Organización de la diversidad genética de los cítricos

Andrés García Lor, Julio 2013

Supervisores: Luis Navarro Lucas, Patrick Ollitrault

Tutor: Ismael Rodrigo Bravo

ivia
instituto valenciano de investigaciones agrarias

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departamento de Biotecnología

Organización de la diversidad genética de los cítricos

Tesis doctoral

Presentada por:

Andrés García Lor

Dirigida por:

Dr. Patrick Ollitrault

Dr. Luis Navarro Lucas

VALENCIA, JULIO 2013

AGRADECIMIENTOS

Bueno, pues parece que ya llegó el final de mi tesis, y ante mí, una nueva página en blanco, los agradecimientos, y por dónde empiezo? y a quién pongo? Tantas cosas han pasado y tanta gente en el camino que es complicado resumirlo. Espero no dejarme nada ni a nadie.

Todo esto empezó aquel día (19/01/2004) en el que emprendí mi aventura en Wageningen (pueblo universitario en Holanda, para los despistados), a donde fui a hacer mi trabajo final de carrera. Es allí donde comencé a sentir el gusanillo de la ciencia y donde se empezó a forjar una vocación que hasta día de hoy sigue en pie. Por aquellas tierras viví muchísimas experiencias y conocí a grandes amigos (inolvidable). Al año siguiente continué con mi vena científica y el espíritu viajero aterrizando en Redhill (pueblo pequeñito al sur de Inglaterra) donde estuve seis meses en los laboratorios de una brewery (centro de investigación de cerveza, "mola") donde también hice muy buenos amigos y aprendí muchas cosas. Además, casualidades de la vida, conocí a la que actualmente es mi novia, Ana.

De vuelta a la terreta mis dos años en el IBMCP. En la universidad empecé mi idilio con los cítricos y compartí laboratorio con grandes compañeros. Además me llevé la amistad de mucha gente, los "ibmceperos".

Después de esta aventura llegué al IVIA. Lo primero, y para no perder las costumbres, quiero agradecer la oportunidad que me brindó Luis Navarro al concederme la beca para poder realizar la tesis doctoral bajo su dirección, y por supuesto, quiero hacer una mención especial a Patrick Ollitrault, que además de codirector de mi tesis, le consideró como mi padre científico. Sin él esta tesis no hubiera sido la misma, muchas gracias por todo. Muchas gracias a mis evaluadores externos y los miembros del tribunal. También quiero dar las gracias a François Luro, quien me acogió en Córcega y me ayudó como a uno más de la gran familia francesa del INRA en San Giuliano (Yann, Gilles, Paul-Eric, Isabelle… et tous les stagiaires). Allí conocí también al que además de un muy buen amigo, es actualmente compañero en el IVIA, Franck. Me gustaría dar las gracias a mis colegas franceses de Montpellier, donde estuve dos semanas, que me enseñaron algunas de las herramientas utilizadas en la tesis.

Durante los últimos cinco años que he pasado en el IVIA, he vivido mil y una experiencias, almuerzos, cenas, paellas de San Isidro, fiestas, alquerías, conversaciones de pasillo, cafés, mudanza de laboratorio (pasamos a denominarnos Francia),… que he compartido con lo que para mí es la gran familia del IVIA, empezando por mis compis de laboratorio que ya no están, Mari Cruz, Rosa, María, Regina, y los que siguen, José (un crack), Franck, Frédérique, Marta, Juan, Gema, Houssem (gracias a todos y cada uno de vosotros por haberme ayudado en algún momento de esta tesis, un pedacito de ella es vuestra). También quiero recordar a los que ahora son "ex-ivias", Esther, Jorge, Giovanni, Inma, Álida, Marta, Águeda, Lucia…. y por supuesto a los que siguen, Jesús, Pablo, Vero, "las leandras" (Ana, Elsa, Nuria, Berta, Montse), Diana, Ezequiel, el equipo de cultivo (Pablo, Toni, Pepe, Vio, Carmen, Juana Mari, Ana, Cloti, Marga,…..). También quiero recordar a aquellos que han tenido un paso más o menos largo por el IVIA (Caroline, Jean Baptiste, Hager…) y al personal del IVIA con el que en un momento u otro también he compartido algunos momentos (investigadores, vigilantes, administración, limpieza). También mencionar a mis amigos de siempre: bankitos, FTK, agrónomos, biotecnólogos, equipo de fútbol... que alguna vez me han preguntado: ¿Qué es lo que haces? ¿Para qué sirve?...

Por último, dedicarle esta tesis a mi familia, mis padres y mi hermana, siempre han estado y estarán ahí. Y por supuesto a mi cítrico favorito, mi media naranja, Ana.

RESUMEN

Citrus es el género de la subfamilia Aurantioideae de mayor importancia económica. Su origen es la región sureste de Asia, en un área que incluye China, India y la península de Indochina y los archipiélagos de los alrededores. Aunque se han realizado múltiples estudios, la taxonomía del género *Citrus* aun no está bien definida, debido al alto nivel de diversidad morfológica encontrado en este grupo, la compatibilidad sexual entre sus especies y la apomixis de muchos genotipos. En la presente tesis doctoral se ha estudiado una amplia diversidad del género *Citrus*, especies relacionadas y otros taxones de la subfamilia Aurantioideae, para poder aclarar su organización y filogenia mediante el empleo de diferentes tipos de marcadores moleculares y métodos de genotipado. Más concretamente, el germoplasma de mandarino juega un papel muy importante en la mejora de variedades y patrones, pero su organización genética no está bien definida. Por lo tanto, se ha realizado un análisis en profundidad de su diversidad y organización genética.

El desarrollo de marcadores moleculares de Inserción-Deleción (indel), por primera vez en cítricos, ha permitido demostrar su utilidad para estudios de diversidad y filogenia en el género *Citrus*. En combinación con los marcadores de tipo microsatélite (SSR), se ha cuantificado la contribución de los tres principales taxones de cítricos (*C. reticulata*, *C. maxima* and *C. medica*) a los genomas de las especies secundarias y cultivares modernos. También se ha definido su estructura genética a partir de los datos obtenidos en la secuenciación de 27 fragmentos de genes nucleares relacionados con la biosíntesis de compuestos que determinan la calidad de los cítricos y genes relacionados con la respuesta de la planta a estreses abióticos. El análisis de la filogenia nuclear ha permitido determinar la relación existente entre la especie *C. reticulata* y *Fortunella*, que se diferencian claramente del grupo formado por las otras dos principales especies de cítricos (*C. maxima* y *C. medica*). Este resultado está en concordancia con el origen geográfico de las especies estudiadas. A partir de este estudio, se han desarrollado marcadores moleculares de tipo SNP con un alto valor filogenético, que han sido transferidos a géneros relacionados de los cítricos. Estos marcadores han dado un resultado muy positivo en el género *Citrus* y serán de gran utilidad para el establecimiento de la huella genética del germoplasma en un nivel de diversidad más amplio.

Se ha estudiado la organización genética dentro del germoplasma mandarino (198 genotipos de tipo mandarino pertenecientes a dos colecciones, INRA-CIRAD e IVIA), así como la introgresión de otros genomas mediante el uso de 50 y 24 marcadores de tipo SSR y indel, respectivamente, además de cuatro marcadores indel mitocondrial (ADNmt). Se ha observado que muchos genotipos, que se creía que eran mandarinos puros, presentan introgresión de otros genomas ancestrales. Dentro del germoplasma de mandarino, se han identificado a nivel nuclear cinco grupos parentales, a partir de los cuales se originaron muchos genotipos, dando lugar a estructuras hibridas complejas. Se ha observado incluso, genotipos con un origen maternal no mandarino, determinado por los marcadores de ADNmt.

La presente tesis doctoral ha aportado nueva información sobre las relaciones filogenéticas entre las especies del género *Citrus*, géneros cercanos, así como de las especies secundarias. Además, se han desarrollado nuevos marcadores moleculares que se complementan entre sí. Se ha establecido una nueva organización genética del germoplasma mandarino y se han caracterizado adecuadamente las dos colecciones de cítricos en estudio. Por lo tanto, todas estas contribuciones, ayudarán a los programas de mejora para la obtención de nuevas variedades de cítricos de alta calidad y permitirán optimizar la conservación y uso de los recursos genéticos existentes, así como su caracterización genética y fenotípica.

RESUM

El gènere *Citrus* és sens dubte el més important de la subfamília Aurantioideae a nivell econòmic. Es creu que s'originà en la regió del sud-est d'Àsia, en una àrea que inclou la Xina, l'Índia, la península d'Indoxina i els arxipèlags dels voltants. Malgrat que s'hagen fet molts estudis, la taxonomia dels cítrics és encara controvertida degut a la gran diversitat morfològica que hi ha dins d'aquest grup, la compatibilitat sexual entre espècies i l'apomixi de molts genotipus. En aquesta tesi doctoral s'ha estudiat una àmplia diversitat dins del gènere *Citrus*, de relatius dels cítrics i d'altres taxa de la subfamília Aurantioideae, per tal d'aclarir la seua organització i filogènia mitjançant la utilització de diferents tipus de marcadors moleculars i vàries plataformes de genotipat. A més a més, el germoplasma de les mandarines juga un paper molt important en la millora genètica de portaempelts i cultivars, però la seua organització genètica no es encara prou coneguda. Per tant, s'ha analitzat a fons la seua diversitat.

Per primera volta en els cítrics, s'han desenvolupat marcadors nuclears "Insertion-Deletion" (indel), que han permès demostrar la seua utilitat per fer estudis de diversitat i filogènia dins del gènere *Citrus*. En combinació amb marcadors del tipus SSR, s'ha pogut quantificar la contribució de les tres taxa més importants dels cítrics (*C. reticulata*, *C. maxima* i *C. medica*) als genomes de espècies secundàries i de cultivars moderns. També s'ha determinat la seua estructura genètica mitjançant les dades obtingudes en seqüenciar 27 gens nuclears responsables de la biosíntesi de compostos relacionats amb la qualitat dels cítrics i de gens involucrats en la resposta a estrès de les plantes. Les anàlisis filogenètiques nuclears han mostrat que *C. reticulata* i *Fortunella* formen una clada clarament diferenciada d'una altra que inclou dos altres taxa bàsics de cítrics cultivats (*C. maxima* i *C. medica*), cosa que està d'acord amb l'origen geogràfic de les espècies que s'han estudiat. Aquest estudi ens ha permès desenvolupar marcadors moleculars de tipus SNP que tenen un gran valor filogenètic i analitzar la seua transferibilitat a altres gèneres relacionats genèticament. Aquests funcionen molt bé dins del gènere *Citrus* i seran molt útils per al "fingerprinting" de germoplasma a un nivell de diversitat molt més ampli.

S'ha estudiat la organització genètica del germoplasma de les mandarines (198 genotipus 'mandarin-like' de dues col·leccions de germoplasma, INRA-CIRAD i IVIA), i la seua introgressió per altres taxa per mitjà de 50 i 24 SSRs i de marcadors indels nuclears respectivament, i quatre indels mitocondrials (ADNmt). S'ha vist que molts genotipus, que es creia que eren mandarines pures, contenen en els seus genomes introgressió d'altres taxa bàsiques. S'han establert cinc grups parentals en el germoplasma analitzat. A més, molts genotipus es deuen haver originat de creuaments entre aquestes mandarines, cosa que ha donat lloc a una estructura híbrida molt complexa. I a més, segons s'ha establert per mitjà de marcadors ADNmt, alguns genotipus de mandarina tenen un origen matern que no és de tipus mandarina.

Aquesta tesi doctoral ha proporcionat nova informació sobre les relacions filogenètiques dels taxa dins del gènere *Citrus* i d'espècies relacionades, com també d'espècies secundaries de tipus comercial. S'han desenvolupat nous grups de marcadors complementaris. S'ha establert l'organització genètica del germoplasma de les mandarines així com una adequada caracterització de dues col·leccions de germoplasma de cítrics. Per tant, aquestes contribucions ajudaran a la millora genètica de nous cultivars de cítrics de qualitat i contribuirà a optimitzar la conservació i caracterització dels recursos genètics de cítrics tant a nivell genètic com fenotípic.

ABSTRACT

Citrus is by far the most economically important genus of the subfamily Aurantioideae. It is believed to have originated in the south-eastern region of Asia, in an area that includes China, India and the Indochinese peninsula and nearby archipelagos. Although many different studies have been done, *Citrus* taxonomy is still controversial due to the large degree of morphological diversity found within this group, the sexual compatibility between the species and the apomixis of many genotypes. In this PhD thesis a broad diversity within the *Citrus* genus, citrus relatives and other taxa from the Aurantioideae subfamily has been studied in order to clarify their organization and phylogeny using different types of molecular markers and different genotyping platforms. The mandarin germplasm plays a major role in citrus rootstock and cultivar breeding, but its genetic organization is still largely unknown. Therefore, an analysis in depth of diversity and organization has been done.

The development of nuclear Insertion-Deletion (indel) markers, for the first time in citrus, has allowed us to demonstrate its utility for diversity and phylogenetic studies in the genus *Citrus*. In combination with SSR markers, the contribution of three basic edible taxa (*C. reticulata*, *C. maxima* and *C. medica*) to the genomes of secondary species and modern cultivars has been quantified. Their mosaic genetic structure has also been determined from the data obtained by sequencing 27 nuclear genes involved in the biosynthesis of quality compounds of citrus and genes involved in plant stress response. Nuclear phylogenetic analysis revealed that *C. reticulata* and *Fortunella* form a clade that is clearly differentiated from the clade that includes two other basic taxa of cultivated citrus (*C. maxima* and *C. medica*), which is consistent with the geographic origin of the species studied. From this study, SNP molecular markers with a high phylogenetic value has been developed and tested for transferability into genetically related genera. They performed very well within the *Citrus* genus and should be useful for germplasm fingerprinting at a much broader diversity level.

The genetic organization within the mandarin germplasm (198 ‘mandarin-like’ genotypes from two germplasm collections, INRA-CIRAD and IVIA), and its introgression by other taxa was studied with 50 and 24 nuclear SSRs and indel markers respectively, and four mitochondrial (mtDNA) indels. It has been shown that many genotypes, believed to be pure mandarins, have introgression from other basic taxa in their genomes. Five parental groups have been determined within the germplasm analysed. Moreover, many genotypes have been originated from the cross between these mandarins, leading to a very complex hybrid structures. Furthermore, some mandarin genotypes have a non-mandarin maternal origin as determined by mtDNA markers.

This PhD thesis has released new information about the phylogenetic relationships of taxa within the *Citrus* genus and relative species, as well as secondary commercial species. New sets of complementary markers had been developed. The genetic organization of the mandarin germplasm was revealed and a proper characterization of two citrus germplasm collections was obtained. These contributions will help in the breeding of new, high-quality citrus cultivars and will contribute to optimizing the conservation and characterization at genetic and phenotypic levels of the citrus genetic resources.

RÉSUMÉ

Le genre *Citrus* est de loin le genre le plus important du point de vue économique de la sous-famille des Aurantioideae (Famille des Rutacées). Les *Citrus* seraient originaires du Sud-Est asiatique, dans une zone comprenant la Chine, l'Inde, la péninsule indochinoise et les archipels voisins. Malgré un grand nombre d'études réalisées, la taxonomie des *Citrus* reste un sujet très controversé du fait de la grande diversité morphologique observée, de la compatibilité sexuelle entre espèces et de l'apomixie de nombreux génotypes. Dans ce travail de thèse une importante diversité du genre *Citrus* et d'autres genres apparentés de la sous-famille des Aurantioideae, ont été étudiés afin de clarifier leur organisation et leur phylogénie en utilisant différents types de marqueurs et de plateformes de génotypage. Par ailleurs, nous avons fait une analyse particulièrement poussée de la diversité génétique des mandarines car, malgré le rôle particulièrement important que joue ce groupe d'agrumes dans la sélection de porte-greffe et l'amélioration variétale, son organisation génétique est encore mal connue.

Le développement, pour la première fois sur agrumes, de marqueurs nucléaires d'insertion-délétion (indel) nous a permis de démontrer leur utilité dans l'étude de la diversité phylogénétique des *Citrus*. Ces marqueurs, en association avec des marqueurs SSR, ont permis de quantifier la contribution des trois taxons de bases (*C. reticulata*, *C. maxima* et *C. medica*) dans le génome des espèces secondaires et des variétés cultivées. Leur structure génétique en mosaïque a également été déterminée à partir des données obtenues par le séquençage de 27 gènes nucléaires impliqués dans la biosynthèse de composés influençant la qualité des agrumes et dans la réponse au stress des plantes. L'analyse phylogénétique nucléaire a révélé que *C. reticulata* et *Fortunella* forment un clade clairement différencié du clade comprenant les autres taxons de base des agrumes cultivés (*C. maxima* et *C. medica*), ce qui est cohérent avec l'origine géographique des espèces étudiées. A partir de cette étude nous avons développé des marqueurs moléculaires SNP à haute valeur phylogénétique et testé leur transférabilité aux autres genres apparentés. Ces marqueurs ont parfaitement fonctionné au sein du genre *Citrus* et devraient être également utiles pour l'identification variétale au sein des collections, sur une diversité beaucoup plus large.

L'organisation génétique des mandarines [198 variétés de type « mandarine » provenant de deux collections d'agrumes : Inra-Cirad (Haute-Corse, France) et Ivia (Valence, Espagne)], et les introgressions d'autres taxons au sein de ces mêmes mandarines ont été étudiées à l'aide de 50 marqueurs SSRs, de 24 indels nucléaires et de quatre indels mitochondriaux (ADNmt). Il a été démontré que de nombreux génotypes, considérés comme des mandarines pures, présentaient en fait des introgressions d'autres taxons de base dans leur génome. Cinq groupes parentaux ont été déterminés parmi les génotypes analysés. De nombreux génotypes sont issus de croisement entre ces différentes mandarines, créant ainsi des structures hybrides très complexes. De plus, certaines mandarines n'ont pas une origine maternelle « mandarine », tel que le démontre les marqueurs d'ADNmt.

Dans le cadre de ce travail de thèse ont été publiées de nouvelles informations sur les relations phylogénétiques entre les différents taxons du genre *Citrus* et apparentés ainsi qu'entre les espèces secondaires cultivées. Des nouveaux sets de marqueurs complémentaires ont été développés. L'organisation génétique des mandarines a été détaillée et une caractérisation fiable des deux collections (France et Espagne) a pu être réalisée. Ces différentes contributions pourront ainsi aider au travail de sélection de nouvelles variétés d'agrumes de haute qualité et permettra d'optimiser la conservation et la caractérisation génétique et phénotypique des ressources génétiques agrumes.

INDEX

INTRODUCCIÓN

ÍNDICE INTRODUCCIÓN

1. IMPORTANCIA ECONÓMICA DE LOS CÍTRICOS.

Los cítricos son el principal cultivo frutal del mundo, con una superficie cultivada superior a 8,6 millones de hectáreas y una producción de casi 124 millones de toneladas en 2010, superando a cultivos como la banana, las manzanas o la vid (FAOSTAT, 2012). China es el mayor productor de cítricos (cerca de 24 millones de toneladas), seguido de Brasil, Estados Unidos, India, Méjico y España con más de 6 millones de toneladas. Las naranjas son los cítricos más cultivados (56% de la producción), le siguen las mandarinas (17%), limones y limas (12%), los pomelos (10%) y otros cítricos (6%). Las naranjas solo son superadas en producción por las bananas y ligeramente por las manzanas.

El 20 % de la producción citrícola mundial procede de la cuenca Mediterránea. Esta es una de las áreas mundiales más importantes en horticultura. Su clima templado, con veranos cálidos y secos, y los inviernos húmedos y suaves, han sido importantísimos en el desarrollo de la horticultura. La industria frutícola ha sido una pieza clave en el desarrollo de la región. Según la FAO, el tamaño de este sector se extiende en alrededor de 6 millones de hectáreas y una producción anual de cerca de 50 millones de toneladas. Los frutos templados forman el grupo más importante, con un 40% del área cultivada y un 50% de la producción, mientras que los frutos secos cubren el 36% del área, aunque solo contribuyen un 4% en la producción. Los cítricos cubren un 18% de área y un 40% de la producción; los frutos tropicales, en cambio, son poco importantes en la cuenca Mediterránea (Fideghelli and Sansavini, 2002).

En España, el cultivo frutal con mayor producción son los cítricos (más de 6 millones de toneladas; 306.000 Ha.), seguido de lejos por los melocotones y nectarinas (1.1 ton.; 80.000 Ha.). Entre los cítricos, las naranjas son las de mayor producción y superficie dentro de los cítricos (52.8%), seguidas por las mandarinas (34.6%), limones (11.4%), pomelos (1.0%) y otros cítricos (0.1%) (http://www.magrama.gob.es). La Comunidad Valenciana es la mayor productora de cítricos en España, seguida de Andalucía y la Región de Murcia. De acuerdo a la FAO, España es el principal exportador de cítricos para fruta fresca en el mundo, destinando más de la mitad de su producción a la exportación.

2. CENTRO DE ORIGEN Y DIFUSIÓN DE LOS CÍTRICOS.

Se han formulado diferentes hipótesis acerca del origen de los cítricos. En general, se está de acuerdo en que las especies de cítricos y géneros afines son originarios de las regiones tropicales y subtropicales del Sureste Asiático y del Archipiélago Malayo, desde donde se distribuyeron a otros continentes donde se han cultivado (Webber *et al.*, 1967; Calabrese 1992).

El centro principal de origen de los cítricos según Tanaka (1954) sería la zona del noroeste de India y Burma, considerando a China como un centro de distribución secundario. Además, propuso una línea teórica que divide el origen de las distintas especies. Esta línea va desde el borde noroeste de India, por encima de Burma, hasta el sur de la isla de Hainan.

Especies como el cidro, el limón, la lima o la zamboa se originaron al sur de esta línea y especies como las mandarinas, *Fortunella* o *Poncirus* al norte de la misma. Swingle and Reece (1967) propuso como centro de origen del género *Citrus* las regiones tropicales y subtropicales de Asia y el Archipiélago Malayo. Calabrese (1998) indicó que el núcleo principal de origen de los cítricos era China, desde donde se empezó a distribuir por la parte oriental y de aquí siguió los pasos de la civilización.

Los primeros datos escritos acerca de los cítricos se remontan alrededor del año 2400 a.C. en China (capítulo del libro "Tribute to Yu"), así como alrededor del año 800 a.C. en la India (texto religioso "Vajaseneyi sambita"). En estos textos se empezó a hablar de las mandarinas de pequeño tamaño, de los kumkuats (*Fortunella*), las zamboas (*C. maxima*) y del "Yuzu" (*C. junos*), a las cuales se les atribuían usos de tipo medicinal e incluso milagroso (Praloran, 1977). Desde este centro indo-chino primitivo, los cítricos se distribuyeron con mayor facilidad hacia el sudeste (Malasia) y el oeste (valle del Indo) que hacia el nordeste. Se cree que los cítricos se pudieron difundir alrededor del tercer milenio a.C. a través de las relaciones comerciales entre la civilización de Mohandjodaro (Indo) y la baja Mesopotamia.

El cidro fue el primer cítrico conocido en Europa, alrededor del año 300 a.C. (Swingle and Reece, 1967). Durante el viaje de Alejandro Magno (334-323 a.C.) los sabios griegos describieron el cidro como "Manzana de Media" o "Manzana de Persia", y no como fruta de Mesopotamia. También existe controversia en la hipótesis de que los egipcios conocían el cidro entre los años 1500 y 1200 a.C. y su paso hacia Europa debido a sus relaciones comerciales. El cultivo del cidro se extendió desde Persia hasta Palestina alrededor del año 136 a.C., utilizado como ofrenda por los judíos en la fiesta de los Tabernáculos. Las colonias judías contribuyeron a su difusión por la cuenca Mediterránea, llegando a Grecia en el siglo III a.C. y a Italia en el siglo I d.C.

En la difusión del resto de cítricos tuvieron un papel importante los árabes. El naranjo amargo existía en Persia en el 1030 y en Sicilia alrededor del 1094. El limonero se introdujo en el siglo XII y las limas en el siglo XIII. La importación del naranjo dulce a Europa fue obra de los genoveses hacia el 1400, así como de los portugueses en 1548 (Zaragoza, 2007). Los cítricos se difundieron desde el mediterráneo por tres vías: los árabes hacia África entre los siglos XI y XIII, Cristóbal Colón los introdujo en Haití en 1493 y los anglo-holandeses los introdujeron en el Cabo en 1654. Con el descubrimiento de América y su conquista paulatina, se produjo la implantación de los cítricos en Méjico (1518), Brasil (1540), Florida (1565), Perú (1609) o Texas (1890). Los colonos de la primera flota llevaron naranjas, limas y limones desde Brasil a Australia en 1769. Las mandarinas no fueron introducidas en Europa hasta el inicio del siglo XIX.

En cuanto a la introducción de los cítricos en España, el cidro es también el primer cítrico del que se tienen noticias (en torno al siglo VII), siendo lo más probable que fuera introducido a través de Italia y cultivado en algunas regiones del litoral mediterráneo español.

Probablemente los comerciantes árabes introdujeron el naranjo amargo en España hacia los siglos X y XI (Zaragoza, 2007).

El limonero se supone que llegó a España al mismo tiempo o poco después que el naranjo amargo. El geópono toledano Ibn Bassal (1048-1075) cita por primera vez al limonero, junto al cidro y el naranjo amargo, en su Libro de Agricultura. Posteriormente, hacia finales del siglo XI o comienzos del siglo XII destacados geóponos andalusíes citan en sus tratados de agricultura la zamboa, diferenciándola claramente del cidro, del naranjo amargo y del limonero (Zaragoza, 2007).

Pese a que se cree que los genoveses introdujeron el naranjo dulce a mediados del siglo XV a través de sus rutas comerciales con Oriente, fueron los portugueses los que contribuyeron a su difusión en la península Ibérica, al importar de China semillas de variedades de naranja dulce de calidad (Zaragoza, 2007). En cuanto a las mandarinas se tienen referencias de su introducción a mediados del siglo XIX desde Italia (Zaragoza, 2007). Por lo que respecta al pomelo, se empieza a cultivar en la primera mitad del siglo XX (Herrero *et al.*, 1996).

3. CLASIFICACIÓN BOTÁNICA Y ORIGEN GENÉTICO DE LOS CÍTRICOS.

3.1. Clasificación botánica de la subfamilia Aurantioideae.

Por norma general los taxonomistas consideran que las especies de cítricos pertenecen al orden Geraniales, la familia Rutaceae y la subfamilia Aurantioideae. Aurantioideae está considerada como un grupo monofilético según varios autores (Scott *et al.*, 2000; Groppo *et al.*, 2008; Morton, 2009). Según Scott *et al.* (2000) y Bayer *et al.* (2009) Ruta parece ser hermana de Aurantioideae. Más aún, Groppo *et al.* (2008) sugieren que Aurantioideae debería ser reconocida como una tribu e incluirla en una subfamilia junto con Rutoideae, Toddalioideae y Flindersioideae. Pese a que se han sido publicados recientemente nuevos datos sobre la clasificación botánica de Aurantioideae (Bayer *et al.*, 2009; Morton, 2009), sigue existiendo una considerable controversia sobre la división en tribus, subtribus, géneros y especies. Según Swingle and Reece (1967) dentro de esta subfamilia existen dos tribus: Clauseneae con cinco géneros y Citreae con 28. La tribu Clauseneae es más primitiva que la Citreae. Dentro de esta última tribu, la subtribu Citrinae está compuesta de tres grupos, siendo el más importante el de los cítricos verdaderos, donde encontramos los seis géneros más cercanos a los cítricos, incluidos estos (*Fortunella*, *Eremocitrus*, *Poncirus*, *Clymenia*, *Microcitrus* y *Citrus*; Tabla 1).

Tabla 1. Clasificación de la subfamilia Aurantioideae (según Swingle and Reece, 1967).

Tribu	Subtribu		Género
Clauseneae	*Micromelinae*		*Micromelium*
	Clauseneae		*Glycosmis* *Clausena* *Murraya*
	Merrillinae		*Merrillia*
Citreae	*Triphasilinae*		*Wenzelia* *Monanthocitrus* *Oxanthera* *Merope* *Tripashia* *Pamburus* *Luvunga* *Paramingnya* *Wenzelia*
	Citrinae	Cítricos primitivos	*Severinia* *Pleiospermium* *Burkillanthus* *Limnocitrus* *Hesperethusa*
		Cítricos próximos	*Citropsis* *Atalantia*
		Cítricos verdaderos	*Fortunella* *Eremocitrus* *Poncirus* *Clymenia* *Microcitrus* *Citrus*
	Balsamocitrinae		*Swinglea* *Aegle* *Afraegle* *Aeglopsis* *Balsamocitrus* *Feronia* *Feroniella*

3.2. Descripción general de los principales géneros de los cítricos verdaderos.

Fortunella se caracteriza por tener frutos de tamaño pequeño con la corteza dulce y comestible, posee de 3 a 7 ovarios loculares y vesículas delgadas; las hojas son duras con muchas glándulas y aceites esenciales. Los árboles tienen hojas perennes; algunos son de tamaño arbustivo y otros pueden tener un tamaño considerable. Son tolerantes al frío debido a su floración más tardía que la de las especies de *Citrus*. Son plantas muy atractivas, por lo que también se cultivan como ornamentales. Su origen es el sureste de China. Está constituido por cuatro especies: *Fortunella margarita* (Lour.) Swing., *F. japonica* (Thunb.) Swing., *F. polyandra* (Ridl.) Tan. y *F. hindsii* (Champ.) Swing. (Krueger and Navarro, 2007).

Eremocitrus es un género monoespecífico (*Eremocitrus glauca* (Lindl.) Swing). Las hojas son de color gris-verdoso, gruesas y con pelos en ambos lados. Las flores están sueltas o en ramilletes, con ovarios entre 3-5 lóculos, similares a *Fortunella.* Los frutos son ovoides o piriformes, con vesículas delgadas. Es nativo de zonas desérticas de Australia y tolerante al frío y a la sequía.

Poncirus (*Poncirus trifoliata* (L.) Raf.) es el único miembro con hojas trifoliadas y caducas, con peciolos alados y brotes florales formados en el principio del verano, que en invierno están protegidos por escamas (Swingle and Reece, 1967). Los ovarios poseen de 6 a 8 lóculos. Los frutos son pubescentes y las vesículas tienen gran cantidad de aceites esenciales. Según Swingle podría representar el ancestro putativo de de los cítricos verdaderos que se difundió hacia el norte de China, adaptando sus características morfológicas y de resistencia a condiciones extremas de frío invernal. Se utiliza principalmente como patrón y como ornamento.

Clymenia es posiblemente el género más primitivo dentro de los cítricos verdaderos. Tiene hojas que se parecen a las de algún género de la subtribu Triphasiinae. Las flores tienen discos alargados y entre 10 y 20 estámenes y sépalos. Las vesículas de la pulpa son de forma subglobosa o piriforme, que en su gran mayoría están adheridas a las paredes radiales de los segmentos del fruto (14-16). Originalmente se consideró dentro del género *Citrus*, pero tanto Swingle (1939) como (Tanaka 1954), lo consideraron fuera de éste. Un estudio reciente (Berhow *et al.,* 2000) sugiere su carácter híbrido, entre *Fortunella* y *Citrus*, basándose en datos bioquímicos y taxonómicos.

Microcitrus posee un follaje dimórfico, con flores pequeñas y unas vesículas de la pulpa subglobosas que los diferencia del género *Citrus*. Posee gotas de aceites en las vesículas de la pulpa. Los árboles son de tamaño pequeño, tipo arbustivo y con frutos generalmente alargados. Es nativo de zonas desérticas de Australia y en consecuencia es semi-xerófito y puede soportar largas sequías. El género *Microcitrus* está constituido por seis especies: *Microcitrus australasica* (F. Muell.) Swing., *M. australis* (Planch.) Swing., *M. garrowayi* (F.M. Bail.) Swing, *M. inodora* (F.M. Bail.) Swing., *M. maideniana* (Domin) Swing. y *M. warburgiana* (F.M. Bail.) Tan. (Krueger and Navarro, 2007).

Citrus presenta un amplio rango de caracteres y a su vez una gran variabilidad dentro de ellos. La maduración de los frutos es desde muy temprana hasta muy tardía en la temporada. El tamaño de los frutos varía desde muy pequeños, como algunas mandarinas (alrededor de 5 cm.), hasta los más grandes como las zamboas o algunos cidros (15-25 cm.). La forma tanto de los frutos como de las hojas, así como el porte y el crecimiento de los árboles y el contenido en semillas son altamente variables.

3.3. Clasificación del género *Citrus*.

Citrus es el género con la taxonomía más complicada y el de mayor importancia económica de la subfamilia Aurantioideae. Las dos clasificaciones de los cítricos más comúnmente empleadas son las de Swingle and Reece (1967) y la de Tanaka (1954, 1961). El primero dividió el género *Citrus* en dos subgéneros, *Citrus* y *Papeda*, que incluían 10 y 6 especies respectivamente. Estos dos subgéneros se separaban por sus características morfológicas y los componentes químicos de sus flores, hojas y frutos. Tanaka publicó en 1954

"Species Problem in Citrus", dividiendo el género *Citrus* en dos subgéneros, *Archicitrus* y *Metacitrus*, 8 secciones, 13 subsecciones, 8 grupos, dos subgrupos, dos microgrupos y 145 especies. Años después, en 1961, añadió dos nuevas subsecciones, otro grupo y 12 nuevas especies, hasta un total de 157 especies. La mayor diferencia entre las dos clasificaciones es que Tanaka realizó una descripción de los cítricos muy exhaustiva, llevándole a dividir las mandarinas en 36 especies. Por el contrario, Swingle incluyó en la especie *C. reticulata* Blanco a todas las mandarinas con excepción de *C. tachibana* (Mak.) Tanaka y *C. indica* Tan.. Hodgson (1967) propuso una nueva clasificación, con 36 especies divididas en 4 grupos: frutos ácidos, grupo de las naranjas, grupo de las mandarinas y otros. Más recientemente, Mabberley (1997) propuso una nueva clasificación de los cítricos comestibles reconociendo 3 especies y 4 grupos híbridos.

Estudios basados en caracteres bioquímicos (Scora, 1975) y morfológicos (Barret and Rhodes, 1976) sugerían que la mayoría de especies del género *Citrus* son probablemente híbridos directos o híbridos sucesivos de tres especies ancestrales (*C. medica* L. -cidro-, *C. reticulata* -mandarinas- y *C. maxima* (Burm.) Merr. –zamboas-). Estudios basados en diversidad morfológica (Ollitrault *et al.*, 2003) y en metabolitos secundarios (Fanciullino *et al.*, 2006a) confirmaron la importancia de estas tres especies en el origen de la mayoría de cítricos comestibles y la contribución mayor de la diferenciación entre estas especies en la diversidad fenotípica global de los cítricos. Además, *C. micrantha* Wester (*Papeda*) es considerado un ancestro de la lima mejicana (*C. aurantifolia* (Christm.) Swing) (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012a).

Los cidros tienen semillas monoembriónicas. Son árboles de tamaño pequeño, de tipo arbusto, sensibles al frío. Las hojas son glabras, elípticas-ovaladas ó ovaladas-lanceoladas, con los márgenes serrados, peciolo no alado (al contrario que las otras especies de cítricos) y no articulado con el resto de la hoja. Las inflorescencias son en racimo con flores de color morado. Los ovarios son cilíndricos con 10-13 lóculos. Los frutos son alargados, oblongos ó ovalados, de superficie lisa o a veces arrugada, con corteza muy gruesa, segmentos pequeños y con bastantes semillas. Su uso más extendido es en confituras, además de utilizarse como encurtidos y sus aceites esenciales destilados.

Las zamboas tienen semillas monoembriónicas. Los árboles tienen de 5 a 15 metros de altura, ramas angulares a menudo pubescentes. Las hojas son alargadas, ovaladas o elíptico-ovaladas con peciolo alado. Las flores son alargadas, crecen individualizadas, en grupos axilares o en inflorescencias subterminales. Tienen ovario globoso con muchos segmentos. Los frutos son de diversos tamaños, formas y colores, tanto interno como externo. Tienen una piel muy gruesa y unas vesículas alargadas no adheridas entre sí.

Las mandarinas son árboles de tamaño variable, con espinas y ramificaciones finas. Las hojas son lanceoladas, las flores se presentan individuales o en inflorescencias no ramificadas. Los frutos son generalmente achatados, con piel fina y suave, fácilmente separable de los segmentos. En los mandarinos podemos encontrar especies con semillas

monoembriónicas y otras con semillas poliembriónicas. A partir de datos existentes en nuestro grupo se ha observado que las mandarinas con semillas monoembriónicas son de origen híbrido, pudiendo por tanto, proceder el carácter de monoembrionía de los parentales de tipo no mandarino introgresados en estos genotipos, siendo el principal candidato *C. maxima*.

El grupo papeda incluye especies silvestres de cítricos. Los peciolos son largos y alados. Las flores son pequeñas, con estambres libres, ya que no hay fusión de los haces del sépalo lateral con la nervadura central del pétalo. En los frutos, las vesículas de la pulpa tienen numerosas gotas de aceite acre, que les hace tener un sabor amargo, por lo que las especies de este grupo no son comestibles.

Como se ha descrito en los párrafos anteriores, se tiene un amplio conocimiento de los taxones básicos del género *Citrus* a nivel morfológico, así como del origen de las especies cultivadas de cítricos. Sin embargo, no se conoce la contribución exacta de las especies ancestrales a las especies cultivadas y además, sus relaciones filogenéticas no están bien definidas.

3.4. Origen genético de las especies cultivadas de cítricos.

Como se ha comentado en el apartado anterior, las tres especies que han dado lugar a la mayoría de cítricos cultivados son: *C. maxima*, *C. medica* y *C. reticulata*, junto con el papeda *C. micrantha* en el caso de la lima.

Trabajos realizados mediante diferentes tipos de marcadores moleculares, como isoenzimas (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003), RFLP (Federici *et al.*, 1998), RAPD, SCAR (Nicolosi *et al.*, 2000), AFLP (Liang *et al.*, 2007), SSR (Luro *et al.*, 2001; Barkley *et al.*, 2006; Ollitrault *et al.*, 2010) o SNPs (Ollitrault *et al.*, 2012a), apoyan las siguientes teorías sobre el origen de las principales especies secundarias:

Los naranjos dulces (*C. sinensis* (L.) Osb.) están emparentados con *C. reticulata*, pero muestran rasgos introgresados en su genoma procedentes del ancestro *C. maxima* (Nicolosi, 2007). La relación más cercana con *C. reticulata* sugiere que los naranjos dulces no son híbridos directos, sino que probablemente sean híbridos retrocruzados de primera o segunda generación con el genoma de mandarina (Barrett and Rhodes, 1976; Nicolosi *et al.*, 2000). Roose *et al.* (2009) sugieren que *C. sinensis* proviene de un retrocruce 1 (BC1) [(*C. maxima* x *C. reticulata*) x *C. reticulata*)]. Esta misma hipótesis se postula en el trabajo de secuenciación del genoma de la naranja (Xu *et al.*, 2013). Sin embargo, un reciente trabajo de nuestro grupo (Garcia-Lor *et al.*, 2013a) contradice estas hipótesis y propone que los dos parentales de la naranja son híbridos inter-específicos.

El naranjo amargo (*C. aurantium*) parece un híbrido natural entre mandarino y zamboa (Scora, 1975; Barrett and Rhodes, 1976; Nicolosi *et al.*, 2000; Uzun *et al.*, 2009).

El pomelo (*C. paradisi* Macf.) es una especie muy cercana a *C. maxima* y pudo resultar del cruce espontáneo entre *C. maxima* y *C. sinensis* (Barrett and Rhodes, 1976; Scora *et al.*, 1982; de Moraes *et al.*, 2007; Ollitrault *et al.*, 2012a).

Diversos análisis del genoma cloroplástico (Green *et al.*, 1986; Nicolosi *et al.,* 2000) y mitocondrial (Froelicher *et al.*, 2011) indicaron que *C. maxima* aportó el cloroplasto y el citoplasma en el origen de estas tres especies (naranjo dulce, naranjo amargo y pomelo).

Existen datos que confirman las relaciones genéticas existentes entre *C. medica* y *C. limon* Osb. (limones) (Froelicher *et al.*, 2011). Marcadores cloroplásticos y nucleares indican qué los genomas de *C. reticulata* y *C. maxima* contribuyen también a la génesis del limón. Nicolosi *et al.* (2000) propuso que esta especie surgió del cruce directo entre *C. aurantium* y *C. medica*. Esta teoría fue apoyada por Gulsen and Roose (2001a) y Ollitrault *et al.* (2012a).

En el caso de la lima 'Mejicana' (*C. aurantifolia*), datos moleculares (Federici *et al.*, 1998; Nicolosi *et al.,* 2000; Ollitrault *et al.*, 2012a) apoyan la idea de Torres *et al.* (1978) de que es un híbrido entre *C. medica* y una variedad de *Papeda*. Nicolosi *et al.* (2000) propuso la hipótesis de que la lima mejicana tiene como parental al papeda *C. micrantha*. El origen de otras limas es desconocido.

Pese a los estudios realizados hasta el momento, hay un escaso conocimiento de la contribución exacta de las especies ancestrales al genoma nuclear de las especies secundarias (*C. sinensis, C. limon, C. aurantium, C. paradisi* y *C. aurantifolia*) y a los híbridos procedentes de los programas de mejora genética del siglo XX.

3.5. La situación particular de la clasificación de los mandarinos.

El germoplasma de mandarino fue clasificado como *C. reticulata* por Swingle and Reece (1967), al igual que Mabberley (1997). Por contra, Webber (1943) clasificó las mandarinas en 4 grupos: king, satsuma, mandarina y tangerina. Tanaka (1954) dividió los mandarinos en 5 grupos que incluían un total de 36 especies, basándose en cambios morfológicos del árbol, hojas, flores y frutos. El grupo 1, incluye a *C. nobilis* Lour. (cultivares tipo King), *C. unshiu* Marc. (satsumas), y *C. yatsushiro* Hort. ex Tanaka; el grupo 2 lo forman *C. keraji* Hort. ex Tanaka, *C. oto* Hort. ex Yuichiro y *C. toragayo* Hort. ex Yuichiro); el grupo 3 contiene 14 especies, incluyendo algunas de las más importantes económicamente: *C. reticulata* ('Ponkan'), *C. deliciosa* Tenore ('Willowleaf' o 'Mandarino común'), *C. clementina* Hort. ex Tanaka (clementinas) y *C. tangerina* Hort. ex Tanaka ('Dancy'); en el grupo 4 incluye a *C. reshni* Hort. ex Tanaka ('Cleopatra'), *C. sunki* Hort ex Tanaka ('Sunki') y *C. tachibana*; y en el grupo 5 incluye las especies *C. depressa* Hayata ('Shekwasha') y *C. lycopersicaeformis* (Lush.) Hort. ex Tanaka. Otro autor que estudió el grupo de las mandarinas fue Hodgson (1967), que sugirió la agrupación en 4 especies: *C. unshiu* (satsuma), *C. reticulata* ('Ponkan', 'Dancy', 'Clementine'), *C. deliciosa* ('Willowleaf'), y *C. nobilis* ('King').

Como se ha comentado anteriormente, el grupo mandarino es considerado uno de los tres principales grupos ancestrales de los cítricos cultivados (Barret and Rhodes, 1976; Nicolosi *et al.*, 2000; Krueger and Navarro, 2007). El centro de diversificación de *C. reticulata* está en Asia, comprendiendo desde Vietnam a Japón. Es un grupo muy polimórfico, como se ha podido observar mediante marcadores moleculares (Coletta Filho *et al.*, 1998; Luro *et al.*, 2004) y caracteres fenotípicos, ya sea la pomología del fruto o la tolerancia a factores bióticos y abióticos. Además, en algunos grupos de cultivares como las 'satsumas' y las 'clementinas', cuya diversidad es debida a acumulaciones de mutaciones somáticas (Cameron and Frost, 1968), la dificultad para su caracterización a nivel molecular es mayor, ya que los marcadores moleculares existentes hasta el momento no permiten diferenciar estos genotipos.

Pese a la gran cantidad de información existente, hay muy pocos datos disponibles con respecto a la organización intraespecífica de *C. reticulata* y los determinantes de su diversidad fenotípica. Estos temas se han abordado en la presente tesis doctoral. Además, esta información es necesaria para optimizar la explotación de los recursos fitogenéticos y la mejora genética de este grupo.

4. MEJORA GENÉTICA DE LOS CÍTRICOS.

La mejora genética de los cítricos se dirige tanto a la obtención de nuevas variedades como a patrones y tiene como objetivos generales la introducción de resistencia o tolerancia a estreses bióticos y abióticos y la mejora de la calidad de los frutos.

Los cítricos están afectados por importantes estreses de tipo abiótico causados por la diversidad climática y de suelos existentes a nivel mundial. La salinidad afecta seriamente al desarrollo vegetativo y reproductivo, así como a la producción (Storey and Walker, 1999). La sequía en áreas templadas, como la cuenca Mediterránea, produce un decrecimiento de los procesos vegetativos, como la caída de hojas (Tudela and Primo-Millo, 1992), se ven afectados el potencial hídrico y la conductancia estomática (Gómez-Cadenas *et al.*, 1996), así como una disminución de la cantidad y calidad del fruto (Yakushiji *et al.*, 1998). Otra de las preocupaciones de la citricultura es la clorosis férrica, que afecta al 20-50 % de los árboles en la cuenca Mediterránea, asociada a suelos calcáreos y básicos (Jaeger *et al.*, 2000). En otras zonas los suelos ácidos son un problema (Ollitrault and Navarro, 2012). Las heladas y las altas temperaturas son también una causa importante de pérdidas en la producción (Krueger and Navarro, 2007).

Los cítricos sufren importantes pérdidas económicas por distintos estreses bióticos causados por patógenos y plagas. Entre los virus más importantes se puede mencionar el *Citrus tristeza virus* (CTV) que produce un decaimiento general de árboles de mandarino, naranjo dulce y pomelo injertados sobre naranjo amargo llegando a producirles la muerte (Moreno *et al.*, 2008), el *Citrus Tatter Leaf Virus* (CTLV) que causa problemas de incompatibilidad en patrones trifoliados o el *Citrus Sudden Death Associated Virus* (CSDAV),

que causa la muerte de naranjos dulces injertados en lima Rangpur. Los cítricos también están afectados por numerosos viroides, entre los que destaca el *Citrus exocortis viroid* causante de la exocortis, que en árboles injertados sobre patrones sensibles produce poco desarrollo general, escamas en la corteza del patrón y también hace perder hojas jóvenes y disminuye el número de brotes y frutos (Duran-Vila *et al.*, 1988).

Las bacterias causan gravísimos daños en los cítricos. La *Candidatus Liberibacter* sp. produce la enfermedad del Huanglongbing, presente en Asia, Brasil y EEUU (Bové, 2006) que es la más grave de las que afectan a los cítricos y que en algunas zonas está impidiendo el cultivo. La *Xanthomonas axonopodis* pv. *citri* (Hasse) produce la cancrosis de los cítricos en Suramérica, algunos estados de EE.UU. y Asia, causando importantes mermas en la producción y problemas de cuarentena en el comercio de frutos (Das, 2003).

En cuanto a enfermedades fúngicas, el oomiceto *Phytophtora* sp., que puede causar la gomosis o el aguado (Cacciola and Lio, 2008), está extendido en muchas regiones. En África encontramos la cercosporiosis (*Phaemularia angolensis* (De Carvalho & O. Mendes) P.M. Kirk), que causa daños significativos en hojas y frutos (Ollitrault and Luro, 2001). La mancha negra de los cítricos es una enfermedad causada por el hongo *Guignardia citricarpa* que tiene lugar en zonas de clima subtropical, causando una reducción de la cantidad de fruta producida y de su calidad y problemas de cuarentena en la comercialización de fruta (Kotzé, 1981). En cuanto a la cuenca Mediterránea, el mal seco (causada por el hongo *Phoma tracheiphila* (Petri) L.A. Kantsch. & Gikaschvili) es una enfermedad fúngica importante en los limoneros y algunos patrones (Perrotta and Graniti, 1988). También es un problema la *Alternaria* en algunos cultivares de tipo mandarino, como Fortune (Vicent *et al.*, 2000; Vicent *et al.*, 2004; Cuenca *et al.*, 2012).

Los cítricos también están afectados por numerosas plagas que causan mermas en la producción, problemas en la comercialización de frutos y su control con pesticidas produce daños en el medio ambiente y residuos en los frutos perjudiciales para el consumidor. Entre las más importantes se pueden citar: algunos arácnidos, insectos como la mosca blanca o la mosca del Mediterráneo (*Ceratitis capitata*), pulgones, cochinillas, etc.

