
| *MPriRCo_14* | C005730 | | | | |
| black | | **0.91** | **0.00** | 0.00 | 0.00 |
| green | | **0.09** | **0.00** | 0.00 | 0.00 |
| cream | | **0.00** | **0.29** | 0.00 | 0.06 |
| white | | **0.00** | **0.71** | 1.00 | 0.94 |
| P | | **0.000†** | | 1.000 | |
| *MPSecRCo_14* | C004187 | | | | |
| dotted speckled | | **0.92** | **0.50** | 0.17 | 0.00 |
| banded | | **0.00** | **0.00** | 0.00 | 0.00 |
| striped | | **0.00** | **0.00** | 0.00 | 0.00 |
| absent | | **0.08** | **0.50** | 0.83 | 1.00 |
| P | | **0.022†** | | 0.163 | |
| *MFCo_16* | C017913 | | | | |
| green | | **0.09** | **0.00** | 0.00 | 0.00 |
| orange | | **0.45** | **0.00** | 0.00 | 0.00 |
| yellow | | **0.27** | **0.67** | 0.00 | 0.00 |
| cream | | **0.09** | **0.00** | 0.15 | 0.00 |
| white | | **0.09** | **0.33** | 0.85 | 1.00 |
| P | | **0.051†** | | 0.222 | |

**A.** Average data for quantitative traits with QTLs displaying significant differences (P> 0.05) between homozygous and heterozygous individuals for linked markers in BC populations. Major QTLs ($R^2$> 25%) are indicated in bold. Data traits with significant differences are indicated in bold. **B.** Frequency for the different categories for qualitative traits with QTLs displaying significant differences between homozygous and heterozygous individuals for linked markers in BC populations and results of the Fisher exact test. Data traits with significant differences are indicated in bold.

[1] Tested markers located in the QTL region (see Figures IV.2-2, IV.2-3, IV.2-4, IV.2-5 and IV.2-6).

[2] Homozygotes for the Zucchini allele of the corresponding marker in BCZ are indicated as a (allele from MU-CU-16), while heterozygotes are indicated as h.

[3] Homozygotes for the Scallop allele of the corresponding marker in BCS are indicated as b (allele from UPV-196), while heterozygotes are indicated as h.

[4] Only the categories represented in the BC populations are included.

[5] Fisher's exact probability test. P ($\alpha$=0.05). P< 0.05 implies association and linkage with the marker, as $H_0$ of independence is rejected (†).
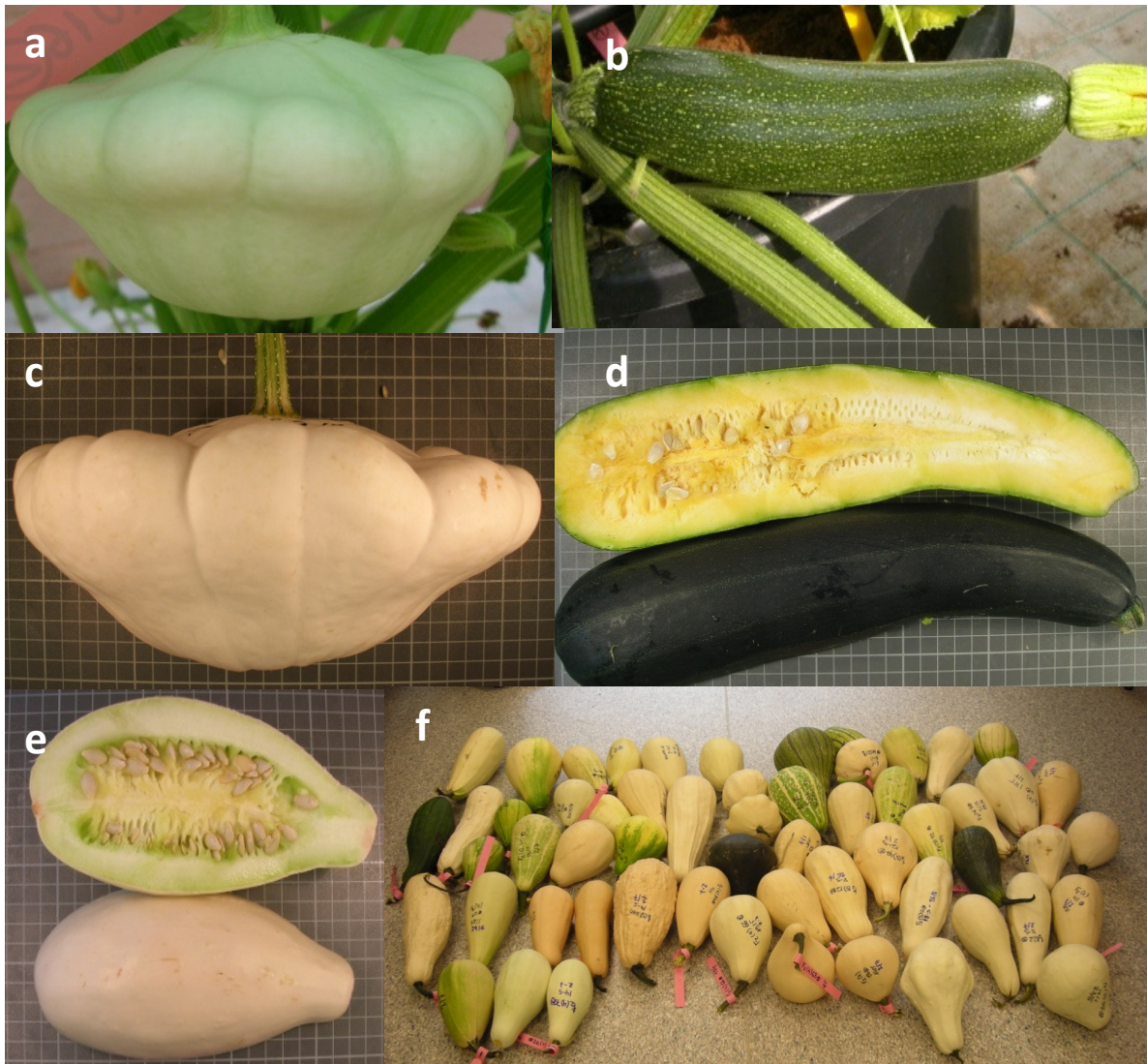
**Figure IV.2-1. Fruit characteristics of the map parentals and derived populations.**

Pictures showing fruit characteristics of Scallop and Zucchini parentals and derived populations.

Legend: Immature fruits of Scallop UPV-196 and Zucchini MU-CU-16 (a and b), mature fruits of Scallop UPV-196 and Zucchini MU-CU-16 (c and d), mature fruits of $F_1$ (e) and a sample of mature fruits of the $F_2$ population (f).

**Figure IV.2-2.**

**LG6**

**LG4**

| | |
|---|---|
| 0.0 | C004993 |
| 22.0 | C013719 |
| 34.1 | C004878 C016414 |
| 34.5 | C000347 |
| 43.1 | C006030 |
| 47.1 | C010187 |
| 51.6 | C019683 |
| 54.6 | C003828 |
| 58.8 | C010321 |
| 61.7 | C003173 |
| 82.6 | C000329 |
| 83.3 | C019143 |
| 84.4 | *C005183* |

**LG5** (LGp6)

| | |
|---|---|
| 0.0 | C013735 |
| 3.0 | C005807 |
| 8.3 | *C008230* |
| 8.7 | *C002947* |
| 11.3 | *C004458* |
| 13.5 | *C011973* |
| 23.2 | *C016718* |
| 30.3 | C014249 |
| 37.8 | C000346 C000405 |
| 40.4 | C007998 |
| 52.7 | C017904 |
| 53.8 | C003102 |
| 58.3 | C007598 |
| 61.7 | C034841 |
| 65.5 | p224 |

MlLoN°_5

| | |
|---|---|
| 0.0 | C002852 |
| 0.4 | C015514 |
| 13.9 | C008686 |
| 18.4 | C015501 |
| 24.1 | C005440 |
| 31.4 | C011991 |
| 31.8 | C006591 |
| 32.6 | C000640 |
| 45.4 | C006873 |
| 50.4 | C003972 |
| 75.2 | C000384 |
| 92.6 | C010469 |
| 93.0 | C003513 C006940 |
| 94.8 | C006249 |
| 110.0 | C002510 |
| 113.9 | C008974 |
| 125.8 | C004289 |
| 128.8 | C006129 |

MFSh_6

MFLe_6

MCaTh_6

MFWi_6

IFLe_6

**LG7** (LGp9)

| | |
|---|---|
| 0.0 | C014849 |
| 6.6 | C000669 |
| 16.4 | C002508 |
| 17.5 | C006159 |
| 20.5 | C003049 |
| 25.8 | p257 |
| 34.3 | C001936 |
| 40.0 | C007994 |
| 43.3 | C001085 |
| 56.7 | C011078 |
| 59.9 | C016360 |
| 66.8 | C003488 |
| 67.5 | C002153 |
| 70.1 | C003838 C002033 |
| 70.5 | C004010 |
| 80.2 | C011611 |
| 81.7 | C018448 |

**Figure IV.2-3.**

312

**Figure IV.2-4.**

**Figure IV.2-5.**

**LG17**

| | |
|---|---|
| 0.0 | C007803  C003719 |
| 14.4 | C001768 |
| 18.1 | C002779 |
| 28.5 | C011310 |

**LG18 (LGp8)**

| | |
|---|---|
| 0.0 | C017202 |
| 0.4 | C021137 |
| 6.5 | *p142* |
| 19.7 | C014772 |
| 24.0 | C003029 |
| 26.7 | C002838 |
| 30.0 | C010047 |
| 34.1 | C011824 |
| 37.4 | C003897 |
| 38.9 | C002195 |
| 39.6 | C005864 |
| 40.7 | C014198 |
| 42.2 | C005102 |
| 42.6 | C014452 |
| 49.0 | C005026 C009536 |
| 55.0 | C003601 |