4.1. Mejora de patrones.

Uno de los principales objetivos de la mejora en patrones es su adaptación a las condiciones ambientales existentes en el área de cultivo (suelos salinos o alcalinos, ácidos, inundados, secos, concentración de caliza, tolerancia al frio) y a los patógenos del suelo.

Algunas de las necesidades comunes en la mayoría de las áreas de cultivo son que los nuevos patrones presenten tolerancia a enfermedades como la tristeza, al oomiceto *Phytophthora sp.* o a los nematodos, principalmente.

Otro de los caracteres importantes en los patrones, es la alta producción de semilla con una elevada poliembrionía, lo que facilita la propagación (reproducción apomíctica que impide

la formación de embriones sexuales) y la uniformidad de las plantas obtenidas en vivero. Además, los patrones tienen que causar una rápida entrada en producción de la variedad y una elevada productividad de fruta de calidad. El control del vigor del árbol es actualmente un importante objetivo en muchos programas con la finalidad de realizar plantaciones muy densas de árboles de pequeño porte (Ollitrault and Navarro, 2012).

4.2. Mejora de variedades.

Los objetivos de la mejora genética de variedades varían ostensiblemente en función de las demandas del mercado, las condiciones ambientales en las áreas de producción y el destino de la fruta producida.

En la industria del zumo es importante una mejora en el contenido en zumo y azúcares, el color y la elevada productividad.

Como preocupaciones más importantes para el comercio de fruta fresca en la actualidad son, la calidad pomológica del fruto (tamaño, color, facilidad de pelado), la calidad organoléptica (aroma, sabor, acidez, azúcares), la ausencia de semillas (autoincompatibilidad gametofítica, esterilidad femenina o masculina, partenocarpia), la extensión del periodo de cosecha, así como la calidad nutricional (vitamina C, carotenoides, compuestos fenólicos) que en la actualidad es considerada como criterio de selección en algunos proyectos de mejora (Ollitrault and Navarro, 2012).

La mejora también va encaminada a la resistencia o tolerancia a diversas enfermedades, como el Huanglonbing, la cancrosis, cercosporiosis, mal seco, o el hongo *Alternaria alternata*, un gran problema aparecido recientemente en la citricultura española, que ha producido graves daños a los cultivos de la variedad 'Fortune' principalmente.

En cuanto a la mejora de las especies secundarias (naranja dulce, limones, pomelos), la falta de diversidad genética y elevada heterocigosis imposibilita la mejora por hibridación sexual y solo queda la selección de mutaciones espontáneas o inducidas o la transformación genética. En el caso de las mandarinas, donde hay una elevada diversidad, además de la selección de mutaciones (clementinas, satsumas), la selección mediante programas de mejora por hibridación sexual ha permitido la obtención de nuevas variedades tanto a nivel diploide como triploide (Russo *et al.*, 2004; Williams and Roose, 2004; Tokunaga *et al.*, 2005; Navarro *et al.*, 2006b; Aleza *et al.*, 2010; Cuenca *et al.*, 2010).

4.3. Problemas en la mejora clásica de los cítricos.

Los cítricos presentan algunas características en su biología reproductiva muy peculiares, como son la apomixis, la incompatibilidad sexual (de algunos genotipos), la esterilidad, la elevada heterocigosis, etc., que dificultan la mejora genética.

La apomixis es la producción de embriones (a partir de la nucela) sin que ocurra meiosis ni fertilización y que dan lugar a plantas genéticamente idénticas a la madre. En los genotipos apomícticos se da un proceso sexual y asexual en el mismo rudimento seminal, formándose semillas con un embrión zigótico y uno o varios nucelares. Habitualmente, los embriones de origen nucelar son más vigorosos que los zigóticos, los cuales no completan su desarrollo y abortan frecuentemente. Este fenómeno, complica la obtención de poblaciones elevadas de híbridos para seleccionar genotipos superiores (Davies and Albrigo, 1994).

Otro problema de la biología reproductiva de los cítricos, es la esterilidad gamética parcial o total en los óvulos y/o en el polen, lo que imposibilita su empleo como parentales en los programas de mejora. También, existen genotipos que presentan incompatibilidad sexual, lo que dificulta la obtención de híbridos (Soost, 1969; Soost and Cameroon, 1975).

La alta heterocigosidad existente en muchas especies de cítricos, provoca una progenie sexual muy variable (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003). Por lo tanto, es difícil reunir en un híbrido los caracteres deseados de los parentales. Además del problema de la depresión por endogamia, que muchas veces se observa en la progenie híbrida (Barrett and Rhodes, 1976).

Otros factores que limitan la mejora genética clásica son, el largo periodo de juvenilidad de los cítricos (4-8 años), el desconocimiento del modo de herencia de la mayoría de caracteres agronómicos de interés, la escasez de marcadores relacionados con estos caracteres y la producción de semillas en la mayoría de híbridos.

Además de todos estos problemas, no disponemos de un conocimiento exhaustivo de la diversidad genética que permitiría la selección de nuevos parentales y facilitaría la planificación de los programas de mejora.

4.4. Calidad del fruto de los cítricos; biosíntesis de metabolitos primarios y secundarios.

La calidad organoléptica (aroma, sabor, acidez, azúcares) y las características pomológicas de las variedades (facilidad de pelado, ausencia de semillas, apariencia externa) son parámetros fundamentales en todos los proyectos de mejora (Navarro *et al.*, 2006a). Además, la calidad nutricional se está empezando a plantear como objetivo en algunos programas en aspectos relacionados con el contenido en vitamina C, carotenoides y compuestos polifenólicos, ya que poseen efectos beneficiosos para la salud humana (Del Caro *et al.*, 2004; Dhuique-Mayer *et al.*, 2005). En los cítricos, se conocen las rutas de biosíntesis que dan lugar a los diferentes compuestos determinantes de la calidad del fruto, pero no se conoce la diversidad existente en los genes implicados en su biosíntesis, así como su posible diferenciación a nivel evolutivo.

Por lo tanto, la secuenciación de genes que codifican para enzimas clave en las rutas de biosíntesis de azúcares, ácidos, flavonoides, y carotenoides en una amplia representación

de la diversidad genética de los cítricos, ayudará a aclarar las relaciones filogéneticas existentes entre las especies del género *Citrus* y afines y podría servir para comprender las diferencias en su acumulación existentes entre ellas.

4.4.1. Azúcares y acidez en los cítricos.

Los principales carbohidratos existentes en el zumo son la sacarosa, la fructosa y la glucosa. En la mayoría de cítricos la sacarosa es el azúcar más abundante (Sanz *et al.*, 2004). Los ácidos orgánicos son los principales responsables de la acidez en los cítricos, siendo el ácido cítrico el más abundante, seguido de ácido málico (Sadka *et al.*, 2001).

A lo largo de la maduración del fruto, se observan habitualmente tres fases de desarrollo (Albertini, 2006): una primera fase de multiplicación celular con un aumento del tamaño del fruto rápido, una segunda fase de crecimiento celular iniciada por una síntesis de azúcares y ácidos orgánicos en el tonoplasto (duración variable en función de la variedad) y la fase final de maduración donde se producen diversas reacciones fisiológicas, como el cambio de color de la piel. El ácido cítrico aumenta muy rápidamente al principio del desarrollo del fruto y disminuye en la maduración. Los azúcares, en cambio, se acumulan mayoritariamente en la segunda y tercera fase del desarrollo del fruto (Erickson *et al.*, 1968). Generalmente, los cítricos se agrupan en dos clases según su acidez: un grupo compuesto por las naranjas, las mandarinas, los zamboas y los pomelos, que son frutos dulces con un poco de acidez; y otro grupo (limones, limas y cidros) que son muy ácidos y contienen pocos azúcares (Webber *et al.*, 1967).

4.4.1.1. Biosíntesis de los azúcares y los ácidos.

La variación en la concentración de azúcares solubles y de los ácidos orgánicos a lo largo del desarrollo del fruto, depende del equilibrio entre la síntesis, la degradación y el desarrollo de estos metabolitos (Tucker, 1993). Por lo tanto, los mecanismos de regulación de la glicólisis, el ciclo de Krebs y el almacenamiento vacuolar son esenciales. Las enzimas que codifican para los genes implicados en estos procesos son las encargadas de la conversión de hexosas en hexosas fosfato, de fructosa-6-fosfato en fructosa-1,6-bifosfato, y de fosfoenolpiruvato (PEP) en piruvato (Plaxton, 1996; Copeland and Turner, 1987), además de las fosfofructoquinasas (PKF) dependientes de ATP y PPi, la fosfoenolpiruvato carboxilasa (PEPC) y la fosfoenolpiruvato carboxikinasa (PEPCK), la malato deshidrogenasa (MDH) y el enzima málico (EMA). En la figura 1 se muestra la biosíntesis de azúcares, que se produce en el citosol y la biosíntesis de ácidos que se produce en las mitocondrias.

El ácido cítrico procedente del ciclo de Krebs está también fuertemente regulado. Las enzimas citosólicas que codifican para el gen aconitasa (ACO) y la isocitrato deshidrogenasa (IDH) están implicadas en el catabolismo del citrato de origen

mitocondrial. El citrato que no es metabolizado, puede ser acumulado en la vacuola, lo cual influye en el pH, que es regulado en los diferentes estadios de desarrollo del fruto. Por medio de los mecanismos de regulación del citrato en la vacuola, el transportador de citrato + H^+ (TRPA) permite el flujo de citrato de la vacuola al citoplasma (Shimada *et al.*, 2006). De esta manera, este gen controla la homeostasis vacuolar del citrato y regula la acidez de los frutos de cítricos. Los azúcares son fuertemente regulados. La sacarosa puede abastecer la glicólisis o ser transportada a la vacuola. También puede ser catabolizada en glucosa y fructosa mediante la ácido invertasa (INVA) (Kubo *et al.*, 2001).

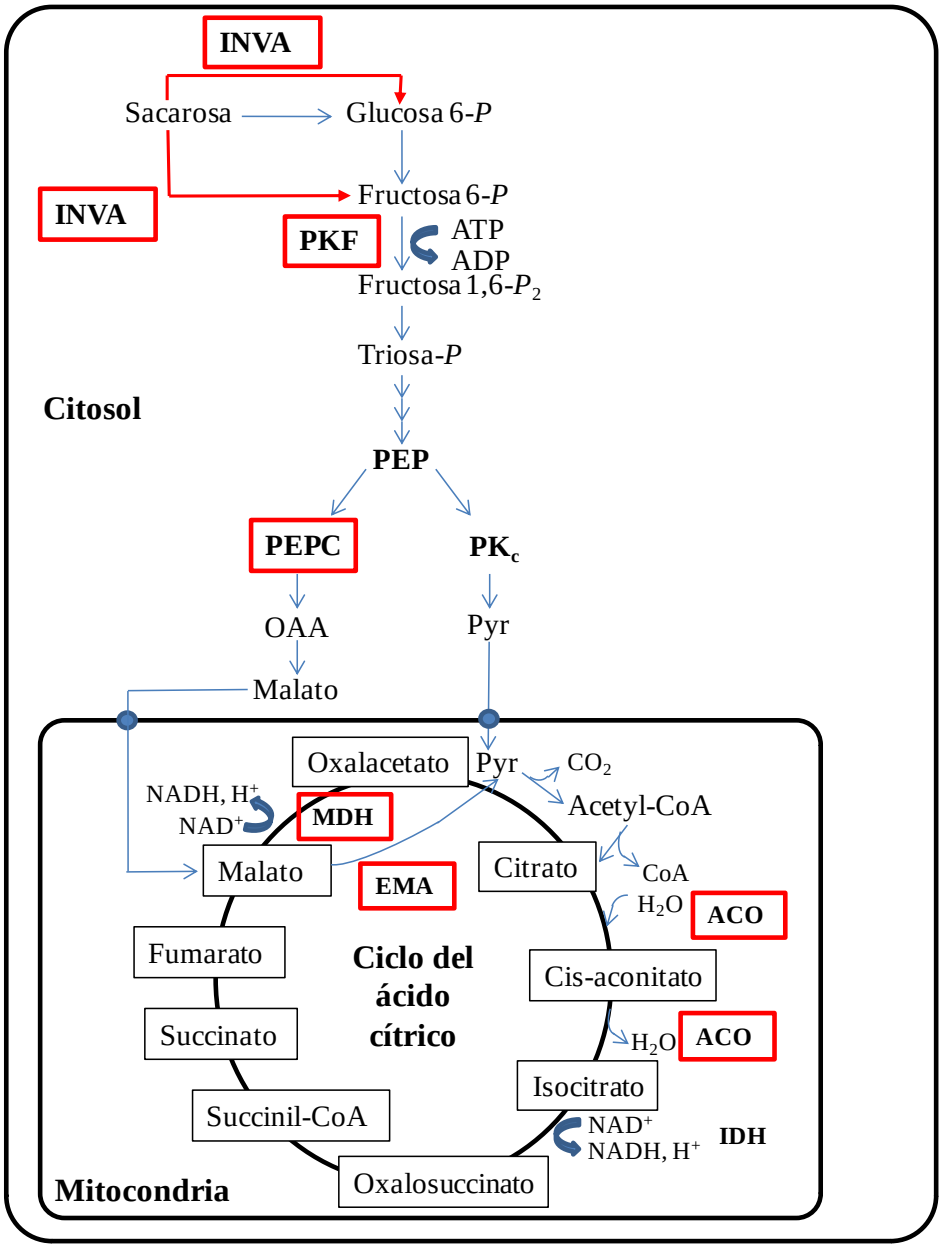


Figura 1. Biosíntesis de azúcares y ácidos. Los genes secuenciados en la tesis codifican para las enzimas recuadradas en rojo.

4.4.2. Flavonoides y antocianos.

Los flavonoides son compuestos que juegan un papel importante en la resistencia de las plantas a la foto-oxidación de la luz ultravioleta, intervienen en el transporte de auxina, son atrayentes de los polinizadores y pueden dar el color a las hojas, frutos, semillas y a las flores (Winkel-Shirley, 2001) y además, afectan el sabor del fruto. En humanos, han mostrado una alta capacidad antioxidante (Kaur *et al.*, 2001), previenen algunos desordenes cardiovasculares (Gross, 2004), tienen actividad antiinflamatoria (HyunPyo *et al.*, 2004) y antialérgica (Middleton y Kandaswami, 1992), entre otras cosas. Por todo esto, se han realizado muchos estudios para modificar su biosíntesis en plantas (Tucker, 2003; Schijlen *et al.*, 2004; Yonekura-Sakakibara and Saito, 2006; Koca *et al.*, 2009). Los frutos de cítricos contienen un amplio rango de flavonoides (principalmente flavanonas y flavonas/oles), que suponen una de las fuentes importantes de compuestos fenólicos en nuestra dieta (Erlund, 2004). La cuantificación de flavonoides ha permitido la diferenciación entre algunos cítricos. Gaydou *et al.* (1987) diferenciaron mandarinas y naranjas, Mouly *et al*. (1994) distinguieron entre el limón, lima, pomelo y naranja dulce, Nogata *et al.* (2006) diferenciaron entre 42 especies y cultivares del género *Citrus*, más dos *Fortunella* y un *Poncirus*.

4.4.2.1. Biosíntesis de los flavonoides y antocianos.

La biosíntesis de flavonoides (Winkel-Shirley, 2001; Bogs *et al.*, 2006) (Figura 2) comienza con la catalización de naringenina chalcona mediante la chalcona sintasa (CHS) y la siguiente conversión en naringenina flavanona a través de la chalcona isomerasa (CHI). A continuación, la adición de grupos hidroxil y/o metil dan lugar a diversas flavanonas, las cuales pueden ser transformadas en flavonoles en varias conversiones enzimáticas; estos compuestos pueden ser glicosilados. Los más importantes en cítricos son las flavanonas, y las que dan el sabor son las glicosiladas (McIntosh *et al.*, 1990). En especies de cítricos como las mandarinas y las naranjas dulces, solo contienen rutinosidos (sin sabor), mientras que zamboa contiene solo flavanonas neohesperidosidas, que le confiere amargura (Kawaii *et al.*, 1999). Frydman *et al.* (2004) aislaron el gen (1,2 ramnosil transferasa) responsable de la biosíntesis de los flavonoides que producen la amargor de los cítricos (zamboas y pomelos).

Por otra parte, encontramos las antocianinas, que son compuestos fenólicos que dan lugar al color rojo en el caso de las naranjas sanguinas (Lo Piero *et al.*, 2005). Estos compuestos comparten parte de la ruta de biosíntesis de los flavonoides que se ramifica para dar lugar a las antocianinas a partir de flavanonas (Figura 2).

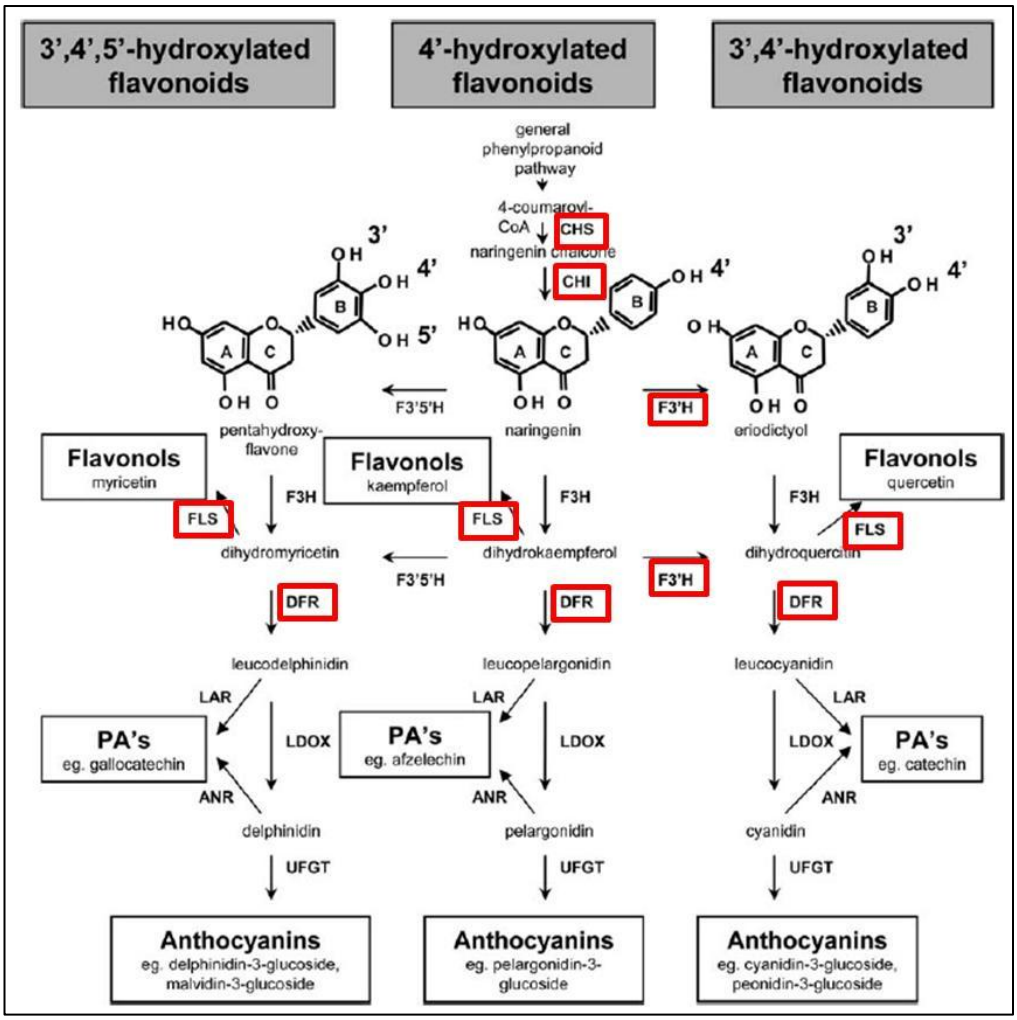


Figura 2. Biosíntesis de flavonoides (Bogs *et al.*, 2006). Los genes secuenciados en la tesis codifican para las enzimas recuadradas en rojo.

4.4.3. Carotenoides.

Los carotenoides son pigmentos sintetizados en plantas, algas y algunas cianobacterias, que juegan un papel muy importante en el aparato fotosintético, protegiéndolas de daños oxidativos producidos por la luz. También participan en el sistema de captación de luz (Goodwin, 1980; Demmig-Adams *et al.*, 1996). En plantas se acumulan en los cromoplastos y juegan un papel importante en la coloración del fruto, la raíz o el tubérculo y en su calidad nutricional. Los carotenoides son utilizados como atrayentes de polinizadores y agentes de dispersión de polen. También sirven de precursores de la vitamina A, esencial en la dieta humana y animal, así como de antioxidantes, los cuales previenen contra ciertas enfermedades cardiovasculares o cáncer (Olson, 1989; Rao and Rao, 2007).

4.4.3.1. Biosíntesis de los carotenoides.

La ruta de biosíntesis de carotenoides (Figura 3) ha sido bien descrita por numerosos trabajos (Sandmann, 2001; Fanciullino, 2007). Los carotenoides son sintetizados en los

plástidos por enzimas codificados en el núcleo. El precursor de los caroteniodes, y también de hormonas como las giberelinas, es el geranilgeranil difosfato (GGPP). La condensación de dos moléculas de GGDP dan lugar al fitoeno, de 40 carbonos (incoloro), reacción catabolizada por la fitoeno sintasa (PSY). A continuación, este sufre 4 desaturaciones, catalizadas por la fitoeno desaturasa (PDS) y la ζ-caroteno desaturasa (ZDS), que lo convierten en licopeno (color rojo). En plantas superiores, la circularización del licopeno en β-caroteno y α-caroteno es un paso crucial en la ramificación de la ruta de biosíntesis (Cunningham *et al.*, 1996; Hirschberg, 2001). Esta reacción es catalizada por una enzima (LCY-b) para obtener β-caroteno en dos pasos.

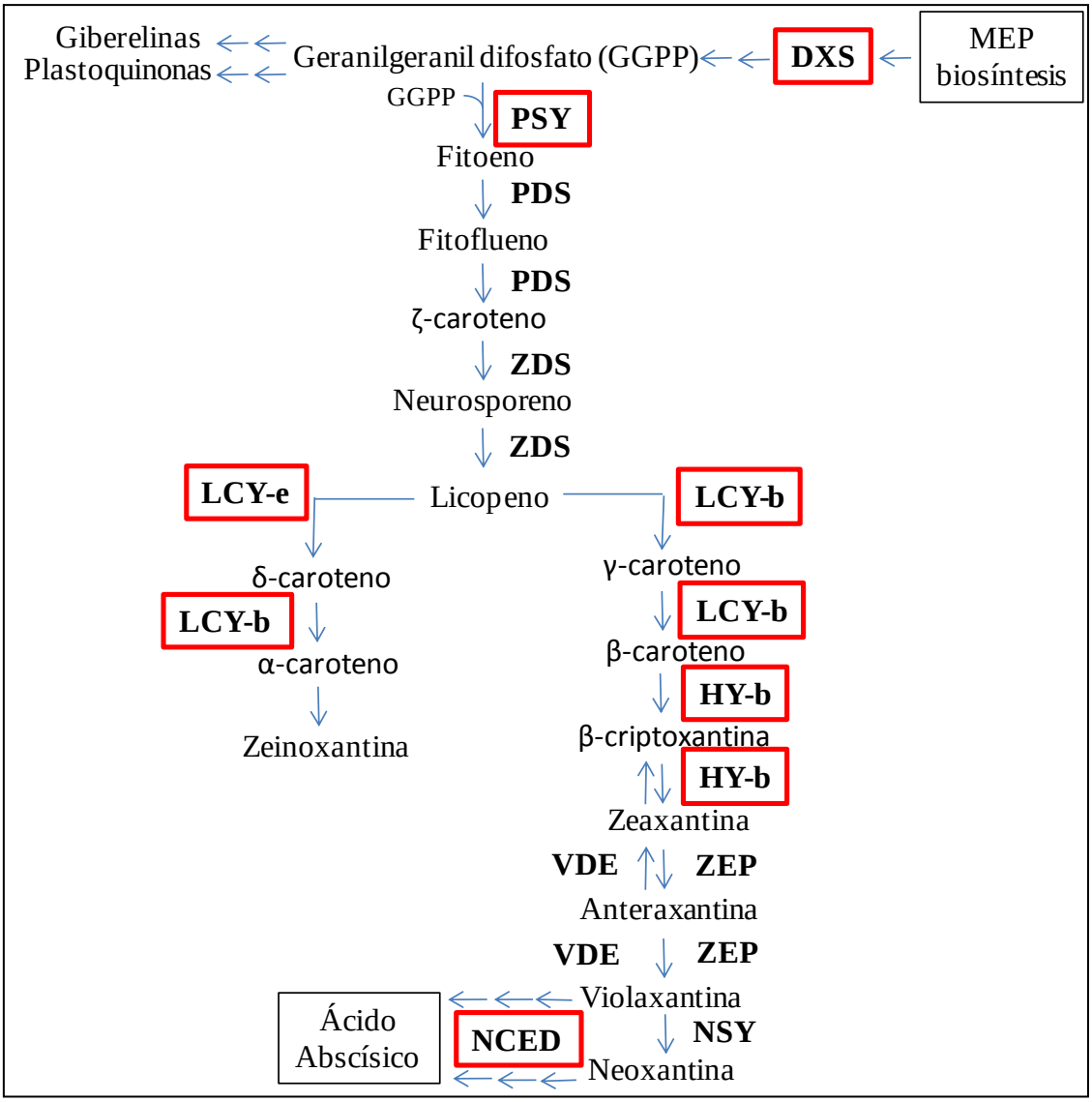


Figura 3. Biosíntesis carotenoides. Los genes secuenciados en la tesis codifican para las enzimas recuadradas en rojo.

En cambio, para obtener α-caroteno se necesitan dos enzimas, la licopeno ε- ciclasa (LCY-e) y la licopeno β-ciclasa (LCY-b). Siguiendo esta parte de la ruta, se obtiene la luteína (xantofilas) tras dos hidroxilaciones catalizadas por la ε-caroteno hidroxilasa

(HY-e) y la β-caroteno hidroxilasa (HY-b). En la otra parte de la ruta, otras xantofilas se producen a partir de la hidroxilación de β-caroteno y la epoxidación catalizada por la zeaxantina epoxidasa (ZEP). La violaxantina puede ser de-epoxidada a zeaxantina a través de la anteraxantina por la violaxantina de-epoxidasa (VDE) (Gilmore and Yamamoto, 1993). A partir de la zeaxantina se puede obtener ácido abcísico en sucesivas reacciones. La biosíntesis de carotenoides y su regulación ha sido estudiada en varias especies de plantas, como Arabidopsis (Hyoungshin *et al.*, 2002), tomate (Isaacson *et al.,* 2002) o pimiento (Bouvier *et al.*, 1998) entre otras. A partir de estudios en tomate, se piensa que la regulación de la ruta es principalmente a nivel transcripcional (Bramley, 2002). Regulación post-transcripcional, por retroalimentación o por hormonas (etileno), se han sugerido como mecanismos para explicar la acumulación de carotenoides. En cítricos, se han realizado diversos estudios en algunas especies para clarificar la regulación de la producción de carotenoides (Rodrigo *et al.*, 2004; Kato *et al.*, 2006; Fanciullino *et al.*, 2008; Alquézar *et al.*, 2009), pero es necesaria más información a nivel intra e interespecífico.

5. RECURSOS FITOGENÉTICOS EN CÍTRICOS.

En el año 2001, se firmó el Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura, el cual señala la importancia que reviste la conservación y el uso sostenible de los recursos fitogenéticos, así como su prospección, recolección, caracterización, evaluación y documentación para garantizar una producción de alimentos diversificada, sostenible y nutricionalmente diversa (FAO, 2001). En este contexto, los bancos de germoplasma (conservación *ex situ*) desarrollan un papel esencial en el que se hace necesario proteger y mantener los recursos fitogenéticos que constituyen las fuentes de variabilidad para la obtención de nuevas variedades en un contexto socio-económico en constante evolución. Los principales Bancos de Germoplasma mundiales del género *Citrus* se encuentran situados en Japón, China, EEUU, Francia y España. En Japón existen seis colecciones donde se mantienen más de 1200 genotipos, destacando el gran número de genotipos de mandarino que se conservan, principalmente del grupo satsumas (Krueger and Navarro, 2007). En el banco de germoplasma de China hay aproximadamente 1000 accesiones mantenidas *ex situ* (Liu and Deng, 2007). La Citrus Variety Collection (http://www.citrusvariety.ucr.edu/) de la Universidad de Riverside en California (EEUU) contiene más de 1000 genotipos del género *Citrus* y afines (Barkley *et al.*, 2006). La colección INRA-CIRAD existente en el *Institut National de la Recherche Agronomique* (INRA), San Giuliano (Córcega, Francia), cuyo germoplasma es uno de los más ricos en variedades del grupo mandarino, consta con alrededor de 1100 accesiones de cítricos y géneros afines, incluyendo *Citrus* (cidros, zamboas, mandarinas y papedas), *Poncirus*, *Fortunella*, *Microcitrus*, *Eremocitrus*... además de híbridos intra- e interespecíficos. La colección existente en el Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada (Valencia, España), incluye alrededor

de 600 genotipos, entre los cuales están la mayoría de los cultivares modernos de mandarina, especies del género *Citrus* y también especies de géneros afines de la subfamilia Aurantioideae. Esta colección también posee los genotipos de la misma dentro de un recinto de malla para evitar la transmisión de enfermedades, la contaminación con los patógenos transferidos por vectores y mantener "árboles iniciales" de los programas de certificación para la propagación comercial de plantas producidas en los viveros comerciales. La colección del IVIA se caracteriza morfológicamente de forma continua con los descriptores del IPGRI y UPOV y esto permite eliminar posibles genotipos duplicados, que suponen un problema frecuente en la mayoría de las colecciones de germoplasma de cítricos.

5.1. Conservación de los recursos fitogenéticos de cítricos.

Las semillas son el método de conservación y distribución más conveniente en los bancos de germoplasma, siendo la conservación de semilla desecada en condiciones de baja humedad y almacenada a baja temperatura la forma más extendida de la colecciones *ex situ*. Las semillas que se pueden conservar de esta manera se les llama "ortodoxas" (Roberts, 1973). La mayoría de especies de plantas se conservan de esta manera. Las condiciones técnicas para el mantenimiento de semillas ortodoxas están descritas por la FAO/IPGRI (Genebank Standards, 1994).

Sin embargo, muchas especies de origen tropical y subtropical (aguacate, mango, cacao, etc.), así como algunas leñosas (género *Quercus*, *Castanea*, *Citrus*, etc.) tienen semillas sensibles al proceso de desecación y conservación en bajas temperaturas, por lo que no pueden ser conservadas mediante este método debido a la pérdida de viabilidad. Este tipo de semillas se denominan "recalcitrantes" (Chin and Roberts, 1980). El problema de estas especies es que la conservación de los recursos genéticos requiere el mantenimiento de plantas.

La selección humana y la propagación vegetativa han llevado a la generación de variedades élite de cítricos, pero ha producido la pérdida de muchos genotipos silvestres originales. Además, la diversidad genética en los centros de origen está en peligro por la pérdida de hábitat debido a la deforestación, presión poblacional, turismo, etc., como sucede en India o China. Por ello, es necesaria una conservación *ex situ* de los recursos fitogenéticos. Debido a que las semillas de los cítricos son de tipo recalcitrante, las colecciones de germoplasma existentes se mantienen mediante plantas en campo y en algunos casos también en recintos de malla, lo cual conlleva unos gastos elevados.

Una alternativa para el mantenimiento de germoplasma de especies con semillas recalcitrantes es la conservación *in vitro* (Engelmann, 1997). En cítricos se han puesto a punto procedimientos de crioconservación de callos embriógenicos y embriones (Duran-Vila, 1995; González-Arnao, 2003) y el Banco de Germoplasma del IVIA mantiene una colección callos embriogénicos crioconservados de unos 60 genotipos. Muy recientemente, se ha abierto la

posibilidad de conservación mediante la criopreservación de ápices y la regeneración de plantas mediante microinjerto de ápices caulinares *in vitro* (Volk *et al.*, 2012).

5.2. Manejo de los recursos fitogenéticos.

Para la conservación de los recursos fitogenéticos de cítricos (Krueger and Navarro, 2007), se han de seguir los siguientes pasos: localización de nuevos genotipos para aumentar la diversidad del banco de germoplasma, introducción, mantenimiento, caracterización y evaluación, documentación y establecimiento de bases de datos.

5.2.1. Localización de nuevos genotipos para aumentar la variabilidad del banco de germoplasma.

Lo primero es la identificación y localización de las fuentes de nuevo material a introducir mediante la exploración de áreas de diversidad, selección de genotipos cultivados o nuevos, o por intercambio entre centros de conservación.

5.2.2. Introducción del material.

Los cítricos pueden verse afectados por hongos, bacterias, plagas y por un alto número de patógenos (virus, viroides) que se transmiten por injerto. Por ello, el movimiento de germoplasma entre distintas áreas geográficas supone un peligro por la posible introducción de plagas y enfermedades. Para evitarlo, la introducción de material cítricos está legalmente regulado en la mayoría de los países y en los más importantes la importación solo se puede realizar a través de Estaciones de Cuarentena (Krueger and Navarro, 2007). Por norma general, se exige un certificado fitosanitario de las autoridades del país de procedencia, una inspección rigurosa del material a la llegada al país de destino y medidas adicionales de aislamiento, análisis de patógenos o procedimientos de cuarentena integrales según los países (Krueger and Navarro, 2007). En el IVIA existen programas de saneamiento (material procedente del mismo país) y cuarentena (material procedente de otros países) basados en técnicas de cultivo *in vitro* (Navarro *et al.*, 1975, 1981; Navarro, 2005) que han permitido establecer unos controles fitosanitarios del material existente en el banco de germoplasma. Para realizar este tipo de programas se requiere personal especializado y unas instalaciones adecuadas (invernaderos, recintos de malla, laboratorios, etc.) que no siempre están disponibles y que en la mayoría de los casos no dependen de los bancos de germoplasma lo que dificulta la introducción de nuevo material.

5.2.3. Mantenimiento de la colección.

Como se ha comentado anteriormente, los cítricos tienen semillas recalcitrantes por lo que los recursos fitogenéticos se conservan generalmente mediante colecciones de plantas. En el caso que sea posible, las colecciones de campo deberían tener una réplica en otra ubicación. Existen también casos en los que hay colecciones duplicadas en recintos de malla, como ocurre en los Bancos de Germoplasma del IVIA y del USDA en Riverside, California. Los métodos existentes de crioconservación aún no son suficientemente efectivos para sustituir a las plantas, por lo que solo en raras ocasiones los bancos de germoplasma poseen una colección crioconservada. Las colecciones en campo deben poseer al menos dos copias de cada accesión para su caracterización y evaluación.

Las colecciones de plantas libres de patógenos mantenidas en el interior de recintos de malla se pueden utilizar como material inicial para la propagación comercial de plantas en viveros en el contexto de programas de certificación (Navarro *et al.*, 2002; Lee *et al.*, 2004; Navarro, 2005). En España, la colección protegida en recintos de malla del banco de germoplasma ha sido el origen de 144 millones de plantones sanos vendidos por los viveros a los agricultores en los últimos 30 años. Este tipo de utilización constituye una fuente de ingresos adicional para los bancos de germoplasma.

5.2.4. Caracterización y evaluación.

Este aspecto es muy importante para una buena utilización de los recursos de un banco de germoplasma. Un primer paso es el establecimiento de un pasaporte para cada accesión, que incluye información acerca de su origen, parentales, método de introducción del material (varetas, semillas, etc.) o nombre científico, de forma que cada genotipo esté bien identificado. Un segundo paso es la caracterización morfológica de los genotipos mediante descriptores adecuados, siendo los más empleados los del International Plant Genetic Resources Institute (IPGRI, 1999), que tienen descriptores de pasaporte (origen, claves de registro,...), de manejo (multiplicación, regeneración,...), de localización y características medioambientales de la colección (clima, tipo suelo, plagas y enfermedades prevalentes,...), de caracterización de los genotipos (caracteres vegetativos, de hojas, flores, frutos y semillas) y de evaluación (susceptibilidad a estreses bióticos y abióticos). La caracterización realizada con estos descriptores tiene un elevado costo por la necesidad de contar con personal especializado y de realizar la evaluación durante varios años para eliminar la influencia de las condiciones climáticas de años concretos. Además, la información proporcionada puede ser criticable debido a los posibles cambios (morfológicos, crecimiento vegetativo, etc.; Reuter and Ríos-Castaño, 1969; Reuther, 1973; Germanà and Sardo, 1988) que se pueden producir por la ubicación geográfica y el clima. Sin embargo, tienen una gran utilidad para el manejo de bancos de germoplasma concretos, ya que permiten comparar las características de los distintos genotipos en un mismo ambiente y para detectar duplicaciones de genotipos. De

hecho, la caracterización con estos descriptores es el único procedimiento fiable para comparar y en su caso descartar genotipos producidos por mutaciones espontáneas en campo, que es el origen de la mayoría de los genotipos de cítricos de las especies secundarias. En la práctica hay muy pocos bancos de germoplasma de cítricos que apliquen esta metodología, pero en el del IVIA está dando unos resultados excelentes.

La evaluación de aspectos como resistencia a factores bióticos y abióticos es prácticamente imposible de realizar en una colección completa (excepto en los casos que se puede hacer una observación directa) porque requiere la realización de experimentos específicos. Este problema podría solventarse con el establecimiento de una colección nuclear para el estudio de este tipo de caracteres.

El uso de marcadores moleculares para la caracterización y el manejo de los bancos de germoplasma está cada vez más implantado (Gulsen and Roose, 2001a, c; Varshney *et al.*, 2005; Wang *et al.*, 2005; Barkley *et al.*, 2006). Mediante la caracterización molecular se pretende clarificar las relaciones taxonómicas entre las accesiones de la colección, estudiar la estructura de la diversidad y establecer las relaciones genéticas entre entradas. Esto nos puede permitir plantear estrategias de recolección (adquisición de nuevo material) o generar una colección nuclear (caracterización y uso). Además, la caracterización molecular nos permite un mejor mantenimiento de la colección a través de la detección de redundancias, evaluación de la erosión genética e identificación de errores de etiquetado, sinonimias o homonimias (Viruel, 2010). En cítricos, el problema reside en el caso de las especies secundarias y algunos mandarinos en las que los genotipos se han originado por mutación espontanea y no se pueden distinguir mediante los marcadores moleculares existentes hasta el momento, y la única forma de distinción es por sus características morfológicas y organolépticas.

En el apartado 6 analizaremos más en profundidad los distintos tipos de marcadores moleculares empleados en cítricos, así como los recursos genómicos existentes en la actualidad.

5.2.5. Documentación y bases de datos.

Los métodos de documentación han ido evolucionando con el tiempo influenciados por la tecnología. Pese al gran uso de los ordenadores en la actualidad, todavía se siguen tomando datos en libretas de campo o papeles, que pueden perderse, deteriorase o producirse errores al pasar los datos al dispositivo electrónico. Por todo ello, se debe tener mucho cuidado en la toma de datos, copias de seguridad del material almacenado, etc. Las bases de datos son importantes para el manejo y funcionamiento de los bancos de germoplasma y también se debe tener mucho cuidado con su almacenamiento y protección. La presente tesis doctoral ha aportado una gran cantidad de información a la base de datos que se está generando en el IVIA.

5.2.6. Utilización de los recursos de los bancos de germoplasma.

El conjunto de genotipos de un banco de germoplasma ofrece un amplio recurso de genes relacionados con resistencia a enfermedades, estreses, producción, calidad del fruto, etc., que son imprescindibles para la mejora genética, la investigación y también para la propagación del material (Krueger and Navarro, 2007). Por ello, es muy importante tener disponible un germoplasma bien caracterizado para una buena planificación de los programas de mejora. Además de la mejora genética, los bancos de germoplasma nutren las investigaciones de otras disciplinas, como la fisiología, biología o la fitopatología, que a su vez pueden repercutir con sus resultados en nuevas vías de investigación para los programas de mejora.

5.2.6.1. Colecciones nucleares.

Como colección nuclear (CN) se entiende un número limitado de muestras que representan, con la menor redundancia posible, la diversidad genética de una especie cultivada (Brown, 1989). Además una CN pretende reducir los costes de mantenimiento y el uso ineficaz de una colección completa (colección base), debido a la existencia de duplicaciones y/o redundancias y la imposibilidad por su elevado costo de analizar en profundidad todos los genotipos de un banco germoplasma (Grenier *et al.*, 2000; van Hintum *et al.*, 2000). Por lo general se considera que una CN debería contener entre un 5-10 % de las accesiones presentes en la colección completa y tener representados al menos el 70% de los alelos, sin redundancias (Brown, 1989).

Antiguamente el establecimiento de CN se realizaba principalmente con datos fenotípicos y de pasaporte de las variedades, pero esto entrañaba ciertos problemas debidos a la falta de información de pasaporte, datos erróneos y los cambios debidos a efectos ambientales (Tanksley and McCouch, 1998; Hu *et al.*, 2000). Actualmente los marcadores moleculares se están convirtiendo en la herramienta más usada para el establecimiento de CN, AFLP (Fajardo *et al.*, 2002; van Treuren *et al.*, 2006), RAPD (Ghislain *et al.*, 1999; Marita *et al.*, 2000), SNP (Mckhann *et al.*, 2004) o SSR (Ellwood *et al.*, 2006; Hao *et al.*, 2006).

Se han propuesto diferentes métodos para la selección de genotipos de una CN, desde el muestreo aleatorio (Brown, 1989), al muestreo estratificado (Peeters and Martinelli, 1989; Johnson and Hodgkin, 1999). Este muestreo estratificado puede basarse en datos morfológicos, fisiológicos y agronómicos (Malosetti and Abadie, 2001), bioquímicos (Grauke *et al.*, 1995) o moleculares (Ghislain *et al.*, 1999).

Para la creación de una CN se pueden seguir, como ejemplo, los pasos descritos por van Hintum *et al.* (2000).

La principal utilidad de una CN es facilitar la caracterización y evaluación de ciertos caracteres que sería muy costoso de realizar en todos los genotipos del banco, si su

número es elevado. Además, las CN pueden estar enfocadas para distintos usos: mantenimiento de la diversidad global de una colección (Escribano *et al.*, 2008), evaluación de caracteres diversos para obtener nuevas fuentes en estudios de mejora (Yan *et al.*, 2009; Agrama and Yan, 2009), resistencias a enfermedades (Pessoa-Filho *et al.*, 2010) o estudios de genética de asociación (Pino Del Carpio *et al.*, 2011). Estos últimos pretenden buscar *loci* asociados a caracteres fenotípicos (resistencias, calidad) a nivel del genoma entero, si existe una buena cobertura de marcadores moleculares y un bajo desequilibrio de ligamiento (Zhang *et al.*, 2009), o en genes candidatos, si no hay suficiente densidad de marcadores (Fournier-Level *et al.*, 2009). Con respecto a las CN, es interesante indicar que hay algoritmos que permiten reducir la estructuración de la población y el desequilibrio de ligamiento entre *loci* asociados a esta estructura poblacional, siendo una situación favorable para realizar estudios de genética de asociación (Breseghello and Sorrells, 2006).

En cítricos, no se ha desarrollado hasta el momento ninguna CN en ninguna de las tres principales especies ancestrales (*C. maxima*, *C. medica*, *C. reticulata*) que han dado lugar a la mayor parte de la variabilidad existente de cítricos cultivados. Únicamente, Bernet *et al*. (2008) realizaron una CN en una especie secundaria (*C. aurantium*) para el estudio de la resistencia al virus de la tristeza. De las especies *C. maxima*, *C. medica* y *C. reticulata* existe una amplia diversidad en las zonas de origen, que aconsejaría el establecimiento de CN para su adecuada gestión. No obstante, la variabilidad existente de *C. maxima* y *C. medica* es escasa en la gran mayoría de los bancos de germoplasma. Por otra parte, la diversidad genética en las especies secundarias, como las naranjas, los pomelos, las limas, los limoneros y algunas mandarinas híbridas (clementinas, satsumas) es muy escasa, ya que la diversidad fenotípica existente se ha generado por mutaciones somáticas espontáneas, que generalmente no se pueden diferenciar por marcadores moleculares. Por ello en estas especies no es posible el establecimiento de CN y tan solo es aconsejable el establecimiento de colecciones basadas en caracteres morfológicos y fenotípicos. En las dos colecciones estudiadas en la presente tesis doctoral (IVIA, INRA/CIRAD), la variabilidad existente de *C. maxima* y *C. medica* es relativamente escasa, mientras que es elevada en *C. reticulata*, y además su mejora está basada en la hibridación sexual, que favorece al aumento continuo de su diversidad. Por lo tanto, en nuestro caso particular es recomendable el establecimiento de una CN de *C. reticulata,* que será posible a partir de los datos moleculares obtenidos en esta tesis doctoral.

5.2.6.2. Genética de asociación.

La genética de asociación se entiende como todo enfoque cuyo objetivo es detectar y/o localizar variables genéticas causales implicadas en la variación de un carácter de interés en un conjunto de individuos o germoplasma (Rafalski, 2010). Para realizar el

estudio de asociación se emplean marcadores moleculares para caracterizar una región de interés, así como medidas fenotípicas y en ocasiones de covariables (como pueden ser diferentes ambientes de evaluación fenotípica). Se trata por tanto, de identificar las zonas del genoma que presenten una diversidad alélica significativamente correlacionada con la variación del carácter (Zhu *et al.*, 2008). La genética de asociación se conoce también como mapeo del desequilibrio de ligamiento (LD), que explota la variación fenotípica y genética presente en una población natural, diferente de lo que se conoce como mapeo de genes que controlan caracteres cuantitativos o QTLs (Quantitative Trait Loci), que están basados en poblaciones segregantes. Estudios de genética de asociación basados en el LD han tenido éxito en diversas especies de plantas cultivadas, como maíz (Thornsberry *et al.*, 2001), *Arabidopsis* (Zhao *et al.*, 2007) o sorgo (Casa *et al.*, 2008), que poseen unas colecciones de germoplasma muy amplias.

Un problema que se presenta en este tipo de estudios es la estructura poblacional (Yu and Buckler, 2006; Abdurakhmonov and Abdukarimov, 2008), que puede provocar asociaciones erróneas, debido a una estratificación fuerte de la población. Factores como los sistemas de mejora y la historia de la domesticación de los cultivos, son determinantes en la estructura poblacional del LD.

La resolución del mapeo por LD en una población, así como la densidad de marcadores moleculares necesarios y los métodos estadísticos a emplear, depende de la diversidad genética, la extensión del desequilibrio y las relaciones existentes entre los individuos de una población (Zhu *et al.*, 2008). En el caso de especies autógamas, la extensión del LD es alto (Arabidopsis, Nordborg *et al.*, 2002; arroz, Garris *et al.*, 2003; y sorgo, Deu and Glaszmann, 2004), lo que conlleva una resolución del mapeo baja, pero la densidad de marcadores necesaria es menor. Por el contrario, en las especies alógamas, con posibilidad de polinización cruzada [maíz, Remington *et al.*, 2001; álamo, Ingvarsson (2005); abeto noruego, Rafalski and Morgante (2004)] el LD decae en distancias cortas, por lo que la resolución del mapeo se espera alta, pero un elevado número de marcadores es necesario.

Por todo ello, es muy importante establecer la distancia a lo largo de la cual permanece el LD en una población, así como su estructura, para conocer la viabilidad de un estudio de genética de asociación. En cítricos, no se tiene un conocimiento previo a la realización de la presente tesis doctoral acerca del alcance del LD en el género *Citrus* ni tampoco a nivel intraespecífico.

6. HERRAMIENTAS MOLECULARES EXISTENTES EN CÍTRICOS.

Como se ha comentado con anterioridad, la caracterización molecular del material vegetal es fundamental tanto para el manejo de los recursos fitogenéticos existentes en un banco de germoplasma, como para ayudar a la mejora genética. Por ello, en los siguientes apartados se explican las distintas herramientas y recursos existentes para el estudio de los cítricos.

6.1. Marcadores moleculares.

En las pasadas dos décadas la evolución en el diseño de marcadores moleculares para estudios de diversidad y de filogenia (entre otros usos) en plantas ha sido importante. En cuanto a los marcadores empleados en cítricos, podemos citar los siguientes ejemplos:

AFLP (Amplification Fragment Lenght Polymorphism). Se basan en la detección de fragmentos de restricción de ADN mediante amplificación por PCR, utilizando cebadores homólogos a la secuencia de los adaptadores y de las dianas de restricción de las enzimas utilizadas previamente para digerir el ADN. Entre las ventajas de los AFLPs destacan su abundancia, generan gran cantidad de bandas por PCR (50 a 100 fragmentos) y no se necesita información previa de secuencia. Su principal problema radica en que son dominantes, los patrones de bandas no siempre son claros y en el caso de que los fragmentos producidos tengan el mismo tamaño, no significa que sean homólogos (Pang *et al.*, 2007). Scarano *et al.*, (2003) los utilizaron en combinación con marcadores SSR para la identificación de plántulas cigóticas de limón. Pang *et al.* (2007) los emplearon para un estudio de filogenia en el género *Citrus* y afines.

IRAP (Inter-Retrotransposon Amplified Polymorphism). Estos marcadores se obtienen por amplificación de un fragmento del genoma situado entre dos retrotransposones (elementos móviles que se encuentran en gran número distribuidos aleatoriamente en el genoma de las plantas) mediante cebadores específicos de secuencia homologa a este. Su gran ventaja es que son muy polimórficos, pero muchos de estos polimorfismos son dominantes. Bretó *et al.* (2001) observaron en cítricos que los polimorfismos basados en elementos transponibles son más abundantes que los basados en marcadores de secuencia aleatoria o en microsatélites. Podrían ser útiles para la diferenciación varietal dentro de grupos como naranja o clementina. Posteriormente, Biswas *et al.* (2010) los emplearon en el análisis genético de 48 variedades del género *Citrus* y géneros afines.

ISSR (Inter Simple Sequence Repeat). Se basan en secuencias repetidas en tándem, microsatélites, en base a los cuales, se diseña un cebador con secuencia homologa y al que se le añaden dos nucleótidos aleatorios extras. De esta manera se amplifica una región situada entre dos microsatélites cercanos que incluyan los nucleótidos complementarios. No es necesario tener información previa de la secuencia y presentan una alta reproducibilidad. Como desventajas, la homología de las bandas es incierta y son marcadores dominantes. Fang *et al.*

(1997, 1998), Gulsen and Roose (2001a, b) y Yang *et al.* (2010), son algunos de los que han empleado estos marcadores en cítricos, entre otras cosas, para estudios de diversidad genética y las relaciones filogenéticas entre especies del género *Citrus*.

RAPD (Random Amplified Polymorphic DNA). Estos marcadores se basan en la amplificación del ADN genómico mediante PCR utilizando un único cebador (10 nucleótidos) de secuencia aleatoria. Se pueden obtener en gran cantidad y en poco tiempo al no necesitar información de secuencia previa. Fueron de los primeros marcadores de ADN usados en cítricos (Luro *et al.*, 1995; Federici *et al.*, 1998) los emplearon en el análisis de 32 accesiones de cítricos y tres de *Microcitrus* y Nicolosi *et al*. (2000) analizó 36 accesiones pertenecientes al género *Citrus* y una de cada uno de los géneros afines, *Poncirus*, *Fortunella*, *Microcitrus* y *Eremocitrus*. También fueron usados para la diferenciación de plantas zigóticas y nucelares de tangerina (Bastianel *et al.*, 1998) y de naranjo amargo (Rao *et al.*, 2008) y en algunos estudios de mapeo genético (de Oliveira *et al.*, 2004; Gulsen *et al.*, 2010). Su mayor problema es la baja reproducibilidad y su dominancia, por lo que actualmente son poco utilizados.

CAPS (Cleaved Amplified Polymorphic Sequences). El polimorfismo se detecta mediante la digestión de un fragmento de ADN amplificado por PCR que puede separarse en un gel de poliacrilamida. Permiten detectar polimorfismos de tipo SNPs o InDels. Son marcadores codominantes y reproducibles, que necesitan poca cantidad de ADN. En cítricos fueron aplicados para estudios de los genomas citoplásmicos (Lotfy *et al.*, 2003) y nucleares (Omura *et al.*, 2000).

RFLP (Restriction Fragment Length polymorphism). Detectan fragmentos de ADN polimórficos que sirven de dianas para enzimas de restricción. Suelen segregar como marcadores codominantes, se encuentran en cualquier región del genoma y son altamente polimórficos. Las desventajas son su elevado coste económico, la necesidad de usar bastante tiempo para analizar los datos y de un conocimiento previo de secuencias. Luro *et al.* (1995) los emplearon junto con los RAPD en la diferenciación entre plántulas de origen cigótico o nucelar, así como para evaluar la variabilidad genética intraespecífica en naranjas y mandarinas. Federici *et al.* (1998) los emplearon en el análisis de 88 accesiones representantes del género *Citrus*, algunos híbridos y especies de géneros afines. Cai *et al.* (1994) los emplearon para establecer los primeros mapas genéticos de cítricos.

SSRs (Simple Sequence Repeats). Estos marcadores, también llamados microsatélites, son secuencias de repeticiones en tándem que se presentan de forma consecutiva en un número variable, por lo que tienen un alto nivel de polimorfismo. Además, se comportan como marcadores codominantes y están dispersos aleatoriamente en el genoma. Todo esto les ha permitido ganar mucha importancia en genética de plantas por su reproducibilidad entre laboratorios. En cítricos y géneros relacionados, los SSRs se han desarrollado a partir de genotecas genómicas (Kijas *et al.*, 1995; Ahmad *et al.*, 2003; Novelli *et al.*, 2006; Froelicher *et al.*, 2008), ESTs (Bausher *et al.*, 2003; Chen *et al.*, 2008; Luro *et al.*, 2008) y secuencias de BACend (Ollitrault *et al*., 2010). Estos marcadores han demostrado ser

muy útiles en estudios de diversidad genética de cítricos en combinación con observaciones fenotípicas (Kijas *et al.*, 1995; Luro *et al.*, 2001, 2008; de Oliveira *et al.*, 2002; Corazza-Nunes *et al.*, 2002; Pang *et al.*, 2003; Golein *et al.*, 2005; Barkley *et al.*, 2006). El principal inconveniente es la necesidad de conocer la secuencia adyacente al microsatélite. Además, Barkley *et al.* (2009) mostraron que la homoplasia puede limitar la utilidad de los marcadores microsatélites en la identificación del origen filogenético de los fragmentos de ADN.

SNPs (Single Nucleotide Polymorphisms). Son polimorfismos de variación en la secuencia de ADN de un solo nucleótido. Su abundancia y distribución a lo largo del genoma (Brookes, 1999) les otorga una ventaja frente a otros marcadores, además de su reproducibilidad entre laboratorios. Son marcadores de tipo codominante, pero requieren un conocimiento previo de la secuencia a analizar. Con las técnicas actuales de secuenciación, cada vez más económicas, los están convirtiendo en unos marcadores muy importantes para el desarrollo de mapas genéticos saturados, identificación de cultivares, detección de asociaciones genotipo/fenotipo o selección asistida por marcadores (Botstein and Risch, 2003; Morales *et al.*, 2004; Xing *et al.*, 2005; Lijavetzky *et al.*, 2007; Ollitrault *et al.*, 2012b). En cítricos, se han realizado diversos trabajos de detección de SNPs por secuenciación en naranja Novelli *et al.* (2006), en clementina (Terol *et al.*, 2008) y satsumas (Dong *et al.*, 2010). Una vez identificados los SNPs se puede hacer un genotipado masivo con micromatrices (Ollitrault *et al.*, 2012a) o genotipados basados en PCR competitiva entre alelos (KASPar). Otras técnicas utilizadas para genotipado son los SSCPs (Single Strand Conformation Polymorphisms; Olivares-Fuster *et al.*, 2007; Simsek *et al.*, 2011) que se basan en las diferencias en la conformación de la estructura terciaria del ADN que el cambio nucleotídico produce. Se pueden analizar un elevado número de muestras en poco tiempo, pero requiere unas condiciones rigurosas de ensayo y una secuenciación previa.

Indel (Inserción o Deleción). Estos marcadores por lo general surgen de la inserción de retrotransposones u otros elementos móviles, por el desfase de una secuencia simple en la replicación o eventos de retrocruzamiento desiguales. Generalmente tienen poca frecuencia de homoplasia, y además, hay poca probabilidad de que dos mutaciones por indel ocurran en el mismo lugar y con la misma longitud, por lo que permiten una identidad descendiente a descendiente (Britten *et al.*, 2003). Otra ventaja es su fácil genotipado mediante PCR y electroforesis (Vasemägi *et al.*, 2010). Este tipo de marcadores ha sido utilizado en estudios genéticos en trigo (Raman *et al.*, 2006), arroz (Hayashi *et al.*, 2006) y poblaciones naturales (Väli *et al.*, 2008), pero no en cítricos. En la presente tesis se ha demostrado su aplicabilidad e interés para estudios filogenéticos en cítricos (Garcia-Lor *et al.*, 2012a) con indels identificados a nivel intra e interespecífico en secuencias de genes. Otros InDels han sido desarrollados recientemente en 'Clemenules' a partir de secuencias de BACend (Ollitrault *et al.*, 2012a).

Marcadores de polimorfismos en secuencia de ADN cloroplástico (ADNcp) (Abkenar *et al.*, 2004; Nicolosi *et al.*, 2000; Jung *et al.*, 2005; de Araújo *et al.*, 2003) y de ADN mitocondrial (ADNmt) (Froelicher *et al.*, 2011) han permitido realizar estudios de filogenia maternal, ya que

estos genomas heredados de la madre son muy conservados. También se han empleado con este propósito los mencionados SSCP (Olivares-Fuster *et al.*, 2007).

Pese a existir una gran información con diversos tipos de marcadores, en la presente tesis doctoral se pretende ampliar los recursos moleculares en cítricos mediante el desarrollo de nuevos marcadores moleculares (SSRs, indels, SNPs) y su aplicación para la caracterización de especies del género *Citrus* y afines.

6.2. Recursos genómicos de cítricos.

Los primeros datos genómicos en cítricos se publicaron en algunas revisiones como las de Gmitter *et al.* (2007), Talon and Gmitter Jr. (2008) and Tadeo *et al.* (2008). Actualmente, los recursos genómicos (http://www.citrusgenome.ucr.edu/, 2004; http://www.citrusgenomedb.org/, 2009) incluyen más de medio millón de ESTs , la mayoría de naranjo dulce (≈90%), seguido por las procedentes de clementina (Forment *et al.*, 2005; Terol *et al.*, 2008), *Poncirus*, satsuma y otras variedades (Shimizu *et al.*, 2009; Delseny *et al.*, 2010); también existen micromatrices de alta densidad en distintas plataformas para estudios de expresión y de genotipado (Shimada *et al.*, 2005; Terol *et al.*, 2007; Martinez-Godoy *et al.*, 2008; Shimizu *et al.*, 2011; Ollitrault *et al.*, 2012a), varias librerías de BACs (Terol *et al.*, 2008), un mapa físico de naranjo dulce y mapas de ligamiento para clementina, naranjo dulce y zamboa entre otros (Ollitrault *et al.*, 2012b; http://www.citrusgenomedb.org/tools/map/cmap).

Además, se ha secuenciado el genoma de un haploide de clementina mediante la tecnología Sanger y un genoma diploide de naranjo dulce mediante la técnica de pirosecuenciación 454 de Roche. Estos recursos se encuentran disponibles en el portal phytozome (2011) del instituto JGI (http://www.phytozome.net/) y en la base de datos del genoma de cítricos (http://www.citrusgenomedb.org/). Paralelamente se ha publicado el genoma de naranjo dulce con la plataforma de Illumina GAII (Xu *et al.*, 2012) y otros genomas (Shimizu *et al.*, 2012; Terol *et al.*, 2012). En la web “http://citrus.hzau.edu.cn/” (2011) se encuentra anotado el genoma de la naranja. Todos estos recursos y herramientas permitirán a los genéticos y mejoradores utilizar más eficazmente distintas características de los cítricos en los programas de mejora.

OBJETIVOS

OBJETIVOS DE LA TESIS DOCTORAL

El conocimiento del origen de las especies cultivadas del género *Citrus* está actualmente bien establecido, considerándose como especies ancestrales a *C. reticulata*, *C. maxima*, *C. medica* y *C. micrantha* (Barret and Rhodes, 1976, Nicolosi *et al.*, 2000; Krueger and Navarro, 2007). Cruzamientos entre estas cuatro especies han dado lugar a la mayoría de especies secundarias e híbridos recientes. Sin embargo, pese a toda esta información, las relaciones filogenéticas entre las especies ancestrales del género *Citrus* y de los géneros afines de los cítricos verdaderos no son bien conocidos. De hecho, los géneros *Citrus*, *Fortunella*, *Poncirus*, *Microcitrus, Eremocitrus* y *Clymenia*, pese a tener una diferenciación morfológica evidente, son sexualmente compatibles, lo que les ha permitido cruzarse y generar nueva variabilidad a lo largo de la historia. Sin embargo, el nivel de diferenciación genética entre y dentro de los taxones básicos del género *Citrus* y los géneros afines no está bien definido.

Tampoco está clara cuál es la organización filogenética del genoma de las especies secundarias y si la organización que resulta de la evolución de los cítricos cultivados es compatible con estudios de asociación entre diversidad fenotípica y polimorfismo molecular basados en el desequilibrio de ligamiento (LD) o en el origen filogenético de genes candidatos.

Como se ha comentado anteriormente existen dos clasificaciones principales de los cítricos, las de Swingle y Tanaka, que tienen una visión muy diferente del grupo mandarino. El germoplasma de mandarino, clasificado como *C. reticulata* por Swingle and Reece (1967), tiene como centro de diversificación Asia, desde Vietnam a Japón. Es un taxon muy polimórfico, tanto con marcadores moleculares (Luro *et al.*, 2004), como caracteres fenotípicos (morfología o tolerancia a factores bióticos y abióticos). Algunos autores suponen que el germoplasma mandarino esta introgresado por otras especies (Barkley *et al.*, 2006). Pese a ello, hay muy pocos datos disponibles respecto a la organización interespecífica del grupo mandarino y los determinantes de su diversidad intraespecífica. Esta información es fundamental para optimizar el manejo y la utilización de los recursos existentes en las colecciones de germoplasma y para establecer en un futuro próximo una colección nuclear, con la finalidad de facilitar la realización de estudios de evaluación de diversos caracteres fenotípicos, de resistencia o tolerancia a estreses bióticos y abióticos y abordar estudios de genética de asociación. También facilitará la selección de parentales para utilizar en los programas de mejora genética.

Para aportar nuevos conocimientos en (1) la organización genética del genero *Citrus* y su compatibilidad con estudios de genética asociación, (2) la filogenia de los cítricos verdaderos y la implicación de la evolución en el polimorfismo de genes candidatos para caracteres de calidad y (3) la estructura genética y el origen del germoplasma mandarino, la tesis doctoral se enfoca a los siguientes objetivos específicos:

1. Estudio de la organización de la diversidad genética en el género *Citrus*.

Considerando los posibles problemas de homoplasia que pueden presentar los SSRs se pretende comparar el valor de marcadores nucleares de tipo microsatélite (SSRs) y de Inserción-Deleción (indels), para estudios de diversidad a nivel inter- e intraespecífico. Los marcadores indel no se han desarrollado hasta el momento en cítricos y se cree que podrían ser muy útiles para estudiar la diversidad genética interespecífica y el origen filogenético de las especies. Una colección de 90 genotipos representativos de tres especies ancestrales (*C. reticulata*, *C. maxima* y *C. medica*) y de especies cultivadas de cítricos será genotipada con estos dos tipos de marcadores, para estimar la contribución de cada especie ancestral al genoma de las especies secundarias. El posicionamiento de los marcadores moleculares en un mapa genético nos permitirá establecer el nivel de desequilibrio de ligamiento (LD) dentro del género *Citrus* tanto a nivel inter- como intracromosómico y por lo tanto la posibilidad de realizar estudios de asociación en este género.

Además, entre todos los marcadores empleados, se pretende seleccionar un pequeño grupo que se encuentren dispersos en el genoma y representen de manera fiable la diversidad existente en los cítricos, para poder realizar genotipados sistemáticos de colecciones de forma rápida y económica.

2. Estimación del nivel de diferenciación genómica entre los cítricos verdaderos y su filogenia nuclear; evolución y herencia de genes candidatos para calidad en las especies cultivadas.

Se pretende identificar polimorfismos mediante la secuenciación de genes candidatos (determinantes de la calidad de los cítricos y algunos involucrados en la respuesta a diferentes estreses) para estudiar la filogenia de los cítricos verdaderos, la estructura de las especies secundarias del género *Citrus*, su origen y su filogenia, los posibles eventos de diferenciación a nivel evolutivo en los genes candidatos y la identificación de un conjunto de SNPs con un fuerte poder de diferenciación filogenético, que sirvan de herramienta para estudios futuros.

La transferibilidad de este conjunto de SNPs, obtenido en un panel de especies reducido, a los cítricos verdaderos y a géneros lejanos (subfamilia Aurantioideae) mediante un método de “PCR competitiva entre alelos específicos” será objeto de estudio en la presente tesis doctoral.

3. Determinación de la estructuración de la diversidad del germoplasma de mandarino.

Con la finalidad de tener un mejor conocimiento del origen y de la organización genética del grupo mandarino, se ha realizado un amplio estudio con marcadores nucleares

(SSR y indel) y mitocondriales de 198 genotipos de mandarino, junto con 25 genotipos representativos de las otras especies de cítricos verdaderos, existentes en los bancos de germoplasma del *Institut National de la Recherche Agronomique* (INRA) y del Instituto Valenciano de Investigaciones Agrarias (IVIA). Se pretende estimar la introgresión de otras especies en el germoplasma mandarino e identificar mandarinos verdaderos.

Una vez definida la estructura del germoplasma de mandarinos, se procederá a la cuantificación de la contribución de los grupos observados (además de los genomas de especies ancestrales) al resto de genotipos en estudio. Los resultados obtenidos se usarán para confirmar o desmentir la clasificación existente en las bases de datos, así como detectar redundancias en las colecciones.

Este estudio de estructuración será la base de futuros análisis de desequilibrio de ligamiento (LD) con perspectivas de realizar estudios de genética de asociación, y además de la implementación de una colección nuclear de mandarinos.

Una vez definidos los objetivos, los resultados de la presente tesis doctoral se han estructurado en los siguientes capítulos, que corresponden a artículos publicados en revistas científicas:

CAPITULO 1: **Comparative use of indel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies.** *Molecular Genetics and Genomics* (2012) 287: 77–94. Objetivo 1.

CAPITULO 2: **A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species.** *Annals of Botany* (2013) 111: 1-19. Objetivo 2.

Annex chapter 2: *Clymenia*'s phylogeny within the 'true citrus fruit trees'.

CAPITULO 3: ***Citrus*** **(Rutaceae) SNP markers based on Competitive Allele-Specific PCR; transferability across the Aurantioideae subfamily.** *Applications in Plant Sciences* (2013) 4: doi:10.3732/apps.1200406. Objetivo 2.

CAPITULO 4: **Genetic diversity analysis and population-structure analysis of mandarin germplasm by nuclear (SSRs, indel) and mitochondrial markers.**

Submitted. Objetivo 3.

CHAPTER 1

Comparative use of indel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies.

Andres Garcia-Lor, François Luro, Luis Navarro and Patrick Ollitrault

Molecular Genetics and Genomics (2012) 287: 77–94

Abstract

Genetic stratification associated with domestication history is a key parameter for estimating the pertinence of genetic association study within a gene pool. Previous molecular and phenotypic studies have shown that most of the diversity of cultivated citrus results from recombination between three main species: *C. medica* (citron)*, C. reticulata* (mandarin) and *C. maxima* (pummelo). However, the precise contribution of each of these basic species to the genomes of secondary cultivated species, such as *C. sinensis* (sweet orange)*, C. limon* (lemon), *C. aurantium* (sour orange)*, C. paradisi* (grapefruit) and recent hybrids is unknown. Our study focused on: (1) the development of Insertion-Deletion (indel) markers and their comparison with SSR markers for use in genetic diversity and phylogenetic studies; (2) the analysis of the contributions of basic taxa to the genomes of secondary species and modern cultivars and (3) the description of the organisation of the citrus gene pool, to evaluate how genetic association studies should be done at the cultivated citrus gene pool level.

Indel markers appear to be better phylogenetic markers for tracing the contributions of the three ancestral species, whereas SSR markers are more useful for intraspecific diversity analysis. Most of the genetic organisation of the *Citrus* gene pool is related to the differentiation between *C. reticulata*, *C. maxima* and *C. medica*. High and generalised LD was observed, probably due to the initial differentiation between the basic species and a limited number of interspecific recombinations. This structure precludes association genetic studies at the genus level without developing additional recombinant populations from interspecific hybrids. Association genetic studies should also be affordable at intraspecific level in a less structured pool such as *C. reticulata*.

INTRODUCTION

Genetic association studies based on linkage disequilibrium (LD) are similar to quantitative trait locus (QTL) mapping. However, whereas QTL mapping considers only variations between two crossed individuals, LD mapping exploits the phenotypic and genetic variation present across a natural population. This method has been successfully applied in studies of cultivated plants (Thornsberry *et al.*, 2001; Casa *et al.*, 2008; Zhu *et al.*, 2008). However, the presence of population stratification and an unequal distribution of alleles within these groups can result in spurious associations (Abdurakhmonov and Abdukarimov, 2008). Breeding systems and domestication history are determinant factors of the LD structure in cultivated species germplasm. The extent of LD is generally higher for species with selfing mating system (Arabidopsis, Nordborg *et al.*, 2002; rice, Garris *et al.*, 2003; and sorghum, Deu and Glaszmann 2004) than for outcrossing organisms (maize, Remington *et al.*, 2001; populus, Ingvarsson 2005; and Norway spruce, Rafalski and Morgante 2004). To our knowledge, no data are available for LD in agamic complexes.

Citrus is one of the most important fruit crops in the world, and its diversity (Krueger and Navarro, 2007) and origin (Webber *et al.*, 1967; Calabrese, 1992) have been widely studied. The taxonomy of citrus remain controversial, due to the conjunction of broad morphological diversity, total sexual interspecific compatibility within the genus and partial apomixis of many cultivars. Fixing complex genetic structures through seedling propagation via apomixis has led some taxonomists to consider clonal families of interspecific origin as new species (Scora 1975). Two major systems are widely used to classify *Citrus* species: the Swingle and Reece (1967) classification that considers 16 species and Tanaka's (1961) one that identifies 156 species. More recently, Mabberley (1997) proposed a new classification of edible citrus recognising 3 species and four hybrid groups. In this paper, we will use the Swingle and Reece (1967) classification system. Indeed, this taxonomic system is widely used in the citrus scientist community and, as mentioned below, mostly agrees with molecular data.

Despite the difficulties involved in establishing a consensual classification of edible citrus, most authors now agree on the origins of most cultivated forms. Early studies by Scora (1975) and Barrett and Rhodes (1976) based on biochemical and morphological polymorphisms, respectively, suggested that most of the cultivated citrus originated from three main species (*C. medica* L., citrons; *C. reticulata* Blanco, mandarins; and *C. maxima* (Burm.) Merr., pummelos). More recent studies involving the diversity of morphological characteristics (Ollitrault *et al.*, 2003) and secondary metabolites (Fanciullino *et al.*, 2006a) confirmed that the majority of the phenotypic diversity of edible citrus results from the differentiation between these three basic taxa. Isoenzymes (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003), RFLP (Federici *et al.*, 1998), RAPD, SCAR (Nicolosi *et al.*, 2000), AFLP (Liang *et al.*, 2007) and SSR (Luro *et al.*, 2001; Barkley *et al.*, 2006) molecular markers generally support the following conclusions for the origin of the other cultivated *Citrus* species (Nicolosi, 2007): (1) *C. sinensis* (L.) Osb. (sweet oranges) and *C. aurantium* L. (sour oranges) are related with *C. reticulata* but display

introgressed traits and markers of *C. maxima.* The closer relation with *C. reticulata* suggests that they are not direct hybrids but are probably backcrossed hybrids of first or second generation crosses with the *C. reticulata* gene pool. Analysis of chloroplastic (Green *et al.*, 1986, Nicolosi *et al.*, 2000) and mitochondrial genomes (Froelicher *et al.*, 2011) indicate a *C. maxima* maternal phylogeny. (2) *C. paradisi* Macf. (grapefruits) is close to *C. maxima*, and could result from hybridization between *C. maxima* and *C. sinensis* (Barrett and Rhodes 1976, Scora *et al.*, 1982, de Moraes *et al.*, 2007). (3) *C. medica* is clearly a progenitor of *C. aurantifolia* (Christm.) Swing (limes) and *C. limon* Osb. (lemons). Chloroplast and nuclear data analysis indicate that the genetic pools of *C. reticulata* and *C. maxima* also contributed to the genesis of *C. limon*. Nicolosi *et al.* (2000) proposed that this species resulted from direct hybridisation between *C. aurantium* and *C. medica*. This assumption is supported by Gulsen and Roose (2001a) and Fanciullino *et al.* (2007). The origin of *C. aurantifolia* is more controversial. However, molecular data (Federici *et al.*, 1998; Nicolosi *et al.*, 2000) support the hypothesis of Torres *et al.* (1978) that the Mexican lime is a hybrid between *C. medica* and a *Papeda* species. Nicolosi *et al.* (2000) proposed that *C. micrantha* might be the parental *Papeda*. These previous molecular studies have provided a better understanding of citrus maternal phylogeny, hybrid origin and parentage determination of many species. However, little is known about the precise contribution of the basic edible species to the nuclear genome constitution of secondary cultivated species (*C. sinensis, C. limon, C. aurantium, C. paradisi* and *C. aurantifolia*) and recent hybrids from twentieth century breeding programs. Furthermore, the impact of this domestication history on global genetic organisation and the extent of linkage disequilibrium (LD) on the *Citrus* gene pool have not been studied. The distance over which LD persists is a fundamental parameter to determine how association studies may be conducted on a gene pool. Regarding the important phenotypic differentiation between the basic taxa and the interspecific origin of most cultivated citrus, a better knowledge of the contribution of the nuclear genome of the basic taxa to the secondary species and modern cultivated citrus, as well as the analysis of the LD extent, appear as prerequisites to undergo association studies in the *Citrus* gene pool.

Among the codominant markers used for citrus genetic studies, simple sequence repeats (SSRs) (Luro *et al.*, 2001, 2008; Gulsen and Roose, 2001a; Barkley *et al.*, 2006; Ollitrault *et al.*, 2010) are regarded as powerful tools because they are highly polymorphic, codominant, generally locus-specific and randomly dispersed throughout the plant genome. Thus, the use of mapped SSR markers should be particularly useful to analyse the extent of LD. However, Barkley *et al.* (2009) showed that homoplasy may limit the usefulness of SSR markers in identifying the phylogenetic origin of DNA fragments in citrus. Insertion or deletion (indel) markers generally have low frequency of homoplasy. Indeed, there is a sufficiently low probability of two indel mutations of exactly the same length occurring at the same genomic position, that shared indels can confidently be related to identity-by-descent. In general, indels arise from the insertion of retroposons or other mobile elements, slippage in simple sequence replication or unequal crossover events (Britten *et al.*, 2003). At the technical level, indels can

be genotyped with simple procedures based on size separation after targeted PCR (Vasemägi *et al.*, 2010). Indels have been used successfully for genetic studies in wheat (Raman *et al.*, 2006), rice (Hayashi *et al.*, 2006) and natural populations (Väli *et al.*, 2008).

Our study focused on three basic species (*C. medica*, *C. reticulata* and *C. maxima*), the secondary species that they generated (*C. sinensis*, *C. aurantium*, *C. paradisi* and *C. lemon*) and some known or putative interspecific hybrids. Twelve indel markers were developed from gene sequencing, and their polymorphism organisation was compared with 50 SSR markers. Next, the complete set of markers was used to answer the following three questions: (1) what is the intraspecific diversity of indel markers and are they more useful than SSRs as tag of DNA fragments in studies of phylogenetic origin? (2) What is the contribution of the three basic edible taxa to the genomes of secondary species and modern cultivars? (3) Are the genetic organisation of the *Citrus* gene pool and the extent of linkage disequilibrium adapted for association genetics? Furthermore, we propose a subset of markers (core markers) for quick and inexpensive systematic germplasm genotyping that maintains most of the organisation and intraspecific polymorphism information.

MATERIALS AND METHODS

Interspecific indel polymorphism research

Plant material and DNA extraction

With the objective to identify indel polymorphism differences between the basic citrus taxa, we selected two cultivars of *C. medica* (Corsican and Buddha's hand citrons), two cultivars of *C. reticulata* (Cleopatra and Willow Leaf mandarins) and two cultivars of *C. maxima* (Chandler and Pink pummelos). High molecular weight genomic DNA was extracted from leaf samples using the DNeasy Plant Mini Kit (Qiagen S.A.; Madrid, Spain) according to the manufacturer's instructions.

Gene sequence amplification and sequencing

Primers were designed from EST sequences corresponding to 16 genes available in public databases. Thirteen genes [chalcone isomerase (CHI), chalcone synthase (CHS), flavonol synthase (FLS), malic enzyme (EMA), malate dehydrogenase (MDH), vacuolar citrate/H+ symporter (TRPA), phosphoenolpyruvate carboxylase (PEPC), phosphofructokinase (PKF), lycopene β-cyclase (LCY2), β-carotene hydroxylase (Hy-b), phytoene synthase (PSY), 1-deoxyxylulose 5-phosphate synthase (DXS) and lycopene β-cyclase (LCYB)] are involved in primary and secondary metabolite biosynthesis pathways that determine the quality of citrus fruit (sugars, acids, flavonoids and carotenoids). In addition, 3 candidate genes for salt tolerance [CAX1 (cation/H+ membrane antiporter), AtGRC (raffinose synthase) and AVP (vacuolar H+ pyrophosphatase)] were used. Primers (Table 1) were designed to amplify fragments with a length between 166 and 1,201 bp. The PCR mixture consisted of 1 ng/μl template DNA, 0.2 mM dNTPs, 0.2 μM forward primer, 0.2 μM reverse primer, 10x PCR buffer (Fermentas), 1.5 mM $MgCl_2$ and 0.027 U/μl *Taq* DNA polymerase (Fermentas), in a final volume of 15 μl. PCR reactions were carried out with the following program: 5 min at 94°C; 40 cycles of 30 s at 94°C, 30 s at 50-58°C and 2 min at 72°C with a final extension of 4 min at 72°C.

Amplicons of the six selected genotypes were sequenced by the Sanger method from the 5' end using dideoxynucleotides labelled by fluorescence (Big Dye Terminator Cycle Sequencing Kit v3.1). The sequencing reaction was carried out in a thermal cycler (ABI GeneAmp PCR System 9700), and the resolution and analysis of the labelled products were performed in a capillary sequencer (ABI 3100).

Table 1. Primers of candidate genes

Process involved	Gene	Primers	AT	High-quality sequence (bp)	EST size (bp)	Genomic size (bp)	Genebank accessions
Flavonoids biosynthesis	Chalcone isomerase	F:TTGTTCTGATGGCCTAATGG R:AAAGGCTGTCACCGATGAAT	55	647	721	721	aCL6103Contig1
	Chalcone synthase	F:GATGTTGGCCGAGTAATGCT R:ATGCCAGGTCCAAAAGCTAA	55	565	659	659	aCL6909Contig1
	Flavonol synthase	F:GGAGGTGGAGAGGGTCCAAG R:GGGCCACCACTCCAAGAGC	55	710	763	763	AB011796
Acids biosynthesis	Malic enzyme	F:ACATGACGACATGCTTCTGG R:CGTAGCCACGCCTAGTTCAT	55	420	166	420	CB417399
	Malate dehydrogenase	F:ATGGCCGCTACATCAGCTAC R:TGCAACCCCCTTTTCAATAC	55	705	1209	1250	DQ901430
	Vacuolar citrate/H+ symporter	F:GGCGCCACTCCTACCTTCCC R:CGGTCATTGAAGAGTGCTCCCC	58	715	987	1300	EF028327
Sugars biosynthesis	Phosphoenolpyruvate carboxylase	F:AGCCAATGGGATTTCTGACA R:GCCAAGCCACACAGGTAAAT	55	669	1201	2000	EF058158
	Phosphofructokinase	F:CGCCGACCTCAGTCCCGTC R:GCTGCACGCCCCATAAGCCG	58	630	807	1650	AF095520
Carotenes biosynthesis	Lycopene β-cyclase	F:GCATGGCAACTCTTCTTAGCCCG R:AGCTCGCAAGTAAGGCTCATTCCC	55	725	850	850	FJ516403
	β-Carotene hydroxylase	F:AGCCCTTCTGTCTCCTCACA R:CCGTGGAATTTATCCGAGTG	55	675	787	1600	AF315289 AF296158
	Phytoene synthase	F:GCTCGTTGATGGGCCTAATGC R:CGGGCGTAAGAGGGATTTTGC	58	560	727	2100	AB037975, AF220218 AF152892
	1-deoxyxylulose 5-phosphate synthase	F:GGCGAGGAAGCGACGAAGATGG R:GGATCAGAACTGGCCCTGGCG	58	590	935	1500	aCL303Contig1
	Lycopene β-cyclase	F:GAATTCTTGCCCCAAGTTCA R:TATGGGCCACAAATCTTTCC	55	710	1206	1500	AY166796, AF152246 AY644699
Salt stress tolerance	Cation/H+ membrane antiporter	F:GTTGCTGATGCTACAGATG R:CCTCTCTCTCTTCTTTACCG	50	840	805	1800	aCL1735Contig1
	Raffinose synthase	F:CATGCGGAAAAGATGTACC R:CAGCAAGGCTGTCCATAAC	52	740	804	1800	aCL3302Contig1
	Vacuolar H+ pyrophosphatase	F:GCATATGCTCCCATCAGTG R:CAGGCTCCTGTCTGTTTGAG	53	800	831	1650	aCL5319Contig1

High-quality sequence resulted from cleaning the alignments. aCLxxxxContig1, sequences were obtained from the Citrus Functional Genomics Project (CFGP), http://bioinfo.ibmcp.upv.es/genomics/cfgpDB/; the rest of the sequences were obtained from the National Center for Biotechnology Information (NCBI).
(AT) Annealing Temperature.

Indel identification and design of new primers for diversity studies

BioEdit (Hall, 1999) was used to align sequences from which indel polymorphisms were identified. For genes with indel polymorphisms, new primer pairs in conserved regions flanking the indel polymorphism were designed using Primer3 software (http://biotools.umassmed.edu/bioapps/primer3) (Table 2) to amplify fragments smaller than 350 bp that were subsequently analysed in a capillary fragment analyser (see below).

Table 2. Characteristics of indel markers

Marker name	Gene	Primers	AT	Fragment size (bp)
IDCHI	Chalcone isomerase	F:TTTCCTCTTGCTTTACGTGT R:GTCACAGGTAACGGATTTTC	55	146-196
IDEMA	Malic enzyme	F:CTCTTTCTGCTTCCTGACATC R:GCCGGTGAATAAAACACAAC	55	263-277
IDTRPA	Vacuolar citrate/H+ symporter	F:CCCTCGTTCTTGGTAGCTTT R:TTATGCATCCACATGCTCAC	55	306-309
IDLCY2	Lycopene β-cyclase	F:CGCAAATAATTGATTCAACA R:GATGATCACGTCATATCGAA	50	220-226
IDHYB1	β-Carotene hydroxylase	F:AAAAACAAAGCACCCAGAT R:GCCACCAGAACCTGTAATAA	53	192-213
IDHYB2	β-Carotene hydroxylase	F:TTTGGCACATTTGCTCTCTCT R:AAAGAAGCATGCCACAGAGC	55	305-307
IDPSY	Phytoene synthase	F:CCTGTCGACATTCAGGTTAG R:CTCATCACATCTTCGGTCTC	55	246-249
IDPEPC1	Phosphoenolpyruvate carboxylase	F:TTTTGAACAATCGGCTAATGG R:TTGCTGGAAGAGAGACTCCAA	55	231-259
IDPEPC2	Phosphoenolpyruvate carboxylase	F:TTGGAGTCTCTCTTCCAGCAA R:GTGAGAGCCACAATGCAAAA	55	128-153
IDCAX	Cation/H+ membrane antiporter	F:TAAGCTGCATTTAACCCTTT R:GCAATTGGGAGATAGTCAAT	55	237-243
IDAtGRC	Raffinose synthase	F:GGCAATGAAAACAATGAGAT R:TTTCAAGATTGTTGGTCCTC	55	208-225
IDAPV	Vacuolar H+ pyrophosphatase	F:CAGCTATTGGAAAGGTTTGT R:GGAGACAGGCATAAAACATC	55	156-163

(AT) Annealing temperature

Diversity analysis

Plant material

Ninety genotypes from the citrus germplasm bank of IVIA (Spain) and INRA/CIRAD (France) were used for the diversity study with SSR and indel markers (Online Resource 1). According to the Swingle and Reece classification system (1967), 45 genotypes belong to the three ancestral species (29 *C. reticulata*, 10 *C. maxima* and 6 *C. medica*) and 11 genotypes represented the secondary species (2 *C. aurantium*, 4 *C. sinensis*, 2 *C. paradisi* and 3 *C. limon*). Seventeen accessions are supposed of interspecific origin from their morphology or previous molecular data (46-50, 53-55, 65-66, 81, 84-89) even some of them were classified by Swingle and Reece (1967) as pure species. The last 17 accessions are hybrids from twentieth century breeding projects (67-80; 82, 83, 90).

Genotyping

Sixty-seven SSR markers were tested on the citrus population selected for our study. Fifty markers presented proper and clear results (Online Resource 2; Kijas *et al.*, 1997; Froelicher *et al.*, 2008; Luro *et al.*, 2008; Aleza *et al.*, 2011; Cuenca *et al.*, 2011; Kamiri *et al.*,

2011) and were used for the diversity study. Forty-seven of them were included in the clementine genetic map (Ollitrault *et al.*, 2012b) and were well distributed between and within all linkage groups. In addition, twelve indel markers were analysed. One of them (TRPA) is located in the clementine genetic map (linkage group 2).

Amplification by polymerase chain reaction (PCR) was performed using wellRED forward oligonucleotides (Sigma-Aldrich; Saint-Louis, USA) for analysis with a capillary genetic fragment analyser (CEQ/GeXP Genetic Analysis Systems; Beckman Coulter; Fullerton, USA). PCR was performed in a final volume of 15 µl. Each PCR reaction consisted of 1 ng/µl template DNA, 0.2 mM dNTPs, 0.2 µM wellRED dye-labelled forward primer, 0.2 µM of non-dye-labeled reverse primer, 10x PCR buffer (Fermentas), 1.5 mM $MgCl_2$ and 0.027 U/µl *Taq* DNA polymerase (Fermentas). PCR reactions were carried out with the following program: 5 min at 94°C; 40 cycles of 30 s at 94°C, 30 s at 55 or 50ºC (depending on the primer) and 1 min at 72°C with a final extension of 4 min at 72°C.

Denaturation and capillary electrophoresis were carried out on a Capillary Gel Electrophoresis CEQ™ 8000 Genetic Analysis System using linear polyacrylamide according to the manufacturer's instructions (Beckman Coulter Inc.). Genetic analysis system software (GenomeLab™ GeXP version 10.0) was used for data collection and analysis. Alleles were sized based on a DNA size standard (400 bp).

Data analysis

Neighbour-joining (NJ) analysis

Population diversity organisation was analysed with DARwin software (Perrier and Jacquemoud-Collet, 2006). For each primer, bands were scored as allelic data to calculate the genetic dissimilarity matrix using the simple matching dissimilarity index (d_{i-j}) between pairs of accessions (units):

$$d_{i-j} = 1 - 1/L \sum_{l=1}^{L} m_l / 2$$

where d_{i-j} is the dissimilarity between units *i* and *j*, *L* is the number of loci and m_l is the number of matching alleles for locus l. From the dissimilarity matrix obtained, a weighted NJ tree (Saitou and Nei, 1987) was computed using the Dissimilarity Analysis and Representation for Windows (DARwin5) software version 5.0.159, and the robustness of branches was tested using 10,000 bootstraps.

To establish the genetic structure with the core set of markers, NJ under topological constraints was used. It is a modified version that forces the *a priori* known topology of a subset of samples and positions additional subsets on the previous organisation. Secondary species and modern cultivars were positioned under the constraint of a tree based on basic taxa.

Severinia buxifolia (Poir.) Ten, a species related to citrus, was used to root NJ trees.

Principal coordinates analysis (PCoA)

It was performed using the software GENEALEX6 (Peakall and Smouse 2006). The data from molecular markers was used to obtain the pairwise genetic distance matrix, which was standardised and used for PCoA analysis.

Population structure

It was inferred with the Structure version 2.3.3 program (http://cbsuapps.tc.cornell.edu/structure), which implements a model-based clustering method using genotype data (Pritchard *et al.*, 2000; Falush *et al.*, 2003). According to the general agreement on the origin of cultivated species (Scora, 1975); Barrett and Rhodes, 1976), we considered an initial structure between three populations (K = 3): mandarin (29 samples), pummelo (10 samples) and citron (6 samples), assuming that the analysed genotypes are derived from these three ancestral taxa. The relative proportion of these ancestral populations in the secondary species and hybrids was assigned based on this assumption of an admixture model. Correlated allele frequencies were determined from the estimates of the three ancestral populations defined in this work. Ten runs of structure were performed with 500,000 steps of burning followed by 1,000,000 Monte Carlo Markov Chain (MCMC) repetitions.

F_{stat} parameters

F_{is}, F_{it} and F_{st} were calculated with the software program GENETIX v. 4.03 based on the parameters of Wright (1969) and Weir and Cockerham (1984).

Linkage disequilibrium

For multiallelic loci, LD between two loci is commonly measured by the *D'* estimate (Gupta *et al.*, 2005). *D'* values for each pair of markers were estimated on the whole data set using the software program PowerMarker v. 3.25 (Liu and Muse, 2005). *D'* values vary from 0 (total random association between alleles of the two considered loci) to 1 (total LD). The p value for obtaining the significance of D' was estimated by the exact test.

Selection of a subset of markers for quick genotyping

The methodology described by Jombart *et al.* (2010) was employed to obtain a small number of markers (core set) with good interspecific and intraspecific differentiation for quick and accurate genotyping. The procedure is based on a discriminant analysis of principal

components (DAPC). Data from molecular markers are transformed with a PCoA, and the matrix obtained is employed to perform a discriminant analysis (DA). These results are used to calculate the allele contribution to the main axes, and the alleles with the highest contribution are selected. Expected heterozygosity was used as an extra parameter to select primers that allow good intraspecific differentiation.

RESULTS

Interspecific indel polymorphism research and indel marker development

For the 16 genes a total of 10,701 bp by genotypes were successfully sequenced and aligned (Table 1), allowing the identification of 12 indel polymorphic loci in 10 genes. Specific indel polymorphisms were encountered in four loci in *C. medica* and another four loci in *C. maxima,* whereas the other indel polymorphisms were detected in different groups.

New primers were designed to analyse the indel diversity of these 12 loci (Table 2). In this diversity study, four loci (IDCHI, IDEMA, IDHYB1 and IDLCY2) had novel alleles not present in the six genotypes initially sequenced. Amplicons of genotypes with these new alleles were sequenced, as described previously, to analyse the origin of this pluri-allelism (Online Resource 3). At locus IDCHI, a new polymorphism was found in heterozygosis in *C. sunki*, another one was found in IDEMA (genotype *C. sunki* and others in heterozygosis), one at IDHYB1 in Cleopatra mandarin and other genotypes in heterozygosis and the last polymorphism was found in homozygosis at locus IDLCY2 in *C. sunki* and other genotypes in heterozygosis. Indel allele sequences of the ten analysed genes are given in Online Resource 3. For multi-allelic loci, the variation of amplicon size is due to variation in size of the same indel (IDCHI, IDHYB1 and IDLCY2) or several indels between the two primer sites (IDCHI, IDHYB2 and IDCAX). Three loci (IDPSY, IDPEPC2 and IDAVP) displayed intra-taxon polymorphisms only in *C. medica,* and the other three loci (IDHYB2, IDPEPC1 and IDATGRC) displayed intra-taxon polymorphisms only in *C. maxima*. Polymorphisms in loci IDTRPA, IDLCY2 and IDHYB1 may be due to copy number variations of SSRs.

Indel analysis

A total of 32 alleles were detected from the indel markers. The average number of alleles per locus was 2.67. Genetic diversity statistics were calculated for each indel marker in the entire population and for different citrus groups, including *C. reticulata*, *C. medica* and *C. maxima* (Online Resource 4). The allele number varied between 2 (for 7 loci) and 5 for IDCAX. IDCAX displayed the highest diversity (H_e = 0.69) related to different alleles in the three ancestral taxa. IDAVP (H_e = 0.12) was the least informative marker, as it differentiated only varieties from the citron subpopulation. The best markers for genotype differentiation within mandarins, pummelo and citron were IDCAX, IDPEPC1 and IDCHI, respectively. F_{stats} parameters (Wright, 1969; Weir and Cockerham, 1984) were estimated to analyse the differentiation between the three ancestral taxa (*C. maxima*, *C. medica* and *C. reticulata*). F_{is} values varied from -0.474 for IDAVP to 0.125 for IDCHI. For four loci, it was not possible to calculate the F_{is} parameter because the loci were monomorphic in each of the ancestral taxa. With the exception of IDAVP, the F_{is} value confirms a situation close to the Hardy-Weinberg equilibrium within each species. In contrast, F_{it} values with a high average (0.730) showed that, in the whole population (of the subset of the three ancestral taxa), the inbreeding coefficient is

higher than within taxa for almost all of the markers, indicating an important organisation between taxa. Only IDTRPA had a low value (-0.149) with two alleles shared by *C. maxima* and *C. reticulata*. The high F_{st} average value (0.766) and the F_{st} value of each locus (excluding IDTRPA) confirms that the inter-taxa differentiation contributes much more to the global inbreeding than does the intra-taxa component. Thus, a large portion of the total variation is explained by the differentiation between populations.

Average data over all indel loci are given in Table 3. The average F_W value (0.433) shows a high deficit of observed heterozygous individuals in the population. Indeed, the whole population had an observed heterozygosity of 0.18, which is 38% lower than the expected heterozygosity (0.29), suggesting an organisation in differentiated sub-gene pools with limited gene flows. Individually, the different taxa had an observed heterozygosity similar to the expected. *C. reticulata* was the most polymorphic (H_e = 0.13) and heterozygous (H_o = 0.14) ancestral taxon, and *C. maxima* was the least polymorphic and heterozygous (H_o = H_e = 0.07) ancestral taxon.

Table 3. Statistical summary of the diversity of indel and SSR markers

Marker type	All citrus accessions				*C. reticulata*			*C. maxima*			*C. medica*			3 basic taxa		
	N	H_o	H_e	F_W	***N***	H_o	H_e	***N***	H_o	H_e	***N***	H_o	H_e	F_{is}	F_{it}	F_{st}
InDel	2.67	0.18	0.29	0.433	1.58	0.14	0.13	1.25	0.07	0.07	1.25	0.09	0.09	-0.148	0.730	0.766
SSR	8.10	0.59	0.71	0.175	5.02	0.56	0.56	3.36	0.50	0.52	1.94	0.17	0.28	0.030	0.454	0.434

Mean values are represented in the table

N Allele number, H_o Heterozygosity observed, H_e Heterozygosity expected, F_w Wright fixation Index over the whole population, F_{is}, F_{it} and F_{st} Weir and Cockerham Index over the subset of *C. maxima*, *C. medica* and *C. reticulata* accessions

SSR analysis

The same genetic diversity parameters were calculated for each individual SSR marker, the entire population and for the different specified Citrus groups (Online Resource 5). A total of 405 alleles were detected with the SSR markers. The average number of alleles and H_e per locus was 8.1 and 0.71, respectively. The allele number varied between 3 (for loci MEST107, CAC15 and CAC23) and 14 (MEST56). TAA41 was the most informative marker with a He of 0.86, and CAC15 was the least informative marker (He = 0.39). Most of the markers (48 out of 50) showed H_e values higher than 0.5. When analysing the organisation among the three basic taxa, F_{is} values varied from -0.114 for CAC23 to 0.594 for mCrCIR05A04. The overall F_{is} value was close to zero (0.030), confirming that few deviations from the Hardy-Weinberg equilibrium occurred within each basic taxon. In contrast, high F_{it} and F_{st} values for almost all markers (averages of 0.454 and 0.434, respectively) are evidence of high differentiation between the 3 basic taxa.