IFLe_18

**LG19**

| | |
|---|---|
| 0.0 | C004893 |
| 5.7 | C026163 |
| 8.3 | C002567 |
| 27.7 | C037292 |
| 44.1 | C002530 |
| 46.3 | C004995 |

**LG20**

| | |
|---|---|
| 0.0 | C010415 |
| 23.0 | C014556  C004091 |
| 27.1 | C008582 |
| 34.3 | C002672 |
| 35.5 | C012979 |
| 40.7 | C003485  C020373 |
| 48.0 | C005014 |
| 53.5 | C011020 |
| 63.9 | C002747 |
| 66.9 | C003461 |
| 69.5 | C007734 |
| 93.1 | C004499 |

IFCo_20

**LG21 (LGp10a)**

| | |
|---|---|
| 0.0 | C005187 C029956 |
| 0.4 | C003408 C002688 |
| 0.8 | C003363 |
| 9.0 | C002528 |
| 28.7 | *p66* |
| 36.7 | C010944 |
| 45.9 | C006595 |
| 46.3 | C004258 |
| 46.7 | C005379 |

**LG22**

| | |
|---|---|
| 0.0 | C002741  C016178 |
| 0.7 | C013426  C011361 |
| 3.3 | C020673  C001423 |
| 12.2 | C014609 |

**LG23 (LGp15)**

| | |
|---|---|
| 0.0 | C002948 |
| 1.1 | *p169* |

**Figure IV.2-6.**

**Figure IV.2-2 to Figure IV.2-6. Genetic map of Zucchini x Scallop F$_2$ population.**

Linkage map and locations of quantitative trait *loci* (QTL) whose effects have been verified in the backcross populations associated with vine development, flowering and fruit quality based on 146 F$_2$ plants derived from a Zucchini x Scallop cross.

Legend: The linkage groups (LGs) have been ordered according to the results obtained in this paper. Group numbers in parenthesis (LGp) correspond to LGs in the map by Gong et al. (2008b). The correspondence between the two linkage groups has been determined according to the common SSR markers between maps (underlined). Markers with distorted segregation in F$_2$ are in italics. QTLs indicated in light grey, grey or black correspond to flowering, immature of mature fruit traits, respectively. QTLs are represented with bars (2-LOD interval) and boxes (1-LOD interval). QTLs for qualitative traits are represented with red lines spanning the region between flanking markers significantly associated with the trait at P< 0.05.
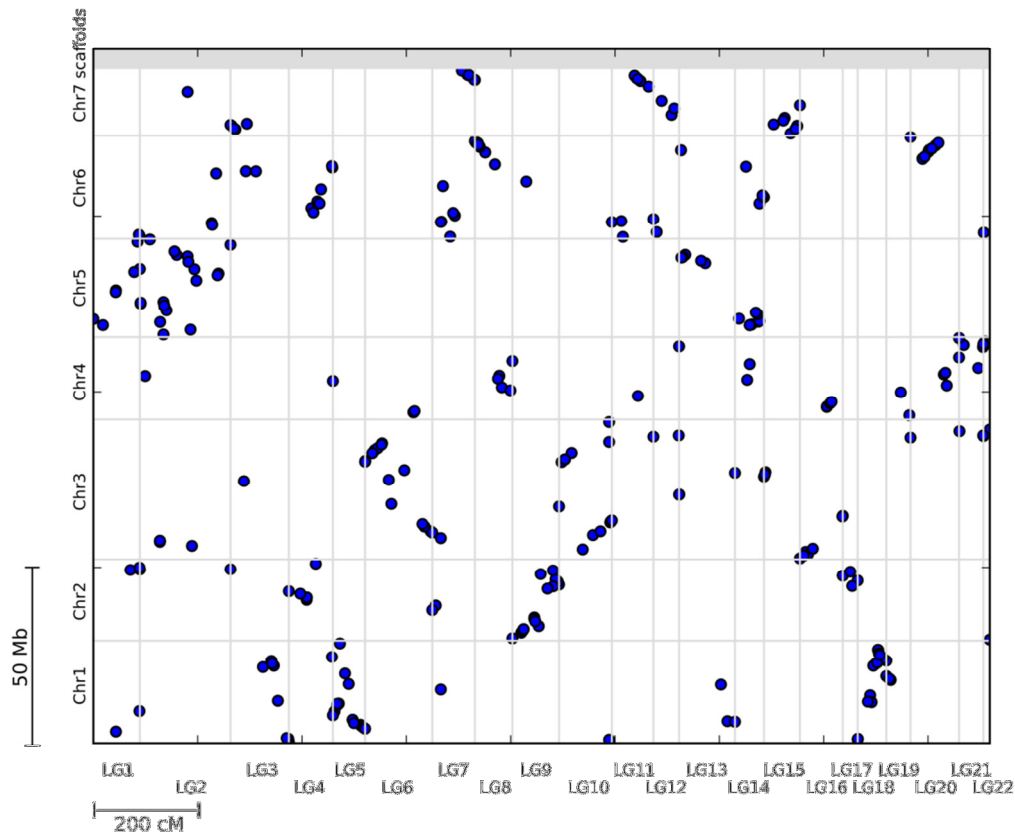
**Figure IV.2-7. Synteny with cucumber.**

Graphical representation of the tBLASTx hits found between the cucumber (Y-axis) chromosomes and the *Cucurbita* linkage groups (X-axis). Both the chromosomes and the linkage groups are numerically sorted in their respective axis and their limits are shown by grey lines. Only the tBLASTx hits with an e-value above $10^{-6}$ are shown in the figure.

**References**

Akhunov E, Nicolet C, Dvorak J: **Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay**. Theor Appl Genet 2009, 119(3):507-17.

Basten CJ, Weir BS, Zeng ZB: **QTL Cartographer, Version 2.5.** Department of Statistics, North Carolina State University, Raleigh, NC; 2005.

Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Picó B: **Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (*Cucurbitaceae*).** BMC Genomics 2011a, 12:104.

Blanca J, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó B: **Melon transcriptome characterization. SSRs and SNPs discovery for high throughput genotyping across the species.** The Plant Genome 2011b, 4(2):118-131.

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari MA, Collin F, Flowers JM, Pitrat M, Purugganan MD, Dogimont C, Bendahmane A: **A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons.** Science 2008, 321(5890):836-838.

Bradeen JM, Staub JE, Wye C, Antonise R, Peleman J: **Towards an expanded and integrated linkage map of cucumber (*Cucumis sativus* L.).** Genome 2001, 44:111-119.

Brown RN, Myers JR: **A genetic map of squash (*Cucurbita* ssp.) with randomly amplified polymorphic DNA markers and morphological markers.** J Am Soc Hortic Sci 2002, 127:568-575.

Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping.** Genetics 1994, 138:963-971.

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K,

Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R: **Development and implementation of high-throughput SNP genotyping in barley.** BMC Genomics 2009, 10:582.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK: **An introduction to markers, quantitative trait *loci* (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts**. Euphytica 2005, 142: 169-196.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for anotation, visualization and analysis in functional genomics research.** Bioinformatics 2005, 21(18):3674-3676.

Cucurbit Genomics Database of the International Cucurbit Genomics Initiative (ICuGI) [http://www.icugi.org].

Dahmani-Mardas F, Troadec C, Boualem A, Léveˆque S, Alsadon AA, Aldoss AA, Dogimont C, Bendahmane A: **Engineering melon plants with improved fruit shelf life using the TILLING approach.** PLoS ONE 2010, 5(12):e15776.

Deleu W, Esteras C, Roig C, González-To M, Fernández-Silva I,  Blanca J, Aranda MA, Arús P, Nuez F, Monforte AJ, Picó MB, Garcia-Mas J: **A set of EST-SNP for map saturation and cultivar identification in melon.** BMC Plant Biol 2009, 9:90.

Deschamps S, Campbell MA: **Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery.** Mol Breed 2010, 25:553-570.

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, Aubert G: **Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea.** Genomics 2010, 11:468.

Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub JE, Zalapa JE, Cuevas HE, Dace G, Oliver M, Boissot N, Dogimont C, Pitrat M, Hofstede R, Koert P, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Schaffer A, Katzir N, Xu Y, Zhang H, Fukino N, Matsumoto S, Garcia-Mas J,  Monforte AJ: **A consensus linkage map for molecular markers and Quantitative Trait *Loci* associated with economically important traits in melon (*Cucumis melo* L.).** BMC Plant Biol 2011, 11**:**111.

Doyle JJ, Doyle JL: **Isolation of plant DNA from fresh tissue.** Focus 1990, 12: 13-15.

Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM, Neale DB: **High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.).** Tree Genetics & Genomes 2009, 5:225-234.

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools.** In Genetics, Genomics and Breeding of Cucurbits. Edited by Behera TK, Wang Y, Kole C. New Hampshire: Science Publishers Inc, Enfield; 2012:140-198.

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS: **Highly parallel SNP genotyping.** Cold Spring Harb Symp Quant Biol 2003, 68:69-78.

Fei, S Guo, H Zhang, Y Ren, Y Zheng, Y Xu: **Genome and transcriptome sequencing of watermelon**. Plant & Animal Genomes XIX Conference: 15-19 January 2011; Town & Country Convention Center. San Diego. CA; 2011.

Ferriol M, Picó B, Fernández de Córdova P, Nuez F: **Molecular diversity of a germplasm collection of squash (*Cucurbita moschata*) determined by SRAP and AFLP markers.** Crop Sci 2004, 44:653-664.

Ferriol M, Pico B, Nuez F: **Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers.** Theor Appl Genet 2003, 107:271-282.

Ferriol M, Picó B: **Pumpkin and Winter Squash.** In Handbook of Plant Breeding. Vegetables I. Part 4. Volume 1. Edited by Prohens J, Nuez, F. Springer; 2008:317-349.

Gong L, Pachner M, Kalai K, Lelley T: **SSR-based genetic linkage map of *Cucurbita moschata* and its synteny with *Cucurbita pepo*.** Genome 2008a, 51:878-887.

Gong L, Stift G, Kofler R, Pachner M, Lelley T: **Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L.** Theor Appl Genet 2008b, 117:37-48.

González M, Xu M, Esteras C, Roig C, Monforte AJ, Troadec C, Pujol M, Nuez F, Bendahmane A, Garcia-Mas J, Picó B: **Towards a TILLING platform for functional genomics in Piel de Sapo melons.** BMC Research Notes 2011, 4:289.

González VM, Rodríguez-Moreno L, Centeno E, Benjak A, Garcia-Mas J, Puigdomènech P, Aranda MA: **Genome-wide BAC-end sequencing of *Cucumis melo* using two BAC libraries.** BMC Genomics 2010, 11:618.

Guo S, Zheng Y, Joung JG, Liu S, Zhang Z, Crasta OR, Sobral BW, Xu Y, Huang S, Fei Z: **Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types.** BMC Genomics 2010, 11:384.

Gupta PK, Rustgi S, Mirr RR: **Array-based high-throughput DNA markers for crop improvement.** Heredity 2008, 101(1):5-18.

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan , Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M,

Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S: **The genome of the cucumber, *Cucumis sativus* L.** Nat Genet 2009, 41:1275-1281.

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB: **High-throughput genotyping with the GoldenGate assay in the complex genome of soybean.** Theor Appl Genet 2008, 116(7):945-52.

International HapMap Consortium I: **The International HapMap Project.** Nature 2003, 426(6968):789-96.

Joehanes R, Nelson JC: **QGene 4.0, an extensible Java QTL-analysis platform.** Bioinformatics 2008, 24:2788-2789.

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, Raadsma H: **A genome wide survey of SNP variation reveals the genetic structure of sheep breeds.** PLoS ONE 2009 4(3):e4668.

Kosambi DD: **The estimation of map distances from recombination values.** Ann Eugen 1944, 12:172-175.

Lebeda A, Widrlechner MP, Staub JE, Ezura H, Zalapa JE, Krístková H: **Cucurbits (*Cucurbitaceae*; *Cucumis* spp., *Cucurbita* spp., *Citrullus* spp.).** In Genetic Resources, Chromosome Engineering and Crop Improvement.Vegetable Crops, Volume 3. Edited by Singh RJ. Taylor & Francis Group, Boca Raton, FL, USA; 2007:271-376.

Lee YH, Jeon HJ, Hong KH, Kim BD: **Use of random amplified polymorphic DNA for linkage group analysis in an interspecific cross hybrid F2 generation of *Cucurbita*.** J Kor Soc Hortic Sci 1995, 36:323-330.

Lepoittevin C, Frigerio J-M, Garnier-Géré P, Salin F, Cervera M-T, Vornam B, Harvengt L, Plomion C: *In vitro vs in silico* **detected SNPs for the development of a genotyping array: What can we learn from a non-model species?** PLoS ONE 2010, 5(6): e11034.

Levi A, Wechter P, Massey L, Carter L, Hopkins D: **An extended genetic linkage map for watermelon based on a testcross and a BC2F2 population**. American Journal of Plant Sciences 2011, 2:93-110.

Li D, Cuevas HE, Yang L, Li Y, Garcia-Mas J, Zalapa J, Staub JE, Luan F, Reddy U, He X, Gong Z, Weng Y. **Syntenic relationships between cucumber (*Cucumis sativus* L.) and melon (*C. melo* L.) chromosomes as revealed by comparative genetic mapping.** BMC Genomics 2011, 12**:**396.

Li Z, Huang S, Liu S, Pan J, Zhang Z, Tao Q, Shi Q, Jia Z, Zhang W, Chen H, Si L, Zhu L, Cai R: **Molecular isolation of the M gene suggests that a conserved-residue conversion induces the formation of bisexual flowers in cucumber plants.** Genetics 2009, 182:1381-1385.

Lin CH, Yeakley JM, McDaniel TK, Shen R: **Medium-to high-throughput SNP genotyping using VeraCode microbeads.** Methods Mol Biol 2009, 496:129-142.

Lincoln S, Daly M, Lander ES: **Constructing genetic maps with MAPMAKER/EXP 3.0: a tutorial and reference manual.** Whitehead Inst Biomed Res Tech Rpt, 3rd edition. Whitehead Institute for Biomedical Research, Cambridge; 1993, pp 97.

Malhi RS, Trask JS, Shattuck M, Johnson J, Chakraborty D, Kanthaswamy S, Ramakrishnan U, Smith DG: **Genotyping single nucleotide polymorphisms (SNPs) across species in Old World Monkeys.** Am J Primatol 2011, 73:1031-1040.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A**: A transposon-induced epigenetic change leads to sex determination in melon**. Nature 2009, 461:1135-1138.

Mascarell-Creus A, Cañizares J, Vilarrasa J, Mora-García S, Blanca J, González-Ibeas D, Saladié M, Roig C, Deleu W, Picó B, López-Bigas N, Aranda MA, Garcia-Mas J, Nuez F, Puigdomènech P, Caño-Delgado A: **An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon** *(Cucumis melo L.).* BMC Genomics 2009, 10:467.

Massman J, Cooper B, Horsley R, Neate S, Dill-Macky R, Chao S, Dong Y, Schwarz P, Muehlbauer GJ, Smith KP: **Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm.** Mol Breed 2011, 27:439-454.

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Neto ED, Gill CA, Gao C, Mannen H, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS: **An assessment of population structure in eight breeds of cattle using a whole genome SNP panel.** BMC Genet 2008, 9:37.

Metzker ML: **Sequencing technologies-the next generation.** Nat Rev Genet 2010, 11:31-46.

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ: **A consensus genetic map of cowpea [***Vigna unguiculata** **(L) Walp.] and synteny based on EST-derived SNPs.** Proc Natl Acad Sci USA 2009, 106(43):18159-64.

Paris H, Kabelka E: **Gene list for** *Cucurbita* **species.** Cucurbit Genet Coop Rep 2009, 31-32:44-69.

Paris H: **Summer Squash.** In Handbook of Plant Breeding. Vegetables I. Part 4. Volume 1. Edited by Prohens J, Nuez, F. Springer; 2008:351-381.

Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J: **Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce.** BMC Genomics 2008, 9:21.

Ren Y, Zhang Z, Liu J, Staub JE, Han Y, Cheng Z, Li X, Lu J, Miao H, Kang H, Xie B, Gu X, Wang X, Du Y, Jin W, Huang S: **An integrated genetic and cytogenetic map of the cucumber genome**. PLoS ONE 2009, 4(6):e5795.

Rostoks N, Ramsay L, Mackenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R: **Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties.** Proc Natl Acad Sci USA 2006, 103:18656-18661.

SISA: Simple Interactive Statistical Analysis. [http://www.quantitativeskills.com/sisa/statistics/ord2.htm]

Tani N, Takahashi T, Iwata H, Mukai Y, Ujino-Ihara T, Matsumoto A, Yoshimura K, Yoshimaru H, Murai M, Nagasaka K, Tsumura Y: **A consensus linkage map for sugi (*Cryptomeria japonica*) from two pedigrees, based on microsatellites and expressed sequence tags.** Genetics 2003, 165:1551-1568.

The Swiss-Prot Database [http://www.uniprot.org/downloads], uniprot_sprot_ release of 2010 04 23.

The TAIR Database: The Arabidopsis Information Resource [http://www.arabidopsis.org/], Tair_9_pep_ release 2009 06 19.

Tindall EA, Petersen DC, Nikolaysen S, Miller W, Schuster SC, Hayes VM: **Interpretation of custom designed Illumina genotype cluster plots for targeted association studies and next-generation sequence validation.** BMC Research Notes 2010, 3:39.

Uniref90 [http://www.ebi.ac.uk/uniref], uniref90_release 2010 04 23.

VassarStats: Website for statistical computation. [http://faculty.vassar.edu/lowry/ fisher3x3.html]

Voorrips RE: **MapChart: Software for the graphical presentation of linkage maps and QTL.** J Hered 2002, 93(1):77-78.

Wechter WP, Levi A, Harris KR, Davis AR, Fei Z, Katzir N, Giovannoni JJ, Salman-Minkov A, Hernandez A, Thimmapuram J, Tadmor Y, Portnoy V, Trebitsh T: **Gene expression in developing watermelon fruit.** BMC Genomics 2008, 9:275.

Weeden NF: **Isozyme studies indicate that the genus *Cucurbita* is an ancient tetraploid.** Cucurbit Genet Coop Rep 1984, 7:84-85.

Wright S: **Evolution and Genetics of Populations.** Vol I, Genetics and Biometric Foundations. The University of Chicago Press, Chicago 1968.

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J: **Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers.** PLoS ONE 2009, 4(12):e8451.

Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouc JH, Xu Y: **High-throughput SNP genotyping with the GoldenGate assay in maize.** Mol Breed 2010, 25:441-451.