Average data over all indel loci are given in Table 3. The population displayed a deficit of average observed heterozygosity (H_o = 0.59) compared with the expected value under Hardy-Weinberg equilibrium (H_e = 0.71). This finding is confirmed by the average F_W value

(0.175). Each of the 3 basic taxa had an observed heterozygosity close to the expected value. *C. reticulata* was the most diverse (H_e = 0.56) and heterozygous (H_o = 0.56) ancestral taxa, but citron was the lowest (H_e = 0.28 and H_o = 0.17).

Comparative diversity structure displayed by indels and SSRs

The genetic parameters for indel and SSR markers, respectively, were as follows: allele number per locus ranged from 2 to 5 and from 3 to 14, observed heterozygosity average was 18 and 59% and the percentage of varieties differentiated among the whole population was 57.78% (52 out of 90) and 91.11% (82 out of 90). The distribution of H_e and F_{st} between the three basic taxa (Figure 1) confirmed that indel markers are less polymorphic than are SSR markers (lower H_e values) but allow a better differentiation between ancestral species (higher F_{st} values). Statistics for the three ancestral groups were calculated for both types of primers (Table 3). Expected and observed heterozygosity were similar for both types of markers but were lower

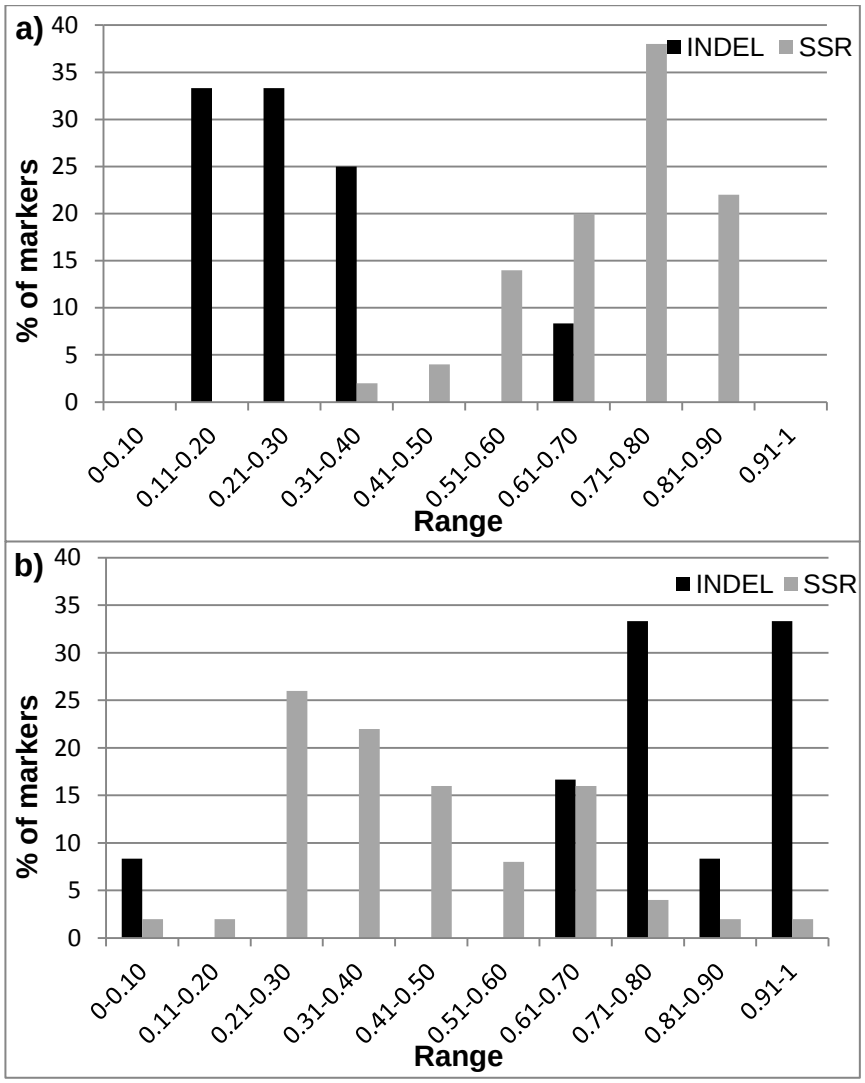


Figure 1. Comparison between indel and SSR markers of the expected heterozygosity (H_e) and the genetic differentiation index (F_{st}) between ancestral taxa. **a)** Expected heterozygosity, **b)** genetic differentiation index

for indels than SSRs within each taxon. With SSR markers, all accessions of *C. medica* and *C. maxima* were fully differentiated, whereas 96.7% of intervarietal differentiation was obtained within *C. reticulata*. The indel intervarietal differentiations were 100, 40 and 53.3% within *C. medica, C. maxima* and *C. reticulata,* respectively. Twelve out of 50 SSR and 7 out of 12 indel markers displayed significant deficits of heterozygous genotypes in the whole sample set (Online resources 4 and 5).

The F_{st} value was estimated for each pair of basic taxa, and it was systematically higher with indel than SSR markers. The least differentiated species were *C. reticulata* and *C. maxima* (F_{st} of 0.373 and 0.422 for SSR and indel, respectively), followed by *C. reticulata*/*C. medica* (0.427 and 0.758) and *C. maxima*/*C. medica* (0.484 and 0.844). All of these data support the conclusion that indel markers yield higher inter-taxa discrimination compared with SSR markers.

Both NJ, figure 2 and principal coordinates analysis PCoA, figure 3 analyses revealed a clear differentiation between the three ancestral citrus taxa for both kinds of markers.

NJ trees (Figure 2) clearly separated *C. medica* and *C. maxima* from *C. reticulata*. For indel markers (Figure 2a), *C. medica* was the best defined group and showed good bootstrap support in all branches of its cluster, and all of the samples were differentiated. The *C. maxima* group formed a well-defined clade, but only four profiles were differentiated among ten accessions. The intraspecific diversity of *C. reticulata* was not well resolved (low bootstrap support), perhaps due to the high number of hybrids (within mandarin) in the sample set. Fourteen genotypes were differentiated among the 29 mandarins.

SSRs allowed a complete intercultivar differentiation for *C. maxima* and *C. medica*, whereas only two *C. reticulata* cultivars (East India SG and Vohangisany Ambodiampoly) were not differentiated (Figure 2b).

NJ analysis confirmed higher intraspecific diversity with SSRs than with indel markers. The lower differentiation obtained with indels may be partly due to the lower number of these markers. However, it is also clearly explained by their lower allelic diversity, which is observed mostly at the interspecific level. Clustering was stronger with indel than with SSR markers, but SSRs allowed a better intra-cluster differentiation between accessions.

PCoA (Figure 3) is more adapted than tree representation in describing the organisation of genetic diversity when hybrids between differentiated groups are frequent in the sample. In our study, PCoA allowed us to have a better idea of the relative contribution of the three basic taxa to the genome constitutions of secondary species and modern hybrids. Almost all of the existing variability (92.10%) is represented in the first two axes for indels (Figure 3a), but only 75.89% variability is represented for SSRs (Figure 3b). This result confirms that higher interspecific organisation is determined using indel markers. For these markers, the *C. medica* group (and its hybrids with citron as one parent) was strongly differentiated from *C. reticulata* (and its hybrids) and *C. maxima* by axis 1, whereas the *C. maxima* group was differentiated from the

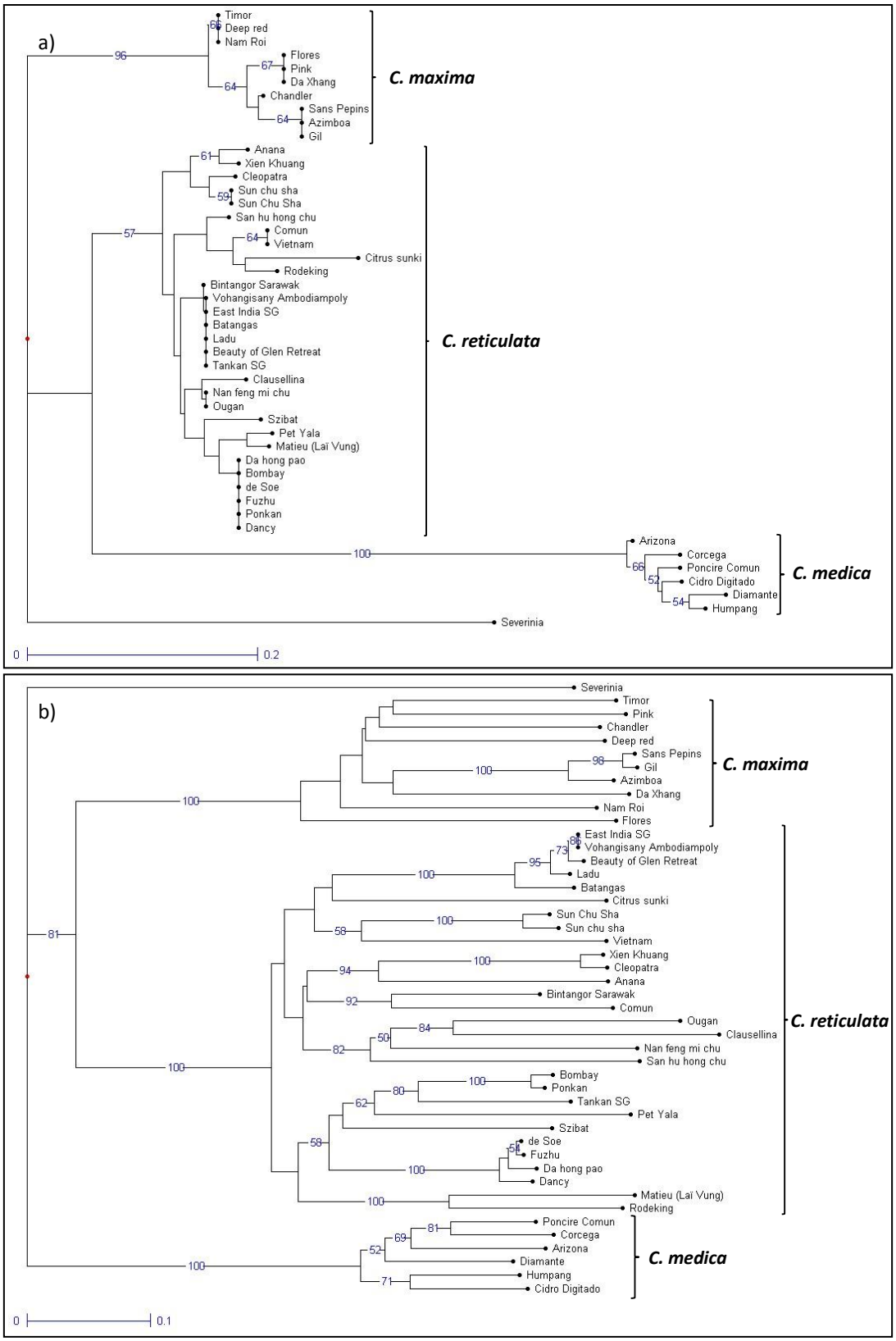


Figure 2. NJ bootstrap consensus trees of 45 accessions of citrus (3 ancestor groups) including one outgroup *Severinia buxifolia*. Numbers are bootstrap values over 50 based on 10,000 resampling. **a)** Indel markers data, **b)** SSR markers data

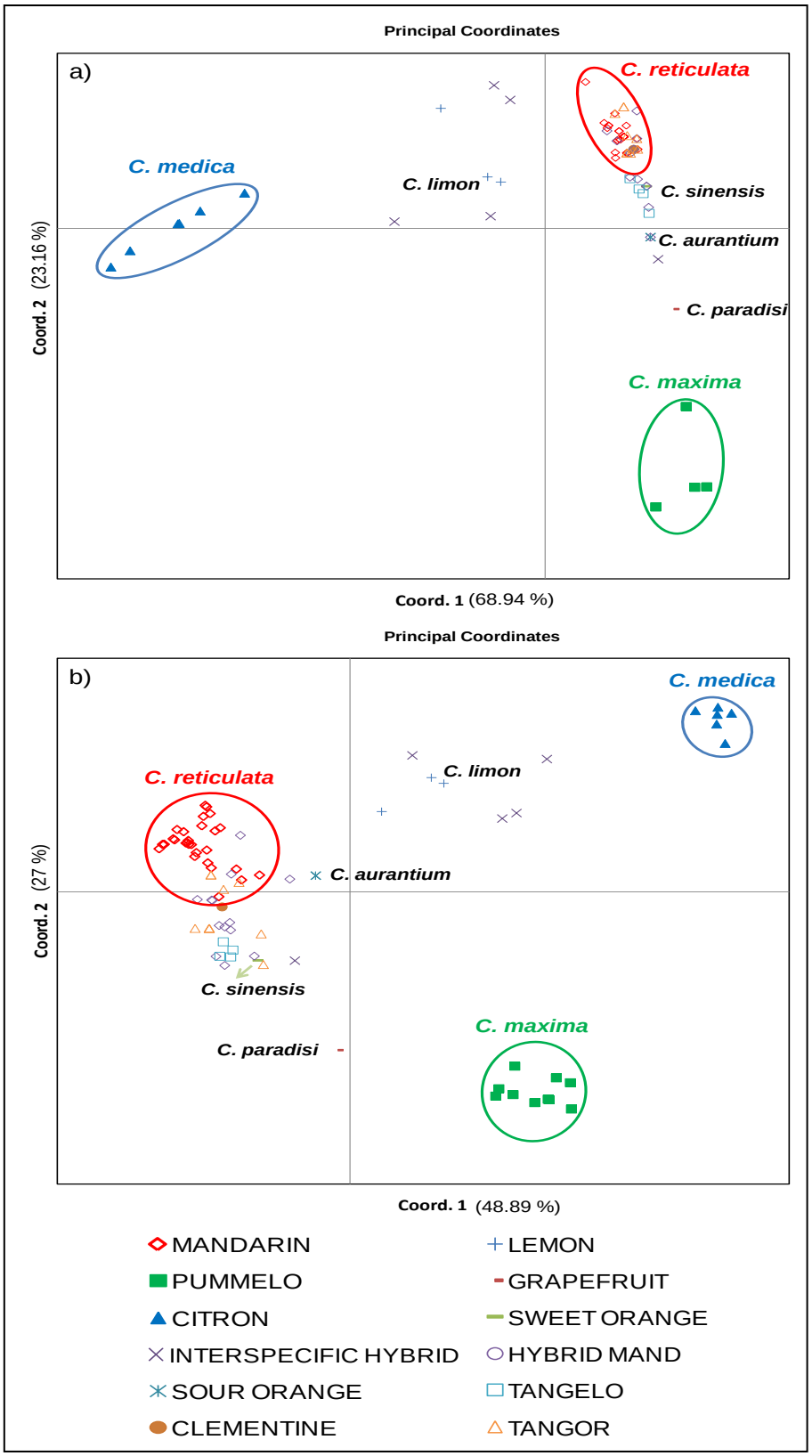


Figure 3. Organization of cultivated Citrus genetic diversity; principal coordinates analysis. **a)** Indel markers data, **b)** SSR markers data. Mandarin (samples 1–29), pummelo (samples 30–39), citron (samples 40–45), interspecific hybrids (samples 46–50), sour orange (samples 51–52), clementine (samples 53–54), lemon (samples 56–58), grapefruit (samples 59–60), sweet orange (samples 61–64), hybrid mandarins (samples 67–76), tangelo (samples 77–80) and tangor (samples 81–90). Sample number assignment can be found in Online Resource 1

other species by axis 2. *C. paradisi* varieties and Bali hybrid, mandarin Suntara and *C. aurantium* (the last two had exactly the same position), in this order, were closer to *C. maxima* with indel markers than with SSR markers. Tangors (mandarin x sweet orange) were closer to the *C. reticulata* cluster and Tangelos (mandarin x grapefruit) were closer to *C. maxima*, as expected from their origin. Clementines were close to *C. reticulata* accessions and some hybrids that have clementines as a parent.

For SSRs, *C. medica* was differentiated from *C. maxima* by axis 1, and the *C. reticulata* group was differentiated from *C. medica* by axis 2. *C. reticulata* accessions were more dispersed around the axis based on SSR markers than with indel markers. As *C. sinensis*, *C. aurantium* appeared much more related to *C. reticulata* than to *C. maxima*, *C. limon* was clearly positioned between the *C. medica* gene pool and *C. aurantium*. Some hybrids derived from *C. medica* (Poncil, Rhobs el Arsa, Kadu Mul and Damas) were positioned in a similar place, suggesting that these hybrids share similar origins as *C. limon*. Tangor was the most dispersed group, Murcott and Umatilla were the closest varieties to *C. reticulata* and Ortanique was the closest to *C. maxima*. Tangelos were similarly distanced between them. Clementines were close to the *C. reticulata* gene pool, whereas *C. paradisi* was the secondary species closest to *C. maxima*.

Contribution of the ancestral taxa to secondary species and modern hybrids; analysis with structure software

PCoA analysis provided some information on the relative contribution of the three basic taxa to the genome constitution of the secondary ones, confirming the status of *C. medica*, *C. reticulata* and *C. maxima* as parental gene pools of the other species and modern hybrids in this study. Assuming an admixture model between the three ancestral species, the relative proportion of ancestral taxa genomes in the secondary species and recent hybrids was inferred using the Structure version 2.3.3 software (Figure 4) with the complete set of data (SSRs + Indels).

Citrus limon and hybrids with *C. medica* as parents (Poncil, Rhobs el Arsa, Kadu Mul and Damas) have the greatest average contribution from *C. medica* (46%). Contributions of *C. medica* lower than 2.5%, which was observed for *C. sinensis*, *C. aurantium*, *C. paradisi*, Bali pummelo, Clementine and Temple, can probably be considered artefacts and related to the relatively low number of representative genotypes of the basic taxa and probable lack of intra-taxa diversity. *Citrus paradisi* is the secondary species with the highest contribution from *C. maxima* (60%), followed by *C. aurantium* (30%), *C. sinensis* (25%), tangelo group (20%), tangor group (10%) and clementines (7%). *Citrus aurantium* varieties displayed seven rare alleles, five of which were shared with Suntara mandarin (two of them were also shared with *C. limon*), one was shared with *C. limon* and another one was only present in *C. aurantium*.

The contributions of the ancestral groups to the secondary species obtained with the Structure software was compared with direct estimations performed with the specific allele from the SSR and indel markers derived from the mandarin, pummelo and citron groups (Table 4). No significant difference was found between the two methods of evaluation. It is interesting to note that no specific allele from *C. medica* was observed in *C. sinensis*, *C. paradisi*, Bali pummelo, Clementine and Temple, which confirms that the low values estimated for the same genotypes with Structure were not significant.

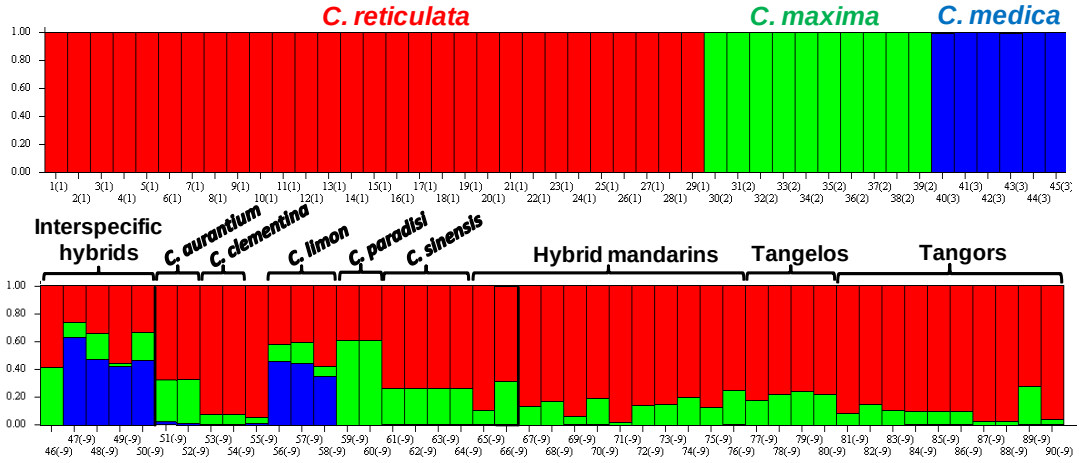


Figure 4. Relative contribution of basic taxa to secondary species and modern cultivars; structure analysis with *K* = 3 as initial hypothesis, considering SSR and indel data. In *parenthesis* are indicated the reference population assignment for the admixture model *1 C. reticulata* population, *2 C. maxima* population, *3 C. medica* population, *-9* population with unknown contribution from ancestors. Sample number assignment can be found in Online Resource 1

Table 4. Contribution of the ancestral taxa to secondary species; comparison between direct estimation from interspecific discriminant allele and the estimation from Structure software (admixture model).

Latin name	Common name	SSR + InDel allele specific from			Total informative alleles	Direct estimation from discriminant alleles			Structure data			
		Re	Ma	Me	SSR + InDel	Re (%)	Ma (%)	Me (%)	Re (%)	Ma (%)	Me (%)	X^2
C. aurantium	Sevillano	32	16	1	49	65.31	32.65	2.04	67.2	30.6	2.2	0.10
C. clementina	Clemenules	49	1	0	50	98	2	0	92	7.1	0.9	2.48
C. limon	Eureka Frost	21	4	22	47	44.68	8.51	46.8	41.6	12.1	46.3	0.61
C. limon	Lisbon Limoneira	20	6	22	48	41.67	12.50	45.8	40.3	14.7	45	0.19
C. paradisi	Marsh	21	22	0	43	48.84	51.16	0	38.6	60.9	0.5	2.05
C. sinensis	Valencia late delta	37	5	0	42	88.10	11.90	0	73.3	25.6	1.1	4.79
x *C. maxima*	Bali	28	18	0	46	60.87	39.13	0	58.4	41.1	0.6	0.37
x *C. medica*	Poncil	14	5	26	45	31.11	11.11	57.8	26.3	10.8	62.9	0.59
x *C. medica*	Rhobs el Arsa	15	9	20	44	34.09	20.45	45.5	33.6	18.7	47.7	0.12
x *C. medica*	Kadu Mul	31	0	23	54	57.41	0	42.6	54.9	2.7	42.3	1.52
x *C. medica*	Damas	11	8	24	43	25.58	18.60	55.8	33.6	18.7	47.7	1.42
x *C. reticulata*	*Citrus daoxianensis*	50	1	0	51	98.04	1.96	0	94.1	4.1	1.8	1.57

Linkage disequilibrium

Based on the data obtained with the 50 SSR markers distributed along the genome, the extent of genome-wide LD was estimated by *D'* for the whole population. indel markers were not selected for this analysis because they were not mapped. *D'* values ranged from 0.11 to 0.9 for interchromosome pairs of loci and from 0.21 to 0.94 for intrachromosome pairs (Figure 5). The average D' estimates for marker pairs within and between chromosomes were 0.56 and 0.51, respectively. For interchromosome and intrachromosome marker pairs, 65.69 and 53.68% of

the *D'* values were over 0.5, respectively. The percentage of significant p values was very high for marker pairs within and between chromosomes: 99.27/99.26% (< 5%) and 97.08/97.89% (< 1%), respectively. When analysing the relation between LD and genetic distances between markers (Figure 6), it appears that there is a high LD even between distant markers with a limited LD decay with increasing distances. The distribution of the interchromosome *D'* is highly similar. The mean value of *D'* was 0.5161 for the whole population and all marker pairs.

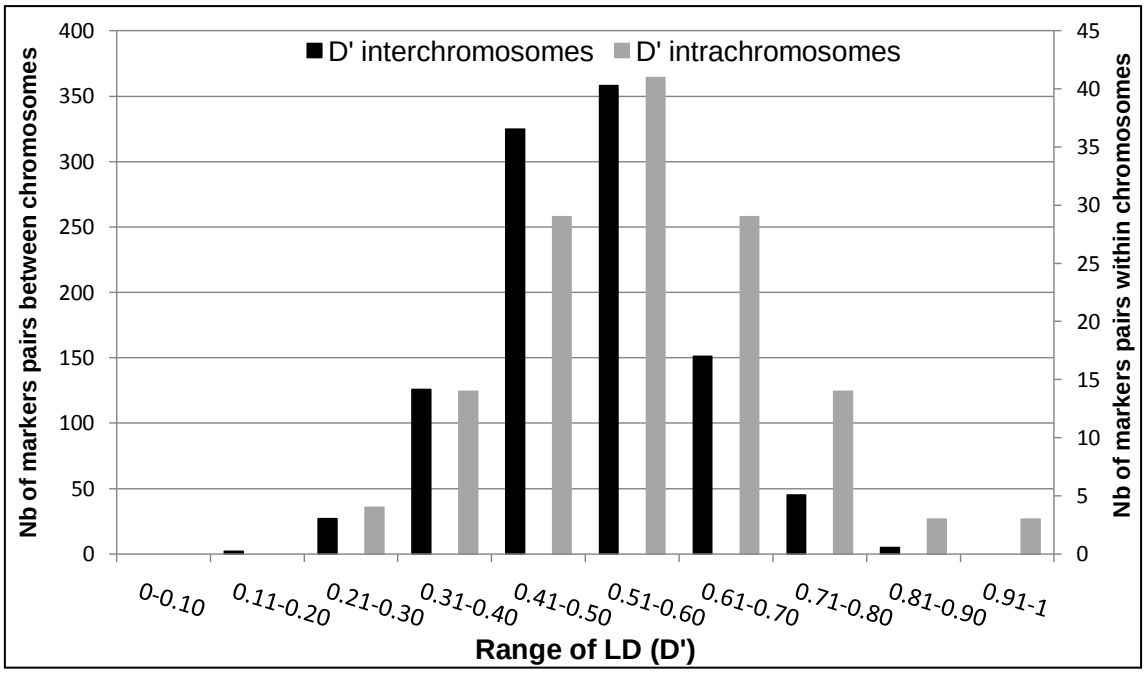


Figure 5. Linkage disequilibrium for marker pairs within a same linkage group (*grey*) and between markers located in different chromosomes (*black*)

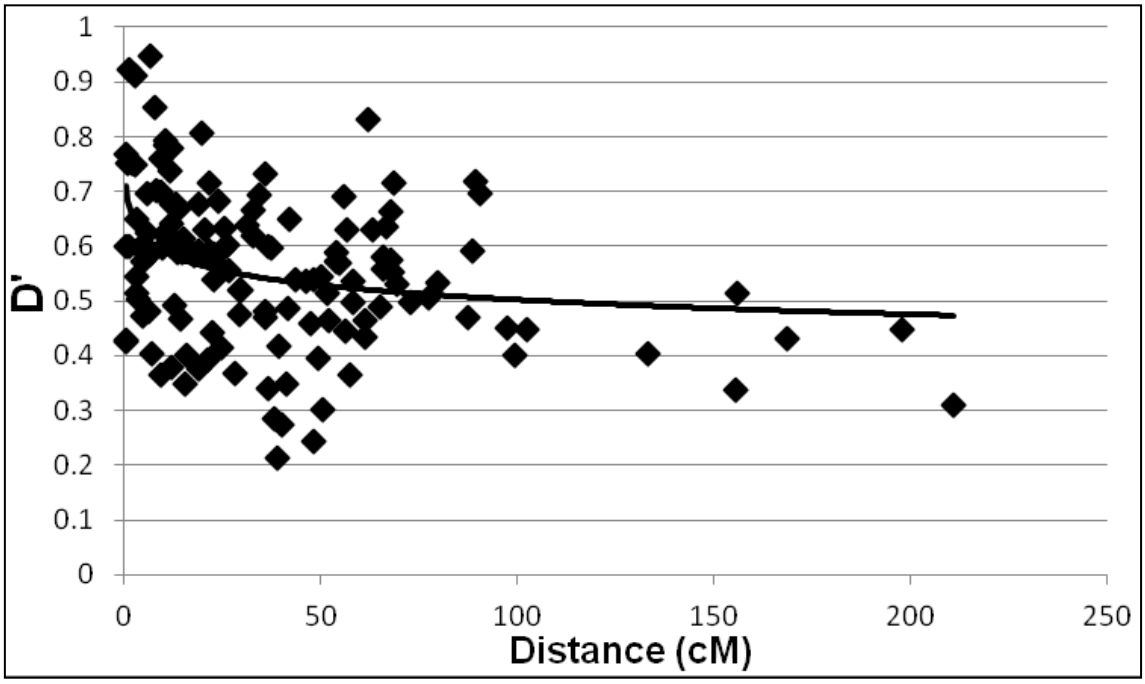


Figure 6. Relation between LD in the population for all markers pairs within chromosomes and genetic distances (Clementine genetic map; Ollitrault *et al.*, 2012b)

Selection of a subset of markers for quick genotyping

Identifying a subset of markers that can differentiate new accessions and study their origin could be useful for quick and inexpensive genotyping. In this study, the parameters used to select the subset of markers were high locus contribution to F1 and F2 coordinates of the PCoA analysis (interspecific organisation), high expected heterozygosity (global diversity displayed by the marker) and limited LD between the selected markers to avoid excessive redundant information between markers (Online Resource 6). A total of nine markers were selected: mCrCI02D04b and MEST431 were selected for their high contribution to the F1 component (which distinguished *C. reticulata* from the other two ancestors), IDCHI and IDCAX have a high contribution to F2 (axis which differentiates between *C. medica* and the other ancestors), mCrCI07F11 and mCrCI07D06 contributed in both axes (it is helpful to distinguish individuals that are intermediate) and MEST488, TAA41 and mCrCI02G12 were selected for their high expected heterozygosity. Six out of the nine linkage groups were represented by the selected marker subset. With these nine markers, the three ancestors groups were clearly differentiated (Online Resource 7). Samples in the *C. medica* group were fully separated, whereas in *C. maxima*, only 'Gil' and 'Sans Pepins' cultivars could not be differentiated. *C. reticulata* within diversity was slightly less resolved than with the whole marker set (6 mandarins were not distinguished). The average observed and expected heterozygosity values were 56 and 64%, respectively, and the F_W was 0.163.

DISCUSSION

Citrus indel markers are less polymorphic but display higher interspecies differentiation than do SSR markers

Indels are generally considered to be interesting polymorphisms for genetic studies. However, despite increasing molecular resources in citrus, such as EST sequence information (Forment *et al.*, 2005; Terol *et al.*, 2007), HarvEST software Version 1.32 of "HarvEST:Citrus" (http://www.harvest-web.org) and genomic sequence information (Terol *et al.*, 2008), no specific study has been conducted prior to the present work to analyse the value of nuclear indels as genetic markers in *Citrus*. We searched for indel polymorphisms in the three basic taxa (*C. reticulata*, *C. maxima* and *C. medica*) by sequencing PCR products obtained from 13 genes. Primers were designed to amplify 150- 350 bp fragments flanking the 12 identified indels, and amplicon size variation was studied by capillary electrophoresis on a sample of 90 genotypes of the *Citrus* genus.

The frequency of indels per kb in citrus was 0.71 and 5.22 in exon and intron sequences, respectively. More sequence polymorphisms were found in non-coding regions than in coding regions. Similar results have been observed in other species. In Brassica, 0.45 and 7.42 indel/kb were found in exons and introns, respectively (Park *et al.*, 2010). In melon, indels occurred less frequently in introns (approximately 0.60/kb) and no indel was found inside coding regions (Morales *et al.*, 2004). In maize, 0.43 and 11.76 indels/kb were found in coding and non-coding regions, respectively (Ching *et al.*, 2002).

The mean number of alleles per locus was 2.83 with a maximum of five alleles at the IDCAX locus. Seven of the twelve markers were diallelic. Retroposon movements, such as *Alu* or the L1 element, are known to generate such diallelic indels (Watkins *et al.*, 2001). In our study, pluri-allelism was caused by differences in indel size or the presence of several indels in the amplified fragments. Indels with a size that is not a multiple of 3 are uncommon in exons but relatively common in introns (Mills *et al.*, 2006; The Arabidopsis Genome Initiative, 2000).

Almost 60% of the whole set of samples were differentiated with the 12 indel markers. A better differentiation may be obtained with more indels; however, the low mean number of alleles per locus may be a limitation compared with techniques using multi-allelic markers, such as SSRs. Indeed, we found a mean value of 8.1 alleles per locus for SSRs. With higher allelic diversity and intra-taxon diversity, SSRs are more informative than indels at the intraspecific level. The number of repeats in microsatellites evolves at a high rate (Weber and Wrong, 1993; Jarne and Lagoda, 1996), which can vary depending on the number of repeats or base composition (Bachtrog *et al.*, 2000). Thus, there are generally good markers for intra-population diversity analysis, as we observed at the intra-taxon level. However, due to this important rate of variation, homoplasy should be relatively frequent, as demonstrated in *Citrus* (Barkley *et al.*, 2009), and should limit the value of SSRs as phylogenetic markers. Our results confirmed this

hypothesis, as we observed that indel markers displayed a much higher differentiation between the three basic taxa than SSRs, with F_{st} value averages of 0.77 and 0.43, respectively. The structure of the whole sample diversity was higher for indels with a fixation index value (F_w; Wright, 1978) of 0.433 and 0.175 for SSRs. Interestingly, the three indel markers (IDTRPA, IDLCY2 and IDHYB1) that may result from variation in copy number of SSRs showed lower F_{st} value than the average. Therefore, these three markers provide less inter-taxa differentiation than the other indels. The PCA also confirmed a higher level of structure of the diversity displayed by indels markers than by SSRs with 92.2 and 75% of the whole diversity, respectively, represented by the first two axes.

Thus, we can conclude that, in the *Citrus* genus, indel markers are less polymorphic than SSRs but display a higher organisation of genetic diversity at the interspecific level. From the 50 SSRs and 12 indels we have selected a core set of 9 markers (2 indels and 7 SSRs) that keep the interspecific structure, as well as a significant part of the intraspecific polymorphism information. These markers should be useful for the rapid and inexpensive assignment of a new germplasm variety to its genetic group or identification of its potential hybrid origin.

Indels play a major role in sequence divergence between closely related DNA sequences in animals, plants, insects and bacteria. Indels are responsible for many more unmatched nucleotides than are base substitutions, and human genetic data suggests that indels are a major source of gene defects (Britten *et al.*, 2003). Indels in coding regions probably have functional roles and are considered to be a significant source of evolutionary change in eucaryotic and bacterial evolution (Britten *et al.*, 2003). Indels in genes with functional diversity between alleles should be highly useful for marker-assisted selection (Raman *et al.*, 2006) or QTL mapping (Vasemägi *et al.*, 2010). Using the increasing amounts of sequence information acquired by new technologies (454-Roche, SOLiD system-Applied biosystems or Solexa-Illumina), the development of PCR-based indel markers will become an important source of genetic markers that are easy and inexpensive to use in phylogenetic and genetic association studies in *Citrus*.

The genetic constitution of secondary species and modern hybrids

In agreement with previous molecular studies (Barkley *et al.*, 2006; Luro *et al.*, 2008), no intercultivar polymorphism was found at intraspecific level for *C. sinensis*, *C. aurantium* and *C. paradisi,* whereas these species are highly heterozygous (H_o values of 0.47, 0.50 and 0.44, respectively). This finding confirms that most of the intervarietal polymorphisms within these secondary species arise from punctual mutation or movement of transposable elements (Bretó *et al.*, 2001). These types of mutations are unlikely to be detected with SSR or indel markers. The three lemon cultivars were differentiated. However, lemons cv ‘Lisbon’ and cv ‘Eureka’ only differed for five markers.

PCA using SSR or indel markers confirmed that the differentiation between *C. reticulata*, *C. maxima* and *C. medica* gene pools was the structuring factor of the analysed edible citrus germplasm. Secondary species and modern tangor and tangelo cultivars (which display higher heterozygosity than *C. reticulata, C. maxima* and *C. medica*) take intermediary positions between the three basic taxa, confirming their hybrid status. Structure analysis with an admixture model considering *C. reticulata, C. maxima* and *C. medica* at the origin of all analysed germplasm allowed us to estimate the contribution of these taxa to the genomes of secondary species, modern cultivars and some genotypes of unclear origin.

Two accessions initially considered as representative of *C. maxima* and *C. medica* (Bali pummelo and Poncil citron, respectively) were discarded by structure analysis from the ancestor species and positioned as hybrids. Bali seemed to be a hybrid between *C. reticulata* and *C. maxima* (genome contributions of 57 and 43%, respectively) and Poncil seemed to be a tri-hybrid from *C. medica* (63%), *C. reticulata* (26%) and *C. maxima* (11%).

As proposed by Roose *et al.* (2009), we found that sweet orange (*C. sinensis*) exhibits close to 75% *C. reticulata* and 25% *C. maxima* contribution and thus should be the result of a backcross 1 (BC1) [(*C. maxima x C. reticulata*) x *C. reticulata*]. These contributions differ from the ones estimated by Nicolosi *et al.* (2000) where *C. sinensis* shared half of its markers with *C. reticulata* and the other half with *C. maxima* and in Barkley *et al.* (2006) where only 6-8% of its genome arose from *C. maxima*.

It is believed that grapefruit (*C. paradisi*) arose from a cross between pummelo and sweet orange in the West Indies where they were introduced after Christopher Columbus discovered the new world (Barrett and Rhodes, 1976; Nicolosi *et al.*, 2000). Grapefruit displays a contribution of 61% from *C. maxima* and 39% from *C. reticulata*, which are values that are close to the theoretical average values (62.5 and 37.5%, respectively) expected for a *C. maxima* x [(*C. maxima x C. reticulata*) x *C. reticulata*] hybrid.

Sour orange (*C. aurantium*) is thought to be derived from hybridisation between *C. maxima* and *C. reticulata* gene pools (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Uzun *et al.*, 2009). Our analysis with Structure suggests that it showed a greater contribution from *C. reticulata* (68%) than did *C. maxima* (30%) and a bit of *C. medica* (2%). Seven rare alleles were found in *C. aurantium* that were not present in the analysed germplasm of the three main ancestors. However, five of them were found in the accession 'Suntara' mandarin. Furthermore, 'Suntara' and *C. aurantium* share the same alleles at most loci. Thus, there is a high probability that *C. aurantium* and 'Suntara' mandarin share parentage, but we do not have sufficient evidences to conclude whether 'Suntara' is a parent or a hybrid from *C. aurantium*. The small contribution of *C. medica* (2%) can probably be considered an artefact by estimation with Structure software, due to an underrepresentation of *C. maxima* and *C. reticulata* diversity. It is likely that *C. aurantium* is a BC1 (*C. maxima* x (*C. maxima x C. reticulata*)).

In agreement with its putative *C. aurantium* x *C. medica* origin (Nicolosi *et al.*, 2000; Gulsen and Roose, 2001a), we found that lemons (*C. limon*) cv. 'Eureka' and 'Lisbon' had a

complex tri-hybrid structure from *C. reticulata* (41%), *C. medica* (45%) and *C. maxima* (13%). The argument that *C. aurantium* is one parent is reinforced by the fact that these two lemons shares three rare alleles with *C. aurantium*.

Mandarin-like varieties are an increasing component of the citrus fresh fruit market and include *C. reticulata* hybrids, known or supposed tangors (hybrids between *C. reticulata* and *C. sinensis*) and tangelos (hybrids between *C. reticulata* and *C. paradisi*). The clementine, a variety selected from a seedling of ”Common mandarin” one century ago in Algeria, is the most popular variety of mandarin in the Mediterranean Basin. Most of its genome is inherited from *C. reticulata*, but it seems to have been introgressed in small part from *C. maxima* (6%). The allelic constitution of clementine is in agreement with the hypothesis of a “Common mandarin” x *C. sinensis* hybridisation (Deng *et al.*, 1996; Nicolosi *et al.*, 2000). In addition, the ‘Temple’, ‘Ellendale’, ‘Murcott’ and ‘King’ varieties have been considered as tangor. These varieties showed close to 90% contribution of the *C. reticulata* genome and 10% contribution of the *C. maxima* genome, as expected for hybrids between *C. reticulata* and *C. sinensis*. Moreover, they shared most of their alleles with these two species. Our results confirm the hypothesis of Swingle (1943), Coletta Filho *et al.* (1998) and Nicolosi *et al.* (2000) regarding the origin of ‘King’. As expected, tangelos had a greater contribution of *C. maxima* than tangors (approximately 20%).

Of the genotypes of uncertain origin, we found that *C. daoxianensis* is mostly of *C. reticulata* origin (94%). This result is in agreement with Li *et al.*, (1992), who considered *C. daoxianensis* to be a wild mandarin. ‘Rhobs el Arsa’ was considered by Federici *et al.,* (1998) to be a cross between *C. aurantium* and *C. medica*, as are lemons. Our results are in agreement with this hypothesis. The origin of ‘Kadu Mul’ has not been reported previously. Our results prompt the hypothesis that ‘Kadu Mul’ arose from a cross *C. medica* x *C. reticulata,* as we found that ‘Kadu Mul’ exhibits 42.3 and 54.9% contribution from *C. medica* and *C. reticulata*, respectively.

This study showed that the ancestral *C. reticulata* group contributes to a great proportion of the genomes of secondary species and recent hybrids. The facultative apomixis exhibited by all secondary species probably arose from the *C. reticulata* germplasm.

Cultivated citrus: a highly structured gene pool with generalised linkage disequilibrium that is not favourable for global association genetic studies

Previous molecular studies (Herrero *et al.*, 1996, Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Luro *et al.*, 2001; Ollitrault *et al.*, 2003; Barkley *et al.*, 2006; Liang *et al.*, 2007) have provided evidence of a strong diversification between the ancestral taxa of all cultivated forms. Therefore, the analysis of the organisation of cultivated citrus and the study of the LD organisation of the genome were necessary to estimate how association studies should be conducted in *Citrus*.

Our analysis of F_{stat} parameters in the subset of the three basic taxa genotypes (*C. reticulata*, *C. medica* and *C. maxima*) with non-significant F_{is} value but high F_{it} and F_{st} values confirms the important structure of the allelic diversity between these taxa. The interspecific differentiation was particularly high using indel markers. Eleven of 50 SSR markers and 7 of 12 indel markers displayed significant deficits of heterozygous genotypes in the whole sample. This indicates a strong population subdivision (Hartl and Clark, 1997) and, therefore, a low gene flow between *C. medica*, *C. reticulata* and *C. maxima*. The differentiation between these sexually compatible taxa can be explained by the foundation effect in three geographic zones and by an initial allopatric evolution. *Citrus maxima* originated in the Malay Archipelago and Indonesia, *C. medica* evolved in North-eastern India and the nearby region of Burma and China and *C. reticulata* diversification occurred over a region including Vietnam, Southern China and Japan (Webber *et al.*, 1967; Scora, 1975). Later on, human activity facilitated migration and hybridization among the differentiated gene pools of the basic taxa. However, the partial apomixis observed in most of the secondary species has strongly limited the interspecific gene flow.

Using 50 mapped SSR markers, we found that the LD decay was very slow as the distance increased in a same linkage group. Moreover, a similar distribution of LD was found when considering LD within or between linkage groups (65.69 and 53.68% of the *D'* values > 0.5, respectively). 99.3% of significant p values (< 0.05) were observed both within and between linkage groups. This LD structure confirms that the history of cultivated *Citrus* (initial allopatric differentiation of basic taxa followed by a limited number of interspecific meiosis) is not a favourable situation for association genetic studies. Indeed, significant LD between polymorphisms on different chromosomes may produce associations between a marker and a phenotype, even though the marker is not physically linked to the locus responsible for the phenotypic variation. Similar population structures exist in many crops where the complex breeding history and limited gene flow found in most wild plants have created complex stratification (Flint-Garcia *et al.*, 2003; Abdurakhmonov and Abdukarimov, 2008). LD between unlinked loci primarily happens due to the occurrence of distinct allele frequencies with different ancestry in an admixed or structured population when predominant parents exist in germplasm groups. This was the case in our sample representative of the cultivated *Citrus* genus. Statistical methodologies have been developed to properly interpret the results of association tests when using such structured populations (Pritchard *et al.*, 2000; Reich and Goldstein, 2001; Price *et al.*, 2006; Yu *et al.*, 2006). However, to be applied properly, these methods require that a significant part of the structured population results from recombination between the ancestral genomes with sufficient meiosis events to reduce the initial extent of LD, whereas the actual cultivated citrus germplasm arises from a limited number of such inter-ancestry meiosis. This result precludes LD-based association study at the genus level without developing additional interspecific hybrids, such as BC1 or F2, between ancestral taxa or hybrids of the secondary species. In addition, the potential use of genetic association studies within basic species should be explored, particularly in *C. reticulata* where useful polymorphisms (resistance to biotic and

abiotic constraints and some quality factors) have been identified. Moreover, markers with a higher rate of identity-by-descent, such as indels or SNPs, should be more useful than SSRs for genetic association studies.

CONCLUSIONS

This work achieves for the first time in c0itrus, the development of indel markers as an important tool for diversity and phylogenetic studies in citrus. Indel markers appear to be better phylogenetic markers for tracing the contributions of the three ancestral species to the secondary species and modern cultivars, whereas SSR markers are more useful for intraspecific diversity analysis. Most of the genetic organisation of the *Citrus* gene pool is related to the differentiation between *C. reticulata*, *C. maxima* and *C. medica*. High and generalised LD was observed, probably due to the initial differentiation between the basic species and a limited number of interspecific meiosis. This structure precludes association genetic studies at the genus level without developing additional recombinant populations from interspecific hybrids. Association genetic studies should also be affordable at intraspecific level in a less structured pool such as *C. reticulata*.