Zamir D: **Improving plant breeding with exotic genetic libraries**. Nature Review Genetics 2001, 2:983-989.

Zraidi A, Lelley T: **Genetic map for pumpkin *Cucurbita pepo* L. using Random Amplified Polymorphic DNA markers.** In Progress in Cucurbit Genetics and Breeding Research, 8th International EUCARPIA Meeting on Cucurbit Genetics and Breeding, Edited byLebeda A, Paris H.Palacky University in Olomouc; 2004:507-514.

Zraidi A, Stift G, Pachner M, Shojaeiyan A, Gong L, Lelley T: **A consensus map for *Cucurbita pepo*.** Mol Breed 2007, 20:375-388.

# V. DISCUSIÓN GENERAL

## V. DISCUSIÓN GENERAL

Los trabajos presentados en este documento han supuesto un avance importante en cuanto a la identificación y desarrollo de marcadores moleculares de alta calidad, tipo SSR y SNP, así como en su aplicación para la construcción de mapas genéticos y estudios de diversidad, en dos especies cucurbitáceas de gran impacto económico como son *Cucumis melo* y *Cucurbita pepo*. Las estrategias empleadas para alcanzar el objetivo planteado al inicio de esta Tesis, el desarrollo de  herramientas genómicas para impulsar la mejora de estos cultivos, han ido cambiando a lo largo de estos años, adaptándose a la disponibilidad de nuevas tecnologías de secuenciación y genotipado de alto rendimiento.

### Melón

En las publicaciones presentadas se hace patente cómo el uso de las nuevas tecnologías de secuenciación aumenta la eficiencia en la detección de nuevos transcritos, involucrados en distintos procesos biológicos, y en la identificación de marcadores de calidad, facilitando su posterior aplicación para genotipado.

En primer lugar, cabe destacar el gran avance realizado en la caracterización del transcriptoma de melón. Con la secuenciación llevada a cabo mediante la tecnología 454 se ha incrementado en casi 690.000 el número de secuencias EST disponibles en melón (Blanca et al., 2011c), pues anteriormente sólo se disponía de unos 126.000 ESTs secuenciados mediante la tecnología de Sanger. La similitud en la distribución de unigenes implicados en los diferentes procesos biológicos y funciones moleculares entre melón y la especie modelo *A. thaliana*, indica que el transcriptoma aquí presentado supone una representación bastante completa de los genes de la especie. La identificación de transcritos no representados anteriormente queda patente en el aumento de genes implicados en ciertos procesos

329

biológicos y funciones moleculares, como los relacionados con la unión a ácidos nucleicos y proteínas, y factores de transcripción. Así pues, el empleo de la secuenciación masiva ha enriquecido de forma importante el transcriptoma de melón, herramienta clave para la anotación del genoma que se publicará próximamente (Garcia-Mas et al., 2012, enviado para publicación).

En cuanto a la generación de nuevos marcadores el avance ha sido claro. La primera colección masiva de SSRs y SNPs generada en 2007 requirió un importante esfuerzo y, a su vez, una cuantiosa inversión económica, de varios grupos de investigación para la generación de colecciones de secuencias (González-Ibeas et al., 2007). La secuenciación, mediante la tecnología Sanger, de 29.604 ESTs, ensamblados en 16.637 unigenes, permitió la identificación de 1.052 SSRs y 356 SNPs de alta calidad (localizados en unos 1.000 y casi 300 unigenes, respectivamente) (González-Ibeas et al., 2007). Posteriormente, en el trabajo aquí presentado de Deleu et al (2009) se validaron e identificaron algunos SNPs más mediante resecuenciación de casi 600 amplicones de la base de datos de ICuGI. Tres años más tarde, con la secuenciación del transcriptoma de la especie mediante tecnología 454 (689.054 ESTs, 53.252 unigenes), se identificaron 3.593 SSRs potenciales (en 3.298 unigenes) y más de 38.587 SNPs (en 14.417 unigenes) (Blanca et al., 2011c), lo que supone triplicar el número de SSRs identificados y multiplicar por 100 los SNPs respecto a la colección de 2007 (González-Ibeas et al., 2007). Por otra parte, y anterior a la publicación del transcriptoma de melón que se presenta en esta Tesis (Blanca et al., 2011c), el esfuerzo inicial en la generación de ESTs por Sanger continuó con una iniciativa internacional que permitió aumentar la colección existente con 93.756 ESTs adicionales que, ensamblados con los aproximadamente 35.000 previamente disponibles, dieron lugar a 24.444 unigenes, de los cuales 3.279 contenían motivos SSRs (4.068) y 1.331 contenían SNPs (3.073) (Clepet et al., 2011). Incluso con este

avance en el descubrimiento de marcadores, el aumento en más de 12 veces del número de SNPs identificados gracias al nuevo transcriptoma obtenido por pirosecuenciación marca una nueva era en la generación de colecciones de secuencias e identificación masiva de marcadores para la especie. Otras estrategias se han llevado a cabo en esta especie para el descubrimiento masivo de marcadores, como el método basado en hibridación de microarrays empleado por Ophir et al. (2010), con el que identificó 6.184 Single Feature Polymorphisms (SFPs), que mediante validación por secuenciación resultaron ser SNPs en su mayoría.

Los primeros SSRs y SNPs obtenidos a partir de las colecciones de ESTs generadas por Sanger fueron de enorme utilidad aplicados a estudios de variabilidad (Monforte et al., 2003; Fernández-Silva et al., 2008) y para la mejora del mapa genético de melón Piel de sapo T111 x Songwhan Charmi (Gonzalo et al., 2005; Fernández-Silva et al., 2008; Deleu et al., 2009). De los 431 SNPs validados en el trabajo presentado en esta Tesis (Deleu et al., 2009), 200 fueron mapados incrementando la densidad del mapa previo de Fernandez-Silva et al. (2008) (4,2 cM/marcador) hasta 2,35 cM/marcador mediante la estrategia de mapado selectivo o "bin-mapping". La mejora de los mapas de melón ha sido clave para avanzar en los estudios de ligamiento e identificación de QTLs (Diaz et al., 2011; Fernandez-Silva et al., 2010). La validación de estos marcadores fue algo costosa y los métodos de genotipado empleados (CAPS, dCAPS, SNaPshot, SCAR y pirosecuenciación) limitaron el tipo y número de marcadores a validar.

Respecto a la validación de los primeros SSRs, procedentes de secuenciación Sanger, Fernández-Silva et al. (2008) obtuvieron un porcentaje del 70% (88% respecto de los amplificados), mientras que un 93% (100% respecto de los amplificados) resultaron polimórficos de entre los validados de la colección procedente de secuenciación 454. Se hace evidente, por tanto, el aumento en la validación y en el número de SSRs identificados

mediante secuenciación masiva frente a la secuenciación convencional, principalmente por el mayor número de unigenes disponibles y la posibilidad de emplear más filtros de calidad para su selección.

En el caso de los SNPs procedentes de Sanger y de secuenciación masiva ocurre algo similar, habiéndose validado porcentajes inferiores, del 51,8% para pSNPs (SNPs putativos con al menos 2 lecturas por alelo) y 21,3% para pSCHs (SNPs putativos que no cumplen la condición anterior) en el trabajo de Deleu et al. (2009) frente a los obtenidos por Blanca et al. (2011c) (89% de polimorfismo validado por secuenciación y 98% de éstos validados *via* CAPS). Esto se debe no sólo al mayor número de lecturas obtenidas con secuenciación masiva, que implica un mayor número de unigenes y mayor cobertura, sino también a la posibilidad de aplicar criterios más exigentes de calidad en la selección de las secuencias que contienen un SNP. La exigencia de cobertura mínima fue menor en el caso de los SNPs de Sanger debido al menor número de secuencias, de hecho los pSCHs son SNPs detectados *in silico* en los que en uno o en los dos alelos sólo se dispone de una lectura, mientras que en los derivados de secuenciación masiva sólo se selecionaron SNPs con al menos dos lecturas por alelo para evitar falsos polimorfismos debidos a errores de secuenciación.

El gran número de marcadores SNPs disponibles tras la publicación del transcriptoma de melón (más de 38.000 SNPs) unido a la implementación de filtros de selección de los mismos por distintos criterios (filtros establecidos por el grupo de Bioinformática y Genómica del COMAV) favorecen la optimización de su uso para distintos objetivos. Los filtros establecidos permiten seleccionar no sólo por calidad de secuencia y cobertura, sino también descartar, si es necesario, aquéllos localizados en regiones no únicas del genoma o hipervariables, cerca de intrones o flanqueados por SNPs adicionales que puedan dificultar el genotipado (Blanca et al., 2011c). Este tipo de criterios son muy útiles, por ejemplo, para

desarrollar plataformas de genotipado masivo. Los filtros además, permiten seleccionar marcadores polimórficos entre ciertos morfotipos de interés, o marcadores adecuados para genotipado mediante CAPS. La implantación de este tipo de filtros va a ser cada vez más habitual debido al gran incremento en información de secuencia que está teniendo lugar en los últimos años. Aunque ya fuera del ámbito de esta Tesis doctoral, se ha establecido una plataforma de genotipado masivo con 768 de estos SNPs de melón, comprobándose la capacidad de estos marcadores para establecer relaciones genéticas entre diversas entradas (resultados coherentes con estudios previos, revisado en Esteras et al., 2012b) y para discriminar tipos muy similares, como las variedades comerciales de tipo Piel de Sapo. Esta plataforma también se ha empleado para incrementar la densidad del mapa existente que ha sido esencial para el anclaje de la secuencia genómica (Garcia–Mas et al., 2012).

Por tanto, el alto porcentaje de validación, tanto de los SSRs como de los SNPs identificados, dota de gran fiabilidad y robustez a estas colecciones de secuencias derivadas de proyectos de secuenciación masiva como fuente principal de marcadores funcionales para melón.

La mejora de las herramientas desarrolladas hasta el momento en melón pasa por ampliar el número de marcadores a partir de la resecuenciacion de un número mayor de genotipos representativos de la variabilidad de toda la especie. La mayoría de los marcadores identificados son de especial utilidad para la mejora de variedades comerciales de tipo cantalupo o *inodorus* con fuentes de variabilidad exótica del grupo *conomon*, cruzamientos empleados frecuentemente en la mejora de este cultivo, pero sería necesario generar nuevos marcadores con el fin de favorecer la transferencia de genes de interés entre otros grupos o morfotipos con características interesantes para la mejora (*agrestis* silvestres, *acidulus* y/o *momordica* de África o Asia), que, sin embargo, no aparecen representados en la secuenciación realizada. La experiencia adquirida y las herramientas desarrolladas hasta la

fecha por el grupo son factores clave para abordar con éxito esta nueva fase de resecuenciación.

**Calabacín**

En *C. pepo* todavía se hace más evidente el avance logrado gracias a las nuevas tecnologías de secuenciación, pues la escasa información de secuencia existente anteriormente impedía el desarrollo masivo de marcadores de naturaleza codominante como SSRs y SNPs. Con la publicación del primer transcriptoma de la especie (Blanca et al., 2011b), se identificaron *in silico* casi 1.935 SSRs en 1.822 unigenes y 19.980 SNPs y 1.174 INDELs en 8.147 unigenes. Los trabajos presentados en esta Tesis suponen la primera aplicación de estos SSRs y SNPs derivados de ESTs en la especie para estudios de variabilidad (Formisano et al., 2011) y para la construcción del primer mapa genético basado en SNPs (Esteras et al., 2012a).

Los EST-SSRs disponibles tras la publicación del transcriptoma fueron tan eficientes como los SSRs genómicos, desarrollados previamente a partir de genotecas enriquecidas por Gong et al. (2008), para establecer relaciones genéticas entre los tipos comerciales y razas locales analizados. La poca diversidad encontrada en los híbridos comerciales, la menor variabilidad del grupo Zucchini, de origen más reciente, y la clasificación de los diferentes genotipos evaluados, es consistente con lo esperado (Esteras et al., 2012b). La detección de variación fue similar por ambos tipos de marcadores en cuanto al grado de polimorfismo, PIC, número promedio de alelos y distancias entre genotipos. La ventaja de estos últimos reside fundamentalmente en su número y en su asociación directa con genes anotados. Así, Gong et al. (2008) detectaron 1.058 clones que contenían motivos SSRs, de los que finalmente validaron 405 (para *C. pepo* y *C. moschata*), mientras que la identificación derivada del proyecto de secuenciación masiva dio lugar a 1.935 marcadores. Recientemente Gong et al.

(2012) han empleado 134 de estos gSSRs para el estudio de las relaciones genéticas y evolutivas de 104 entradas, pertenecientes a las tres subespecies de *C. pepo*.

Del estudio de variabilidad presentado (Formisano et al., 2011) se desprende que este conjunto de microsatélites, desarrollados a partir del transcriptoma generado con dos genotipos, uno Zucchini (*C. pepo* ssp. *pepo*) y otro Scallop (*C. pepo* ssp. *ovifera*), puede ser transferido a un conjunto diverso de cultivares pertenecientes a distintos morfotipos de las dos subespecies. El hecho de que muchos de los EST-SSRs se encuentran en regiones ORF, más conservadas entre genotipos, aumenta la conservación de los cebadores utilizados en su amplificación y mejora la transferencia de los mismos. Un ejemplo de la dificultad de transferencia de los SSRs genómicos lo tenemos en la imposibilidad de transferir los SSRs mapados en el mapa de Gong et al. (2008) a nuestro mapa, en parte debido a la falta de conservación de las secuencias flanqueantes entre los distintos genotipos utilizados.

Los resultados obtenidos con el conjunto de híbridos comerciales y razas locales tiene utilidad para orientar la utilización de estos recursos en la mejora del calabacín. La existencia de bastantes alelos exclusivos de las variedades tradicionales españolas, y en especial de los tipos Zucchini, pone de manifiesto que nuestras variedades tradicionales no están tan íntimamente relacionadas con estos cultivares comerciales y que, por tanto, suponen un reservorio genético por explotar en los programas de mejora. De hecho, la caraterización fenotípica también señala a algunas de estas variedades como genotipos con alto potencial para la mejora por poseer características deseables para el mercado, como precocidad o tendencia femenina de la planta, a la vez que una mayor riqueza alélica que los tipos comerciales.

Sin embargo, a pesar de la gran utilidad de estos marcadores SSRs, la gran ventaja que supone la automatización de los SNPs y su abundancia, los convierte en los marcadores más requeridos para determinados objetivos. La construcción de una plataforma de genotipado

masivo basada en la tecnología GoldenGate es la primera aplicación en el género *Cucurbita* de este tipo de marcadores (Esteras et al., 2012a) y supone un gran impulso para la mejora de esta especie y otras cercanas del género *Cucurbita*. Este tipo de herramientas se están convirtiendo en imprescindibles y son ya habituales en otras especies (Hyten et al., 2008; Close et al., 2009; Deulvot el al., 2010), para la saturación de mapas, el desarrollo de líneas de introgresión y poblaciones de mapeo, y la caracterización molecular de tipos de interés y colecciones previamente evaluadas con marcadores menos informativos y reproducibles.

La construcción de un mapa entre dos tipos "summer squash", Zucchini x Scallop, con fenotipos contrastantes para muchos caracteres de interés, con 302 SNPs, supone un avance importante para la obtención de mapas genéticos saturados y, en un futuro, de mapas físicos. Con este mapa se han podido identificar por primera vez QTLs relacionados con la precocidad en la floración, la tendencia femenina de la planta, la morfología y coloración del fruto, su firmeza, así como ciertas características vegetativas. Algunos QTLs ya se han validado en las poblaciones de retrocruzamiento derivadas de los mismos parentales, pero además su consistencia en distintos fondos genéticos se ha visto confirmada por su identificación en una segunda población de mapado derivada de un cruce Pumpkin (*C. pepo* ssp. *pepo*) x Scallop (trabajo en fase de redacción). El grupo ha construido además una población de RILs (la primera disponible en la especie), derivada del cruce Zucchini x Scallop, que se empleará para incrementar la saturación de mapa y realizar un mapado de mayor precisión de los caracteres de interés comercial.

Las colecciones de marcadores, el mapa y los QTLs identificados suponen herramientas sin precedentes que impulsarán la mejora asistida por marcadores en esta especie y en las especies relacionadas *C. moschata* y *C. maxima,* en las que las herramientas moleculares son aún más escasas. El gran avance alcanzado en *C. pepo* debido a la caracterización de su

transcriptoma se hace evidente al comparar con los últimos trabajos de cartografiado en otras especies de *Cucurbita* como *C. maxima* para la que recientemente Singh et al. (2011) han publicado un mapa de ligamiento empleando RAPDs, marcadores dominantes poco reproducibles, en una población de retrocruzamiento *C. maxima* x *C. ecuadorensis*. Por otra parte, el desarrollo de marcadores SNPs y SSRs en *C. moschata* se verá impulsado próximamente con la resecuenciación del transcriptoma de *Cucurbita* que incluye entradas de esta especie (Blanca et al., 2011a) y que favorecerá estudios de sintenia más detallados.

**Referencias**

Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Esteras C, Picó B: **High throughput SNPs identification in *C. pepo* and *C. moschata* by Solexa transcriptome resequencing.** 9 Plant Genomics European Meetings (Plant Gem). Istambul (Turkey), 4 -7 mayo 2011a.

Blanca J, Cañizares J, Roig C, Ziarsolo P, Nuez F, Picó B: **Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (*Cucurbitaceae*).** BMC Genomics 2011b, 12:104.

Blanca J, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó B: **Melon transcriptome characterization. SSRs and SNPs discovery for high throughput genotyping across the species.** The Plant Genome 2011c, 4(2):118-131.

Clepet C, Joobeur T, Zheng Y, Jublot D, Huang MY, Truniger V, Boualem A, Hernandez-Gonzalez ME, Dolcet-Sanjuan R, Portnoy V, Mascarell-Creus A, Cano-Delgado AI, et al.: **Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon.** BMC Genomics 2011, 12:252, doi: 10.1186/1471-2164-12-252.

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R: **Development and implementation of high-throughput SNP genotyping in barley.** BMC Genomics 2009, 10:582.

Deleu W, Esteras C, Roig C, Gonzalez-To M, Fernandez-Silva I, Gonzalez- Ibeas D, Blanca J, Aranda MA, Arus P, Nuez F, Monforte AJ, Picó MB, Garcia- Mas J: **A set of EST-**

**SNPs for map saturation and cultivar identification in melon.** BMC Plant Biol 2009, 9:90.

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, Aubert G: **Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea.** Genomics 2010, 11:468.

Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub JE, Zalapa JE, Cuevas HE, Dace G, Oliver M, Boissot N, Dogimont C, Pitrat M, Hofstede R, Koert P, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Schaffer A, Katzir N, Xu Y, Zhang H, Fukino N, Matsumoto S, Garcia-Mas J, Monforte AJ: **A consensus linkage map for molecular markers and Quantitative Trait *Loci* associated with economically important traits in melon (*Cucumis melo* L.).** BMC Plant Biol 2011, 11:111.