ONLINE RESOURCES CHAPTER 1

Online Resource 1. Genotypes classification

Common name	Group	Swingle system	Accession	Collection
Comun	Mandarin	*C. reticulata* Blanco	154	IVIA
Anana	Mandarin	*C. reticulata* Blanco	390	IVIA
Ponkan	Mandarin	*C. reticulata* Blanco	482	IVIA
Sun chu sha	Mandarin	*C. reticulata* Blanco	483	IVIA
Dancy	Mandarin	*C. reticulata* Blanco	434	IVIA
Clausellina	Mandarin	*C. reticulata* Blanco	19	IVIA
Fuzhu	Mandarin	*C. reticulata* Blanco	571	IVIA
Cleopatra	Mandarin	*C. reticulata* Blanco	385	IVIA
Citrus sunki	Mandarin	*C. reticulata* Blanco	239	IVIA
Bintangor Sarawak	Mandarin	*C. reticulata* Blanco	0100683	INRA/CIRAD
Vohangisany Ambodiampoly	Mandarin	*C. reticulata* Blanco	0100437	INRA/CIRAD
San hu hong chu	Mandarin	*C. reticulata* Blanco	0100769	INRA/CIRAD
Nan feng mi chu	Mandarin	*C. reticulata* Blanco	0100839	INRA/CIRAD
Vietnam	Mandarin	*C. reticulata* Blanco	0100800	INRA/CIRAD
Rodeking	Mandarin	*C. reticulata* Blanco	0100431	INRA/CIRAD
Ladu	Mandarin	*C. reticulata* Blanco	0100595	INRA/CIRAD
Batangas	Mandarin	*C. reticulata* Blanco	0100057	INRA/CIRAD
Bombay	Mandarin	*C. reticulata* Blanco	0100518	INRA/CIRAD
East India SG	Mandarin	*C. reticulata* Blanco	0100414	INRA/CIRAD
Xien Khuang	Mandarin	*C. reticulata* Blanco	0100868	INRA/CIRAD
Ougan	Mandarin	*C. reticulata* Blanco	0100680	INRA/CIRAD
de Soe	Mandarin	*C. reticulata* Blanco	0100653	INRA/CIRAD
Pet Yala	Mandarin	*C. reticulata* Blanco	0100694	INRA/CIRAD
Szibat	Mandarin	*C. reticulata* Blanco	0100596	INRA/CIRAD
Beauty of Glen Retreat	Mandarin	*C. reticulata* Blanco	0100261	INRA/CIRAD
Da hong pao	Mandarin	*C. reticulata* Blanco	0100591	INRA/CIRAD
Tankan SG	Mandarin	*C. reticulata* Blanco	0100524	INRA/CIRAD
Mathieu (Laï Vung)	Mandarin	*C. reticulata* Blanco	¿?	INRA/CIRAD
Sun Chu Sha	Mandarin	*C. reticulata* Blanco	0100786	INRA/CIRAD
Azimboa	Pummelo	*C. maxima (L.) Osb.*	420	IVIA
Deep red	Pummelo	*C. maxima (L.) Osb.*	277	IVIA
Pink	Pummelo	*C. maxima (L.) Osb.*	275	IVIA
Chandler	Pummelo	*C. maxima (L.) Osb.*	207	IVIA
Gil	Pummelo	*C. maxima (L.) Osb.*	321	IVIA
Da Xhang	Pummelo	*C. maxima (L.) Osb.*	589	IVIA
Nam Roi	Pummelo	*C. maxima (L.) Osb.*	590	IVIA
Flores	Pummelo	*C. maxima (L.) Osb.*	0100673	INRA/CIRAD
Timor	Pummelo	*C. maxima (L.) Osb.*	0100707	INRA/CIRAD
Sans Pepins	Pummelo	*C. maxima (L.) Osb.*	0100710	INRA/CIRAD
Arizona	Citron	*C. medica* L.	169	IVIA
Corcega	Citron	*C. medica* L.	567	IVIA
Cidro Digitado	Citron	*C. medica* L.	202	IVIA
Diamante	Citron	*C. medica* L.	560	IVIA
Poncire Commun	Citron	*C. medica* L.	0100701	INRA/CIRAD
Humpang	Citron	*C. medica* L.	0100722	INRA/CIRAD
Bali	Hybrid	x *C. maxima* L.	663	INRA/CIRAD
Poncil	Hybrid	x *C. medica* L.	151	IVIA
Rhobs el Arsa	Hybrid	*C. aurantium* L. x *C. medica* L.	0110244	INRA/CIRAD
Kadu Mul	Hybrid	*C. medica* L. x *C. maxima* (L.) Osb.	0100717	INRA/CIRAD
Damas	Hybrid	*C. aurantium* L. x *C. medica* L.	0100837	INRA/CIRAD
Sevillano	Sour orange	*C. aurantium* L.?	117	IVIA
Bouquet de Fleurs	Sour orange	*C. aurantium* L.?	139	IVIA
Clemenules	Clementine	*C. reticulata* Blanco	22	IVIA
Oronules	Clementine	*C. reticulata* Blanco	132	IVIA
Citrus daoxianensis	Mandarin	*C. reticulata* Blanco	359	IVIA
Eureka Frost	Lemon	*C. limon* (L.) Burm. f.	297	IVIA
Lisbon Limoneira	Lemon	*C. limon* (L.) Burm. f.	214	IVIA
Lemon meyer	Lemon	*C. limon* (L.) Burm. f.	145	IVIA
Marsh	Grapefruit	*C. paradisi* Macf.	176	IVIA
Star Ruby	Grapefruit	*C. paradisi* Macf.	197	IVIA
Shamouti	Sweet orange	*C. sinensis* (L.) Osb.	270	IVIA
Valencia late delta	Sweet orange	*C. sinensis* (L.) Osb.	363	IVIA
Lane late	Sweet orange	*C. sinensis* (L.) Osb.	198	IVIA
Sanguinelli	Sweet orange	*C. sinensis* (L.) Osb.	34	IVIA
Temple	Mandarin	Tangor?	81	IVIA
Suntara	Mandarin	*C. reticulata* Blanco	0110251	INRA/CIRAD
C-54-4-4	Hybrid	(*C. reticulata* x (*C.reticulata* x *C.sinensis*))	453	IVIA
Fairchild	Hybrid	(*C. reticulata* x (*C. paradisi* x *C. reticulata*))	83	IVIA
Fallglo	Hybrid	(*C. reticulata* x (*C. paradisi* x *C. reticulata*) x *C. reticulata*)	466	IVIA
Fortune	Hybrid	(*C. reticulata* x *C. reticulata*)	80	IVIA
Kara	Hybrid	(*C. reticulata* x *C. reticulata*)	218	IVIA
Nova	Hybrid	(*C. reticulata* x (*C. paradisi* x *C. reticulata*)	74	IVIA
Osceola	Hybrid	(*C. reticulata* x (*C. paradisi* x *C. reticulata*))	573	IVIA
Page	Hybrid	(*C. paradisi* x *C. reticulata*) x *C. reticulata*)	79	IVIA
Sunburst	Hybrid	((*C. reticulata* x (*C. paradisi* x *C. reticulata*)) x ((*C. reticulata* x (*C. paradisi* x *C. reticulata*))	200	IVIA
Wallent	Hybrid	Chance seedling	404	IVIA
Mapo	Tangelo	(*C. deliciosa* Ten. x *C. paradisi* Macf.)	190	IVIA
Minneola	Tangelo	(*C. paradisi* Macf. x *C. tangerina* Hort. ex Tan.)	84	IVIA
Orlando	Tangelo	(*C. paradisi* Macf. x *C. tangerina* Hort. ex Tan.)	101	IVIA
Seminole	Tangelo	(*C. paradisi* Macf. x *C. tangerina* Hort. ex Tan.)	348	IVIA
King	Tangor	*C. reticulata* Blanco	477	IVIA
Avasa 9	Tangor	(*C. reticulata* x *C. sinensis*)	405	IVIA
Dweet	Tangor	(*C. reticulata* x *C. sinensis*)	165	IVIA

Ellendale	Tangor	(*C.reticulata* x *C. sinensis*)	194	IVIA
Ellendale leng	Tangor	(*C.reticulata* x *C. sinensis*)	353	IVIA
Ellendale taranco	Tangor	(*C.reticulata* x *C. sinensis*)	575	IVIA
Murcott	Tangor	(*C.reticulata* x *C. sinensis*)	196	IVIA
Murcott sin semillas	Tangor	(*C.reticulata* x *C. sinensis*)	371	IVIA
Ortanique	Tangor	(*C.reticulata* x *C. sinensis*)	276	IVIA
Umatilla	Tangor	(*C. reticulata* x *C. sinensis*)	100	IVIA

(IVIA) *Banco de Germoplasma de Cítricos del Instituto Valenciano de Investigaciones Agrarias* (IVIA) *Apartado Oficial* 46113 Moncada, Valencia, Spain.

(INRA/CIRAD) INRA/CIRAD Citrus Germplasm, Centre Inra de Corse, F-20230 San Giuliano, France

Online Resource 2. SSRs markers characteristics

Marker name	Reverse sequence	Forward sequence	AT	Size	EMBL Accession	Published
MEST121	TCCCTATCATCGGCAACTTC	CAATAATGTTAGGCTGGATGGA	55	177-189	DY275927	Luro *et al*. 2008
MEST1	CAAGCCTCTCTCTTTAGTCCCA	AGTTCTTTGGTGCTTCAGGC	55	170-190	DY262452	New primer
MEST431	GAGCTCAAAACAATAGCCGC	CATACCTCCCCGTCCATCTA	55	331-345	DY291553	New primer
TAA15	GAAAGGGTTACTTGACCAGGC	CTTCCCAGCTGCACAAGC	55	164-204		Kijas *et al*. 1997
mCrCIR02D09	AATGATGAGGGTAAAGATG	ACCCATCACAAAACAGA	55	226-240	FR677569	Cuenca *et al*. 2011
TAA41	AGGTCTACATTGGCATTGTC	ACATGCAGTGCTATAATGAATG	55	127-162		Kijas *et al.* 1997
mCrCIR06B07	CGGAACAACTAAAACAAT	TGGGCTTGTAGACAGTTA	50	96-108	AM489745	Froelicher *et al.* 2008
mCrCI01C07	GTCACTCACTCTCGCTCTTG	TTGCTAGCTGCTTTAACTTT	55	260-298	AJ567394	Froelicher *et al.* 2008
mCrCIR05A05	CGGAACAACTAAAACAAT	TGGGCTTGTAGACAGTTA	50	144-179	FR677580	Cuenca *et al.* 2011
mCrCIR04H06	GGACATAGTGAGAAGTTGG	CAAAGTGGTGAAACCTG	55	184-196	FR677579	Cuenca *et al.* 2011
MEST46	GAACCAGAATCAGAACCCGA	GGTGAGCATCTGGACGACTT	55	230-256		New primer
CAC15	TAAATCTCCACTCTGCAAAAGC	GATAGGAAGCGTCGTAGACCC	55	168-180		Kijas *et al.* 1997
mCrCIR03C08	CAGAGACAGCCAAGAGA	GCTTCTTACATTCCTCAAA	55	200-225	FR677576	Cuenca *et al.* 2011
CAC23	ATCACAATTACTAGCAGCGCC	TTGCATTGTAGCATGTTGG	55	240-260		Kijas *et al.* 1997
MEST256	CATTAAAATATCCGTGCCGC	GAGCAAGTGCGTTGTTGTGT	55	200-230	DY290355	New primer
mCrCI02G12	AAACCGAAATACAAGAGTG	TCCACAAACAATACAACG	55	240-260	FR677575	Froelicher pers com
MEST131	TACCTCCACGTGTCAAACCA	GCTGTCACGTTGGGTGTATG	55	120-150	DY276912	New primer
mCrCI03D12a	GCCATAAGCCCTTTCT	CCCACAACCATCACC	50	240-280	FR677577	Aleza *et al.* 2011
mCrCI02D04b	CTCTCTTTCCCCATTAGA	AGCAAACCCCACAAC	50	199-229	FR677564	Kamiri *et al.* 2011
mCrCIR03G05	CCACACAGGCAGACA	CCTTGGAGGAGCTTTAC	50	199-228	FR677578	Cuenca *et al.* 2011
mCrCIR07D06	CCTTTTCACAGTTTGCTAT	TCAATTCCTCTAGTGTGTGT	55	164-197	FR677581	Cuenca *et al.* 2011
MEST15	TTATTACGAAGCGGAGGTGG	GCCTCGCATTCTCTTGACTC	55	192-210	FC912829	New primer
MEST104	CCTTATCTTCATCACCTCCGTC	TAAAAAGATGGGGCCTTGTG	55	240-260	DY273697	New primer
mCrCIR01F08a	ATGAGCTAAAGAGAAGAGG	GGACTCAACACAACACAA	50	128-156	AM489737	Froelicher *et al.* 2008
MEST88	GCCTGTTTGCTTTCTCTTTCTC	ATGAGAGCCAAGAGCACGAT	55	99-130	DY271576	New primer
mCrCI05A04	AAACGAGACAAGACCAAC	TATCAAACTCCCCTCACT	55	245-268	FR692372	Froelicher pers com
mCrCIR07E12	TGTAGTCAAAAGCATCAC	TCTATGATTCCTGACTTTA	50	106-146	AM489750	Froelicher *et al.* 2008
MEST115	CCCCCTCTTCTTTCACACAA	GGTGAGCAGCCATCTTCTTC	55	147-167	DY274953	New primer
mCrCIR06A12	CCCAACAAACTCAAACTTC	TTTTTATTTCGGTCTCCTT	50	84-102	AM489742	Froelicher *et al.* 2008
MEST56	AGTCCGCCTTTGCTTTTTCT	GGTGCAAAAGAGAGCGAGAG	55	129-145	DY267791	Aleza *et al.* 2011
mCrCI04H12	TTCCTCTACAACTACAACCA	ATTATCCTCAACCTCCAA	50	179-194	FR692371	Froelicher pers com
MEST192	CGCGGATCATCTAGCATACA	CTTGGCACCATCAACACATC	55	200-240	DY283129	Aleza *et al*. 2011
mCrCIR02F12	GGCCATTTCTCTGATG	TAACTGAGGGATTGGTTT	55	116-144	FR677570	Cuenca *et al.* 2011
Ci01D11	GCAAAACAAGCAGACTACAAAT	AGGACAGATGACCCAGATGACA	55	214-230	AJ567397	Froelicher *et al.* 2008
MEST488	CACGCTCTTGACTTTCTCCC	CTTTGCGTGTTTGTGCTGTT	55	133-164	DY297637	New primer
mCrCIR01E02	TGAATGGTACGGGAAATGC	CAGGGTCGGTGGAGAGGAT	55	172-184	AM489735	Froelicher *et al.* 2008
mCrCIR01C06	GGACCACAACAAAGACAG	TGGAGACACAAAGAAGAA	50	131-170	AJ567393	Cuenca *et al.* 2011
TAA1	GACAACATCAACAACAGCAAGAGC	AAGAAGAAGAGCCCCCATTAGC	55	161-180		Kijas *et al.* 1997
MEST107	GCTGAGATGGGGATGAAAGA	CCCCATCCTTTCAACTTGTG	55	183-201	DY274062	New primer
mCrCIR03B07	CACCTTTCCCTTCCA	TGAGGGACTAAACAGCA	55	263-279	FR677573	Cuenca *et al.* 2011
Ci07C07	TATCCAGTTTGTAAATGAG	TGATATTTGATTAGTTTGG	50	243-258	AJ567409	Froelicher *et al.* 2008
mCrCIR02A09	ACAGAAGGTAGTATTTTAGGG	TTGTTTGGATGGGAAG	55	151-177	FR677568	Cuenca *et al.* 2011
mCrCIR02G02	CAATAAGAAAACGCAGG	TGGTAGAGAAACAGAGGTG	55	110-138	FR677572	Cuenca *et al.* 2011
Ci02F07	GCAGCGTTTGTTTTCT	TGCTGGTTTTCAGATACTT	55	188-215	AJ567406	Froelicher *et al.* 2008
mCrCIR07B05	TTTGTTCTTTTTGGTCTTTT	CTTTTCTTTCCTAGTTTCCC	50	218-254	AM489747	Froelicher *et al.* 2008
mCrCIR01F04a	AAGCATTTAGGGAGGGTCACT	TGCTGCTGCTGTTGTTGTTCT	55	190-228	AM489736	Froelicher *et al.* 2008
MEST86	CCAACTGACACTAATCCTCTTCC	CCTCTCTGGCTTCTGGATTG	55	110-128	DY271447	New primer
Ci07C09	GACCCTGCCTCCAAAGTATC	GTGGCTGTTGAGGGGTTG	55	258-274	AJ567410	Froelicher *et al.* 2008
mCrCIR07F11	ACTATGATTACTTTGCTTTGAG	GAAGAAACAAGAAAAAAAAAT	50	146-176	FR677567	Kamiri *et al.* 2011
Ci02B07	CAGCTCAACATGAAAGG	TTGGAGAACAGGATGG	50	178-212	AJ567403	Froelicher *et al.* 2008

(AT) Annealing temperature

Online Resource 3. InDel sequence alignment found in candidate genes

```
385I-CHI      TTTCCTCTTGCTTTACGTGTAATAATAATAAATTAACAATACAGGTGCATTAAATATTTA 60
154I-CHI      TTTCCTCTTGCGTTACGTGTAATGATAATAAATTAACAATACAGGTGCATTAAATATTTA 60
239I-CHI      TTTCCTCTTGCGTTACGTGTAATA-----------------CAGGTGCATTAAATATTTA 43
207I-CHI      TTTCCTCTTGCGTTACGTGTAATAATAATAAATTAACAATACAGGTGCATTAAATATTTA 60
275I-CHI      TTTCCTCTTGCGTTACGTGTAATAATAATAAATTAACAATACAGGTGCATTAAATATTTA 60
567I-CHI      TTTCCTCTTGCGTTACGTGTAATA---------------------------ATAA---TA 30
202I-CHI      TTTCCTCTTGCGTTACGTGTAATA---------------------------ATAAATTTA 33
              ********** **********                               * *    **

385I-CHI      AATTCACACTATCCGTATGGGAATCCTTTCCCGTCATAAACGCTGCTTAAAGAGTAGTCA 120
154I-CHI      AATTCACACTATCCGTATGGGAATCCTTTTCCGTCATAAACGCTGCTTAAAGAGTAGTGA 120
239I-CHI      AATTCACACTATCCGTATGGGAATCCTTTCCCGTCATAAACGCTGCTTAAAGAGTAGTCA 103
207I-CHI      AATTCACACTATCCGTATGGGAATCCTTTTCCGTCATAAACGCTGCTTAAAGAGTAGTCA 120
275I-CHI      AATTCACACTATCCGTATGGGAATCCTTTTCCGTCATAAACGCTGCTTAAAGAGTAGTCA 120
567I-CHI      AATTAGCAATACAGGTG-----------------CATTAAA--TATTTAAAGAGTAGTCA 71
202I-CHI      AATTCCCACTATCCGTATGAGAATCCTTTTCCGTCATAAACGCTGCTTAAAGAGTGGTCA 93
              ****  ** **   **                  *** **   *  ********* ** *

385I-CHI      ACGTCAGTACTACACTCAAAATCTAAAACAGAATCCAACAGAAGACACCAACGGCGAAAA 180
154I-CHI      ACGTCAGTACTTCACTCAAAATCTAAAACAGAATCCAACAGAAGAAACCAACGGCGAAAA 180
239I-CHI      ACGTCAGTACTACACTCAAAATCTAAAACAGAATCCAACAGAAGACACCAACGGCGAAAA 163
207I-CHI      ACGTCAGTACTNCACTCAAAATCTAAAACAGAATCCAACAGAAGAAACCAACGGCGAAAA 180
275I-CHI      ACGTCAGTACTNCACTCAAAATCTAAAACAGAATCCAACAGAAGAAACCAACGGCGAAAA 180
567I-CHI      ACGTCAGCACTTCAATCAAAATCTAAAACAGAATCCAACAGAAGAAACCAACGGCGAAAA 131
202I-CHI      ACGTCAGTACTTCACTCAAACTCTAAAACAGAATCCAACAGAAGAAACCAACGGCGAAAA 153
              ******* *** ** ***** ********************** **************

385I-CHI      TCCGTTACCTGTGAC 195
154I-CHI      TCCGTTACCTGTGAC 195
239I-CHI      TCCGTTACCTGTGAC 178
207I-CHI      TCCGTTACCTGTGAC 195
275I-CHI      TCCGTTACCTGTGAC 195
567I-CHI      TCCGTTACCTGTGAC 146
202I-CHI      TCCGTTACCTGTGAC 168
              ***************
```

```
385I-EMA      CTCTTTCTGCTTCCTGACATCTAAATTATATGAATAGGCTTTTGTTG----------TCA 50
154I-EMA      CTCTTTCTGCTTCCTGACATCTAAATTATATGAATAGGCTTTTGTTG----------TCA 50
207I-EMA      CTCTTTCTGCTTCCTGACATCTAAATTATATGAATAGGCTTTTGTTG----------TCA 50
275I-EMA      CTCTTTCTGCTTCCTGACATCTAAATTATATGAATAGGCTTTTGTTG----------TCA 50
567I-EMA      CTCTTTCTGCTTCCTGACATCTAAATTATATGAATAGGCTTTTGTTGTCAATTGTTGTCA 60
202I-EMA      CTCTTTCTGCTTCCTGACATCTAAATTATATGAATAGGCTTTTGTTGTCAATTGTTGTCA 60
              ********************************************          ***

385I-EMA      AATGGACTGAAATAATTAGGATGCAACAGAAATTAACTGCATGTTGCACCACCATTTAAG 110
154I-EMA      AATGGACTGAAATAATTAGGATGCAACAGAAATTAACTGCATATTGCACCACCATTTAAG 110
207I-EMA      AATGGACTGAAATAATTAGGATACAACAGAAATTAACTGCATATTGCACCACCATTTAAG 110
275I-EMA      AATGGACTGAAATAATTAGGATACAACAGAAATTAACTGCATATTGCACCACCATTTAAG 110
567I-EMA      AATGGACTGAAATAATTAGGATACAACAGAAATTAACTGCATATTGCACCACCATTTAAG 120
202I-EMA      AATGGACTGAAATAATTAGGATACAACAGAAATTAACTGCATATTGCACCACCATTTAAG 120
              **********************.******************.*****************

385I-EMA      AACAGTTTGTTACAATGTGAACAAGTCCACTGGAAAAATCCATTAACAAATATTTGAATT 170
154I-EMA      AACAGTTTGTTACAATGTGAACAAGTCCACTGGAAAAATCCATTAACAAATATTTGAATT 170
207I-EMA      AACAGTTTGTTACAATGTGAACAAGTCCACTGGAAAAATCCATTAACAAATATTTGAATT 170
275I-EMA      AACAGNTTGTTACAATGTGAACAAGTCCACTGGAAAAATCCATTAACAAATATTTGAATT 170
567I-EMA      AACAGTTTGTTACAATGTGAACAAGTCCACTGGAAAAATCCCTTGACAAAAATTTGAATT 180
202I-EMA      AACAGTTTGTTACAATGTGAACAAGTCCACTGGAAAAATCCCTTGACAAAAATTTGAATT 180
              *****.*********************************.**.*****:*********
```

```
385I-EMA      AGCCGTGAACGTAAGTGTTCTCTTGGCAAACGTGTAAAATCNTTAGAGCTTGTTTACTTG 230
154I-EMA      AGCCGTGAACGTAAGTGTTCTCTTGGCAAACGTGTAAAATCATTAGAGCTTGTTTACTTG 230
207I-EMA      AGCCGTGAACGTAAGTGTTCTCTTGGAAAACGTGTAAAATCGTTAGAGCTTGTTTACTTG 230
275I-EMA      AGCCGTGAACGTAAGTGTTCTCTTGGAAAACGTGTAAAATCGTTAGAGCTTGTTTACTTG 230
567I-EMA      AGCCGTGAACGTAAGTGTTCTCTTGGAAAACGTGTAAAATCGTTAGAGCTTGTTTACTTG 240
202I-EMA      AGCCGTGAACGTAAGTGTTCTCTTGGAAAACATGTAAAATCGTTAGAGCTTGTTTACTTG 240
              **************************.****.********* ******************

385I-EMA      GTGATTGATAAACTAGTTGTGTTTTATTCACCGGC 265
154I-EMA      GTGATTGATAAACTAGTTGTGTTTTATTCACCGGC 265
207I-EMA      GCGATTGATAAACTAGTTGTGTTTTATTCACCGAC 265
275I-EMA      GCGATTGATAAACTAGTTGTGTTTTATTCACCGAC 265
567I-EMA      GCGATTGATAAACTAGTTGTGTTTTATTCACCGAC 275
202I-EMA      GCGATTGATAAACTAGTTGTGTTTTATTCACCGAC 275
              * ******************************.*
```

```
385I-TRPA     CCCTCGTTCTTGGTAGCTTTATTCTTGCTCTCGCCGTTGAGCACTACAACATTCACAGAA 60
154I-TRPA     CCCTCGTTCTTGGTAGCTTTATTCTTGCTCTCGCCGTTGAGCACTACAACATTCACAGAA 60
207I-TRPA     CCCTCGTTCTTGGTAGCTTTATTCTTGCTCTTGCCGTTGAGCACTACAACATTCACAAAA 60
275I-TRPA     CCCTCGTTCTTGGTAGCTTTATTCTTGCCCTTGCCGTTGAGCACTACAACATTCACAAAA 60
567I-TRPA     CCCTCGTTCTTGGTAGCTTTATTCTTGCTCTTGCCGTTGAGCACTACAACATTCACAAAA 60
202I-TRPA     CCCTCGTTCTTGGTAGCTTTATTCTTGCTCTTGCCGTTGAGCACTACAACATTCACAAAA 60
              **************************** ** ************************ **

385I-TRPA     GATTGGCCTTAAATGTAAGTTCCCATAATGCATCATCATCATCATGTCATTAATCGTTAC 120
154I-TRPA     GATTGGCCTTAAATGTAAGTTCCCATAATGCATCATCATCATCATGTCATTAATCGTTAC 120
207I-TRPA     GATTGGCCTTAAATGTAAGTTCCCGTAATGCATCATCATCATCATGTCATTAATTGTTAC 120
275I-TRPA     GATTGGCCTTAAATGTAAGTTCCCGTAATGCATCATCATCATCATGTCATTAATTGTTAC 120
567I-TRPA     GATTGGCCTTAAATGTAAGTTCCCGTAATGCATCATCATCATCATGTCATTAATTGTTAC 120
202I-TRPA     GATTGGCCTTAAATGTAAGTTCCCGTAATGCATCATCATCATCATGTCATTAATTGTTAC 120
              ************************ ***************************** *****

385I-TRPA     GATTTCTTTTTCAGAAAAATTATCAGTGACAAAAGATGAATTAATTATGTATGGACAATC 180
154I-TRPA     GATTTCTTTTTCAGAAAAATTATCAGTGACAAAAGATGAATTAATTATGTATGGACAATC 180
207I-TRPA     GATTTCTTTTTCAGAAAAATTATCAGTGACAAAAGATGAATTAATTATGTATGGACAATC 180
275I-TRPA     GATTTCTTTTTCAGAAAAATTATCAGTGACAAAAGATGAATTAATTATGTATGGACAATC 180
567I-TRPA     GATTTCTTTTTCAGAAAAATTATCAGTGACAAAAGATGAATTAATTATGTATGGACAATC 180
202I-TRPA     GATTTCTTTTTCAGAAAAATTATCAGTGACAAAAGATGAATTAATTATGTATGGACAATC 180
              ************************************************************

385I-TRPA     CTATACCATATAATATATATTAATAACTACAGATAACTATTCTATTCTGTGGAGAGCCAA 240
154I-TRPA     CTATACCATATAATATATATTAATAACTACAGATAACTATTCTATTCTGTGGAGAGCCAA 240
207I-TRPA     CNATACCATATAATATATATTAATAACTACCGATAACTATTCTATTCTGTGGANAGCCAA 240
275I-TRPA     CTATACCATATA---NANNNTANTAACNACTNATNACTANTNTANTCTGCGGAGANCNNA 237
567I-TRPA     CTATACCATATAATTTATATTAATAACTACAGATAACTATTCTATTCTGTGGAGAGCCAA 240
202I-TRPA     CTATACCATATAATTTATATTAATAACTACAGATAACTATTCTATTCTGTGGAGAGCCAA 240
              * **********    *   ** **** **  ** **** * ** **** *** * *  *

385I-TRPA     TGAATCCGCCCTTGCTGCTTCTTGGGATATGTGGCACGACAGCATTCGTGAGCATGTGGA 300
154I-TRPA     TGAATCCGCCCTTGCTGCTTCTTGGGATATGTGGCACGACAGCATTCGTGAGCATGTGGA 300
207I-TRPA     TGAATCCGCCCTTGCTGCTTCTTGGGATATGTGGCACGACAGCATTCCTGAACATGTGGA 300
275I-TRPA     NGAAGNNNCCCNNGCTGCTNCTTGGGNNAAGAGGNGNGACNACATNCTTNATNATGNGGA 297
567I-TRPA     TGAATCCGCCCTTGCTGCTTCTTGGGATATGTGGCACGACAGCATTCGTGAGCATGTGGA 300
202I-TRPA     TGAATCCGCCCTTGCTGCTTCTTGGGATATGTGGCACGACAGCATTCGTGAGCATGTGGA 300
               ***    ***  ****** ******  * * **   ***  *** * * *  *** ***

385I-TRPA     TGCATAA 307
154I-TRPA     TGCATAA 307
207I-TRPA     TGCATAA 307
275I-TRPA     NGNTNNA 304
567I-TRPA     TGCATAA 307
202I-TRPA     TGCATAA 307
               *    *
```

```
385I-PEPC1    TTTTGAACAATCGGCTAATGGTAGATATTGTACCAACTTTTTATATGTAATATGAAATTT 60
154I-PEPC1    TTTTGAACAATCGGCTAATGGTAGATATTGTACCAACTTTTTATATGTAATATGAAATTT 60
207I-PEPC1    TTTTGAACAATCGGCTAATGGTAGATATTGTACCAACTTTTTATATGTAATATGAAATTT 60
275I-PEPC1    TTTTGAACAATCGGCTAATGGTAGATATTGTACCAACTTTTTATATGTAATATGAAATTT 60
567I-PEPC1    TTTTGAACAATCGGCTAATGGTAGACATTGTACCAACTTTTTATATGTAATATGAAATTT 60
202I-PEPC1    TTTTGAACAATCGGCTAATGGTAGACATTGTACCAACTTTTTATATGTAATATGAAATTT 60
              ************************* ********************************

385I-PEPC1    TGGTTATTTAT--------------------------GTAGCCTTATTTATTGGAAAGT 93
154I-PEPC1    TGGTTATTTAT--------------------------GTAGCCTTATTTATTGGAAAGT 93
207I-PEPC1    TGGTTATTTATATGTAATATGAAATTTTGGTTATTTTTGTAGTCTTATTTATTGGAAAGT 120
275I-PEPC1    TGGTTATTTATATGTAATATGAAATTTTGGTTATTTTTGTAGTCTTATTTATTGGAAAGT 120
567I-PEPC1    TGGTTATTTAT--------------------------GTAGCCTTATTTATTGGAAAGT 93
202I-PEPC1    TGGTTATTTAT--------------------------GTAGCCTTATTTATTGGAAAGT 93
              ***********                          **** *****************

385I-PEPC1    GCATTTAAGAACTGAGAAGGCATAGAATATTCCATTAGGTTTGAAGAAATTCATTGCTCT 153
154I-PEPC1    GCATTTAAGAACTGAGAAGGCATAGAATATTCCATTAGGTTTGAAGAAATTCATTGCTCT 153
207I-PEPC1    GCATTTAAGAACTGAGAAGGCATAGAATATTCCACTAGGTTTGAAGAAATTCATTGCTCT 180
275I-PEPC1    GCATTTAAGAACTGAGAAGGCATAGAATATTCCACTAGGTTTGAAGAAATTCATTGCTCT 180
567I-PEPC1    GCATTTAAGAACTGAGAAGGCATAGAATATTCCACTAGGTTTGAAGGAATTCATTGCTCT 153
202I-PEPC1    GCATTTAAGAACTGAGAAGGCATAGAATATTCCACTAGGTTTGAAGGAATTCATTGCTCT 153
              ********************************** *********** *************

385I-PEPC1    TTAAGTCAGCTTTAAGTGAATATCCTTGTTATAAACTTTAGTGAGAGTGAATGCATTGGA 213
154I-PEPC1    TTAAGTCAGCTTTAAGTGAATATCCTTGTTATAAACTTTAGTGAGAGTGAATGCATTGGA 213
207I-PEPC1    TTAAGTCAGCTTTAAGTGAATATCCTTGTTATAAACTTTAGTGAGAGTGAATGCATTGGA 240
275I-PEPC1    TTAAGTCAGCTTTAAGTGAATATCCTTGTTATAAACTTTAGTGAGAGTGAATGCATTGGA 240
567I-PEPC1    TTAAGTCAGCTTTAAGTGAATATCCTTGTTATAAACTTTAGTGAGAGTGAATGCATTGGA 213
202I-PEPC1    TTAAGTCAGCTTTAAGTGAATATCCTTGTTATAAACTTTAGTGAGAGTGAATGCATTGGA 213
              ************************************************************

385I-PEPC1    GTCTCTCTTCCAGCAA 229
154I-PEPC1    GTCTCTCTTCCAGCAA 229
207I-PEPC1    GTCTCTCTTCCAGCAA 256
275I-PEPC1    GTCTCTCTTCCAGCAA 256
567I-PEPC1    GTCTCTCTTCCAGCAA 229
202I-PEPC1    GTCTCTCTTCCAGCAA 229
              ****************
```

```
385I-PEPC2    TTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTTTCCCACAACAGAC 60
154I-PEPC2    TTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTTTCCCACAACAGAC 60
207I-PEPC2    TTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTTTCCCATAACAGAC 60
275I-PEPC2    TTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTTTCCCATAACAGAC 60
567I-PEPC2    TTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTTTCCCATAACAGAC 60
202I-PEPC2    TTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTTTCCCATAACAGAC 60
              ************************************************* *******

385I-PEPC2    TAGCTGAGCTTCAATTTTGATTTTCTTTTCTGAATGAGTTTTGAAAATATTCGATAGGAC 120
154I-PEPC2    TAGCTGAGCTTCAATTTTGATTTTCTTTTCTGAATGAGTTTTGAAAATATTCGATAGGAC 120
207I-PEPC2    TAGCTAAGCTTCAATTTTGATTTTCTTTTCTGAATGAATTTTGAAAATATTCGATAGGAC 120
275I-PEPC2    TAGCTAAGCTTCAATTTTGATTTTCTTTTCTGAATGAATTTTGAAAATATTCGATAGGAC 120
567I-PEPC2    TAGCTAAGCTTCAATTTTGA------------------------AAATATTCGATAGGAC 96
202I-PEPC2    TAGCTAAGCTTCAATTTTGA------------------------AAATATTCGATAGGAC 96
              ***** **************                        ****************

385I-PEPC2    AATACTGAAATTTTTGCATTGTGGCTCTCAC 151
154I-PEPC2    AATACTGAAATTTTTGCATTGNGGCTCTCAC 151
207I-PEPC2    AATACTGAAATTTTTGCATTGTGGCTCTCAC 151
275I-PEPC2    AATACTGAAATTTTTGCATTGTGGCTCTCAC 151
567I-PEPC2    AATACTGAAATTTTTGCATTGTGGCTCTCAC 127
202I-PEPC2    AATACTGAAATTTTTGCATTGTGGCTCTCAC 127
              ********************* *********
```

```
385I-LCY2     CGCAAATAATTGATTCAACATCATCACATTCATTTTCGCTATTTCCATTAGGCCGCCAAA 60
154I-LCY2     CGCAAATAATTGATTCAACATCATCACCTTCATTTTCCCTATTTCCATTAGGCCGCCAAA 60
239I-LCY2     CGCAAATCATTGATTCAACATCGTCCCATTCATTTTCCCTATTTCCATTAGGCCGCCAAA 60
207I-LCY2     CGCAAATAATTGATTCAACATCATCACCTTCATTTTCCCTATTTCCATTAGGCCGCCAAA 60
275I-LCY2     CGCAAATAATTGATTCAACATCATCACCTTCATTTTCCCTATTTCCATTAGGCCGCCAAA 60
567I-LCY2     CGCAAATAATTGATTCAACATCATCACATTCATTTTCGCTATTTCCATTAGGCCGCCAAA 60
202I-LCY2     CGCAAATAATTGATTCAACATCATCACATTCATTTTCGCTATTTCCATTAGGCCGCCAAA 60
              ******* ************** ** * ********* ***********************

385I-LCY2     ATGCATGTTCAAGAAAGGCGGATCATCATCATCAT---CACAGGATCCGGACAAGCAAGT 117
154I-LCY2     ATGCATGTTCAAGAAAGGCGGATCATCATCATCAT---CACAGGATCCGGACAAGCAAGT 117
239I-LCY2     ATGCGTGTTCAAGAAAGGCGGGTCGTCATCATCGC------AGGATCCGGACAAGCAAGT 114
207I-LCY2     ATGCATGTTCGAGAAAGGCGGATCATCATCATCAT---CACAGGATCCGGACAAGCAAGT 117
275I-LCY2     ATGCATGTTCNAGAAAGGCGGATCATCATCATCAT---CACAGGATCCGGACAAGCAAGT 117
567I-LCY2     ATGCATGTTCAAGAAAGGCGGATCATCATCATCATCATCACAGGATCCGGACAAGCAAGT 120
202I-LCY2     ATGCATGTTCAAGAAAGGCGGATCATCATCATCATCATCACAGGATCCGGACAAGCAAGT 120
              **** ***** ********** ** ********        ******************

385I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGAACCTGAATTCTTAGTCTTTGATCTCC 177
154I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGAACCTGAATTCTTAGNCTTTGATCTCC 177
239I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGAACCTGAATTGTTAGACTTTGATCTCC 174
207I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGTACCTGAATTCTTAGACTTTGATCTCC 177
275I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGTACCTGAATTCTTAGACTTTGATCTCC 177
567I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGAACCTGAATTCTTAGACTTTGATCTCC 180
202I-LCY2     TTGGTAACTTCCTAGAGTTGACACCGGAGTCGGAACCTGAATTCTTAGACTTTGATCTCC 180
              ********************************* ********* **** ***********

385I-LCY2     CCTGGTTTCATCCGTCCGATCGTATTCGATATGACGTGATCATC 221
154I-LCY2     CCTGGTTTCATCCGTCCGATCGTATTCGATATGACGTGATCATC 221
239I-LCY2     CCTGGTTTCATCCATCCGATCGTATTCGATATGACGTGATCATC 218
207I-LCY2     CCTGGTTTCATCCGTCCGATCGTATTCGATATGACGTGATCATC 221
275I-LCY2     CCTGGTTTCATCCGTCCGATCGTATTCGATATGACGTGATCATC 221
567I-LCY2     CCTGGTTTCATCCGTCCGATCGTATTCGATATGACGTGATCATC 224
202I-LCY2     CCTGGTTTCATCCGTCCGATCGTATTCGATATGACGTGATCATC 224
              ************* ****************************
```

```
385I-HYB1     AAAAACAAAGCACCCAGATCGAGACTTTCACGGACGAGGAGGAGGAGGAG---GNGGANN 57
154I-HYB1     AAAAACAAAGCACCCAGATCGAGACTTTCACGGAGGAGGAGGAGGAGGAG---TCGGGTA 57
207I-HYB1     AAAAACAAAGCACCCAGATCGAGACTTTCACGGAGGAGGAGGAGGAGGAG---TCGGGTA 57
275I-HYB1     AAAAACAAAGCACCCAGATCGAGACTTTCACGGAGGAGGAGGAGGAGGAG---TCGGGTA 57
567I-HYB1     AAAAACAAAGCACCCAGATCGAGACTTTCACGGAGGAGGAGGAGGAGGAGGAGTCGCGTA 60
202I-HYB1     AAAAACAAAGCACCCAGATCGAGACTTTCACGGAGGAGGAGGAGGAGGAGGAGTCGCGTA 60
              ********************************** ***************     *

385I-HYB1     NCCCGNNCNCCNCTGNNGCCCCCGNGGCCCACANGNTGGAGAGANNGACAANCAAGAGGT 117
154I-HYB1     CCCAGATCTCGACTGCTGCCCGCGTGGCCGAGAAATTGGCGAGAAAGAGATCCGAGAGGT 117
207I-HYB1     CCCAGATCTCGACTGCTGCCCGCGTGGCCGAGAAATTGGCGAGAAAGAGATCCGAGAGGT 117
275I-HYB1     CCCAGATCTCGACTGCTGCCCGCGTGGCCGAGAAATTGGCGAGAAAGAGATCCGAGAGGT 117
567I-HYB1     CCCAGATCTCGACTGCTGCCCGCGTGGCCGAGAAATTGGCGAGAAAGAGATCCGAGAGGT 120
202I-HYB1     CCCAGATCTCGACTGCTGCCCGCGTGGCCGAGAAATTGGCGAGAAAGAGATCCGAGAGGT 120
               ** *  * *  ***  **** ** **** * *   *** ****  ** *  * ******

385I-HYB1     NCAATNNNCTCNCTNCTNCCGTCGNGNCCAGNTTGGNTATCNCTNNCATGGCTGCCATGG 177
154I-HYB1     TCACTTATCTCGTTGCTGCCGTCATGTCTAGTTTTGGTATCACTTCCATGGCTGTCATGG 177
207I-HYB1     TCACTTATCTCGTTGCTGCCGTCATGTCTAGTTTTGGTATCACTTCCATGGCTGTCATGG 177
275I-HYB1     TCACTTATCTCGTTGCTGCCGTCATGTCTAGTTTTGGTATCACTTCCATGGCTGTCATGG 177
567I-HYB1     TCACTTATCTCGTTGCTGCCGTCATGTCTAGTTTTGGTATCACTTCCATGGCTGTCATGG 180
202I-HYB1     TCACTTATCTCGTTGCTGCCGTCATGTCTAGTTTTGGTATCACTTCCATGGCTGTCATGG 180
               ** *   ***  * ** *****  * * ** ** * **** **  ******** *****

385I-HYB1     CTGTNNNGNNCGTGTNCTGGTGGN 201
154I-HYB1     CTGTTTATTACAGGTTCTGGTGGC 201
207I-HYB1     CTGTTTATTACAGGTTCTGGTGGC 201
275I-HYB1     CTGTTTATTACAGGTTCTGGTGGC 201
567I-HYB1     CTGTTTATTACAGGTTCTGGTGGC 204
202I-HYB1     CTGTTTATTACAGGTTCTGGTGGC 204
              ****      *  ** *******
```

```
385I-HYB2     TTTGGCACATTTGCTCTCTCTGTTGGCGCTGCCGTAAGTTCAATCACCTTCTTCCTTACA 60
154I-HYB2     TTTGGCACATTTGCTCTCTCTGTTGGTGCTGCTGTAAGTTCAATCACCTTCTTCCTTACA 60
207I-HYB2     TTTGGCACATTTGCTCTCTCTGTTGGTGCTGCTGTAAGTTCAATCACCTTCTTCCTTACA 60
275I-HYB2     TTTGGCACATTTGCTCTCTCTGTTGGTGCTGCTGTAAGTTCAATCACCTTCTTCCTTACA 60
567I-HYB2     TTTGGCACATTTGCTCTCTCTGTTGGCGCTGCCGTAAGTTCAATCACCTTCTTCCTTACA 60
202I-HYB2     TTTGGCACATTTGCTCTCTCTGTTGGCGCTGCCGTAAGTTCAATCACCTTCTTCCTTACA 60
              ************************** ***** **************************

385I-HYB2     ATGATTTGAAAACAAGACTAGAATTTTGGTTCTAATAGGAGCCGCGGTGGGGATGTTACA 120
154I-HYB2     ATGATTTGAAAACAAGACTAGAATTTTGGTTCTAATAGGAGCCGCGGTGGGGATGTTACA 120
207I-HYB2     ATGATTTGAAAACAAGACTAGAATTTTGGTTCTAATAGGAGCCGCNGTGGGGATGTTACA 120
275I-HYB2     ATGATTTGAAAACAAGACTAGAATTTTGGTTCTAATAGGAGCCGCNGTGGGGATGTTACA 120
567I-HYB2     ATGATTTGAAAACAAGACTAGAATTTTGGTTCTAATAGGAGCCGCGGTGGGGATGTTACA 120
202I-HYB2     ATGATTTGAAAACAAGACTAGAATTTTGGTTCTAATAGGAGCCGCGGTGGGGATGTTACA 120
              ********************************************** **************

385I-HYB2     AACTTGATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTTTGAATTTTCAAT 180
154I-HYB2     AACTTGATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTGTGAATTTTCAAT 180
207I-HYB2     AACTTGATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTGTGAATTTTCAAT 180
275I-HYB2     AACTTGATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTGTGAATTTTCAAT 180
567I-HYB2     AACTTGATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTTTGAATTTTCAAT 180
202I-HYB2     AACTTGATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTTTGAATTTTCAAT 180
              ********************************************** ************

385I-HYB2     GTGAAAGCCTTTTCTGGCAAATTATATGATGATGATTCGCATTGGGTCCCTTTTTTTTTC 240
154I-HYB2     GTGAAGGCCTTTTCTGGCAAATTATATGATGATGATTCGCATTGGGTACCTTTTTTTTTC 240
207I-HYB2     GTGAAGGCCTTTTATG-CAAATTATGTGTTGATGATTAGCATTGGGTACCTTTTTTTT-C 238
275I-HYB2     GTGAAGGCCTTTTATG-CAAATTNTGTGTTGATGATTAGCATTGGGTACCTTTTTTTT-C 238
567I-HYB2     GTGAAAGCCTTTTCTGGCAAATTATATGATGATGATTCGCATTGGGTCCCTTTTTTTTTC 240
202I-HYB2     GTGAAAGCCTTTTCTGGCAAATTATATGATGATGATTCGCATTGGGTCCCTTTTTTTTTC 240
              ***** ******* ** ****** * ** ******** ********* ********** *

385I-HYB2     ATTTGCAGGTGGGCATGGAGTTTTGGGCACGATGGGCTCATAAAGCTCTGTGGCATGCTT 300
154I-HYB2     ATTTGCAGGTGGGCATGGAGTTTTGGGCACGATGGGCTCATAAAGCTCTGTGGCATGCTT 300
207I-HYB2     ATTTGCAGGTGGGCATGGAGTTTTGGGCACGATGGGCTCATAA-GCTCTGTGGCATGCTT 297
275I-HYB2     ATTTGCAGGTGGGCATGGAGTTTTGGGCACGATGGGCTCATAA-GCTCTGTGGCATGCTT 297
567I-HYB2     ATTTGCAGGTGGGCATGGAGTTTTGGGCACGATGGGCTCATAAAGCTCTGTGGCATGCTT 300
202I-HYB2     ATTTGCAGGTGGGCATGGANTTTTGGGCACGATGGGCTCATAAAGCTCTGTGGCATGCTT 300
              ******************* ********************** ****************

385I-HYB2     CTTT 304
154I-HYB2     CTTT 304
207I-HYB2     CTTT 301
275I-HYB2     CTTT 301
567I-HYB2     CTTT 304
202I-HYB2     CTTT 304
              ****
```

```
385I-PSY      CCTGTCGACATTCAGGTTAGACTATGTTTTCAAGATCAAATTATATTTTAACAAAATGGT 60
154I-PSY      CCTGTCGACATTCAGGTTAGACTATGTTTTCAAGATCAAATTAGATTTTAACAAAATGGT 60
207I-PSY      CCTGTCGACATTCAGGTTAGACTATGTGTTCAAGATCAANTTAGATTTTAACAAAATGGT 60
275I-PSY      CCTGTCGACATTCAGGTTAGACTATGTTTTCAAGATCAAATTAGATTTTAACAAAATGGT 60
567I-PSY      CCTGTCGACATTCAGGTTAGACTATGTTTTCAAGATCAAATTAGATTTTAACAAAATGGT 60
202I-PSY      CCTGTCGACATTCAGGTTAGACTATGTTTTCAAGATCAAATTAGATTTTAACAAAATGGC 60
              ************************** *********** *** ***************

385I-PSY      TGTTATAGTACTCTCTCTACTCTCTTAAGTGTACTTGTATTAAATTAAAATAAGGAACAA 120
154I-PSY      TGTTATAGTACTCTCTCTACTCTCTTAAGTGTACTTGTATTAAATTAAAATAAGGAACAA 120
207I-PSY      TGTTATAGTACTCTCTCTACTATCTTAAGTGTACTTGTATTAAATTAAAATAAGGAACAA 120
275I-PSY      TGTTATAGTACTCTCTCTACTATCTTAAGTGTACTTGTATTAAATTAAAATAAGGAACAA 120
567I-PSY      TGTTATAGTACTCTCTCTACTATCTTAAGTGTACTTGTATTAAATTACAATAAGGAACAA 120
202I-PSY      TGTTATAGTACTCTCTCTACTATCTTAAGTATACTTGTATTAAATTAAAATAAGGAACAA 120
              **********************.********.****************.************
```

```
385I-PSY      CTTCTGCTTTCTAATTGGTTTTTAAAACATTAAGCCTTGATGCATAATGACAGACCTTAT 180
154I-PSY      CTTCTGCTTTCTAATTGGTTTTTAAAACATTAANCCTTGATGCATAATGACAGACCTTAT 180
207I-PSY      CTTCTGCTTTCTAATTGGTTTTTAAAACATTAAGCCTTGATGCATAATGACAGACCTTAT 180
275I-PSY      CTTCTGCTTTCTAATTGGTTTTTAAAACATTAAGCCTTGATGCATAATGACAGACCTTAT 180
567I-PSY      CT---GCTTTCTAATTGGTTTTTAAAACATTAAGCCTTGATGCATAATGATAGACCTTAT 177
202I-PSY      CT---GCTTTCTAATTGGTTTTTAAAACATTAAGCCTTGATGCATAATGATAGACCTTAT 177
              **   ****************************.*************** *********

385I-PSY      TTACATTTAATTGAGTCATACCATTTTTGCATTTTCAATTTATCCAGGAGACCGAAGATG 240
154I-PSY      TTACATTTAATTGAGTCATGCCATTTTTGCATTTTCAATTTATCCNNGAGACCGAAGATG 240
207I-PSY      TTACATTTAATTGAGTCATGCCATTTTTGCATTTTCAATTTATCCTAGAGACCGAAGATG 240
275I-PSY      TTACATTTAATTGAGTCATGCCATTTTTGCATTTTCAATTTATCCTAGAGACCGAAGATG 240
567I-PSY      TTACATTTAATTGAGTCATGCCATTTTTGCATTTTCAATTTATCCTAGAGACCGAAGATG 237
202I-PSY      TTACATTTAATTGAGTCATGCCATTTTTGCATTTTCAATTTATCCTAGAGACCGAAGATG 237
              ******************.************************  *************

385I-PSY      TGATGAG 247
154I-PSY      TGATGAG 247
207I-PSY      TGATGAG 247
275I-PSY      TGATGAG 247
567I-PSY      TGATGAG 244
202I-PSY      TGATGAG 244
```

```
385I-IDCAX    TAAGCTGCATTTAACCCTTTTTTTTTGGTTGGGTCTTTTCCGCCATTCAGTTTGAAGTTC 60
154I-IDCAX    TAAGCTGCATTTAACCCTTTTTTTTTGGTTGGGTCTTTTCCGCCATTCAGTTTGAAGTTC 60
207I-IDCAX    TAAGCTGCATTTAACCCTTTTTTTTTGGTTGGGTCTTTTCCGCCATTCAGTTTGAAGTTC 60
275I-IDCAX    TAAGCTGCATTTAACCCTTTTTTTTTGGTTGGGTCTTTTCCGCCATTCAGTTTGAAGTTC 60
567I-IDCAX    TAAGCTGCATTTAACCCTTTTTTTTTGGTTGGGTCTTTTCCGCCATTCAGTTTGAAGTTC 60
202I-IDCAX    TAAGCTGCATTTAACCCTTTTTTTTTGGTTGGGTCTTTTCCGCCATTCAGTTTGAAGTTC 60
              ************************************************************

385I-IDCAX    CTCTGTTTTACTCAGTGTCTAATTTAGCTTTATTTTATCTTTNNNNNNNNNNNNNNNNNN 120
154I-IDCAX    CTCTGTTTTACTC--NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 118
207I-IDCAX    CTCTGTTTTACTCAGTGTCTAATTTAGCTTTATTTTATCTTTATCTTCTCANATATATGC 120
275I-IDCAX    CTCTGTTTTACTCAGTGTCTAATTTAGCTTTATTTTATCTTTATCTTCTCANATATATGC 120
567I-IDCAX    CTCTGTTTTACTCAGTGTCTAATTTAGCTTTATTTTATCTTTGTCTTCTCAGATATATGC 120
202I-IDCAX    CTCTGTTTTACTCAGTGTCTAATTTAGCTTTATTTTATCTTTGTCTTCTCAGATATATGC 120
              ************  ... .  ... . ... .... . ... . .. .  . . . ..

385I-IDCAX    NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
154I-IDCAX    NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178
207I-IDCAX    TGTTTCAATTACCATTCGTATTGTGGTAAGTGCACCAAATATATTCCAAAAATCTCC--- 177
275I-IDCAX    TGTTTCAATTACCATTCGTATTGTGGTAAGTGCACCAAATATATTCCAAAAATCTCC--- 177
567I-IDCAX    TGTTTCAATTACCATTCGTATTGTGGTAAGTGTACCAAATATATTCCAAAA-TCTCCGTA 179
202I-IDCAX    TGTTTCAATTACCATTCGTATTGTGGTAAGTGTACCAAATATATTCCAAAA-TCTCCGTA 179
              .....   ..    .. .. .......  ...       . . ..       . .

385I-IDCAX    NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 238
154I-IDCAX    NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 236
207I-IDCAX    --ATTGATAATCGAAGCCTGCCAATCTCAATCATTAATATTGACTATCTCCCAATTGC 233
275I-IDCAX    --ATTGATAATCGAAGCCTGCCAATCTCAATCATTAATATTGACTATCTCCCAATTGC 233
567I-IDCAX    CTATTGATAATCGAAGCCTGCCAATCTCAATCATTAATATTGACTATCTCCCAATTGC 237
202I-IDCAX    CTATTGATAATCGAAGCCTGCCAATCTCAATCATTAATATTGACTATCTCCCAATTGC 237
                 ... .  . .  .  ..    . .   .  ..  . ...  . . .     ...
```

```
385I-IDATGRC  GGCAATGAAAACAATGAGATAGAGATCTGCCTTGAGAGTGGTGAGTAGATCGTGGCCAAC 60
154I-IDATGRC  GGCAATGAAAACAATGAGATACAGATCTGCCTTGAGAGTGGTGAGTAGATCGTGGCCAAC 60
207I-IDATGRC  GGCAATGAAAACAATGAGATAGAGATCTGCCTTGAGAGTGGTGAGTAGATCGTGGCCAAC 60
275I-IDATGRC  GGCAATGAAAACAATGAGATAGAGATCTGCCTTGAGAGTGGTGAGTAGATCGTGGCCAAC 60
567I-IDATGRC  GGCAATGAAAACAATGAGATAGAGATCTGCCTTGAGAGTGGTGAGTAGATCGTGGCCAAC 60
202I-IDATGRC  GGCAATGAAAACAATGAGATAGAGATCTGCCTTGAGAGTGGTGAGTAGATCGTGGCCAAC 60
              ********************* ************************************
```

```
385I-IDATGRC  AAAAAATTATATGGTCTAAGCTACAT-----------------GCATCTGGCCTGCCATA 103
154I-IDATGRC  AAAAAATTATATGGTCTAAGCTACAA-----------------GCATCTGGCCTGCCATA 103
207I-IDATGRC  AAAAAATTATATGGTCTAAGCTACATGCATCTCTAAGCTACATGCATCTGGCCTGCCATA 120
275I-IDATGRC  AAAAAATTATATGGTCTAAGCTACATGCATCTCTAAGCTACATGCATCTGGCCTGCCATA 120
567I-IDATGRC  AAAAAATTATATGGTCTAAGCTACAT-----------------GCATCTGGCCTGCCATA 103
202I-IDATGRC  AAAAAATTATATGGTCTAAGCTACAT-----------------GCATCTGGCCTGCCATA 103
              *************************:                ****************

385I-IDATGRC  TCATTATCTAGTGCCTTATCTATGTTCATTCAATCGCATCCTGCCTGCGTTAATTGAGTC 163
154I-IDATGRC  TCATTATCTAGTGCCTTATCTATGTTCATTCAATCGCATCCTGCCTGCGTTAATTGAGTC 163
207I-IDATGRC  TCATTATCTAGTGCCTTATCTATGTTCATTCAATCGCATCCTGCCTGCGTTAATTGAGTC 180
275I-IDATGRC  TCATTATCTAGTGCCTTATCTATGTTCATTCAATCGCATCCTGCCTGCGTTAATTGAGTC 180
567I-IDATGRC  TCATTATCTAGTGCCTTATCTATGTTCATTCAATCGCATCCTGCCTGCGTTAATTGAGTC 163
202I-IDATGRC  TCATTATCTAGTGCCTTATCTATGTTCATTCAATCGCATCCTGCCTGCGTTAATTGAGTC 163
              ************************************************************

385I-IDATGRC  CCATTTCATCGGCCATTAATGAGGAGGACCAACAATCTTGAAA 206
154I-IDATGRC  CCATTTCATCGGCCATTAATGAGGAGGACCAACAATCTTGAAA 206
207I-IDATGRC  CCATTTCATCGGCCATTAATGAGGAGGACCAACAATCTTGAAA 223
275I-IDATGRC  CCATTTCATCGGCCATTAATGAGGAGGACCAACAATCTTGAAA 223
567I-IDATGRC  CCATTTCATCGGCCATTAATGAGGAGGACCAACAATCTTGAAA 206
202I-IDATGRC  CCATTTCATCGGCCATTAATGAGGAGGACCAACAATCTTGAAA 206
              *****************************************
```

```
385I-IDAPV    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTCGAACTTGTATCTGTTGAT 60
154I-IDAPV    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTCGAACTTGTATCTGTTGAT 60
207I-IDAVP    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTCGAACTTGTATCTGTTGAT 60
275I-IDAVP    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTCGAACTTGTATCTGTTGAT 60
567I-IDAVP    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTNGAACTNGTATCTGTTGAT 60
202I-IDAVP    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTNGAACTNGTATCTGTTGAT 60
560I-IDAVP    CAGCTATTGGAAAGGTTTGTAAAATTGTTTACACTTAAATTTGAACTCGTATCTGTTGAT 60
              ************************************ ***** ************

385I-IDAPV    ------GCATGTCTTGTATCAGCTGCTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 114
154I-IDAPV    ------GCATGTCTTGTATCAGCTGCTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 114
207I-IDAVP    ------GCATGTCTTGTATCAGCTACTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 114
275I-IDAVP    ------GCATGTCTTGTATCAGCTACTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 114
567I-IDAVP    ------GCATGTCTTGTATCAGCTACTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 114
202I-IDAVP    ------GCATGTCTTGTATCAGCTACTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 114
560I-IDAVP    GCTGATGCATGTCTTGTATCAGCTGCTTTTCCATATTGTTGCTTTGAGAAATATTAGATC 120
                    ******************.***********************************

385I-IDAPV    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 154
154I-IDAPV    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 154
207I-IDAVP    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 154
275I-IDAVP    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 154
567I-IDAVP    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 154
202I-IDAVP    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 154
560I-IDAVP    TTCATCCAAATAACTTGAGAGATGTTTTATGCCTGTCTCC 160
              *****************************************
```

Online Resource 4. Statistics data of InDel markers diversity

	All citrus accessions					*C. reticulata*			*C. maxima*			*C. medica*			3 basic taxa		
Marker name	***N***	H_o	H_e	F_W	X^2	***N***	H_o	H_e	***N***	H_o	H_e	***N***	H_o	H_e	F_{is}	F_{it}	F_{st}
IDCHI	4	0.10	0.23	0.572	6.87*	2	0.03	0.03	1	0	0	3	0.50	0.62	0.125	0.762	0.728
IDEMA	3	0.19	0.28	0.321	2.60	2	0.18	0.17	1	0	0	1	0	0	-0.062	0.762	0.776
IDTRPA	2	0.34	0.33	-0.049	0.07	2	0.34	0.29	2	0.30	0.27	1	0	0	-0.168	-0.149	0.015
IDPEPC1	2	0.19	0.35	0.466	6.92*	1	0	0	1	0	0	1	0	0	-	1	1
IDPEPC2	2	0.07	0.20	0.664	7.89*	1	0	0	1	0	0	1	0	0	-	1	1
IDLCY2	3	0.30	0.38	0.210	1.51	1	0.48	0.41	1	0	0	1	0	0	-0.175	0.541	0.609
IDHYB1	3	0.16	0.27	0.407	4.03*	2	0.17	0.16	2	0.30	0.27	1	0	0	-0.090	0.680	0.706
IDHYB2	2	0.07	0.25	0.732	11.98*	1	0	0	1	0	0	1	0	0	-	1	1
IDPSY	2	0.06	0.19	0.707	8.63*	1	0	0	1	0	0	1	0	0	-	1	1
IDCAX	5	0.56	0.69	0.198	2.45	3	0.62	0.51	1	0	0	1	0	0	-0.200	0.636	0.697
IDAtGRC	2	0.07	0.20	0.664	7.89*	1	0	0	2	0.30	0.27	1	0	0	-0.143	0.876	0.892
IDAVP	2	0.08	0.12	0.326	1.10	1	0	0	1	0	0	2	0.67	0.48	-0.474	0.655	0.766

(*N*) Allele number; (H_o) Observed heterozygosity; (H_e) Expected heterozygosity; (F_w) Wright fixation Index over the whole population; (F_{is}, F_{it} and F_{st}) Weir and Cockerham Index over the subset of *C. maxima, C.medica* and *C. reticulata* accessions. (-) Not possible to calculate. X^2 confirmed no significant differences between the expected and observed heterozygosity, F_w (α=0.05; X^2<3,84). * Means significant differences

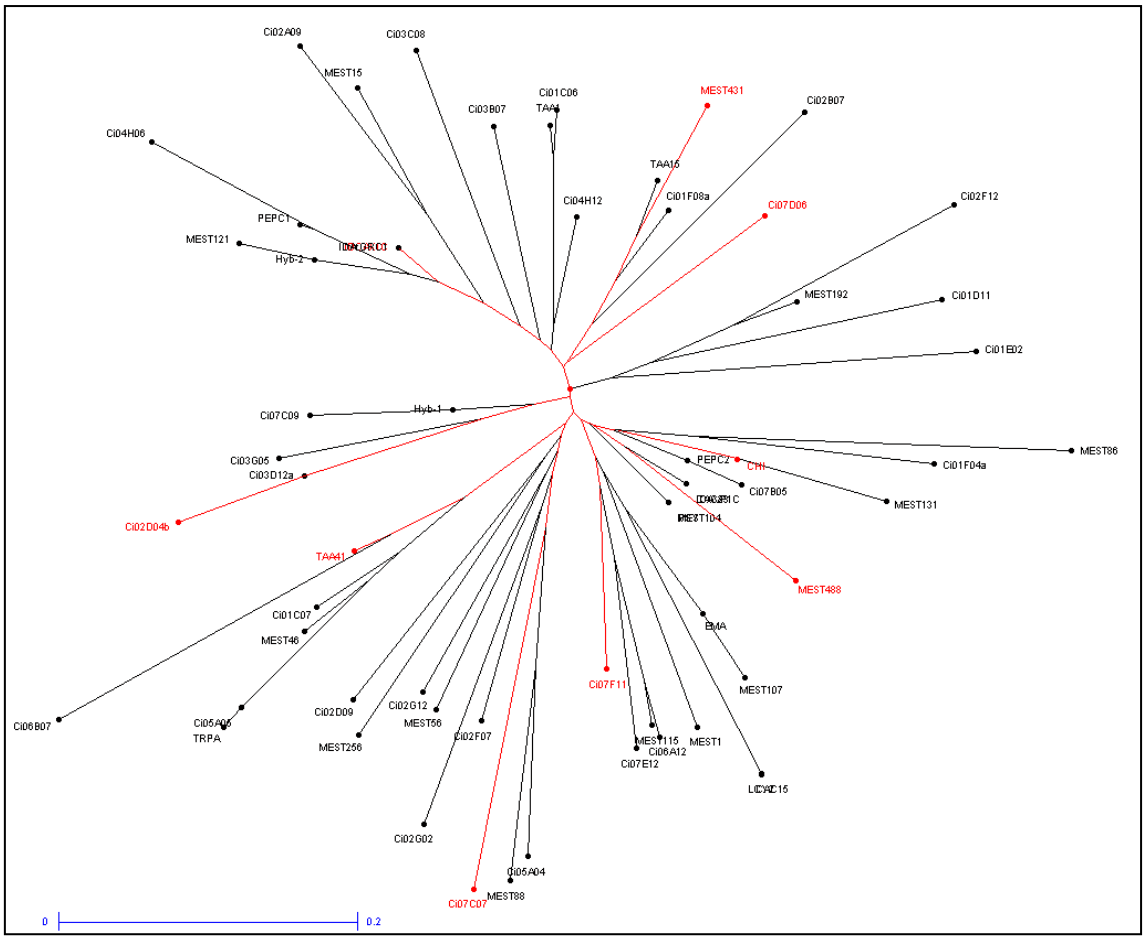
Online Resource 5. Statistics data of SSR markers diversity

		All citrus accessions				*C. reticulata*			*C. maxima*			*C. medica*			3 basic taxa		
Marker name	***LG***	***N***	H_o	H_e	F_W	***N***	H_o	H_e	***N***	H_o	H_e	***N***	H_o	H_e	F_{is}	F_{it}	F_{st}
MEST121	1	7	0.58	0.62	0.074	2	0.59	0.44	5	0.40	0.76	2	0	0.44	0.101	0.409	0.343
MEST1	1	10	0.60	0.70	0.140	4	0.62	0.62	5	0.60	0.71	2	0.50	0.38	0.033	0.345	0.322
MEST431	1	5	0.44	0.57	0.215	3	0.34	0.30	2	0.10	0.10	3	0.33	0.61	0.056	0.699	0.681
TAA15	1	9	0.63	0.77	0.179	9	0.69	0.74	1	0	0	1	0	0	0.083	0.547	0.505
mCrCIR02D09	2	11	0.55	0.74	0.252	7	0.86	0.73	5	0.40	0.67	2	0	0.32	0.037	0.296	0.269
TAA41	2	12	0.72	0.86	0.164	8	0.72	0.79	2	0.30	0.45	3	0.17	0.54	0.235	0.459	0.293
mCrCIR06B07	2	5	0.48	0.64	0.251	3	0.41	0.49	3	0.80	0.55	3	0.67	0.49	-0.011	0.389	0.396
mCrCI01C07	2	9	0.81	0.83	0.025	4	0.79	0.62	4	0.80	0.59	2	0.17	0.49	-0.175	0.286	0.392
mCrCIR05A05	2	13	0.57	0.80	0.295	9	0.52	0.76	6	1	0.78	5	0.33	0.53	0.209	0.351	0.18
mCrCIR04H06	2	6	0.44	0.68	0.347	4	0.48	0.52	1	0	0	2	0.33	0.28	0.135	0.655	0.601
MEST46	2	7	0.70	0.76	0.077	5	0.79	0.63	4	0.20	0.47	1	0	0	-0.075	0.425	0.466
CAC15	2	3	0.44	0.39	-0.148	2	0.31	0.26	2	0.80	0.50	1	0	0	-0.314	0.116	0.327
mCrCIR03C08	2	10	0.73	0.85	0.138	6	0.86	0.75	5	0.20	0.62	3	0.50	0.63	0.217	0.406	0.242
CAC23	3	3	0.62	0.56	-0.105	2	0.41	0.33	2	0.80	0.48	1	0	0	-0.373	0.351	0.527
MEST256	3	9	0.31	0.66	0.527	6	0.28	0.68	2	0.50	0.38	2	0.50	0.38	0.385	0.567	0.296
mCrCI02G12	3	10	0.67	0.80	0.166	7	0.89	0.77	4	0.10	0.69	2	0.60	0.42	0.096	0.317	0.245
MEST131	3	5	0.60	0.67	0.104	5	0.72	0.65	2	0.30	0.26	1	0	0	-0.100	0.378	0.434
mCrCI03D12a	4	8	0.81	0.82	0.011	5	0.86	0.74	3	0.60	0.45	1	0	0	-0.158	0.355	0.443
mCrCI02D04b	4	10	0.82	0.81	-0.018	5	0.79	0.73	4	0.70	0.53	5	0.67	0.72	-0.068	0.214	0.264
mCrCIR03G05	4	6	0.62	0.80	0.218	5	0.59	0.74	2	0.40	0.32	1	0	0	0.172	0.529	0.432
mCrCIR07D06	4	9	0.69	0.80	0.141	4	0.69	0.66	3	0.60	0.45	2	0.17	0.49	0.032	0.432	0.413
MEST15	5	5	0.72	0.75	0.040	5	0.79	0.62	1	0	0	1	0	0	-0.241	0.515	0.609
MEST104	5	7	0.71	0.75	0.057	5	0.76	0.60	3	0.90	0.54	1	0	0	-0.311	0.281	0.452
mCrCIR01F08a	5	6	0.37	0.50	0.273	2	0.03	0.03	3	0.80	0.61	1	0	0	-0.248	0.771	0.816
MEST88	5	7	0.33	0.72	0.540	6	0.69	0.74	2	0	0.42	1	0	0	0.243	0.441	0.261
mCrCI05A04	5	4	0.20	0.50	0.608	2	0.38	0.31	2	0	0.44	1	0	0	-0.053	0.644	0.662
mCrCIR07E12	5	8	0.61	0.66	0.072	6	0.31	0.38	3	0.50	0.51	1	0	0	0.134	0.558	0.49
MEST115	5	4	0.42	0.50	0.149	3	0.10	0.10	3	0.20	0.58	1	0	0	0.451	0.825	0.681
mCrCIR06A12	5	6	0.54	0.63	0.132	3	0.14	0.13	3	0.70	0.57	3	0.17	0.57	0.100	0.697	0.663
MEST56	5	14	0.64	0.82	0.213	8	0.62	0.72	5	0.80	0.63	3	0	0.67	0.136	0.358	0.257
mCrCI04H12	6	8	0.46	0.60	0.244	3	0.14	0.19	5	0.90	0.69	1	0	0	-0.020	0.683	0.689
MEST192	6	13	0.69	0.81	0.151	7	0.83	0.77	5	0.89	0.67	1	0	0	-0.089	0.310	0.367
mCrCIR02F12	6	8	0.70	0.75	0.070	5	0.55	0.57	4	1	0.74	3	0.67	0.5	-0.061	0.246	0.29
Ci01D11	6	8	0.33	0.74	0.555	5	0.52	0.61	1	0	0	3	0	0.63	0.306	0.663	0.515
MEST488	6	11	0.84	0.82	-0.035	8	0.90	0.79	5	0.90	0.68	1	0	0	-0.137	0.198	0.295
mCrCIR01E02	6	7	0.57	0.72	0.201	4	0.55	0.62	4	0.78	0.67	1	0	0	0.081	0.377	0.323
mCrCIR01C06	6	12	0.82	0.82	-0.007	5	0.69	0.62	6	0.90	0.75	3	0.33	0.57	-0.045	0.253	0.285
TAA1	6	6	0.50	0.59	0.150	2	0.24	0.21	2	0.50	0.38	2	0.17	0.15	-0.174	0.674	0.722
MEST107	7	3	0.32	0.43	0.257	3	0.07	0.07	1	0	0	1	0	0	-0.007	0.905	0.905
mCrCIR03B07	7	9	0.51	0.69	0.262	4	0.38	0.33	2	0.20	0.42	3	0.17	0.40	0.141	0.667	0.612
Ci07C07	7	7	0.59	0.76	0.222	7	0.69	0.80	3	0.75	0.53	3	0.67	0.50	0.105	0.178	0.081
mCrCIR02A09	8	7	0.66	0.75	0.124	6	0.61	0.59	1	0	0	2	0	0.28	0.104	0.536	0.482
mCrCIR02G02	8	13	0.73	0.78	0.066	7	0.79	0.73	6	0.60	0.67	4	0.67	0.58	-0.016	0.193	0.206
Ci02F07	8	10	0.31	0.58	0.469	4	0.14	0.16	4	0.38	0.73	1	0	0	0.366	0.827	0.726

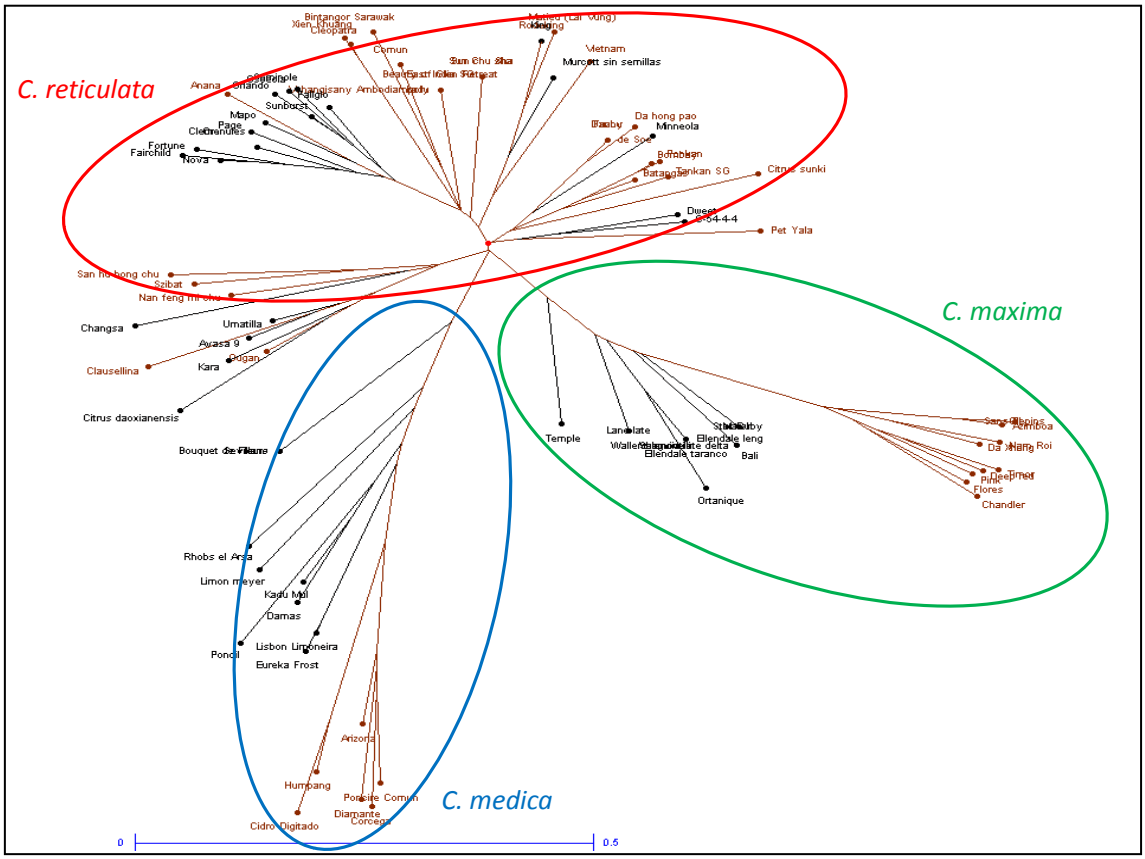
mCrCIR07B05	8	10	0.49	0.78	0.370	6	0.41	0.75	5	0.70	0.65	2	0.17	0.15	0.336	0.514	0.268
mCrCIR01F04a	8	10	0.80	0.84	0.055	7	0.86	0.80	5	0.89	0.72	1	0	0	-0.073	0.252	0.302
MEST86	8	6	0.53	0.73	0.274	3	0.72	0.63	5	0.30	0.67	1	0	0	0.051	0.410	0.379
Ci07C09	9	8	0.69	0.70	0.016	6	0.45	0.43	4	0.70	0.57	2	0.17	0.15	-0.076	0.508	0.542
mCrCIR07F11	9	12	0.77	0.83	0.079	6	0.66	0.59	3	0.80	0.65	2	0.50	0.49	-0.104	0.325	0.389
Ci02B07	9	9	0.71	0.78	0.083	8	0.72	0.72	5	0.30	0.42	2	0.33	0.28	0.061	0.377	0.337

(*LG*) Linkage group; (*N*) Allele number. (H_o) Observed heterozygosity; (H_e) Expected heterozygosity; (F_w) Wright fixation Index over the whole population; (F_{is}, F_{it} and F_{st}) Weir and Cockerham Index over the subset of *C. maxima*, *C.medica* and *C. reticulata* accessions

Online Resource 6. LD in the whole data set (InDel and SSR markers). Subset of primers selected in red



Online Resource 7. NJ tree under topological constraints. Structure of the three ancestor groups was used as initial tree and the rest of the population was positioned (constraint) on this previous topology. Samples in brown were used as initial structuration



CHAPTER 2

A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the ‘true citrus fruit trees’ group (Citrinae, Rutaceae) and the origin of cultivated species.

Andres Garcia-Lor, Franck Curk, Hager Snoussi-Trifa, Raphael Morillon, Gema Ancillo, François Luro, Luis Navarro and Patrick Ollitrault

Annals of Botany (2013) 111: 1-19

Abstract

- *Background and Aims*

Despite differences in morphology, the genera representing 'true citrus fruit trees' are sexually compatible, but their phylogenetic relationships remain unclear. Most of the important commercial 'species' of *Citrus* are believed to be of interspecific origin. By studying polymorphisms of 27 nuclear genes, the average molecular differentiation between species was estimated, and some phylogenetic relationships between 'true citrus fruit trees' were clarified.

- *Methods*

Sanger sequencing of PCR-amplified fragments from 18 genes involved in metabolite biosynthesis pathways and nine putative genes for salt tolerance, was performed for 45 genotypes of *Citrus* and relatives of *Citrus* to mine SNP and indel polymorphisms. Fifty nuclear SSRs were also analysed.

- *Key results*

A total of 16238 kb of DNA was sequenced for each genotype, and 1097 single nucleotide polymorphisms (SNPs) and 50 indels were identified. These polymorphisms were more valuable than SSRs for inter-taxon differentiation. Nuclear phylogenetic analysis revealed that *Citrus reticulata* and *Fortunella* are joined in a cluster that is differentiated from the clade that includes three other basic taxa of cultivated citrus (*C. maxima*, *C. medica* and *C. micrantha*). These results confirm the taxonomic subdivision between the subgenera Metacitrus and Archicitrus. A few genes displayed positive selection patterns within or between species, but most of them displayed neutral patterns. The phylogenetic inheritance patterns of the analysed genes were inferred for commercial *Citrus* species.

- *Conclusions*

Numerous molecular polymorphisms (SNPs and indels), which are potentially very useful for the analysis of interspecific genetic structures, have been identified. The nuclear phylogenetic network for *Citrus* and its sexually compatible relatives was consistent with the geographic origins of these genera. The positive selection observed for a few genes will orient further works to analyse the molecular basis of the variability of the associated traits. This study presents new insights into the origin of *C. sinensis*.

INTRODUCTION

Aurantioideae (Rutaceae) are considered to be a monophyletic group (Scott *et al*., 2000; Groppo *et al*., 2008; Morton, 2009) and *Ruta* appears to be sister to Aurantioideae (Scott *et al*., 2000; Bayer *et al.*, 2009). Furthermore, Groppo *et al*. (2008) suggest that Aurantioideae should be recognized as a tribe and be included in a subfamily together with Rutoideae, Toddalioideae and Flindersioideae. Although new insights in the circumscription of the tribes of Aurantioideae have been recently released (Bayer *et al*., 2009; Morton, 2009), it remains unresolved. In the classification of Swingle and Reece (1967), which remains the most used by citrus researchers, Aurantioidae is divided in two tribes, Clauseneae and Citreae. Citreae includes, among others, subtribe Citrinae, which is in turn divided into six genera (*Fortunella*, *Eremocitrus*, *Poncirus*, *Clymenia*, *Microcitrus* and *Citrus*) that comprise the important 'true citrus fruit trees' group (Swingle and Reece, 1967; Krueger and Navarro, 2007).

Among these genera, *Citrus* is by far the most economically important. It is believed to have originated in south-eastern Asia, in an area that includes China, India and the Indochinese peninsula and nearby archipelagos (Krueger and Navarro, 2007). *Citrus* taxonomy is still controversial due to the large degree of morphological diversity found within this group, the sexual compatibility between the species and the apomixis of many genotypes. Two major classification systems based on morphological and phenotypic data are currently used, i.e., those of Swingle and Reece (1967) and Tanaka (1977), who recognized 16 and 162 species, respectively. Here we adopt the classification system of Swingle and Reece (1967), which is more in line with the main clustering system derived from the molecular analysis described in this report. More recently, Mabberley (1997) proposed a new classification system for edible citrus that recognizes three species and four hybrid groups. In agreement with a pioneering numerical taxonomic study (Barret and Rhodes, 1976), the classification system of Mabberley confirms that three main taxa [*C. medica* L. (citrons), *C. maxima* (Burm.) Merr. (pummelos) and *C. reticulata* Blanco (mandarins)] were the ancestors of cultivated *Citru*s. However, the subdivision into four hybrid groups remains questionable, and relatively few authors have adopted the classification system of Mabberley. More recent studies involving the diversity of morphological characteristics (Ollitrault *et al*., 2003) and the analysis of primary metabolites (Luro *et al*., 2011) and secondary metabolites (Fanciullino *et al*., 2006a), have indicated that the phenotypic diversity of edible citrus species primarily resulted from the initial differentiation between these three basic taxa. Molecular marker studies using restriction fragment length polymorphism (RFLP; Federici *et al*., 1998), random amplified polymorphic DNA (RAPD) and sequence characterized amplified regions (SCAR; Nicolosi *et al*., 2000), simple sequence repeats (SSRs; Luro *et al*., 2001; Barkley *et al*., 2006), SSR and insertion-deletion (indels; Garcia-Lor *et al*., 2012a) and single nucleotide polymorphisms (SNPs; Ollitrault *et al*., 2012a) have confirmed the central role played by these three taxa but also pointed out that *C. micrantha*, a member of subgenus *Papeda*, is a potential parent of some limes (*C. aurantifolia* Christm.). Swingle and Reece (1967) differentiated between the subgenera *Papeda* and *Citrus*. The genome of most of the important commercial *Citrus* spp. (secondary species) can be

considered to be a mosaic of large DNA fragments of the ancestral species that resulted from a few inter-specific recombination events (Garcia-Lor *et al*., 2012a).

Fortunella is generally considered to be a separate genus (Swingle and Reece, 1967), but it closely resembles *Citrus*. According to Swingle (1943), this genus includes four species (*F. margarita*, *F. japonica*, *F. hindsii* and *F. polyandra*) and Fantz (1988) included two hybrid taxa in *Fortunella* (*F. obovata*, *F. Crassifolia*). Its origin is northern China, and it is well adapted to cold areas due to its propensity for prolonged winter dormancy and late flowering. Its fruits, commonly called kumquats, are edible, and *Fortunella* trees are appreciated for their ornamental qualities. *Poncirus* is another genus that originated in northern China, and for a long time, it was considered to be monotypic (*P. trifoliata*). However, a new species belonging to this genus, *P. polyandra*, was found in Yunnan (China) in the 1980s (Ding *et al*., 1984). *Poncirus trifoliata* is the only species of 'true citrus fruit trees' with deciduous leaves. It is highly tolerant to cold and resistant to several citrus pathogens. It is therefore an important source of germplasm for citrus rootstock breeding.

According to Krueger and Navarro (2007), *Microcitrus* includes five species that originated in Australia (*M. australis*, *M. australasica*, *M. inodora*, *M. garrowayii* and *M. maindeniana*) and two from Papua New Guinea (*M. papuana* and *M. warbugiana*). *Microcitrus australasica*, the finger lime, is cultivated on a small scale for its fruit with aromatic, spherical juice vesicles. *Eremocitrus* is a monospecific genus (*E. glauca*) that is native to the Australian desert. This genus is cold-tolerant and xerophytic. *Eremocitrus* and *Microcitrus* are closely related, morphologically and molecularly (Swingle and Reece, 1967; Bayer et al., 2009). They are graft-compatible with *Citrus* and other related genera.

Despite considerable morphological differentiation, *Citrus*, *Fortunella*, *Poncirus*, *Microcitrus* and *Eremocitrus* are sexually compatible genera (Krueger and Navarro, 2007). Studies based on plastid sequences (Abkenar *et al*., 2004; Morton, 2009; Bayer *et al*., 2009) concur that the six genera of the tribe Citrinae (*Fortunella*, *Eremocitrus*, *Poncirus*, *Clymenia*, *Microcitrus* and *Citrus*) form a monophyletic clade. However, these authors did not fully agree on the organization within this clade. *Clymenia polyandra* appeared in the same subclade as *Citrus*, *Fortunella* and *Poncirus* in the analysis of Abkenar *et al*. (2004), but Morton (2009) and Bayer *et al*. (2009) found it in the *Eremocitrus* and *Microcitrus* subclade. Moreover, Bayer *et al*. (2009) included *Oxanthera* and *Feroniella* in the clade of the 'true citrus fruit trees'. To analyse the gene pool of 'true citrus fruit trees', with potential gene flow between sexually compatible taxa, some of which share the same diversification area, phylogenetic analysis based on nuclear sequences should be more informative than the analysis of maternally inherited plastid sequences (Ramadugu *et al*., 2011). However, the phylogenetic relationships between 'true citrus fruit trees' based on the analysis of nuclear genomes have not been clearly elucidated.

In genetic studies of *Citrus*, SSR analysis (Gulsen and Roose, 2001a; Luro *et al*., 2001, 2008; Barkley *et al*., 2006; Ollitrault *et al*., 2010) is seen as a powerful tool because SSRs are co-dominant, randomly dispersed throughout the plant genome, generally highly polymorphic

and locus-specific. However, Barkley *et al*. (2009) showed that homoplasy might limit the usefulness of SSRs as tags to elucidate the phylogenetic origin of specific DNA fragments in citrus. Moreover, the high mutation rate of SSRs can often lead to an underestimation of subpopulation divergence (Coates *et al*., 2009). In recent studies, Garcia-Lor *et al*. (2012a) and Ollitrault *et al*. (2012) analysed the value of nuclear indels as genetic markers in *Citrus*. These studies showed that indels are more suitable than SSRs for differentiating between the three basic taxa of cultivated *Citrus*. However, the relatively low frequency of indels limits their utility.

SNPs are the most abundant type of DNA sequence polymorphism (Brookes, 1999). Due to the high frequency of occurrence of SNPs and their relatively dense and uniform distribution in genomes, SNPs are an important source of variability and are therefore useful for many applications, including the development of saturated genetic maps, cultivar identification, detection of genotype/phenotype associations and marker-assisted breeding (Botstein and Risch, 2003; Morales *et al*., 2004; Xing *et al*., 2005; Lijavetzky *et al*., 2007). The frequency of occurrence of SNPs in the genomes of eukaryotes depends on the domestication and breeding history, mating system and frequency of mutation, recombination and other features (Buckler and Thornsberry, 2002; Rafalski and Morgante, 2004). Although individual SNPs are less informative than other marker types for population genetic studies because of their biallelic nature, they have several advantages over other marker types due to the high frequency of SNP occurrence, the easy automation of SNP genotyping, the low-scoring error rates and the high levels of reproducibility of SNP analysis results between laboratories (Morales *et al*., 2004; Helyar *et al*., 2011).

Many efforts have been made to detect SNPs in plants. SNPs have been used to perform comparative diversity analysis and genotyping, to reveal genetic structures and to assess molecular evolutionary patterns in many plant species including Norway spruce (Heuertz *et al*., 2006), sunflower (Kolkman *et al*., 2007), grapevine (Lijavetzky *et al*., 2007), European aspen (Ingvarsson, 2005) and eucalyptus (Külheim *et al*., 2009). Some studies have been performed in *Citrus*, but these studies were generally limited due to narrow genetic basis of the discovery panel. Novelli *et al*. (2004) searched for SNPs among several sweet orange lines. Terol *et al.* (2008) identified 6617 putative SNPs from Nules clementine BAC end sequences, from which 622 were successfully transferred to the entire genus using GoldenGate array technology (Ollitrault *et al*., 2012a). Dong *et al*. (2010) mined SNPs from sweet orange and satsuma mandarin expressed sequence tag (EST) databases.

The ascertainment bias associated with a low genetic basis of the discovery panel has been widely discussed for humans and animals (Clark *et al*., 2005; Rosenblum and Novembre, 2007; Albrechtsen *et al*., 2010), and it was observed for *Citrus* at the genus level when diversity studies with SNP markers mined in clementine were performed (Ollitrault *et al*., 2012a). It is therefore important to develop a good sampling strategy for SNP discovery (Garvin *et al*., 2010; Helyar *et al*., 2011) that would help to elucidate the true differentiation level between basic taxa and related genera at the nuclear level.

In the present study, we searched for SNP and indel polymorphisms in 45 accessions of *Citrus*, *Poncirus*, *Fortunella*, *Microcitrus* and *Eremocitrus*, with *Severinia buxifolia* employed as outgroup, using Sanger sequencing of amplified DNA fragments from 18 genes involved in primary and secondary metabolite biosynthetic pathways that determine citrus fruit quality (sugars, acids, flavonoids and carotenoids) and nine putative salt tolerance genes. In addition to the identification of useful intra- and interspecific SNP and indel markers, this study addresses the following questions: (1) What are the phylogenetic relationships at the nuclear level between different *Citrus* spp. and between genera? (2) What is the level of intra- and interspecific diversity between the *Citrus* taxa at the origin of the cultivated forms? (3) Did the evolution of genes involved in different metabolic pathways and some putative stress adaptation genes follow a similar neutral pattern regarding the history and reproductive biology of *Citrus*, or did some genes experience selective evolution? (4) What is the phylogenetic inheritance pattern of the analysed genes in secondary *Citrus* spp.?

MATERIALS AND METHODS

Plant material

Leaf material from 44 true citrus accessions and one relative (*Severinia buxifolia*) used as the outgroup [Supplementary Information 1] was collected, and DNA was extracted using the DNeasy plant mini kit (Qiagen S.A., Madrid, Spain). The samples represented all major *Citrus* species (seven *C. reticulata*, five *C. maxima*, five *C. medica*, and four representatives of the subgenus *Papeda*) and five *Fortunella* spp., two *Microcitrus* spp., one *Eremocitrus* sample and three *Poncirus trifoliata*, all of which are sexually compatible with *Citrus*. These eight groups are considered to be ancestral populations. Some representatives of secondary species were added (two diploid clementines and one haploid clementine, two *C. sinensis*, two *C. aurantium*, one *C. paradisi*, one *C. limon* and one *C. aurantifolia*) and two hybrids, including one tangor (*C. reticulata* x *C. sinensis*) and one tangelo (*C. paradisi* x *C. reticulata*). These 12 genotypes are known to be hybrids derived from the ancestral populations and are economically important cultivars. Haploid clementine (Aleza *et al*., 2009) is currently used by the International Citrus Genome Consortium to establish the whole genome reference sequence of citrus. It was used in the present study to test whether some genes were duplicated. Forty-two accessions were obtained from the IVIA Citrus Germplasm Bank of pathogen-free plants (Navarro *et al*., 2002), and three were obtained from the INRA/CIRAD collection. All accessions were used for Sanger sequencing of gene fragments and indel and SSR genotyping.

Gene fragment sequencing

Eighteen genes involved in primary and secondary metabolite biosynthesis pathways that determine citrus fruit quality (sugars, acids, flavonoids and carotenoids) and nine putative salt tolerance genes were selected. The selection of the 27 gene fragments was based on the quality of sequencing chromatograms for all genotypes. Primers were designed using Primer3 (Rozen and Skaletsky, 2000) to amplify fragments ranging from 190 to 941 bp, according to the ESTs available in GenBank (http://www.ncbi.nlm.nih.gov/genbank/; Table 1). PCR amplifications were performed using a Mastercycler Ep Gradient S thermocycler (Eppendorf) in a final volume of 25 µL containing 0.027 U/µl *Taq* DNA polymerase (Fermentas), 1 ng/µL of genomic DNA, 10x PCR buffer (Fermentas), 0.2 mM of each dNTP, 1.5 mM MgSO4 and 0.2 µM of each primer. The following PCR program was applied: denaturation at 94°C for 5 min and 40 cycles of 30 s at 94°C, 1 min at 55°C or 60°C (according to the melting temperature of the primers), 2 min at 72°C, and a final elongation step of 4 min at 72°C. PCR product purification was done using a QIAquick PCR Purification Kit (Qiagen S.A.). Amplicons of the 45 genotypes were sequenced using the Sanger method from the 5' end using fluorescently labelled dideoxynucleotides (Big Dye Terminator Cycle Sequencing Kit v3.1).

Table 1. Primer sequences for the genes that were studied

Biosyntetic pathway	Gene	Primers	AT	GBA
Flavonoids	Chalcone isomerase (CHI)	F:TTGTTCTGATGGCCTAATGG	58	DY263683
		R:AAAGGCTGTCACCGATGAAT	58	
	Chalcone synthase (CHS)	F:GATGTTGGCCGAGTAATGCT	60	CV885475
		R:ATGCCAGGTCCAAAAGCTAA	59	
	Flavonol synthase (FLS)	F:GGAGGTGGAGAGGGTCCAAG	59	AB011796
		R:GGGCCACCACTCCAAGAGC	61	
	Flavonoid 3'-hydroxylase (F3'H)	F:CTCGAGCCTTCCTCAAAACC	60	HQ634392
		R:AACAAGCACAATCCCCATTC	57	
	Dihydroflavonol 4-reductase (DFR)	F:CTGGGTTTATCGGTTCATGG	60	DQ084722
		R:TCCACAGCACCTGTGAACAT	60	
Acids	Malic enzyme (EMA)	F:ACATGACGACATGCTTCTGG	58	CB417399
		R:CGTAGCCACGCCTAGTTCAT	60	
	Malate dehydrogenase (MDH)	F:ATGGCCGCTACATCAGCTAC	60	DQ901430
		R:TGCAACCCCCTTTTCAATAC	59	
	Aconitase (ACO)	F:AAGCCATGGGTCAAAACAAG	59	AF073507
		R:GATTTCCCAGTGTCGGTTGT	59	
	Vacuolar citrate/H+ symporter (TRPA)	F:GGCGCCACTCCTACCTTCCC	62	EF028327
		R:CGGTCATTGAAGAGTGCTCCCC	60	
Sugars	Acid invertase (INVA)	F:ATTGCGGATGTGAAGAAAGG	56	AB074885
		R:TTTGCCATGCTTTGAGTGAG	56	
	Phosphoenolpyruvate carboxylase (PEPC)	F:AGCCAATGGGATTTCTGACA	60	EF058158
		R:GCCAAGCCACACAGGTAAAT	60	
	Phosphofructokinase (PKF)	F:CGCCGACCTCAGTCCCGTC	63	AF095520
		R:GCTGCACGCCCCATAAGCCG	64	
Carotenes	1-deoxyxylulose 5-phosphate synthase (DXS)	F:GGCGAGGAAGCGACGAAGATGG	62	DN959423
		R:GGATCAGAACTGGCCCTGGCG	62	
	Phytoene synthase (PSY)	F:GCTCGTTGATGGGCCTAATGC	59	AB037975
		R:CGGGCGTAAGAGGGATTTTGC	59	
	β-Carotene hydroxylase (HYB)	F:AGCCCTTCTGTCTCCTCACA	59	AF315289
		R:CCGTGGAATTTATCCGAGTG	59	
	Lycopene β-cyclase 2 (LCY2)	F:GCATGGCAACTCTTCTTAGCCCG	60	FJ516403
		R:AGCTCGCAAGTAAGGCCATTCC	61	
	Lycopene β-cyclase (LCYB)	F:GAATTCTTGCCCCAAGTTCA	60	AY16696
		R:TATGGGCCACAAATCTTTCC	59	
	9-cis-epoxy hydroxy carotenoid dyoxygenase 3 (NCED3)	F:GCAGTCAAATTCAACAAAGG	55	DQ309332
		R:AATCCCAAATCTTGACACCT	55	
Aldarate and ascorbate metabolism	Ascorbate oxydase (AOC)	F:TCAGTGAGAACCCTAAAGC	58	DY293375
		R:CAGTACAACCCCAGTAAGC	60	
	Ascorbate peroxidase (LAPX)	F:CAGCGGGGACTTATGACG	58	EU719653
		R:GCCCTCCGGTAACTTCAAC	59	
Cellular Detoxification	MRP-like ABC transporter (MRP4)	F:AGAAGCAGCATGGAAGATGG	60	CD574223
		R:CCGATCGGTTGGCATACTC	62	
	Cation chloride cotransporter (CCC1)	F:GCAGCTTGCTACCTACATTGAC	63	FN662480
		R:ACTGAACTCCACATCCCAAAAG	61	
	High-affinity K+ transporter 1 (HKT1)	F:GTCCATGGAGAAAAAGAACC	58	DY297409
		R:TGCTAGTGTCCGTGAAGAAG	60	
	NADH kinase (NADK2)	F:TGCAGAGACAAGATATTCCC	58	DN619491
		R:ATGTGAGGTGAGAAATCCC	58	
Salt tolerance	Aquaporin PIP1A (PIP1)	F:GACACTCGGCCTGTTCTTG	62	CK938271
		R:TCCGGTAATTGGGATGGTAG	60	
	Salt overly sensitive 1 (SOS1)	F:ACCAGTCAGACAACCATTTG	55	DN959478
		R:CCAATTAGCACCTCATAGAGAC	58	
Sucrose and starch metabolism	Tréhalose-6-phosphate synthase (TSC)	F:TGCAGAACCTGTAATGAAGC	58	FC875388
		R:CTGGTAGGATGCCGACTTAG	62	

(AT) Annealing temperature ºC; (GBA) GenBank accession number.

Sequence polymorphism analysis

Sequences were aligned using BioEdit (Hall, 1999), SeqMan version 7.0. (http://www.dnastar.com) and SATé-II (Liu *et al*., 2012). The homogeneity of the alignment obtained with the three software programs was checked and heterozygosity or homozygosity of

all genotypes verified visually in the chromatogram for all SNP positions. Estimates of nucleotide polymorphisms (segregating sites, *S*, nucleotide diversity, π) and between-species divergences were obtained using DnaSP v. 5.10.01 (http://www.ub.es/dnasp). The genomic DNA sequences were subjected to blast analysis using the protein databases (blastx) at NCBI to identify the coding and non-coding regions. DnaSP was also used to calculate the statistical test of neutrality, Tajima's *D* (Tajima, 1989a, b).

For each target gene fragment, the haplotype number and the haplotype diversity were estimated with DnaSP software using coalescent process simulations. Unbiased expected heterozygosity, observed heterozygosity, fixation index value (F_w; Wright, 1978) and the F_{stat} parameter (F_{st}) were calculated using GENETIX v. 4.03 software (Belkhir *et al*., 2002).

Indel marker development

Primer pairs for 12 indel markers are already available for the 'true citrus fruit trees' group (Garcia-Lor *et al*., 2012a). New primer pairs for genes with indel polymorphisms were designed with Primer3 in conserved regions flanking the indel polymorphism (http://biotools.umassmed.edu/bioapps/primer3; Supplementary Information 2) to amplify fragments smaller than 400 bp that were subsequently subjected to fragment size polymorphism analysis in a capillary fragment analyser as described by Garcia-Lor *et al*. (2012a).