Esteras C, Gómez P, Monforte AJ, Blanca J, Vicente-Dólera N, Roig C, Nuez F, Picó B: **High-throughput SNP genotyping in *Cucurbita pepo* for map construction and quantitative trait *loci* mapping.** BMC Genomics 2012a, 13:80, doi: 10.1186/1471-2164-13-80.

Esteras C, Nuez F, Picó B: **Genetic diversity studies in Cucurbits using molecular tools.** In Genetics, Genomics and Breeding of Cucurbits. Edited by Behera TK, Wang Y, Kole C. New Hampshire: Science Publishers Inc, Enfield; 2012b:140-198.

Fernandez-Silva I, Eduardo I, Blanca J, Esteras C, Pico B, Nuez F, Arus P, Garcia-Mas J, Monforte AJ: **Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.).** Theor Appl Genet 2008, 118:139-150.

Fernandez-Silva I, Moreno E, Essafi A, Fergany M, Garcia-Mas J, Martín-Hernandez AM, Álvarez JM, Monforte AJ: **Shaping melons: agronomic and genetic**

**characterization of QTLs that modify melon fruit morphology**. Theor Appl Genet 2010, 121:931–940, doi:10.1007/s00122-010-1361-2.

Formisano G, Roig C, Esteras C, Ercolano MR, Nuez F, Monforte AJ, Pico B: **Genetic diversity of Spanish *Cucurbita pepo* landraces: an unexploited resource for summer squash breeding**. Genet Resour Crop Evol 2011, doi:10.1007/s10722-011-9753-y.

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutiérrez S, Blanca J, Cañizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodríguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Melé M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Picó B, Gabaldón T, Roma G, Guigó R, Casacuberta JM, Arús P, Puigdomènech P: **The genome of melon (*Cucumis melo* L.). Genome amplification in the absence of recent duplication in an old widely cultivated species.** (en fase de publicación)

Gong L , Paris HS, Nee MH, Stift G, Pachner M, Vollmann J, Lelley T: **Genetic relationships and evolution in *Cucurbita pepo* (pumpkin, squash, gourd) as revealed by simple sequence repeat polymorphisms.** Theor Appl Genet 2012, 124(5):875-891, doi: 10.1007/s00122-011-1752-z.

Gong L, Stift G, Kofler R, Pachner M, Lelley T: **Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L.** Theor Appl Genet 2008, 117:37-48.

Gonzalez-Ibeas D, Blanca J, Roig C, González-To M, Picó B, Truniger V, Gómez P, Deleu W, Caño-Delgado A, Arús P, Nuez F, Garcia-Mas J, Puigdomènech P, Aranda MA:

**MELOGEN: an ESTdatabase for melon functional genomics.** BMC Genomics 2007, 8:306, doi:10.1186/1471-2164-8-306.

Gonzalo MJ, Oliver M, Garcia-Mas J, Monfort A, Dolcet-Sanjuan R, Katzir N, Arús P, Monforte AJ: **Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.).** Theor Appl Genet 2005, 110:802–811.

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, NelsonRL, Shoemaker RC, Young ND, Cregan PB: **High-throughput genotyping with the GoldenGate assay in the complex genome of soybean.** Theor Appl Genet 2008, 116(7):945-52.

Monforte AJ, Garcia-Mas J, Arús P: **Genetic variability in melon based on microsatellite variation.** Plant Breed 2003, 122:153–157.

Ophir R, Eshed R, Harel-Beja R, Tzuri G, Portnoy V, Burger Y, Uliel S, Katzir N, Sherman A: **High-throughput marker discovery in melon using a self-designed oligo microarray.** BMC Genomics 2010, 11:269

Singh AK, Singh R, Weeden NF, Robinson RW, Singh NK: **A linkage map for *Cucurbita maxima* based on Randomly Amplified Polymorphic DNA (RAPD) markers.** Indian J Hortic 2011, 68(1):44-50.

# VI.   CONCLUSIONES

# VI. CONCLUSIONES

La aplicación de nuevas tecnologías de secuenciación y genotipado en melón y calabacín ha supuesto la generación y uso de grandes colecciones de secuencias y de marcadores de alta calidad, SSRs y SNPs, ampliando significativamente la cantidad y calidad de los recursos previamente obtenidos mediante secuenciación por Sanger y genotipado convencional.

## Melón

1. La colección de SNPs desarrollada inicialmente a partir de los ESTs secuenciados por Sanger fue de utilidad, validándose 431 SNPs mediante resecuenciacion. Sin embargo, los porcentajes de validación de estos SNPs fueron muy bajos (21-52%), lo que contrasta con un porcentaje de validación muy superior (90%) en el caso de SNPs identificados mediante secuenciación masiva. Esto se debe principalmente al mayor número de secuencias generadas y la posibilidad de aplicación de criterios de selección más exigentes en estas últimas colecciones.

2. Aproximadamente la mitad de los SNPs validados (199) pudieron mapearse, empleando la población Piel de Sapo T111 x PI161375, y distintas técnicas manuales como CAPS, dCAPS, SNaPshot, SCAR y pirosecuenciación, permitiendo el desarrollo del primer mapa de melón basado en SNPs y contribuyendo significativamente al aumento de densidad del mapa (del mapa de "bins" previo con 296 marcadores y una densidad de 4,2 cM/marcador se incrementó a 528 marcadores y una densidad de 2,35 cM/marcador). Estos marcadores se han empleado posteriormente para la construcción del primer mapa consenso para la especie.

3. Se ha demostrado cómo las nuevas tecnologías de secuenciación (454 en este caso) han contribuido en gran medida a la mejora del transcriptoma de melón,

previamente basado en secuencias Sanger. La mejora en el incremento del número y tamaño de las lecturas entre los dos sistemas probados GS FLX y GS FLX Titanium, pone de manifiesto el rápido avance en estas técnicas. Se ha aumentado en más de 5 veces el número de secuencias ESTs disponibles y en más del doble el número de unigenes con respecto a la versión del transcriptoma anterior, pasándose de 24.444 a 53.252 unigenes.

4.     La detallada anotación provista en el transcriptoma de melón, clave para la anotación del genoma, y la amplia colección de marcadores de alta calidad, tipo SSR y SNP (3.593 y 38.587, respectivamente), así como las herramientas informáticas desarrolladas para su manejo y aplicación eficiente han resultado de gran utilidad, habiéndose validado un alto porcentaje de marcadores con diferentes criterios de mejora (polimórficos entre ciertos morfotipos de interés) y de aplicación práctica (genotipado mediante CAPS o plataformas de alto rendimiento).

**Calabacín**

5.     Los nuevos EST-SSRs, obtenidos a partir del primer transcriptoma de *C. pepo* generado por secuenciación masiva, combinados con un conjunto de gSSRs (previamente obtenidos a partir de genotecas enquiquecidas por Sanger), han resultado de utilidad para el estudio de la diversidad de una amplia colección de variedades tradicionales y cultivares e híbridos comerciales. Este estudio es el primero en aplicar marcadores codominates en un conjunto amplio de genotipos que incluyen tipos comerciales de esta especie. Las variedades tradicionales, portadoras de alelos exclusivos, se perfilan como potenciales fuentes de variabilidad en futuros programas de mejora, especialmente aquellas que han mostrado buenas características

comerciales. La eficacia de ambos sistemas de marcadores, gSSRs y EST-SSRs, para la detecccción de las relaciones genéticas entre las variedades estudiadas, ha resultado ser similar, aunque estos últimos presentan claras ventajas por su asociación directa a genes anotados (marcadores funcionales).

6.      La primera aplicación de una plataforma de genotipado masivo (GoldenGate) en el género *Cucurbita* ha presentado un alto nivel de éxito (90%), similar a ensayos llevados a cabo en otras especies como guisante, soja o maíz. Esto supone además la primera aplicación en la especie de marcadores tipo SNP y su uso para la construcción del primer mapa basado exclusivamente en marcadores de alta calidad (302 SNPs y 11 SSRs). El mapa Zucchini x Scallop presentado, con una cobertura de 1.741 cM y una densidad de 5.6 cM/marcador, localiza por primera vez diversos QTLs relacionados con la precocidad de la floración, tendencia femenina de la planta y características de calidad como la forma y coloración del fruto de gran interés para la mejora de esta especie.

# VII. ANEXOS

**ANEXO 1**

# Towards a TILLING platform for functional genomics in Piel de Sapo melons

González M, Xu M, Esteras C, Roig C, Monforte AJ, Troadec C, Pujol M, Nuez F, Bendahmane A, Garcia-Mas J, Picó B: **Towards a TILLING platform for functional genomics in Piel de Sapo melons.** BMC Research Notes 2011, 4:289

## Abstract

Background: The availability of genetic and genomic resources for melon has increased significantly, but functional genomics resources are still limited for this crop. TILLING is a powerful reverse genetics approach that can be utilized to generate novel mutations in candidate genes. A TILLING resource is available for *cantalupensis* melons, but not for *inodorus* melons, the other main commercial group.

Results: A new ethyl methanesulfonate-mutagenized (EMS) melon population was generated for the first time in an andromonoecious non-climacteric *inodorus* Piel de sapo genetic background. Diverse mutant phenotypes in seedlings, vines and fruits were observed, some of which were of possible commercial interest. The population was first screened for mutations in three target genes involved in disease resistance and fruit quality (Cm-PDS, Cm-eIF4E and Cm-eIFI(iso)4E). The same genes were also tilled in the available monoecious and climacteric *cantalupensis* EMS melon population. The overall mutation density in this first Piel de sapo TILLING platform was estimated to be 1 mutation/1.5 Mb by screening four additional genes (Cm-ACO1, Cm-NOR, Cm-DET1 and Cm-DHS). Thirty-three point mutations were found for the seven gene targets, six of which were predicted to have an impact on the function of the protein.