SSR markers

The 50 SSRs markers used for the diversity analysis within *Citrus* by Garcia-Lor *et al*. (2012a) were used to complete the genotyping for the accessions of the other genera. The list of primers that were used, the PCR conditions that were employed and the method used for capillary electrophoresis can be found in Garcia-Lor *et al*. (2012a).

Phylogenetic analysis

Phylogenetic analysis was performed by joining the 27 sequences together for each genotype (eight taxa), creating a sequence with a total length of 16238 bp. Indels were excluded from the analysis. Several analyses were performed to determine which model best matched our data using the Phylemon 2.0 website (http://phylemon.bioinfo.cipf.es; Sánchez *et al*., 2011), which integrates different tools for molecular evolution, phylogenetics, phylogenomics and hypothesis testing. PhyML Best AIC Tree (v. 1.02b) software, which uses a model test program (Posada and Crandall, 1998) that performs hierarchical Likelihood Ratio Tests (LRTs) in an ordered way using Akaike's information criteria (AIC), was used to select the model that most closely fitted the data (lowest AIC value), taking into account the nucleotide substitution

model, the proportion of invariable sites (*I*), the nucleotide frequency (*F*) and the gamma distribution (*G*).

The construction of the maximum-likelihood (ML) tree was performed using 1000 bootstraps to assess the branch support using the SH-like approximate likelihood ratio test (ranges from 0 to 1), assuming uniform rates among sites and deleting gaps and missing sites. Trees obtained in Phylemon (newick format) were drawn using the TreeDyn 198.3 tool found at www.phylogeny.fr (Dereeper *et al*., 2008).

Neighbour-Joining (NJ) analysis

Population diversity organisation based on the SNP data was analysed with DARwin software (Perrier and Jacquemoud-Collet, 2006) as explained by Garcia-Lor *et al*. (2012a).

Principal Coordinate Analysis (PCoA)

PCoA was performed using GENEALEX6 software (Peakall and Smouse, 2006). The genomic sequence data were used to obtain a pairwise genetic distance matrix, which was standardised and used for PCoA.

RESULTS

SNP and indel discovery and analysis of polymorphic loci diversity

SNPs

SNPs were not encountered in any of the 27 genetic sequences of the haploid clementine. This confirms that there were no duplicated genes in our sample of 27 genes. A total of 16238 bp were sequenced for each of the accessions analysed, from which 10427 bp were coding regions and 5811 bp were non-coding sequences (Table 2). A total of 1097 SNPs were found in the 'true citrus fruit trees' samples. Another 262 SNPs were found in the outgroup, *Severinia buxifolia*. 'true citrus fruit trees' had an average of 52.89 SNPs/kb for coding regions and 98.39 SNPs/kb for non-coding regions. Considering only *Citrus*, 28.96 SNPs/kb were found in coding regions and 51.45 SNPs/kb were found in non-coding regions. In the 'true citrus fruit trees', most of the SNP loci were biallelic, while 21 (1.86 %) revealed three alleles. Among the polymorphisms described, 59.18 % were transitions (A/G ≈ C/T) and 40.82 % were transversions (A/C ≈ A/T > G/T > C/G). For the 'true citrus fruit trees', but excluding secondary *Citrus* spp., the average polymorphism rate was 51.76 SNPs/kb for coding regions and 95.43 SNPs/kb for non-coding regions, with a total of 1066 SNP loci. Among the basic *Citrus* taxa, *Papeda* had 252 polymorphic loci (12.18 SNP/kb in coding region and 21.51 SNP/kb in non-coding region), followed by *C. reticulata* (236, 15.15 SNP/kb in coding region and 13.94 SNP/kb in non-coding region), *C. maxima* (107, 4.70 SNP/kb in coding region and 9.98 SNP/kb in non-coding region) and *C. medica* (70, 2.21 SNP/kb in coding region and 8.09 SNP/kb in non-coding region). Large differences in the number of polymorphic loci were observed among close relatives including *Fortunella* (227), *Microcitrus* (171), *Eremocitrus* (93) and *Poncirus* (53). Among the secondary species and hybrids, *C. aurantium* had 211 polymorphic sites, *C. limon* had 173, *C. sinensis* had 162, *C. aurantifolia* had 158, *C. paradisi* had 115 and clementine had 119. Interestingly, among the 31 alleles found exclusively in the secondary species (not present in any other true citrus species), 15 were heterozygous in *C. aurantium*. Four of these alleles (found in the genes INVA, LCY2, DXS and AOC) were shared with *C. limon*.

The average rate of heterozygosity observed in the eight ancestral taxa was very low (H_o = 0.051), and 27.79% of the SNPs detected were homozygous in all individuals (H_o = 0). The most heterozygous site was at locus F3'H (SNP51), with a H_o = 0.39. We estimated the average rates of inter-accession polymorphism (SNPs/kb) within and between the ancestral taxa (Table 3). Considering only *Citrus* spp., the average rates of intra- and inter-taxon polymorphisms were 1.76 SNPs/kb and 11.31 SNPs/kb, respectively. Intra-taxon SNP rates varied from 0.65 for *C. maxima* to 3.37 for *Papeda* (*C. hystrix*, *C. inchangensis*, *C. micrantha*). Interspecific rates in *Citrus* varied from 8.56 between *C. reticulata* and *Papeda* to 14.43 between *C. medica* and *Papeda*. The SNP rate between *C. reticulata and C. maxima*, the two

Table 2. Polymorphisms of nucleotide sequences of genes for all samples analysed

Gene	CS	TS	GS	SC	SNC	SNPc	Freq.	SNPnc	Freq.	π_{nonsyn}/π_{syn}	indelc	Freq.	indelnc	Freq.
CHI	652	721	721	206	446	11	53.40	68	152.47	1.38	0	0	8	17.94
CHS	565	659	659	574	0	20	35.40	-	-	0.06	0	0	-	-
FLS	473	763	763	419	54	41	97.85	6	111.11	0.12	0	0	3	55.56
F3'H	783	1000	1400	569	214	40	70.30	20	93.46	0.55	0	0	3	14.02
DFR	421	1017	1650	171	250	7	40.94	26	104.00	0.25	0	0	3	12.00
EMA	428	166	450	131	297	7	53.44	27	90.91	2.27	1	7.63	4	13.47
MDH	712	1209	1250	712	0	28	39.33	-	-	1.06	0	0	-	-
ACO	695	1196	2000	250	445	5	20.00	39	87.64	0.02	0	0	2	4.49
TRPA	795	987	1300	657	138	40	60.88	15	108.70	0.43	0	0	1	7.25
INVA	908	679	1100	515	393	36	69.90	38	96.69	0.23	0	0	1	2.54
PEPC	694	1201	2000	61	633	2	32.79	51	80.57	0.00	0	0	4	6.32
PKF	775	807	1650	406	369	16	39.41	31	84.01	0.88	0	0	3	8.13
DXS	722	935	1500	327	395	13	39.76	37	93.67	0.29	0	0	3	7.59
PSY	606	727	2100	97	509	5	51.55	40	78.59	0.39	0	0	2	3.93
HYB	680	787	1600	379	301	19	50.13	27	89.70	0.91	1	2.638	2	6.64
LCY2	738	850	850	738	0	65	88.08	-	-	0.27	5	6.77	-	-
LCYB	941	1206	1500	941	0	37	39.32	-	-	0.13	0	0	-	-
NCED3	560	650	650	560	0	22	39.29	-	-	0.39	0	0	-	-
AOC	675	801	800	675	0	37	54.81	-	-	0.12	0	0	-	-
MRP4	774	782	900	363	411	14	38.57	24	58.39	0.29	0	0	1	2.43
CCC1	762	805	850	762	0	33	43.31	-	-	0.06	0	0	-	-
HKT1	238	1003	1200	116	122	10	86.21	9	73.77	0.17	0	0	1	8.20
LAPX	282	321	400	145	137	11	75.86	8	58.39	0.19	0	0	-	-
NADK2	339	787	1200	65	274	3	46.15	25	91.24	2.12	0	0	1	3.65
PIP1	190	346	500	103	87	5	48.54	21	241.38	0.01	0	0	0	0.00
SOS1	495	579	1000	358	137	22	61.45	12	87.59	0.18	0	0	1	7.30
TSC	335	505	800	136	199	7	51.47	17	85.43	0.58	0	0	0	0.00
Total	16238			10427	5811	556	52.89	541	98.39		7	0.66	43	7.58

(CS) Cleaned sequence (bp); (TS) Theoretical size EST; (GS) Genomic size; SC (Sequence Coding region); SNC (Sequence Non-coding region); (SNPc) SNPs in the coding region; (Freq) SNPs frequency per Kb; (SNPnc) SNPs in the non-coding region; (π_{nonsyn}/π_{syn}) average nonsynonymous/synonimous substitution rate; (indelc) indels in coding region; (indelnc) indels in non-coding region.

See Table 1 for gene abbreviations.

species believed to have given rise to *C. sinensis, C. aurantium, C. paradisi* and clementine, was 10.16 SNPs/kb. Comparing genera, the lowest density of SNPs was found in *Poncirus trifoliata* (0.55 SNPs/kb), but the highest level of inter-species differentiation was found between the latter and *C. medica* (18.18 SNPs/kb).

The average number of SNPs per kb that were specific to one taxon (observed at least in one genotype of the considered taxon but not in other taxa) was very similar for *C. reticulata, C. medica*, *Papeda*, *Fortunella* and *Poncirus*, with an average of 6.6, while lower rates were observed for *Microcitrus* (4.93), *C. maxima* (4.25) and *Eremocitrus* (3.3). No polymorphisms were observed between accessions of the same secondary species when two cultivars per species were studied (clementine, *C. sinensis*, *C. aurantium*).

Table 3. Inter accession polymorphism levels within and between taxa, and frequency of SNPs found in only a single taxon. Diagonal: average dissimilarities between two accessions within taxa (SNP/kb). Intersection: average dissimilarities between two accessions between taxa (SNP/Kb). Last lane: frequency of SNP found only in one taxon (SNP/Kb)

SNP/Kb	*C. reticulata*	*C. maxima*	*C. medica*	*Papeda*	*Fortunella*	*Microcitrus*	*Eremocitrus*	*Poncirus*
C. reticulata	1.54							
C. maxima	10.16	0.65						
C. medica	13.92	11.13	1.50					
Papeda	8.56	9.66	14.43	3.37				
Fortunella	8.70	7.95	12.27	5.71	6.04			
Microcitrus	9.99	10.09	13.77	9.74	8.74	2.41		
Eremocitrus	9.62	9.96	13.17	10.24	8.82	2.85	-	
Poncirus	13.37	13.17	18.18	13.85	13.00	14.90	14.98	0.55
Specific SNPs	6.77	4.25	6.28	6.47	6.65	4.93	3.33	6.84

Diagonal: average dissimilarities between two accessions within taxa (SNP per kb). Intersection: average dissimilarities between two accessions between taxa (SNP per kb). Last lane: frequency of SNPs found only in one taxon (SNP per kb). (*CR*) *C. reticulata*; (*CMAX*) *C. maxima*; (*CMED*) *C. medica*; (*PAP*) *Papeda*; (*FOR*) *Fortunella*; (*M*) *Microcitrus*; (*E*) *Eremocitrus*; (*P*) *Poncirus*.

Indels

Fifty indel polymorphisms were found. The average indel frequency in coding regions was 0.66 per kb, while the non-coding regions contained an average of 7.58 per kb. The most frequent indel was a mononucleotide (20 out of 50), but di-, tri-, tetra- and hexa-nucleotides were also abundant (20 out of 50 in total). Larger indels were less common. The largest indel, which contained 56 bp, was found in the PKF gene.

Comparison of diversity revealed at the intra- and inter-taxa level by SNPs, indel*s and SSRs*

We compared the diversity structures revealed by the identification of SNPs, indel markers defined from mined indel polymorphisms and 50 SSRs markers [previously used by Garcia-Lor *et al*. (2012b) to describe the genetic structure within *Citrus*]. Among the 50 indel sites identified, 25 were selected to develop indel markers. Twelve indel markers were published by Garcia-Lor *et al*. (2012b), and the primers for the 13 remaining markers can be found in Supplementary Information 2.

Average data for all of the SNP, indel and SSR loci analysed in this study are presented in Table 4. The lowest average number of alleles (*N*), and the observed (H_o) and expected heterozygosity (H_e) in the combined eight taxa, were revealed in the SNP markers (*N* = 2.008, H_o = 0.045, H_e = 0.173). SSR markers had the highest values (*N* = 11.080, H_o = 0.486, H_e = 0.822), and indel markers displayed intermediate values (*N* = 3.308, H_o = 0.125, H_e = 0.317). At the interspecific level in *Citrus*, an increasing order of H_e values was observed for *C. medica*, *C. maxima*, and *C. reticulata* in all markers types (SNP, indel, SSR). However, the relative values were variable. For example, the ratios between *C. maxima* and *C. reticulata* were 0.54 or 0.92 for SNPs and SSRs, respectively.

The average F_w values (excluding secondary species) for the three types of markers showed that there was a large deficit of heterozygous individuals observed in the population ($F_{w\ SNP}$ = 0.741, $F_{w\ indel}$ = 0.605, $F_{w\ SSR}$ = 0.409), which points to a high level of differentiation between the taxa. The Fst values of the differentiation between taxa (excluding secondary species) ($F_{st\ SNP}$ = 0.644; $F_{st\ indel}$ = 0.596; $F_{st\ SSR}$ = 0.392) were similar to F_w values, indicating that the taxon subdivision represents most of the genetic stratification. SNPs and indels revealed a higher inter-taxon structure than SSRs. At the intraspecific level, the only taxon that showed a consistently higher level of heterozygosity than was expected for all three marker types was *Poncirus trifoliata*.

Table 4. Statistical summary of the diversity of SNP, indel and SSR markers

	SNP				Indel				SSR			
	He	**Ho**	**Fw**	**N**	**He**	**Ho**	**Fw**	**N**	**He**	**Ho**	**Fw**	**N**
C. reticulata	0.067	0.061	0.091	1.212	0.225	0.245	-0.093	1.615	0.586	0.569	0.029	3.680
C. maxima	0.036	0.034	0.050	1.097	0.083	0.096	-0.155	1.231	0.540	0.549	-0.016	2.900
C. medica	0.022	0.006	0.737	1.059	0.027	0.031	-0.124	1.077	0.268	0.179	0.331	1.860
Papeda	0.088	0.048	0.450	1.223	0.113	0.051	0.545	1.308	0.775	0.480	0.380	3.520
Fortunella	0.075	0.065	0.140	1.207	0.260	0.231	0.112	1.923	0.616	0.575	0.067	3.674
Microcitrus	0.082	0.069	0.163	1.150	0.077	0.077	0.000	1.077	0.713	0.610	0.145	2.700
Eremocitrus	0.085	0.085	0.000	1.085	0.000	0.000	0.000	1.000	0.563	0.563	0.000	1.563
Poncirus	0.024	0.034	-0.416	1.049	0.046	0.077	-0.665	1.077	0.309	0.440	-0.423	1.660
Total AT	0.173	0.045	0.741	2.008	0.317	0.125	0.605	3.308	0.822	0.486	0.409	11.080
Whole dataset	0.166	0.072	0.568	2.036	0.317	0.172	0.457	4.154	0.814	0.554	0.320	11.560

Mean values are represented in the table. (H_e) Unbiased expected heterozygosity; (H_o) Heterozygosity observed; (F_w)Wright fixation Index; (*N*) Allele number; (AT) Ancestral taxa

Statistical test of neutrality and haplotype structure in the 'true citrus fruit trees' excluding secondary cultivated citrus species and hybrid cultivars.

The nucleotide variation observed for the gene sequences analysed is summarised for each taxon in Table 5, and the data presented for each gene is provided in [Supplementary Information 3]. Average total nucleotide diversity (π_T) was 0.012 for the entire sample set, ranging from 0.003 for citron to 0.009 for the *Papeda* group. Nucleotide diversity in silent and synonymous substitution sites was similar between the taxa and for the entire population, but non-synonymous nucleotide diversity was 3.52 times lower than the synonymous one (average π_{nonsyn} = 0.006). The non-synonymous substitution rate varied from 0.000 (PEPC, ACO and PIP1) to 0.010 (CHI, PSY, NADK2), and the ratio of non-synonymous to synonymous diversity ranged from 0.000 at PEPC (high conservative selection) to 2.273 at the EMA locus, which

suggests that selective constraints and/or the history of adaptive evolution vary between genes. The average non-synonymous/silent substitution rate was 0.345 for all of the genes and the entire population, indicating purifying selection. Within taxa, only the *C. reticulata* group at the HYB locus (π_{nonsyn}/π_{syn} = 1.421) and the F3'H locus (π_{nonsyn}/π_{syn} = 1.767) displayed higher non-synonymous than synonymous diversity. There were some groups with null synonymous mutations in the exons, so the π_{nonsyn}/π_{syn} ratio was not possible to calculate. In the entire sample set, several loci displayed a non-synonymous/synonymous ratio > 1, including CHI (π_{nonsyn}/π_{syn} = 1.377), EMA (π_{nonsyn}/π_{syn} = 2.273) and NADK2 (π_{nonsyn}/π_{syn} = 2.117). Taking into account only the basic taxa (excluding secondary species and recent hybrids), four loci showed values > 1, including CHI (π_{nonsyn}/π_{syn} = 1.381), EMA (π_{nonsyn}/π_{syn} = 1.511), PSY (π_{nonsyn}/π_{syn} = 3.533) and NADK2 (π_{nonsyn}/π_{syn} = 2.043). The PKF locus had a π_{nonsyn}/π_{syn} value of 0.883 for the entire population and 1.072 for the ancestral taxa group. For the entire population MDH and HYB loci had a π_{nonsyn}/π_{syn} value of 1.065 and 0.914 respectively.

The level of differentiation between the taxa (evaluated by F_{st}; Supplementary Information 3) was relatively homogenous among the genes. Highest and lowest values were found for SOS1 (F_{st} = 0.814) and PIP1 (F_{st} = 0.438), respectively, with an average of 0.644 +/- 0.036.

No significative Tajima's *D* value was found in any of the genes in the entire population [Supplementary Information 3].

Table 5. Summary of nucleotide diversity and divergence within and between species

Taxa	*S*	π_T	π_{sil}	π_{syn}	π_{nonsyn}	π_{nonsyn}/π_{syn}	N_h	H_d	H_d (SD)
C. reticulata	8.926	0.005	0.008	0.010	0.003	0.411	4.407	0.593	0.096
C. maxima	3.926	0.004	0.005	0.004	0.001	0.191	3.222	0.521	0.116
C. medica	2.815	0.003	0.004	0.004	0.001	0.256	2.037	0.296	0.068
Fortunella	8.481	0.006	0.009	0.008	0.003	0.285	5.185	0.683	0.097
Papeda	9.630	0.009	0.015	0.014	0.003	0.292	4.519	0.871	0.126
Microcitrus	5.889	0.006	0.009	0.011	0.003	0.184	2.926	0.760	0.198
Eremocitrus	3.407	0.006	0.009	0.013	0.004	0.154	1.778	0.772	0.380
Poncirus	2.407	0.003	0.005	0.003	0.000	0.088	2.148	0.469	0.099
Main taxa	39.667	0.013	0.021	0.020	0.006	0.555	23.074	0.926	0.016
Whole Pop	40.926	0.012	0.021	0.020	0.005	0.495	28.333	0.901	0.015
max	9.630	0.009	0.015	0.014	0.004	0.411	5.185	0.871	0.380
min	2.407	0.003	0.004	0.003	0.000	0.088	1.778	0.296	0.068

(*S*) Segregating sites, (π_T) Nucleotide diversity total, (π_{sil}) Nucleotide diversity silent sites, (π_{syn}) Nucleotide diversity synonymous sites, ($\pi_{nonsyn/syn}$) Ratio nucleotide diversity nonsynonymous/synonymous sites, (N_h) Number of haplotypes, (H_d) Haplotype diversity, (SD) Standard deviation. Max and min: maximum and minimum values within the basic taxa

The average number of haplotypes per locus in the entire population was 28.33, with a maximum value of 5.185 haplotypes in *Fortunella* and a minimum value of 1.778 in *Eremocitrus*. Regarding the four main ancestors in *Citrus*, *Papeda* had the highest number of haplotypes (4.519), followed by *C. reticulata* (4.407), *C. maxima* (3.222) and *C. medica* (2.037). At intra-

taxon level, haplotype diversity ranged from 0.871 for the *Papeda* to 0.296 for *C. medica* (Table 5).

Phylogenetic analysis

Among all of the models tested via Phylemon website (, the model with the best fit was TVM+I+G+F (with SH-like branch supports alone). This model takes into account the nucleotide substitution model TVM 'Transitional model' (five substitution classes: AC, AT, CG, GT, AG = CT), the proportion of invariable sites (I), the nucleotide frequency (F) and the gamma distribution (G). The phylogenetic relationships between *Citrus* species and their relatives inferred from ML method using this model are represented in Figure 1. Branch support (BS) is given in all branches. The different 'true citrus fruit trees' genotypes were rooted using *Severinia buxifolia* as outgroup.

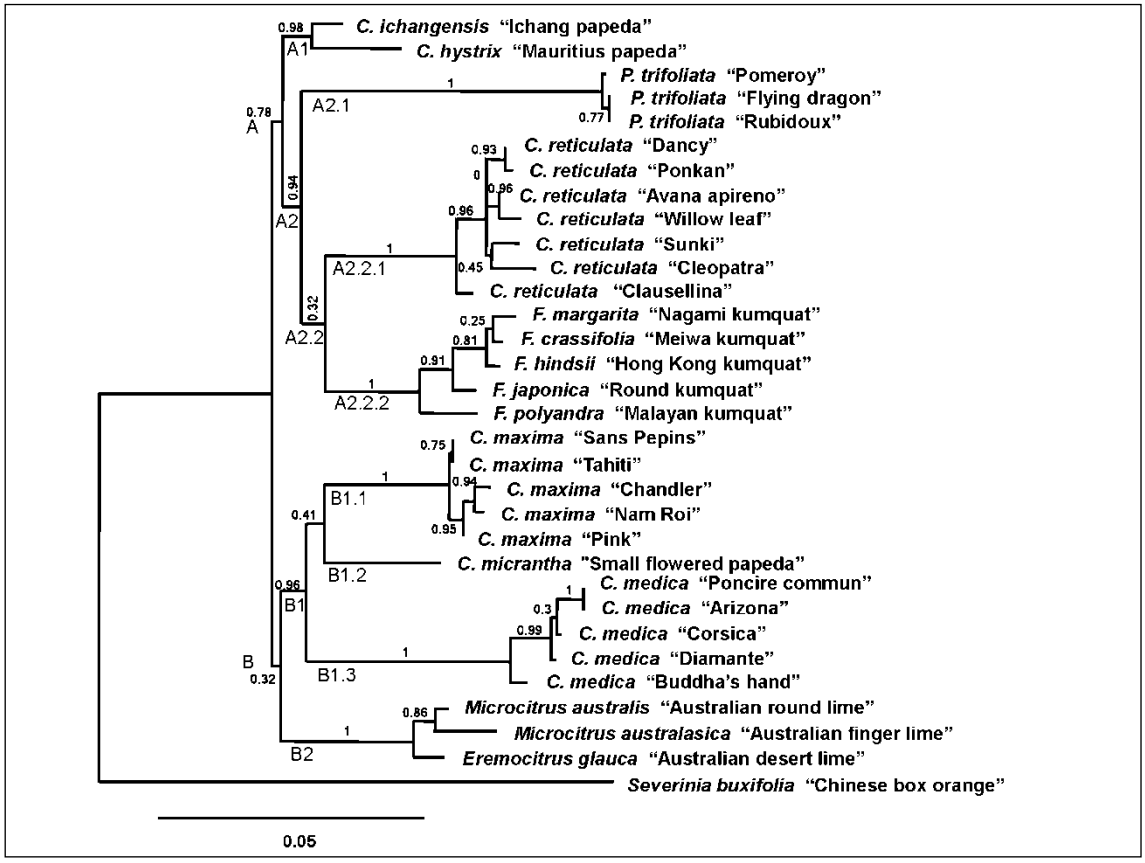


Figure 1. Phylogenetic relationship between *Citrus* ancestral taxa (*C. reticulata*, *C. maxima*, *C. medica*, *Papeda*) and relatives (*Fortunella*, *Microcitrus*, *Eremocitrus*, *Poncirus trifoliata*). Phyml Best AIC Tree (v. 1.02b), model TVM+I+G+F (with SH-like branch supports alone)

The first two clades (A and B) are each divided in two subclades. The clade A has a medium BS (0.78), joining a subclade A1 (BS = 0.98) of two *Papeda* species (*C. hystrix* and *C. ichangensis*) and a strong subclade A2 (BS = 0.94) including all *Poncirus trifoliata* (monospecific subclade A2.1, BS = 1), all the *C. reticulata* accessions (monospecific subclade A2.2.1, BS = 1) and all *Fortunella* accessions (monogeneric subclade A2.2.2, BS = 1). *Fortunella* and *C.*

reticulata are joined in a subclade A2.2 with a low BS (0.32). On the other side of the tree, clade B (low BS = 0.32) includes two groups. The first group, B1 (BS = 0.96), is divided into three specific subclades highly supported, *C. maxima* accessions (B1.1; BS = 1), *C. micrantha* (B1.2; only one accession) and *C. medica* (B1.3; BS = 1) accessions. The second subclade B2 (BS = 1) includes *Microcitrus* and *Eremocitrus*, two strongly associated genera of Australian origin. *Papeda* is the only group that does not display a monophyletic structure, the accessions of each of the other groups (*Poncirus*, *C. reticulata*, *Fortunella*, *C. maxima* and *C. medica*, *Microcitrus* and *Eremocitrus*) are all joined in specific clades clearly differentiated from the other taxa.

This phylogenetic structure is similar, for several strong grouping, to the structure observed using Neighbour Joining (NJ) analysis based on SNP data (Figure 2). In the NJ tree, the association between *C. reticulata* and *Fortunella* (BS = 0.96) is maintained, as are the *C. maxima* / *C. medica* (BS = 0.8) and *Microcitrus* / *Eremocitrus* (BS = 1) associations. The *Papeda* group is shifted from one group to the other. *Poncirus trifoliata* appears as the most distant species, it is the first one that separates from the others. This in agreement with the high differentiation level of *Poncirus* with all other taxa (Table 3).

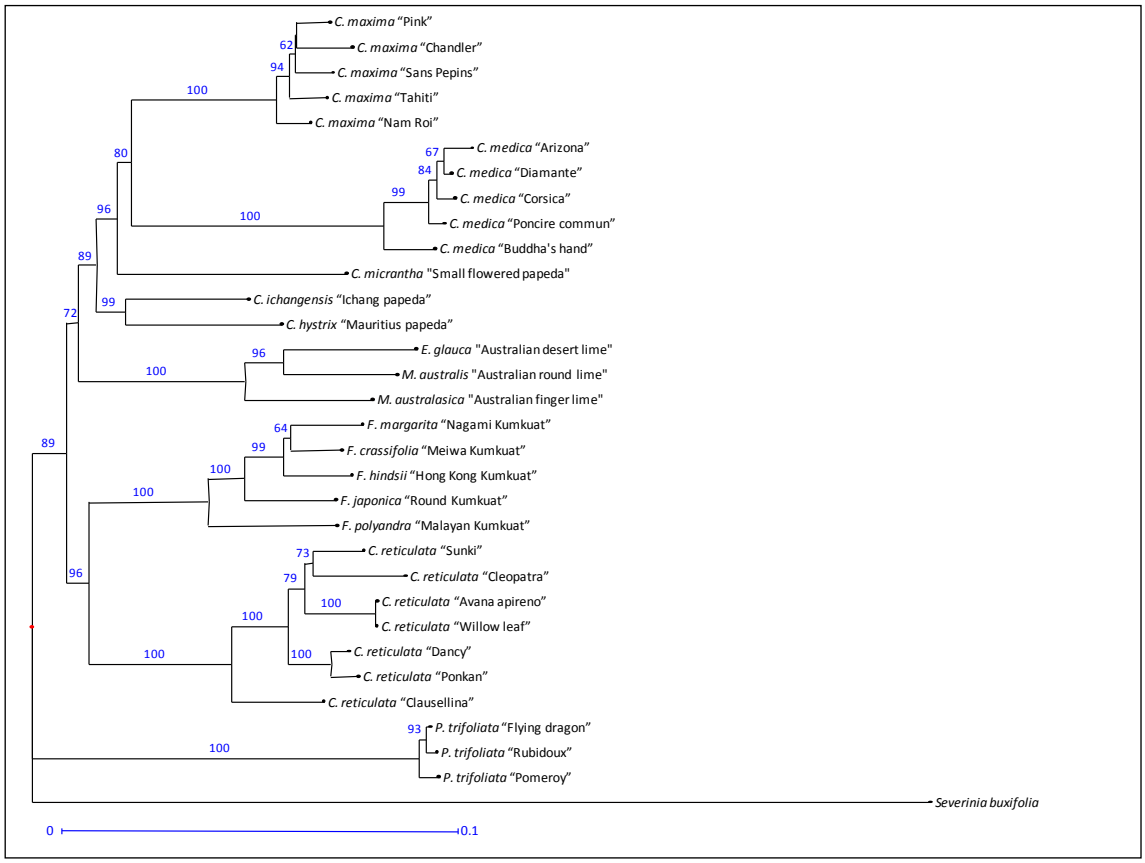


Figure 2. NJ tree with 1097 SNP markers in the ancestral *Citrus* species and relatives (1000 bootstraps performed). Branch support over 50% are shown

When the secondary species and interspecific hybrids were added to the analysis [Supplementary Information 4], the NJ representation was modified, the relationships described before are not maintained. *Citrus reticulata* appears to be more closely related to *C. maxima*

than to *Fortunella*, and *C. medica* is not so closely related to *C. maxima* as was suggested by the Phylemon and Darwin analysis that excluded the hybrid genotypes.

Genome structure of citrus secondary species and hybrids

We used factorial analysis to examine the potential contribution of the ancestral species to the inheritance of 27 genes in secondary cultivated species (Table 6). For the SNPs of these 27 genes, almost 70% of the diversity in *Citrus* species is explained by the first two axes (Figure 3). The basic *Citrus* taxa are clearly distinguished. Secondary species are positioned between their putative parental gene pools: *C. sinensis* is between *C. maxima* and *C. reticulata*, *C. paradisi* is between *C. sinensis* and *C. maxima*, *C. limon* is between *C. aurantium* and *C. medica* and *C. aurantifolia* between *C. medica* and *C. micrantha* (Figure 3). With the goal of performing a gene-by-gene analysis of the phylogenetic inheritance in the secondary species, we performed a PCoA for each gene using the basic taxa of cultivated citrus as active individuals, and we projected the secondary species genotypes onto the defined axes. The phylogenetic inheritance was inferred from the position of the secondary species in the PCoA relative to the ancestral species and the analysis of SNP allelic locus configurations. The genetic structure of the FLS locus (Figure 4) is presented as an example of phylogenetic assignation. Grapefruit, sweet orange, sour orange, tangor 'King' and tangelo 'Orlando' are in an intermediate position between the *C. reticulata* (mandarin; M) and *C. maxima* (pummelo; P) groups. It was therefore assumed that these species should have inherited one allele of this gene from each of these ancestral groups (interspecific heterozygosity MP). This was confirmed by examining the allelic configuration at each SNP locus. Using the same approach, lemon appears to be heterozygous (MC) for the *C. reticulata* and *C. medica* (citron; C) alleles, while clementine appears to have inherited two *C. reticulata* alleles (MM).

For most genes (18/27) clementines appear to have inherited *C. reticulata* alleles in phylogenetic homozygosity. However, nine genes appear to be heterozygous between *C. reticulata* and *C. maxima*. For all the genes analysed, the estimated contribution of *C. reticulata* was 83.3%, while the estimated contribution of *C. maxima* was 16.7%.

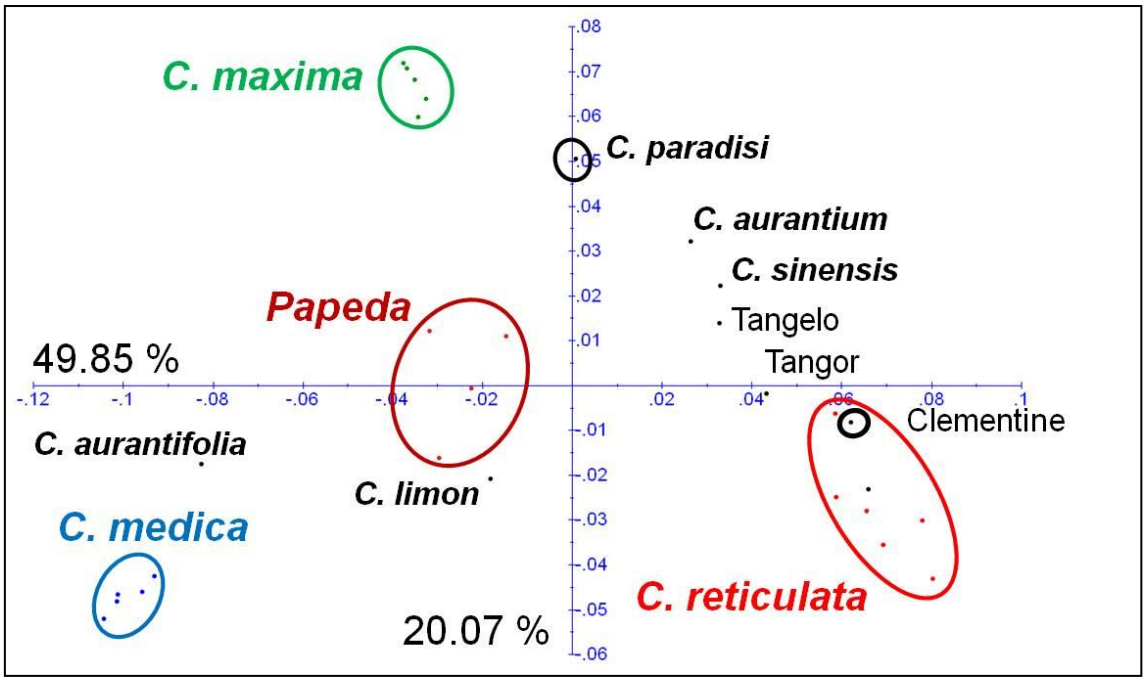


Figure 3. Genetic relationship between secondary *Citrus* species and basic taxa (factorial analysis; axes 1/2)

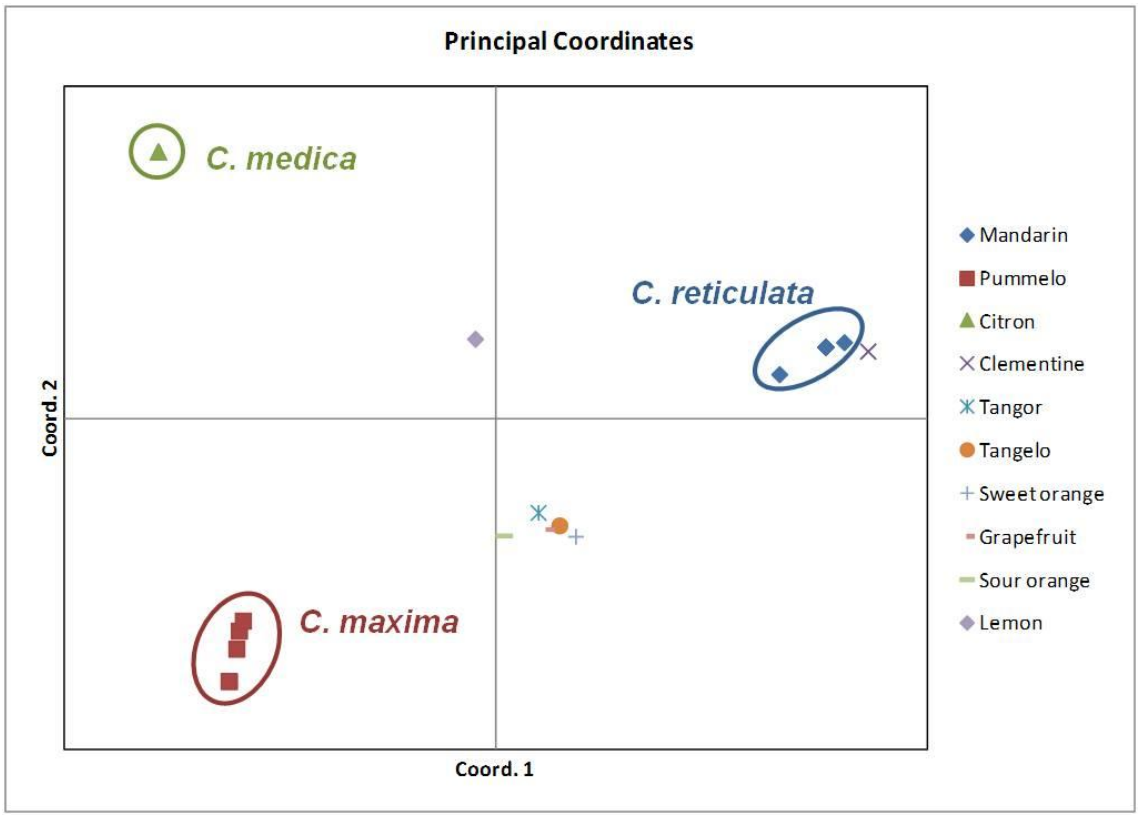


Figure 4. Genetic organizational analysis (principal co-ordinates) of secondary species and recent hybrids (flavonoid synthase gene)

Table 6. Phylogenetic origins of genes of secondary species and hybrids

Gene	Clem		CS		CP		CA			CAU		CL	TK		TO	
CHI	M	P	M	P	P	P	M	P	C	C	C	M/P?	M	M	M	P
CHS	M	M	M	M	M	P	M	P	C	PAP	C	M	M	M	M	M
FLS	M	M	M	P	M	P	M	P	C	PAP	C	M	M	P	M	P
F3'H	M	M	M	M	P	P	M	P	C	PAP	C	M	M	M	M	M
DFR	M	P	M	P	P	P	M	P	C	PAP	C	M	M	M	M	P
EMA	M	M	M	M	M	P	M	P	C	PAP	C	M	M	M	M	M
MDH	M	M	M	P	P	P	M	P	C	PAP	C	P	M	P	M	P
ACO	M	M	M	M	M	P	M	P	?	?	C	M	M	M	M	M
TRPA	M	P	P	P	P	P	M	P	C	PAP	C	M	M	P	M	P
INVA	M	M	M	P	P	P	M	P	C	PAP	C	P	M	P	M	P
PEPC	M	M	M	P	P	P	M	P	C	PAP	C	P	M	M	M	P
PKF	M	P	M	P	P	P	?	P	C	PAP	C	M	M	P	M	P
DXS	M	M	M	P	M	P	M	P	C	PAP	C	M	M	P	M	P
PSY	M	M	M	P	M	P	M	P	C	PAP	C	M	M	M	M	M
HYB	M	M	M	M	M	P	M	P	?	?	C	M	M	M	M	M
LCY2	M	P	M	P	M	P	M	P	C	PAP	C	M	M	P	M	M
LCYB	M	P	M	P	M	P	M	P	C	PAP	C	M	M	M	M	M
NCED3	M	P	P	P	P	P	M	P	C	PAP	C	P	M	P	M	P
AOC	M	M	M	M	M	P	M	P	C	PAP	C	M	M	M	M	M
MRP4	M	M	M	M	M	P	M	P	C	PAP	C	M	M	M	M	M
CCC1	M	P	P	P	M	P	M	P	C	PAP	C	M	M	P	P	P
HKT1	M	P	M	P	M	P	M	P	C	PAP	C	M	M	P	M	M
LAPX	M	M	M	P	M	P	M	P	C	PAP	C	M	?	P	M	P
NADK2	M	M	M	P	P	P	M	P	C	PAP	C	M	M	M	P	P
PIP1	M	M	M	M	M	P	?	P	?	?	C	P	M	M	M	M
SOS1	M	M	M	P	P	P	M	P	C	PAP	C	P	M	P	M	P
TSC	M	M	M	P	M	P	M	P	C	PAP	C	M	M	M	M	P

(Clem) Clementine; (CS) *C. sinensis*; (CP) *C. paradisi*; (CA) *C. aurantium*; (CAU) *C. aurantifolia*; (CL) *C. limon*; (TK) Tangor 'King'; (TO) Tangelo 'Orlando'; (M) Mandarin, (P) Pummelo, (C) Citron, (PAP) Papeda, (?) Origin not known.

Citrus sinensis appears to contain more alleles from *C. reticulata* (59.3%) than from *C. maxima* (40.7%). It inherited two alleles from *C. maxima* (PP) for three genes and two alleles from *C. reticulata* (MM) for eight genes. The remainder of the genes appear to be in phylogenetic heterozygosity from both gene pools (MP).

Citrus paradisi has 11 genes that were solely inherited from *C. maxima*, while the rest of the genes were heterozygously inherited from *C. maxima* and *C. reticulata*. The contributions from the parental lines were therefore 70.4% for *C. maxima* and 29.6% for *C. reticulata*.

Citrus aurantium contains two loci with parental origins that were not possible to define due to the presence of specific alleles at the SNP loci. The other loci were heterozygous *C. maxima* / *C. reticulata* (MP). Therefore, for the loci with complete phylogenetic assignation, the contributions of *C. maxima* and *C. reticulata* were each 50%.

Citrus aurantifolia contains three genes with phylogenetic origins that were not possible to infer. Most of the other genes showed interspecific heterozygosity between *C. medica* and *Papeda*. However, CHI appeared to be homozygous for *C. medica* alleles (CC). Therefore, for the 24 genes that could be analysed, the contributions of *C. medica* and *Papeda* were 53 and 47%, respectively.

Citrus limon showed the most diverse combination of parental contribution patterns. Twenty genes resembled a combination of *C. medica* and *C. reticulata* genes, six genes resembled a combination of *C. medica* and *C. maxima* genes, and one locus could not clearly be identified. For the genes that could be identified, *C. medica* contributed 50%, *C. reticulata* contributed 38.5 % and *C. maxima* contributed 11.5% to the *C. limon* genome.

'King', which is assumed to be a tangor (*C. reticulata* x *C. sinensis*), and tangelo 'Orlando' (*C. paradisi* x *C. reticulata*) contained some genes that exhibited interspecific heterozygosity (*C. reticulata* and *C. maxima*; MP) and some that displayed monospecific inheritance (MM or PP). The relative contributions of the *C. reticulata* and *C. maxima* gene pools were, respectively, 75.93 and 24.07% for 'King' and 66.67 and 33.33% for 'Orlando'.

DISCUSSION

SNP and indel discovery and analysis of the relative utility of these markers compared to SSRs for use in diversity and phylogenetic studies.

In 'true citrus fruit trees', the average number of SNPs per kb in non-coding regions is almost two times higher than in coding regions. This value is high compared to the value obtained for *Eucalyptus* spp. (1.5 times higher; Külheim *et al*., 2009). The mean frequency of SNPs/kb found in exons was 28.96 for *Citrus*, which is higher than in other species such as *Populus tremula*, with 16.7 SNPs/Kb (Ingvarsson, 2005), and in maize, with 23.25 SNPs/Kb (Yamasaki *et al*., 2005). Regarding the SNP frequency in *Citrus* spp. the values were lower [*C. reticulata* (15.15 SNP/kb), *C. maxima* (4.70 SNP/kb), *C. medica* (2.21 SNP/kb)]. Moreover, the value is lower than that found in *Quercus crispula*, with 40 SNPs/Kb (Quang *et al*., 2008) and *Eucalyptus camaldulensis*, with 47.62 SNPs/Kb (Külheim *et al*., 2009). The percentage of transition and transversion events are similar to those found in other species, such as oil palm (0.58 and 0.42, respectively; Riju *et al*., 2007). In *Citrus*, these results are in agreement with results reported by Dong *et al*. (2010), Terol *et al*. (2008) and Novelli *et al*. (2004). In contrast, the transition fraction was found to be substantially higher in poplar (70%; Tuskan *et al*., 2006).

The nucleotide diversity value observed in the 'true citrus fruit trees' and in *C. reticulata* ($\pi = 0.005$) was similar to the values observed in grapevine ($\pi = 0.005$; Lijavetzky *et al*., 2007), maize ($\pi = 0.006$; Ching *et al*., 2002) and rye ($\pi = 0.006$; Li *et al*., 2011), while the value was approximately five times higher than those observed in soybean ($\pi = 0.00097$; Zhu *et al*., 2003) and human ($\pi = 0.001$; Sachidanandam *et al*., 2001). Compared with the diversity data within *Citrus* obtained with SNPs mined in clementine (Ollitrault *et al*., 2012a), it appears that the relative diversity levels of the three basic taxa were quite different. Indeed, the Nei diversity values (H_e) of *C. maxima* and *C. medica* over *C. reticulata* were 0.23 (0.063 / 0.279) and 0.20 (0.057 / 0.279), respectively, while the values obtained in the present study were 0.53 (0.036 / 0.067) and 0.33 (0.022 / 0.067), respectively, confirming the conclusion of *Ollitrault et al*. (2012a) that the ascertainment bias due to the scarcity and specificity of the discovery panel of the SNPs mined in clementine resulted in an overestimation of the relative diversity within *C. reticulata*. Analysis of the average inter-accession polymorphism within and between species reveals that for the three basic taxa of cultivated *Citrus* (*C. reticulata*, *C. maxima*, *C. medica*), the ratios between and within species were high. For example, within *C. reticulata* and between *C. reticulata* and *C. maxima*, the ratio was close to 6.6 (10.16 / 1.54). Therefore, the analysis of SNP density along the genome should help differentiate between genomic regions with interspecific heterozygosity (MP for example) and those that result from intraspecific inheritance (MM or PP, for example) in the genomes of secondary species.

The information obtained by studying the allelic diversity of the analysed genes will allow us to optimise molecular tools for both genomic and transcriptomic studies. The identification of conserved areas can be used to develop primers or hybridization sequences to

limit sources of bias such as null alleles or differential allelic PCR competition or hybridisation. Identification of the different alleles of these genes also opens the way for allele-specific expression studies.

The frequencies of indels per kb in the 'true citrus fruit trees' species were 0.66 and 7.58 in exon and intron sequences, respectively. These frequencies are comparable to values reported for other species such as maize (18 genes studied, 6935 bp), where 0.43 and 11.76 indels/kb were found in coding and non-coding regions, respectively (Ching *et al*., 2002), and *Brassica* (557 clone sequences, 1 396 498 bp), with 0.45 and 7.42 indel/kb in coding and non-coding regions, respectively (Park *et al*., 2010). In melon (34 ESTs sequenced, ± 15000 bp), indels occurred less frequently in introns (approximately 0.60/kb), and no indels were found inside coding regions (Morales *et al*., 2004). In grapevine (230 gene fragments sequenced, > 1Mb), very low levels of indel polymorphism were found, with 0.11 and 2.25 indel/kb in coding and non-coding regions, respectively (Lijavetzky *et al*., 2007).

Considering the eight basic taxa together, the fixation index (F_w) values and the differentiation index values (F_{st}) between taxa obtained using three types of markers (SSRs, SNPs, indels) confirmed the high degree of stratification in differentiated taxa with limited gene flows. However, the levels of diversity revealed by the three types of markers were quite different. The indel markers developed in this study confirmed that indels are very efficient tools for inter-specific differentiation, as was demonstrated by Garcia-Lor *et al*. (2012a) and Ollitrault *et al*. (2012). The indel markers developed in this study had an average F_{st} value of 0.596, similar to that obtained using SNP markers (F_{st} = 0.644), whereas with 50 SSR markers analysed for the same accessions, the F_{st} value was only 0.392. In contrast, the SNP loci and indels mined from our much diversified interspecific panel appeared, on average, to be less polymorphic to describe intraspecific polymorphism. However, in our study, which includes several genotypes for each species, we also identified numerous SNP loci that revealed intraspecific diversity that should be useful for germplasm characterisation and management. Unlike SSRs and indel sequences, SNPs can be employed in high-throughput screening and in relatively low-cost genotyping methods. Their utility is limited, however, due to the fact that they are usually present only as diallelic polymorphisms.