The genotype/phenotype correlation was demonstrated for a loss-of-function mutation in the Phytoene desaturase gene, which is involved in carotenoid biosynthesis.

Conclusions: The TILLING approach was successful at providing new mutations in the genetic background of Piel de sapo in most of the analyzed genes, even in genes for which natural variation is extremely low. This new resource will facilitate reverse genetics studies in non-climacteric melons, contributing materially to future genomic and breeding studies.

**ANEXO 2**

# Phenotyping a highly diverse core melon collection to be screened using Ecotilling.

Among the *Cucurbitaceae*, melon (*Cucumis melo*) is one of the most important cultivated cucurbits. Under the framework of the ERA-PG project "Understanding the climateric vs non-climateric fruit ripening mechanisms in melon using transcriptomic, metabolomic and reverse genetic approaches" (GEN2006-27773-C2-2-E) we have established an Ecotilling platform for melon. The choice of the genetic material is a key factor for the success of the approach. We have selected a collection of 212 accessions representing the variability of the species. This set includes wild relatives, landraces, and breeding lines from 54 countries in the world, having representatives from the two subspecies (*agrestis* and *melo*) and all reported botanical groups (*cantalupensis*, *reticulatus*, *inodorus*, *flexuosus*, *ameri*, *tibish*, *momordica*, *dudaim*, *acidulus*, *conomon*, *makuwa*, *chinensis* and *agrestis*). The full set of melon accessions has been genotyped with AFLP markers. Clustering analysis shows the relationships between the two subspecies, *agrestis* and *melo*, and among the different groups. All accessions were also extensively characterized for plant and fruit traits related to fruit quality. A high variation has been found in sexual systems (andromonoecious, monoecious, gynoecious, hermaphroditic), and fruit traits such as size, shape, rind and flesh colour, firmness, climacteric/nonclimacteric behaviour, fruit abscission, etc. Biochemical analysis of sugars, organic acids and volatiles has also been performed. Glucose,

fructose, sucrose and malic acid content were analyzed for the total collection, as well as the volatile compounds which constitute the rind and flesh aroma for a selected group. Few previous studies have assessed the rind aroma and this trait shows a large variability among accessions. The aroma profile can be as useful as other markers to distinguish and characterize melon varieties. Results support the accepted view of the complexity of the fruit ripening process in melons, including the existence of intermediate types and the involvement of many genes. This characterized EcoTILLING collection is being used for high-throughput analyses targeting genes involved in fruit quality and ripening such as the ethylene synthesis and signaling pathway genes. The deep characterization presented here will allow us to correlate phenotypes with mutations.

**ANEXO 3**

# Use of Ecotilling to identify natural allelic variants of melon candidate genes involved in fruit ripening

The development of an EcoTILLING platform to detect polymorphisms in natural diversity of *Cucumis melo* L. is one of the objectives of the project ´´Understanding the climacteric vs non-climacteric fruit ripening mechanisms in melon using transcriptomic, metabolomic and reverse genetic approaches´´ (GEN2006- 27773-C2-2-E) funded by the ERA-PG programme. After the first application in *Arabidopsis*, this reverse genetics approach has been used to target natural variation in a few genera: *Populus*, *Vigna*, *Solanum*, *Oryza*, and *Cucumis*. In *Cucumis* spp., allelic variation for disease resistance was screened. Melon is an adequate model plant for elucidating the ripening process due to the presence of climacteric and non-climacteric types within the species. Due to the importance of this trait in melon quality, the understanding of the underlying processes is nowadays a main objective in this species. It is known that climacteric fruits show a typical respiration burst followed by an autocatalytic ethylene synthesis. Ethylene plays an important role in this kind of fruits. However, much less is known about non-climacteric ones. We applied EcoTILLING to target melon candidates involved in fruit ripening, using a collection of 210 melon accessions, representing the variability of the species (different geographical origins, sexual systems, wild and cultivated, climacteric, non-climacteric, and intermediate types). To date, two genes involved in the ethylene pathway (CmACO1 and CmNOR) have been screened. We have found 15 mutations in

CmACO-1, 8 exonic, defining 7 haplotypes. From the 3 missense mutations discovered, two are likely to affect the protein function according to SIFT analysis. Ten mutations have been found in CmNOR, 3 in intronic and 7 in exonic regions, defining 6 haplotypes. Four of them imply amino acid change, although they are predicted to be tolerated. One mutation leads to the loss of an amino acid and is present in African melons of *agrestis* type. More studies are needed to elucidate the relationship among these mutations and the protein function. Using this platform, other genes are being analyzed related to this purpose using this platform.

**ANEXO 4**

# *Cucumis* SSRs Markers Applied to the Study of the Genetic Diversity in the *Cucurbita* genus

B. Picó, A. Sifres, C. Esteras, and F. Nuez
Instituto de Conservación y Mejora de la Agrodiversidad (COMAV). Department of Biotechnology. Polytechnic University of Valencia. Camino de Vera, 46022, Valencia, Spain.

The genus *Cucurbita* L. includes five cultivated species. *C. pepo* L.*, C. maxima* Duchesne, and *C. moschata* Duchesne are the three most economically important. A high level of morphological and agronomical variability exists within these species. Large collections of *Cucurbita* landraces with a high potential value for breeding are maintained elsewhere. The COMAV´s Genebank, at the Polytechnic University of Valencia, holds the largest collection of Spanish *Cucurbita* landraces. This collection also includes many accessions from the areas of origin and diversification of the different species (3).

Molecular tools, and particularly molecular markers, are very useful for the management of large germplasm collections. However, although new biotechnologies are being increasingly applied in *Cucurbita* breeding, it is still in the early stages when compared to other major Cucurbits. Only a few linkage maps with a low marker density have been reported. These maps are mainly based on dominant molecular markers, such as RAPDs, ISSRs or AFLPs. This type of markers has been mostly used to study the genetic variation in *Cucurbita* (3, 7).

SSRs are *locus*-specific markers assumed to be of high quality (codominant, highly informative and evenly dispersed throughout genomes). Identification of SSR markers can be attempt by the screening of genomic or cDNA libraries or by searching DNA sequence databases. None of these genomic tools are available for *Cucurbita* species. SSR markers have been largely reported in the *Cucumis* L. genus, also belonging to the Cucubitaceae family. Katzir *et al*. (5) demonstrate that primers specific to SSRs of melon and cucumber also amplified DNA from the *Cucurbita* genus. Seven (14%) of the 50 SSR *Cucumis* primers tested (1) were found to be functional and polymorphic among a collection of *C. pepo* accessions, but their utility in other species of *Cucurbita* was not tested (6, 7). Since these studies, the density of SSRs in the melon genetic map has been significantly increased (4), providing new SSRs potentially useful in *Cucurbita*.

We have studied the cross homology of 60 *Cucumis* SSRs in the *Cucurbita* genus. These markers were selected between those mapped (from 3 to 10 per chromosome) in the last version of the Spanish reference genetic map (Piel de sapo x PI 161375) (4). These included both, SSRs derived from genomic libraries (gSSRs) and derived from EST sequences (EST-SSRs). Some of the *loci* are conserved across the two main species of the *Cucumis* genus. Accessions representative of the phenotypic diversity of each *Cucurbita* species (8, 12 and 15 for *C. maxima, C. moschata* and *C. pepo,* respectively) were used in the assays. This sample included Spanish landraces and some landraces from South and Central America. In the case of *C. pepo,* the eight edible cultivars groups were represented.

A high percentage (63.3%) of the SSRs assayed did not amplified in any species, even varying the PCR conditions described for melon. Most of the amplified SSRs were monomorphic, but a small sub-set was polymorphic in one, two, or the three species. Of the 20 EST-SSRs assayed, 2 amplified in *C. pepo* (TJ2 and TJ31), 1 in *C. maxima* (TJ24), and 2 in the three species (CSWTA02 and CSTCC813), but only CSTCC813 was polymorphic in *C. moschata* and *C. pepo* (Table 1). Of the 40 gSSRs assayed, 28 were developed by Danin-Poleg *et al.* (1) and the remaining 12 were among the newly described in Gonzalo *et al.* (4). Of the first subset, 1 amplified in *C. pepo* (CSWCT01), 1 amplified in *C. moschata* and *C. maxima* (CMGA165), being both monomorphic, and 7 amplified in the 3 species. CMGA128 and CMGA15 were monomorphic and the remaining five (CMTC51, CMTC123, CMAG59, CMGA172, CMTC168) were polymorphic within and between species (Table 1, Figure 1). The *loci* CSGA057, CMCT160a and CSTTT15a, previously reported as polymorphic in *C. pepo* (6), did not amplified in our assay. Of the second subset, 1 amplified in *C. pepo* and *C. maxima* (CMTCN56), being polymorphic between species, and 3 amplified in *C. pepo* and *C. moschata* (CMAGN75, monomorphic between and within species, and CMGAN21 and CMTAAN100, polymorphic between species). Three *loci* amplified in the 3 species, but only two were polymorphic within and between species (CMTGN17 and CMAGN73).

Results agree with previous studies in Cucurbitaceae and other families, indicating limited conservation of SSRs *loci* between *Cucumis* and *Cucurbita*. Despite this limited success, the selected SSRs markers were informative and allow us to reach some conclusion about the analysed germplasm. In *C. maxima* a similar level of variation was detected among the Spanish landraces and the South American accessions. However, accessions of *C. moschata* from South and Central America displayed a higher level of variation than Spanish landraces. *C. pepo* was the most variable species and alleles unique of subspecies were detected.

The density of EST-SSR markers in the melon genetic map is currently being increased using a large collection of ESTs obtained in the Spanish Melon Genome Initiative (2). The usefulness of these EST-SSRs for genetic studies of *Cucurbita* is being now investigated.

**Literature cited:**

1. Danin-Poleg, Y., N. Reis, G. Tzuri, and N. Katzir. 2001. Development and characterization of microsatellite markers in *Cucumis*. Theoretical and Applied Genetics 102:61-72.
2. Fernandez-Silva, I., J. Blanca, B. Picó, F. Nuez, P. Arus, and A.J. Monforte. 2006. Mapeo de marcadores microsatélites a paritr de librerías de ESTs de melón (*Cucumis melo*) mediante genotipado selectivo. Actas de Horticultura 54:109-110.
3. Ferriol, M., B. Picó, and F. Nuez. 2003. Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. Theoretical and Applied Genetics 107:271-282.

4.    Gonzalo, M.J., M. Oliver, J. García-Mas, A. Monfort, A. Dolcet-Sanjuan, N. Katzir, P. Arus, and A.J. Monforte. 2005. Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). Theoretical and Applied Genetics 110:802-811.

5.    Katzir, N., Y. Danin-Poleg, G. Tzuri, Z. Karchi, U. Lavi, and P.B. Cregan. 1996. Length polymorphism and homologies of microsatellites in several Cucurbitaceae species. Theoretical and Applied Genetics 93:1282-1290.

6.    Katzir, N, Y. Tadmor, G. Tzuri, E. Leshzeshen, N. Mozes-Daube, Y. Danin-Poleg, and H.S. Paris. 2000. Further ISSR and preliminary analysis of the relationships among accessions of *Cucurbita pepo*. Acta Horticulturae 510:433-439.

7.    Paris, H.S., N. Yonash, V. Portnoy, N. Mozes-Daube, G. Tzuri, and N. Katzir. 2003. Assesment of genetic relationships in *Cucurbita pepo* (Cucurbitaceae) using DNA markers. Theoretical and Applied Genetics 106:971-978.

Table 1. SSR *loci*, derived from two *Cucumis* species, which detected polymorphism between or within species in a collection of *Cucurbita* germplasm.

| SSR *locus* | WS [z] | | | LG [y] | BS [x] | Origin | Cm/Cs [w] | Ref. |
|---|---|---|---|---|---|---|---|---|
| | *C. maxima* | *C. moschata* | *C. pepo* | | | | | |
| CSTCC813 | 1/nt | 2/nt | 2/ni | G12 | NO | Cucumber cDNA library | ns/2 | 1,7,4 |
| CMTC51 | 3/nt | 4/nt | 2/1-null | G4 | YES | Non-coding región of EMBL CMACO1 | 1/2 | 1,6,7,4 |
| CMTC123 | 3/nt | 1/nt | 3/nt | G12 | YES | Melon genomic library | 2/1 | 1,4 |
| CMAG59 | 1/1 | 1-null/1 | 2-null/3 | G1 | YES | Melon genomic library | 5/ns | 5,6,7,4 |
| CMGA172 | 1/nt | 2/nt | 1/nt | G9 | NO | Melon genomic library | 3/2 | 1,4 |
| CMTC168 | 1/nt | 1/nt | 1-null/nt | G10 | YES | Melon genomic library | 4/ns | 1,4 |
| CMTGN17 | 1/ni | 3/ni | 2-null/ni | nm | YES | Melon genomic library | ni | 7 |
| CMAGN73 | 1/ni | 1/ni | 3/ni | G10 | YES | Melon genomic library | ni | 6,4 |

[z] Variability within species, the first number indicates the number of alleles identified in our assay (null, null allele), and the second number the alleles previously reported (nt: not previously tested, ni: previously reported as polymorphic, but without information about the number of alleles)

[y] Linkage group according to Gonzalo *et al.* (4), nm: no mapped

[x] Variability between species

[w] Number of alleles reported in *Cucumis melo* and *Cucumis sativus*, ns: no signal, ni: no information
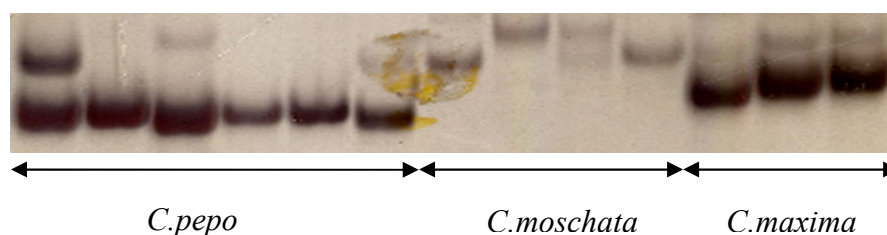


Figure 1. Example of the SSR *locus* CMTC51 that detects different alleles in *C. maxima*, *C. moschata* and *C. pepo*.

**ANEXO 5**

# Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.)

**Abstract**

We report the development of 158 primer pairs thanking SSR motifs in genomic (gSSR) and EST (EST-SSR) melon sequences, all yielding polymorphic bands in melon germplasm, except one that was polymorphic only in *Cucurbita* species. A similar polymorphism level was found among EST-SSRs and gSSRs, between dimeric and trimeric EST-SSRs, and between EST-SSRs placed in the open reading frame or any of the 5´- or 3´-untranslated regions. Correlation between SSR length and polymorphism was only found for dinucleotide EST-SSRs located within the untranslated regions, but not for trinucleotide EST-SSRs. Transferability of EST-SSRs to *Cucurbita* species was assayed and 12.7% of the primer pairs amplified at least in one species, although only 5.4% were polymorphic. A set of 14 double haploid lines from the cross between the cultivar "Piel de sapo" and the accession PI161375 were selected for the bin mapping approach in melon. One hundred and twenty-one SSR markers were newly mapped. The position of 46 SSR *loci* was also verified by genotyping the complete population. A final bin-map was constructed including 80 RFLPs, 212 SSRs, 3 SNPs and the Nsv *locus*, distributed in 122 bins with an average bin length of 10.2 cM and a maximum bin length of 33 cM. Map density was 4.2 cM/marker or 5.9 cM/SSR.

**ANEXO 6**

# Variabilidad de *Cucurbita maxima* en su zona de origen: un recurso de interés para la mejora de esta hortaliza

## Resumen

Se ha llevado a cabo una caracterización molecular, empleando marcadores AFLP, de la diversidad genética de un conjunto de 38 entradas sudamericanas de *C. maxima*, estudiando su relación con el ancestro silvestre. Los resultados indican la existencia de distintos grados de diversidad genética en el germoplasma de esta especie recolectado en distintos países de su zona de origen y diversidad. Las entradas procedentes de Argentina son las más variables y las más parecidas al ancestro silvestre, lo que apunta a su primitivismo y/o a la ocurrencia de cruzamientos con poblaciones de la subespecie *andreana*, con las que coexiste en la actualidad. Estos resultados apoyan las suposiciones previas que señalan a Argentina como centro de origen de la especie y refuerzan la prioridad de colectar y conservar las entradas de este país. Sin embargo, encontramos discrepancias con referencias previas que indican una elevada diversidad genética en Bolivia y Perú. En estos países se observa una reducida diversidad, que puede ser consecuencia de la ocurrencia de un proceso de erosión genética intenso en los últimos años. Por otro lado, nuestros resultados muestran que en Ecuador todavía existe una diversidad importante de tipos, no descrita con anterioridad, que es recomendable conservar antes de que se pierdan.

**ANEXO 7**

# High throughput SNPs identification in *C. pepo* and *C. moschata* by Solexa transcriptome resequencing

The *Cucurbita* genus, of American origin, is one of the most variable within the *Cucurbitaceae* family. *C. pepo* is the main economically important species. This species include many morphotypes, used as winter or summer squashes, grouped into two subspecies. Recently, we have obtained the first transcriptome of *C. pepo* by 454 sequencing. The assembly of approximately 500.000 ESTs from multi-tissue libraries of the two main commercial genotypes, Zucchini (ssp. *pepo*) and Scallop (ssp. *ovifera*) yielded 49.610 unigenes, and allowed the identification of a large set of SNPs, useful for detecting variation between both subspecies. Illumina Solexa technology has been now applied for sequencing a cDNA collection from a pool of 13 genotypes, representing the variability within ssp. *pepo*. We have also resequenced cDNA from a pool of genotypes of the related species *C. moschata*, the most rustic among the cultivated *Cucurbita*. It is a reservoir of resistance and quality genes and acts as a genetic bridge to transfer favorable traits from wild *Cucurbita* into *C. pepo*. A set of 20 genotypes representing the wide variability found in this species has been used. After cleaning the raw reads, we have analyzed 25,557,203 and 27,250,208 sequences, with an average length of 37 bp, for the pool of *C. pepo* and *C. moschata* respectively (a total sequence length of 1,9 Gb).The processed reads were mapped against the *C. pepo*

transcriptome for SNPs mining. A total of 532,811 variants were identified in 37,275 unigenes (483,486 SNPs and 49,325 INDELs). Different filters have been applied to select those variants polymorphic within each pool, and those monomorphic in each pool, but polymorphic between both subspecies. These new set of SNPs will be suitable for high throughput genotyping in genetic diversity studies and in breeding programs conducted using different varieties of the ssp. *pepo* or introgressing traits from *C. moschata* into *C. pepo*.