Evolution of citrus genes

In 'true citrus fruit trees', the average ratio of non-synonymous to silent SNP rates per site (π_{nonsyn}/π_{sil}) was 0.345. Within *Citrus* spp. similar values were found in *C. reticulata* (0.385) and *C. medica* (0.339), but higher in *C. maxima* (0.577). This is higher than the 0.17 and 0.21 ratios observed in white spruce (Pavy *et al*., 2006) and in *Arabidopsis thaliana* (in a study of 242 genes; Zhang *et al*., 2002), respectively. These relatively low values indicate that, on average, white spruce open reading frames and nuclear genes in *A. thaliana* are probably under higher purifying selection pressure than the genes of 'true citrus fruit trees'. This can probably be attributed to the wide diversity encompassed by 'true citrus fruit trees' and the high genetic and

phenotypic differentiation between the different taxa that have experienced allopatric evolution (even if they are still sexually compatible). The minimum value of π_{nonsyn}/π_{sil} in our entire data set was 0 at the PEPC locus, and the maximum value was 1.09 at the NADK2 locus. The non-synonymous substitution rate varied from 0.000 in PEPC to 0.010 in CHI, which suggests that selective constraints vary between loci (Fu *et al*., 2010).

In the carotenoid biosynthetic pathway, different key steps have been found to be associated with differentiation between cultivated *Citrus* spp. (Kato *et al*., 2004; Fanciullino *et al*., 2006a, 2007). Several studies have tried to clarify the regulation of carotenoid biosynthesis (Rodrigo *et al*., 2004; Kato *et al*., 2004; Kim *et al*., 2001), but this regulation has not yet been fully elucidated.

PSY drives the formation of phytoene, the first product in the carotenoid biosynthetic pathway and a major step in the differentiation between cultivated basic taxa (Fanciullino *et al*., 2006, 2007). Considering the eight taxa studied, it appears that PSY is under positive selection (π_{nonsyn}/π_{syn} = 3.533) and is associated with a high level of allelic differentiation between the taxa (F_{st} = 0.750), which is higher than the average. There were nine sites with SNP polymorphisms between *C. reticulata* and the other taxa that produced changes in the amino acid composition that may be responsible for their differentiation. In contrast, in *C. reticulata*, no changes were found (excepted for one heterozygous change in the cultivar 'Ponkan'). Further functional analysis of the different alleles of this gene should provide insights into the molecular basis of phenotypic differentiation.

LCYB is a key enzyme required for the conversion of lycopene into β-carotenoids (Fanciullino *et al*., 2006a; Alquézar *et al*., 2009). Fanciullino *et al*. (2007) proposed that allelic variation at this locus should strongly limit this biosynthetic step in *C. maxima*. The numerous amino acid changes observed in *C. maxima* compared with *C. reticulata* might be associated with this limitation due to changes in the functionality of the pummelo allele.

HYB also plays a major role in the carotenoid biosynthetic pathway (Fanciullino *et al*., 2006a) by catalysing the transformation of β-carotene into β-cryptoxanthin and zeaxanthin. *Citrus reticulata* produces these compounds, while *C. maxima* do not converts β-carotene into β-cryptoxanthin and zeaxanthin and *C. medica* only convert β-carotene into β-cryptoxanthin. Within *C. reticulata*, the ratio between non-synonymous/synonymous substitutions was higher than one (positive selection) at the HYB locus, which might be related to the significant variation in β-cryptoxanthin levels found among *C. reticulata* cultivars (Fanciullino *et al*., 2006a). The β-cryptoxanthin content greatly enhances fruit colour and has probably been under human-induced selection during domestication.

Regarding the flavonoid pathway, positive selection was found to occur in *C. reticulata* at the F3'H locus, which belongs to the cytochrome P450 family and catalyses the hydroxylation of flavonoids at the 3' position of the B-ring, leading to the production of hydroxylated flavonols, proanthocyanidins (condensed tannins) and anthocyanins (Winkel-Shirley, 2001). This gene plays an important role in flavonoid biosynthesis in *Arabidopsis* (Schoenbohm *et al*., 2000) and

grapevine (Bogs *et al*., 2006) and was previously isolated in clementine by Garcia-Lor *et al*. (2012b). Schoenbohm *et al*. (2000) demonstrated that in yeast, this enzyme could convert naringenin or dihydrokaempferol into eriodictyol or dihydroquercetin, respectively. Therefore, the changes in non-synonymous amino acid composition in the mandarin group (*C. reticulata*) may be associated with the different flavonol compositions found in some studies (Gattuso *et al*., 2007). At the CHI locus, a greater number of non-synonymous vs. synonymous substitutions were not found to have occurred in the eight subpopulations studied, but at the interspecific level, the ratio was higher than 1, meaning that the gene was probably subjected to positive selection during the interspecific differentiation process. This gene controls the second step of the flavonoid biosynthetic pathway (Winkel-Shirley, 2001), and it was shown that it can alter flavonoid levels in citrus leaves (Koca *et al*., 2009). Understanding F3'H and CHI regulation and allelic functionality could be important for the analysis of molecular determinants of flavonoid composition in citrus fruits.

In the biosynthesis of acidic compounds, EMA displayed non-synonymous/synonymous ratios greater than one (π_{nonsyn}/π_{syn} = 2.273) and evidenced positive selection at the interspecific level. EMA is involved in the last steps of the citric acid cycle, catalysing the transformation of malate into pyruvate, the precursor of citrate formation (Kay and Weitzman, 1987). Malic enzyme is activated by the accumulation of citric acid cycle intermediates, allowing excess intermediates to leave the cycle and re-enter as acetyl groups, producing more citric acid. Citric acid content is strongly differentiated between *Citrus* taxa and ranges from 0.005 mol/L for oranges and grapefruits to 0.30 mol/L for lemons and limes (Penniston *et al*., 2008).

None of the sugar biosynthesis genes exhibited positive selection. It is well known that the total concentration of sugars increases throughout maturation in all *Citrus* spp. (Albertini *et al*., 2006). The null level of non-synonymous divergence at PEPC is consistent with strong selection for conserved amino acid sequences in this gene, which plays a crucial role in such important processes as C4 and Crassulacean acid metabolism (CAM) photosynthesis.

In the entire sample set, taking into account only the eight ancestral taxa (excluding secondary species and recent hybrids), NADK2 displayed a non-synonymous/synonymous ratio greater than 1 (π_{nonsyn}/π_{syn} = 2.117 and π_{nonsyn}/π_{syn} = 2.043, respectively). NADK (NAD kinase) catalyses the ATP-dependent phosphorylation of NAD(H) (Berrin *et al*., 2005). In *A. thaliana*, there are three isoforms of NADK. Two isoforms, NADK1 and NAD(H)K3, are cytosolic and one, NADK2, is found in the chloroplast (Turner *et al*., 2004, 2005; Chai *et al*., 2005, 2006). These isoforms play an essential role in the phosphorylation of NAD(H) and have been linked to plant stress response. Chai *et al*. (2005) showed that manipulation of *At*NADK2 levels affected plastid NADPH levels, and null mutants were stunted, with a pale yellow colour, and were hypersensitive to abiotic stress.

Differences found in the coding regions of NADK2, and thus variations in amino acid sequences between the taxa, might affect the responses of these genotypes to abiotic stresses. Full sequencing of this gene and functional analysis of the different alleles could greatly

increase our understanding of the role that this gene plays in increasing stress tolerance in *Citrus* and its relatives.

For all of the genes discussed here, the sequence data highlight amino acid variability of corresponding proteins that were probably subjected to selection. Therefore, these genes are good candidates for further complete sequencing studies (including promoter sequencing) and allelic functional studies to decipher the molecular basis of the phenotypic variability in the species examined.

Despite the previous discussion concerning the possible selective pressure exerted on some of the genes studied, the genetic organization of *Citrus* obtained from the SNP data (Figure 1) is similar to the genetic organization elucidated in previous SSR studies (Ollitrault *et al*., 2010, Garcia-Lor *et al*., 2012a). This suggests that the same basic type of evolutionary components led to the diversity structures of both types of markers. Therefore, a predominantly neutral selection pattern can be assumed for most of the current SNP markers. The minimum F_{st} value was 0.438 at the PIP1 locus and the maximum value was 0.814 at the SOS1 locus for the differentiation of the eight taxa analysed in this work, *i.e. C. reticulata*, *C. maxima*, *C. medica*, *Papeda*, *Fortunella*, *Microcitrus*, *Eremocitrus* and *Poncirus trifoliata*. This study sheds light on the important differentiation between the taxa and demonstrates that SNP markers are efficient tools for phylogenetic studies and inheritance analysis of secondary species.

Phylogenetic relationships

For a biologically complex crop such as citrus, information obtained from nuclear gene sequences is more useful than the information gleaned from maternally inherited plastid sequences (Ramadugu *et al*., 2011; Puritz *et al*., 2012) due to the possibility of gene flow between sexually compatible species and the fact that the species belong to the same area of diversification. Previous phylogenetic molecular analyses using plastid markers showed that all 'true citrus fruit trees' species constitute a clade that is differentiated from other genera (de Araújo *et al*., 2003; Bayer *et al*., 2009).

In our study, all accessions of the same species form a clade with mainly high branch support values. Two species in the *Papeda* group, *C. hystrix* and *C. ichangensis*, are closely related. The other species of subgenus *Papeda*, *C. micrantha*, is separated from the two previous ones, possibly due to its geographical origin and distribution. The origin of *C. micrantha* is believed to be in the Philippines, whereas *C. hystrix* and *C. ichangensis* are of continental origin, in Burma, Thailand and Indo-China (Tanaka, 1954). Therefore, Swingle and Reece's (1967) subdivision of the genus into subgenera *Papeda* and *Citrus* seems to be inadequate.

An important observation maintained through the ML phylogenetic trees and the NJ cluster analysis is that *C. reticulata* and *Fortunella* form a cluster clearly differentiated from another cluster including *C. maxima*, *C. medica* and *C. micrantha*. The close relationship

between *C. reticulata* and *Fortunella* matches the results obtained by Penjor *et al.* (2010) that were based on the analysis of *rcbL* plastid gene sequences, but it differs from the results obtained from the analysis of amplified fragment length polymorphism (AFLP) molecular markers (Pang *et al*., 2007) and SSR markers (Barkley *et al*., 2006) and Swingle and Reece's (1967) treatment of *Fortunella*. In the ML phylogenetic analysis, *P. trifoliata* was found to belong to the same clade as *C. reticulata* and *Fortunella* with strong branch support (0.94). However, in the NJ analysis *P. trifoliata* appears as the more distant to all the 'true citrus fruit trees' taxa analysed, in agreement with our estimation of the inter-taxon differentiations. The strongly supported clade (B1; BS = 0.96) including *C. medica*, *C. maxima* and *C. micrantha* of subgenus *Papeda* is also observed in the NJ analysis. However, our results are in contrast to information derived from other studies, including the analysis of nine plastid markers by Bayer et al. (2009), the analysis of SSR, SRAP and (CAPS)-SNP markers (Amar *et al*., 2011), SSRs (Barkley *et al*., 2006) and RAPD, SCAR and plastid markers (Nicolosi *et al*., 2000). All of these studies suggested that *C. maxima* and *C. reticulata* share a clade and are separated from *C. medica*. The inconsistency with previous nuclear studies may be due to the inclusion of secondary species of interspecific origin in these previous studies, which might have led to the artefactual clustering of the *C. maxima* and *C. reticulata* gene pools due to the numerous accessions resulting from hybridisation between these gene pools. Our phylogenetic ML analysis (Figure 1) and the NJ analysis done with the SNPs in the absence of secondary species (Figure 2) are consistent, while the NJ tree that includes the secondary species [Supplementary Information 4] displays clustering of *C. maxima* and *C. reticulata* with low branch support. This illustrates the bias associated with the inclusion of genotypes of inter-taxon origin in NJ cluster analyses. Another source of bias in molecular studies might be the choice of molecular marker type and the genotype panel used for its development. In our study, using Sanger sequencing, all SNPs from all accessions are revealed, so there was no bias towards any of the ancestral species.

The consistent clades observed in the ML phylogenetic study are in agreement with the geographical distribution of species divided by the 'Tanaka line' (Tanaka, 1954). *Fortunella*, *Poncirus* and *C. reticulata* (clade A2) share the same area of diversification, where subgenus *Metacitrus* predominates (East Asiatic floral zone) (Tanaka, 1954), whereas the *C. medica* and *C. maxima* clade (B1) is in agreement with the area of distribution where the subgenus *Archicitrus*, described by Tanaka (1954), predominates (Indo-Malayan floral zone). Some phenotypic traits differentiate these two clades. For example, *Fortunella*, *Poncirus* and *C. reticulata* are facultative apomictic species with high carotenoid contents, while *C. maxima* and *C. medica* are monoembryonic non-apomictic species, which have strong limitations in the carotenoid pathway. The speciation between *Fortunella*, *Poncirus* and *C. reticulata* might be explained by their different flowering periods (precocious in *Poncirus* and late in *Fortunella*). However, gene flow probably occurred by accidental, out-of-time flowering. Despite sharing the Indo-Malayan floral zone (Tanaka, 1954), *C. maxima* and *C. medica* were geographically separated, with a more intertropical specialization for *C. maxima*.

Eremocitrus and *Microcitrus* were found to be associated in all our analyses. This result is consistent with the conclusions of Barrett and Rhodes (1976), based on morphological traits, and also with previous molecular phylogenetic analysis (e.g. Bayer *et al*., 2009). The phylogenetic placement of these Australian genera within the 'true citrus fruit trees' remains unclear, due to the lack of branch support for the deeper branches in the phylogenetic trees.

Secondary species structure

The origin of secondary species and many recent hybrids formed by interspecific hybridisation between the basic *Citrus* taxa (*C. maxima*, *C. reticulata*, *C. medica* and *C. micrantha*) has been well documented in several molecular studies (Nicolosi *et al*., 2000; Barkley *et al*., 2006; Garcia-Lor *et al*., 2012a; Ollitrault *et al*., 2012a), and the relative contribution of the ancestral taxa to their genomes was estimated by Barkley *et al*. (2006) and Garcia-Lor *et al*. (2012a). However, these two studies were based on SSRs and these estimations could be biased by the frequent homoplasy observed for these markers (Barkley *et al.*, 2009). The genomes of secondary species can be considered to be mosaics of large DNA fragments of ancestral species that resulted from a few interspecific recombination events (Garcia-Lor *et al*., 2012a). However, the phylogenetic structures of secondary species in concrete points of the genome remain obscure. For *C. sinensis*, *C. aurantium*, *C. paradisi* and clementine, previous molecular studies (Nicolosi *et al*., 2000; Barkley *et al*., 2006; Garcia-Lor *et al*., 2012a; Ollitrault *et al*., 2012a) also showed that intra-taxon diversity resulted only from mutation and/or epigenetic variation without further sexual recombination events. Therefore, these species generally present very low or null molecular intercultivar diversity in genetic markers such as SSRs or SNPs. Such low molecular diversity was confirmed in this work for secondary taxa for which two cultivars were sequenced (*C. sinensis*, *C. aurantium* and Clementine). Due to this intra-secondary taxon diversification history, most of the conclusions about the mosaic structure inferred from one or two genotypes should be extended to other cultivars of the same secondary species.

Clementine is believed to have resulted from a cross between mandarin 'Willowleaf and sweet orange (Nicolosi *et al*., 2000; Ollitrault *et al*., 2012a), which means that there were contributions from both the *C. reticulata* and *C. maxima* gene pools (Garcia-Lor *et al*., 2012a). From the analysis of 27 genes, the observation that there was a majority of mandarin/mandarin phylogenetic homozygosity and very little mandarin/pummelo heterozygosity is in agreement with this hypothesis. The proportion of the pummelo genome estimated from these 27 sequences (16.7%) is higher than the one estimated from SSR markers (7%) by Garcia-Lor *et al*. (2012a).

Several hypotheses have been proposed for the origin of *C. sinensis*. According to Barrett and Rhodes (1976), Torres *et al*. (1978), Scora (1988), Nicolosi *et al*. (2000) and Moore (2001), sweet orange should be a direct interspecific hybrid between a pummelo (*C. maxima*) and a mandarin (*C. reticulata*), whereas Roose *et al*. (2009) and Garcia-Lor *et al*. (2012a)

suggested that *C. sinensis* resulted from a backcross 1 (BC1) [(*C. maxima* x *C. reticulata*) x *C. reticulata*]. The identification of interspecific phylogenetic heterozygosity MP and phylogenetic homozygosity PP and MM (Table 6) in the *C. sinensis* genome contradicts these two models. Indeed, the presence of both types of phylogenetic homozygosity (reported for the first time for pummelo homozygosity) implies that both parents of sweet orange were of interspecific origin. The presence of intraspecific heterozygous SNPs for some genes in phylogenetic homozygosity (EMA and HYB; data not shown) also contradicts the hypothesis that *C. sinensis* resulted from an F_2 interspecific hybrid (self-fecundation of an interspecific F_1).

Sour orange (*C. aurantium*) is thought by some authors to be a natural hybrid of a mandarin and a pummelo (Scora, 1975; Barrett and Rhodes, 1976; Nicolosi *et al*., 2000; Uzun *et al*., 2009). The interspecific heterozygosity (MP, Table 6) observed for all interpretable loci is in agreement with this hypothesis. However, specific SNP alleles were found in *C. aurantium*, indicating that the parental pummelo or mandarin was not part of the germplasm analysed and that sweet orange and sour orange were not related as considered by some authors.

Grapefruit (*C. paradisi*) is thought to have arisen from a natural hybridization between *C. maxima* and *C. sinensis* in the Caribbean after the discovery of the New World by Christopher Columbus (Barrett and Rhodes, 1976, de Moraes *et al*., 2007, Ollitrault *et al*., 2012a). The results obtained in this study help to confirm this theory, as many loci were homozygous for the *C. maxima* genome and other loci showed interspecific heterozygosity (MP, Table 6). Nicolosi *et al*. (2000) proposed that Mexican lime (*C. aurantifolia*) is a hybrid between *C. medica* and *C. micrantha*. This theory fits with our data for 23 out of 27 genes. For, three genes, it was not possible to decipher the mosaic structure and for the gene leading to a CC conclusion it should be supposed that PCR competition resulted in an apparent *Papeda* null allele (C0). The tri-hybrid origin (*C. medica*, *C. reticulata*, *C. maxima*) accepted for *C. limon* (Nicolosi *et al*., 2000; Barkley *et al*., 2006; Garcia-Lor *et al*., 2012a) was confirmed by our sequence data for the lemon cultivar 'Eureka', which has contributions from its ancestors (*C. medica*: 50%, *C. reticulata* 38.46% and *C. maxima* 11.54%, Table 6) that are similar to those described by Garcia-Lor *et al*. (2012a). Moreover, the systematic presence of a *C. medica* allele, and the fact that lemon shares heterozygosity with some rare sour orange alleles support the hypothesis proposed by Nicolosi *et al.* (2000) that lemon resulted from a direct hybridisation between *C. medica* and *C. aurantium*.

Both tangors (*C. reticulata* x *C. sinensis*) and tangelos (*C. paradisi* x *C. reticulata*) were bred from recombination between the *C. reticulata* and *C. maxima* gene pools. The SNP pattern for tangelo 'Orlando', originated from a controlled cross between a grapefruit and a 'Dancy' mandarin (Hodgson, 1967), with both mandarin and pummelo allele inheritance is logical. Our results also confirm that the tangor 'King' classified by Tanaka (1977) as *C. nobilis* is most probably a tangor with at least one mandarin allele for each gene and MP heterozygosity inheritance for some genes.

With the next release of the pseudo-chromosome sequence assembly of the reference haploid clementine genome (Gmitter, 2012), the assignation of the phylogenetic origin of these 27 genes will contribute to the deciphering of the interspecific mosaic genome structure of the secondary species. Moreover, this allelic assignation in genotypes of interspecific origin, coupled with further analysis of functionality of the alleles of the different ancestral species, will provide a very promising pathway for understanding the molecular basis of phenotypic variability in this highly stratified gene pool in which the organization of phenotypic and molecular diversity is closely linked.

CONCLUSION AND PERSPECTIVES

Sanger sequencing of 27 nuclear gene fragments for 45 genotypes resulted in the identification of a great number of molecular polymorphisms (1097 SNPs and 50 indels). For the indels, half of the mined polymorphisms have been used to define new markers. A significant number of the mined SNP loci should be converted into efficient markers to perform high throughput genotyping studies that will be important for the management of *Citrus* collections and marker/trait association studies. The nuclear phylogenetic analyses of *Citrus* and its sexually compatible relatives showed coherence with the geographic distribution and differentiation proposed by Tanaka (1954), with *C. reticulata* and *Fortunella* appearing to be closely related. A cluster that joins *C. medica*, *C. maxima* and the *Papeda* species *C. micrantha* was consistently revealed.

In the near future, by using the entire *Citrus* genome as a reference and resequencing data from the main secondary species, the resulting estimations of the relative levels of within- and between-taxon differentiation will be useful for deciphering the interspecific mosaic structure of the *Citrus* secondary cultivated species and modern cultivars. The present study has allowed us to assign a phylogenetic inheritance of the genes that were examined for most of the genotypes of interspecific origin under study. One of our major results concerns *C. sinensis*, which has alleles of three genes that appear to have been inherited solely from the *C. maxima* gene pool and alleles of eight genes that appear to have been inherited from *C. reticulata*. This result contradicts the hypothesis that *C. sinensis* originated directly from F_1 or by BC1 hybridization between the *C. maxima* and *C. reticulata* gene pools. However, our study confirms previous hypotheses concerning the origins of the other secondary species.

Positive selection was observed for a few genes within or between the species studied, suggesting that these genes may play a key role in phenotypic differentiation. These genes are therefore major candidates for future studies, including complete gene sequencing and functional analysis of different alleles to analyse the molecular basis of the phenotypic variability of corresponding traits.

SUPPLEMENTARY INFORMATION CHAPTER 2

Supplementary Information 1. Genotypes used in this study

Group	Scientific name (Swingle)	Scientific name (Tanaka)	Cultivar	Ref.*
Mandarin	*C.reticulata* var. austera	*C. reshni* Hort. ex Tan.	Cleopatra	385I
Mandarin	*C.reticulata* Blanco	*C. deliciosa* Ten	Willow leaf	154I
Mandarin	*C.reticulata* Blanco	*C.reticulata* Blanco	Ponkan	482I
Mandarin	*C.reticulata* var. austera	*C. sunki* Hort. ex Tan.	Sunki	239I
Mandarin	*C.reticulata* Blanco	*C. tangerina* Hort. ex Tan.	Dancy	434I
Mandarin	*C.reticulata* Blanco	*C unshiu* (Mak.) Marc.	Clausellina	19I
Mandarin	*C.reticulata* Blanco	*C. deliciosa* Ten	Avana apireno	189I
Pummelo	*C. maxima* (Burm.) Merr.	*C. maxima* (Burm.) Merr.	Chandler	207I
Pummelo	*C. maxima* (Burm.) Merr.	*C. maxima* (Burm.) Merr.	Pink	275I
Pummelo	*C. maxima* (Burm.) Merr.	*C. maxima* (Burm.) Merr.	Nam Roi	590I
Pummelo	*C. maxima* (Burm.) Merr.	*C. maxima* (Burm.) Merr.	Tahiti	727C
Pummelo	*C. maxima* (Burm.) Merr.	*C. maxima* (Burm.) Merr.	Sans Pepins	710C
Citron	*C. medica* L.	*C. medica* L.	Corsica	567I
Citron	*C. medica* L.	*C. medica* L.	Buddha's hand	202I
Citron	*C. medica* L.	*C. medica* L.	Diamante	560I
Citron	*C. medica* L.	*C. medica* L.	Arizona	169I
Citron	*C. medica* L.	*C. medica* L.	Poncire commun	701C
Papeda	*C. micrantha* Wester	*C. micrantha* Wester	Small flowered papeda	626I
Papeda	*C. hystrix* DC.	*C. hystrix* DC.	Mauritius papeda	178I
Papeda	*C. ichangensis* Swing.	*C. ichangensis* Swing.	Ichang papeda	358I
Papeda	*C. macroptera* Montr.	*C. macroptera* Montr.	Melanesian papeda	279I
Fortunella	*F. hindsii* (Champ.) Swing.	*F. hindsii* (Champ.) Swing.	Hong Kong kumkuat	281I
Fortunella	*Fortunella* hybrid	*F. crassifolia* Swing.	Meiwa kumkuat	280I
Fortunella	*F. japonica* (Thunb.) Swing.	*F. japonica* (Thunb.) Swing.	Round kumkuat	381I
Fortunella	*F. polyandra* (Ridl.) Tan	*F. polyandra* (Ridl.) Tan.	Malayan kumquat	375I
Fortunella	*F. margarita* (Lour.) Swing.	*F. margarita* (Lour.) Swing.	Nagami kumkuat	38I
Microcitrus	*Microcitrus australasica* (F. Muell.) Swing.	*Microcitrus australasica* (F. Muell.) Swing.	Australian finger lime	150I
Microcitrus	*Microcitrus australis* (F. Muell.) Swing.	*Microcitrus australis* (F. Muell.) Swing.	Australian round lime	313I
Eremocitrus	*Eremocitrus glauca* (Lindl.)	*Eremocitrus glauca* (Lindl.)	Australian desert lime	346I
Poncirus	*Poncirus trifoliata* (L.) Raf.	*Poncirus trifoliata* (L.) Raf.	Pomeroy	374I
Poncirus	*Poncirus trifoliata* (L.) Raf.	*Poncirus trifoliata* (L.) Raf.	Rubidoux	217I
Poncirus	*Poncirus trifoliata* (L.) Raf.	*Poncirus trifoliata* (L.) Raf.	Flying dragon	537I
Haploid	*C.reticulata* Blanco	*C. clementina* Hort. ex Tan.	Haploid	HapClem
Clementine	*C. reticulata* Blanco	*C. clementina* Hort. ex Tan.	Clemenules	22I
Clementine	*C. reticulata* Blanco	*C. clementina* Hort. ex Tan.	Arrufatina	58I
Tangelo	*C. reticulata* x *C. paradisi*	*C. reticulata* x *C. paradisi*	Orlando	101I
Tangor	*C. reticulata* x *C. sinensis*	*C. nobilis* Lour.	King	477I
Sweet orange	*C. sinensis* (L.) Osb	*C. sinensis* (L.) Osb	Valencia Late Delta	363I
Sweet orange	*C. sinensis* (L.) Osb	*C. sinensis*(L.) Osb	Salustiana	125I
Grapefruit	*C. paradisi* Macf	*C. paradisi* Macf	Marsh	176I
Sour orange	*C. aurantium* L.	*C. aurantium* L.	Sevillano	117I
Sour orange	*C. aurantium* L.	*C. aurantium* L.	Bouquet de Fleurs	139I
Lime	*C. aurantifolia* (Christm.) Swing.	*C. aurantifolia* (Christm.) Swing.	Mexican	164I
Lemon	*C. limon* (L.) Burm	*C. limon* (L.) Burm	Eureka	297I
Severinia	*Severinia buxifolia* (Poir.) Tenore	*Severinia buxifolia* (Poir.) Tenore	Chinese box orange	147I

*(I) IVIA germplasm; (C) INRA/CIRAD germplasm.

Supplementary Information 2. New InDel primers developed from polymorphisms found during sequencing of the candidate genes

PRIMER	GBA	SEQUENCE	Lenght	Tm	PCR	Product Size
IDCHI2	DY263683	F:AATCAATTATTTTCCACATT	20	48.91	50	94-96
		R:ATTACACGTAACGCAAGA	18	53.2		
IDFLS1	AB011796	F:GATCATCTCTTCCACAGG	18	50.64	50	144-158
		R:GAAAATAAATTATTTATACATTTTGTTT	28	52.86		
IDFLS2	AB011796	F:AAACAAAATGTATAAATAATTTATTTTC	28	52.86	50	184-204
		R:AGCATGTACTCAATGTCG	18	49.76		
IDF3'H1	HQ634392	F:AAAGGCTCACCATCACCAAC	20	59.97	55	180-196
		R:AAAATGAACAACACAAAGAAAGACC	25	55.2		
IDDFR1	DQ084722	F:CCACGCCTATGGACTTTGAG	20	60.65	55	181-192
		R:TCAATGTTATGCGGCTGTTC	20	59.69		
IDDFR2	DQ084722	F:ACTGTTCGCGATCCTGGT	18	59.21	55	140-156
		R:GCAACTCCAGCAAATGTTTC	20	58.35		
IDINVA1	AB074885	F:GAGCTCCCCTTTTGCTTAAT	20	57.58	55	218-220
		R:AGTAGCTGAGCCAACATCAA	20	56.09		
IDINVA2	AB074885	F:CCTTCTGGTTCTTGCAGAT	19	55.35	55	233-237
		R:TATTGACATCATTTGCCTCA	20	55.01		
IDINVA3	AB074885	F:TTCTGAGGCAAATGATGTCAA	21	59.26	55	203-206
		R:CGAATGATCCACCTGCAAAT	20	60.86		
IDPEPC3	EF058158	F:TTTGTGATGTTCCACAAATG	20	55.3	55	130-133
		R:CTACCATTAGCCGATTGTTC	20	54.93		
IDPFK1	AF095520	F:AAAACCCTTTCAAAATCGTC	20	55.85	55	246-248
		R:CCGATTTTCAACTTCTCATC	20	54.84		
IDPSY2	AB037975	F:TTGAGTCATGCCATTTTTGC	20	59.67	55	347-364
		R:ATTGGGTTAAGGGTCCACTG	20	58.76		

(GBA) Genebank accession.

Supplementary Information 3. Nucleotide diversity and divergence for each gene and taxa. Mand (*C. reticulata*), Pum (*C. maxima*), Cit (*C. medica*), For (*Fortunella*), Pap (Papeda, wild citrus), Mic (*Microcitrus*), Ere (*Eremocitrus*), Pon (*Poncirus trifoliata*), AncTaxa (*C. reticulata*, *C. maxima*, *C. medica*, wild citrus). (Pop) Population, (S) Segregating sites, (π_T) Total nucleotide diversity, (π_{sil}) Nucleotide diversity silent sites, (π_{syn}) Nucleotide diversity synonymous sites, (**π_{nonsyn}**) Nucleotide diversity nonsynonymous sites, ($\pi_{nonsyn/syn}$) Ratio Nucleotide diversity nonsynonymous/synonymous sites, (**π_{nonsyn}/π_{sil}**) Ratio Nucleotide diversity nonsynonymous/silent sites, (D_{tajima}) Tajima's D neutrality test, (N_h) Number of haplotypes, (H_e) Haplotype diversity, (SD) Standard deviation, (F_{st}) Wright's differentiation index. See Table 1 for locus abbreviations

		Polymorphism							Haplotype diversity			
Locus	**Taxa**	**S**	**π_T**	**π_{sil}**	**π_{syn}**	**π_{nonsyn}**	**π_{nonsyn}/π_{syn}**	**π_{nonsyn}/π_{sil}**	**D_{tajima}**	**N_h**	**H_e**	**(SD)**	**F_{st}**
CHI	Mand	20	0.009	0.010	0.011	0.006	0.522	0.541		7	0.833	0.072	
	Pum	6	0.003	0.002	0.000	0.007	-	3.036		5	0.756	0.130	
	Cit	21	0.013	0.017	0.007	0.002	0.320	0.136		3	0.622	0.138	
	For	12	0.007	0.007	0.000	0.006	-	0.848		5	0.667	0.163	
	Pap	22	0.014	0.016	0.000	0.007	-	0.425		5	0.933	0.122	
	Mic	8	0.008	0.010	0.000	0.004	-	0.448		2	0.667	0.204	
	Ere	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Pon	1	0.001	0.001	0.000	0.000	-	0.000		2	0.600	0.129	
	AncTaxa	73	0.026	0.032	0.008	0.011	1.381	0.361	-0.300	28	0.959	0.010	0.757
	Whole Pop	76	0.024	0.030	0.008	0.010	1.377	0.350	-0.407	34	0.958	0.008	
	SD		0.001										
CHS	Mand	1	0.001	0.003	0.003	0.000	0.000	0.000		2	0.440	0.112	
	Pum	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	4	0.001	0.005	0.005	0.000	0.096	0.096		3	0.378	0.181	
	Pap	5	0.004	0.009	0.009	0.002	0.207	0.207		5	0.933	0.122	
	Mic	4	0.003	0.008	0.008	0.002	0.289	0.289		2	0.500	0.265	
	Ere	1	0.002	0.008	0.008	0.000	0.000	0.000		2	1.000	0.500	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	20	0.004	0.016	0.016	0.001	0.079	0.079	-1.256	14	0.885	0.015	0.698
	Whole Pop	20	0.004	0.015	0.014	0.001	0.065	0.064	-1.258	15	0.857	0.018	
	SD		0.000										
FLS	Mand	14	0.008	0.015	0.019	0.002	0.442	0.145		7	0.817	0.073	
	Pum	4	0.004	0.007	0.005	0.003	0.567	0.372		5	0.844	0.080	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	7	0.007	0.017	0.018	0.003	0.151	0.167		6	0.889	0.075	
	Pap	14	0.014	0.040	0.033	0.004	0.137	0.113		6	1.000	0.096	
	Mic	8	0.010	0.025	0.030	0.004	0.121	0.142		4	1.000	0.177	
	Ere	10	0.022	0.046	0.064	0.012	0.191	0.267		2	1.000	0.500	
	Pon	4	0.004	0.008	0.011	0.003	0.240	0.332		5	0.933	0.122	
	AncTaxa	45	0.021	0.055	0.059	0.007	0.122	0.131	-0.329	34	0.958	0.014	0.608
	Whole Pop	47	0.020	0.055	0.062	0.007	0.120	0.135	-0.211	48	0.960	0.011	
	SD		0.001										
F3'H	Mand	10	0.004	0.003	0.003	0.005	1.767	1.714		6	0.747	0.111	
	Pum	4	0.002	0.001	0.000	0.004	-	6.066		4	0.733	0.101	
	Cit	6	0.003	0.005	0.000	0.002	-	0.383		4	0.778	0.091	
	For	21	0.012	0.011	0.013	0.014	1.048	1.312		7	0.933	0.062	
	Pap	10	0.006	0.005	0.010	0.006	0.619	1.136		6	1.000	0.096	
	Mic	10	0.007	0.012	0.019	0.006	0.327	0.534		3	0.833	0.222	
	Ere	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Pon	1	0.001	0.002	0.000	0.000	-	0.000		2	0.600	0.129	
	AncTaxa	55	0.010	0.013	0.016	0.008	0.515	0.612	-1.285	33	0.970	0.009	0.574
	Whole Pop	60	0.009	0.012	0.014	0.008	0.554	0.643	-1.474	36	0.949	0.013	
	SD		0.001										
DFR	Mand	1	0.000	0.000	0.000	0.001	-	-		2	0.143	0.119	
	Pum	1	0.000	0.001	0.000	0.000	-	0.000		2	0.200	0.154	
	Cit	4	0.004	0.006	0.025	0.000	0.000	0.000		5	0.822	0.097	
	For	7	0.007	0.010	0.006	0.000	0.000	0.000		6	0.867	0.085	
	Pap	7	0.008	0.012	0.009	0.000	0.000	0.000		5	0.933	0.122	
	Mic	4	0.005	0.008	0.000	0.000	-	0.000		3	0.833	0.222	
	Ere	4	0.010	0.011	0.000	0.007	-	0.660		2	1.000	0.500	
	Pon	3	0.004	0.005	0.000	0.000	-	0.000		3	0.733	0.155	
	AncTaxa	32	0.013	0.017	0.024	0.005	0.194	0.280	-0.893	26	0.928	0.020	0.675
	Whole Pop	32	0.011	0.015	0.019	0.005	0.248	0.308	-0.949	34	0.898	0.026	
	SD		0.001										
EMA	Mand	3	0.004	0.005	0.000	0.000	-	0.000		3	0.667	0.075	
	Pum	1	0.000	0.001	0.000	0.000	-	0.000		2	0.200	0.154	
	Cit	1	0.001	0.002	0.000	0.000	-	0.000		2	0.533	0.095	
	For	10	0.007	0.008	0.006	0.002	0.308	0.246		7	0.911	0.077	
	Pap	2	0.007	0.000	0.011	0.003	0.255	-		3	0.733	0.155	

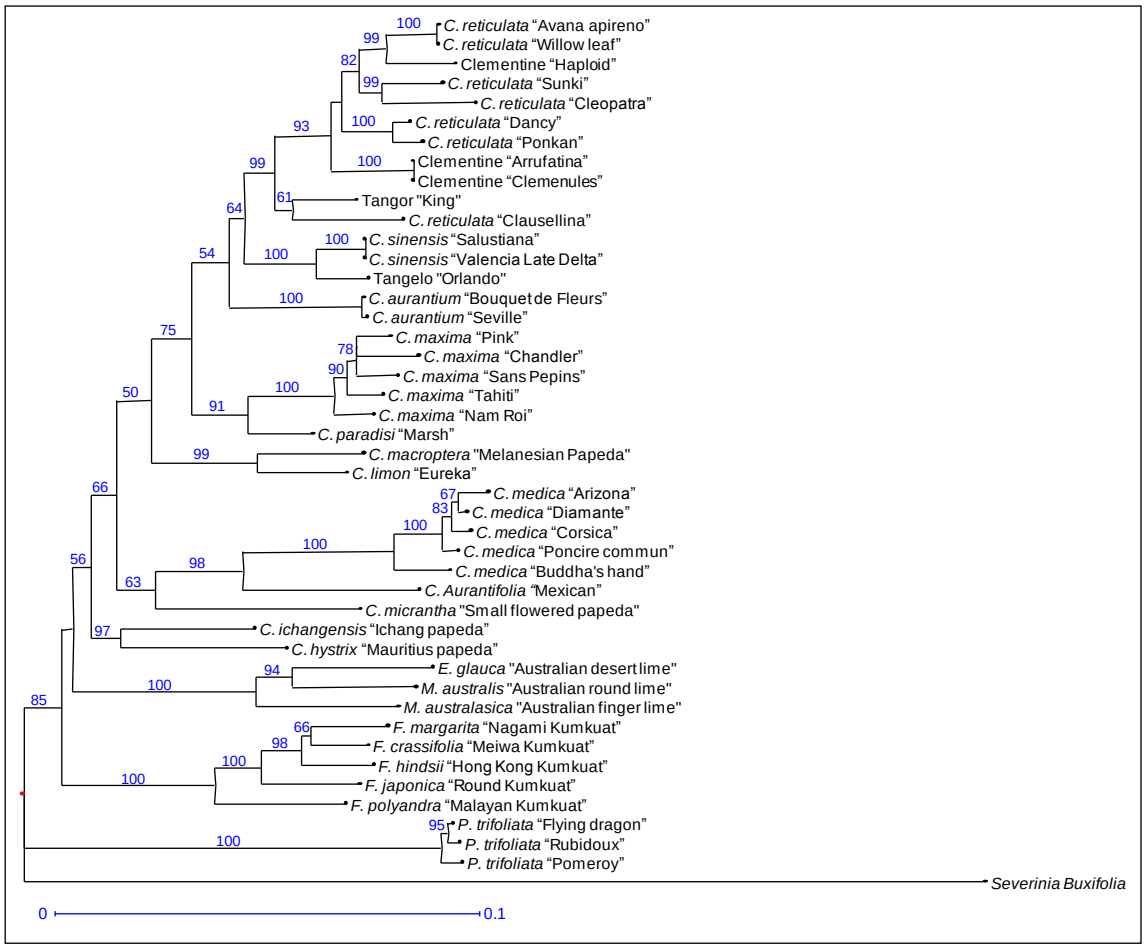
	Mic	8	0.010	0.013	0.016	0.000	0.000	0.000		3	0.833	0.222	
	Ere	3	0.007	0.010	0.032	0.000	0.000	0.000		2	1.000	0.500	
	Pon	5	0.005	0.007	0.000	0.000	-	0.000		3	0.600	0.215	
	AncTaxa	34	0.014	0.017	0.003	0.005	1.511	0.289	-0.732	21	0.921	0.018	0.677
	Whole Pop	34	0.013	0.016	0.002	0.005	2.273	0.323	-0.622	27	0.914	(0.016)	
	SD		0.001										
MDH	Mand	6	0.002	0.004	0.004	0.001	0.179	0.179		3	0.908	0.115	
	Pum	2	0.002	0.000	0.000	0.002	-	-		2	0.538	0.075	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.556	0.000	
	For	10	0.005	0.011	0.011	0.004	0.338	0.338		7	0.000	0.077	
	Pap	6	0.004	0.007	0.007	0.003	0.418	0.418		4	0.867	0.129	
	Mic	5	0.004	0.011	0.011	0.002	0.235	0.235		3	0.821	0.222	
	Ere	4	0.006	0.000	0.000	0.007	-	-		2	0.833	0.500	
	Pon	1	0.001	0.000	0.000	0.001	-	-		2	1.000	0.129	
	AncTaxa	31	0.007	0.014	0.014	0.005	0.356	0.356	-0.768	21	0.929	0.015	0.677
	Whole Pop	31	0.006	0.012	0.012	0.005	1.065	0.394	-0.835	23	0.600	0.014	
	SD		0.000										
ACO	Mand	3	0.001	0.001	0.000	0.001	-	0.778		3	0.473	0.136	
	Pum	9	0.007	0.010	0.008	0.000	0.000	0.000		2	0.556	0.075	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	7	0.004	0.005	0.000	0.000	-	0.000		7	0.911	0.077	
	Pap	6	0.004	0.006	0.000	0.000	-	0.000		3	0.733	0.155	
	Mic	14	0.012	0.016	0.018	0.000	0.000	0.000		4	0.031	0.177	
	Ere	6	0.009	0.012	0.031	0.000	0.000	0.000		2	1.000	0.500	
	Pon	4	0.003	0.004	0.000	0.000	-	0.000		4	0.867	0.129	
	AncTaxa	44	0.009	0.012	0.006	0.000	0.030	0.014	-1.184	26	0.935	0.016	0.554
	Whole Pop	45	0.008	0.011	0.005	0.000	0.024	0.010	-1.173	30	0.908	0.019	
	SD		0.001										
TRPA	Mand	14	0.005	0.008	0.011	0.003	0.286	0.423		3	0.385	0.149	
	Pum	5	0.001	0.002	0.003	0.001	0.327	0.414		3	0.378	0.181	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	8	0.003	0.005	0.002	0.002	0.779	0.341		7	0.911	0.077	
	Pap	10	0.005	0.017	0.014	0.006	0.430	0.355		5	0.933	0.122	
	Mic	15	0.010	0.012	0.017	0.008	0.486	0.682		4	1.000	0.177	
	Ere	7	0.009	0.013	0.025	0.006	0.242	0.469		2	1.000	0.500	
	Pon	4	0.002	0.006	0.004	0.000	0.000	0.000		3	0.733	0.155	
	AncTaxa	56	0.014	0.020	0.025	0.010	0.411	0.514	-0.705	26	0.918	0.019	0.688
	Whole Pop	57	0.012	0.018	0.022	0.009	0.430	0.518	-0.745	32	0.883	0.022	
	SD		0.001										
INVA	Mand	22	0.008	0.010	0.022	0.004	0.190	0.401		8	0.867	0.060	
	Pum	11	0.004	0.004	0.010	0.004	0.397	0.848		6	0.778	0.137	
	Cit	6	0.002	0.003	0.003	0.002	0.614	0.656		3	0.622	0.138	
	For	9	0.004	0.006	0.015	0.001	0.091	0.226		5	0.800	0.100	
	Pap	22	0.012	0.019	0.028	0.004	0.148	0.219		6	1.000	0.096	
	Mic	13	0.008	0.011	0.017	0.003	0.177	0.268		4	1.000	0.177	
	Ere	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Ponc	2	0.001	0.001	0.004	0.002	0.345	1.462		2	0.733	0.155	
	AncTaxa	69	0.014	0.017	0.039	0.008	0.197	0.449	-0.640	34	0.970	0.009	0.593
	Whole Pop	72	0.014	0.019	0.031	0.007	0.226	0.368	-0.514	49	0.975	0.006	
	SD		0.000										
PEPC	Mand	3	0.001	0.002	0.000	0.000	-	0.000		3	0.001	0.138	
	Pum	3	0.001	0.002	0.000	0.000	-	0.000		4	0.001	0.152	
	Cit	1	0.001	0.001	0.000	0.000	-	0.000		2	0.001	0.095	
	For	12	0.006	0.006	0.000	0.000	-	0.000		4	0.006	0.101	
	Pap	5	0.003	0.004	0.000	0.000	-	0.000		4	0.800	0.172	
	Mic	10	0.008	0.008	0.033	0.000	0.000	0.000		3	0.833	0.222	
	Ere	4	0.006	0.006	0.066	0.000	0.000	0.000		2	1.000	0.500	
	Ponc	5	0.004	0.004	0.000	0.000	-	0.000		3	0.733	0.155	
	AncTaxa	52	0.014	0.015	0.016	0.000	0.000	0.000	-0.821	25	0.945	0.013	0.714
	Whole Pop	53	0.013	0.014	0.012	0.000	0.000	0.000	-0.757	33	0.950	0.011	
	SD		0.000										
PKF	Mand	9	0.003	0.004	0.003	0.003	0.914	0.681		4	0.659	0.120	
	Pum	1	0.000	0.000	0.000	0.000	-	0.000		2	0.200	0.154	
	Cit	2	0.001	0.000	0.000	0.002	-	-		2	0.356	0.159	
	For	6	0.002	0.003	0.000	0.000	-	0.000		3	0.378	0.181	
	Pap	12	0.008	0.010	0.000	0.005	-	0.517		4	0.867	0.129	
	Mic	2	0.002	0.000	0.000	0.004	-	-		3	0.833	0.222	
	Ere	4	0.005	0.007	0.000	0.003	-	0.492		2	1.000	0.500	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	38	0.009	0.012	0.005	0.006	1.072	0.491	-0.855	21	0.925	0.014	0.647
	Whole Pop	44	0.009	0.011	0.006	0.006	0.883	0.514	-1.058	27	0.937	0.010	
	SD		0.000										
DXS	Mand	14	0.006	0.009	0.009	0.002	0.227	0.228		6	0.767	0.084	
	Pum	4	0.003	0.004	0.007	0.000	0.000	0.000		4	0.733	0.101	
	Cit	2	0.001	0.001	0.006	0.001	0.235	1.219		3	0.689	0.104	

	For	7	0.004	0.006	0.005	0.001	0.301	0.230		7	0.911	0.077	
	Pap	17	0.012	0.016	0.015	0.007	0.466	0.434		6	1.000	0.096	
	Mic	4	0.003	0.004	0.006	0.002	0.304	0.534		3	0.833	0.222	
	Ere	7	0.010	0.013	0.026	0.004	0.153	0.296		2	1.000	0.500	
	Pon	4	0.003	0.005	0.000	0.000	-	0.000		2	0.533	0.172	
	AncTaxa	51	0.015	0.021	0.018	0.006	0.348	0.289	-0.321	32	0.967	0.009	0.659
	Whole Pop	52	0.014	0.020	0.019	0.005	0.285	0.271	-0.394	39	0.947	0.014	
	SD		0.001										
PSY	Mand	15	0.010	0.012	0.012	0.010	0.886	0.886		8	0.850	0.075	
	Pum	3	0.002	0.000	0.000	0.003	-	-		3	0.711	0.086	
	Cit	3	0.002	0.004	0.004	0.002	0.495	0.495		2	0.467	0.132	
	For	6	0.005	0.009	0.009	0.004	0.443	0.443		4	0.778	0.091	
	Pap	8	0.007	0.008	0.025	0.000	0.000	0.000		4	0.867	0.129	
	Mic	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Ere	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Pon	3	0.003	0.010	0.010	0.001	0.119	0.119		3	0.733	0.155	
	AncTaxa	45	0.013	0.014	0.003	0.011	3.533	0.775	-0.646	23	0.953	0.010	0.707
	Whole Pop	45	0.013	0.026	0.026	0.010	0.389	0.389	-0.526	28	0.950	0.011	
	SD		0.001										
HYB	Mand	10	0.006	0.004	0.005	0.008	1.421	1.704		5	0.780	0.085	
	Pum	4	0.002	0.002	0.002	0.001	0.534	0.633		4	0.733	0.120	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	25	0.013	0.014	0.000	0.013	-	0.918		10	1.000	0.045	
	Pap	5	0.003	0.001	0.000	0.006	-	4.657		4	0.800	0.172	
	Mic	7	0.007	0.010	0.000	0.002	-	0.221		4	1.000	0.177	
	Ere	4	0.006	0.005	0.000	0.007	-	1.337		2	1.000	0.500	
	Pon	4	0.003	0.002	0.000	0.004	-	2.697		3	0.733	0.155	
	AncTaxa	52	0.014	0.015	0.018	0.012	0.669	0.795	-0.821	25	0.945	0.013	0.624
	Whole Pop	47	0.012	0.015	0.010	0.009	0.914	0.618	-0.368	37	0.956	0.009	
	SD		0.001										
LCY2	Mand	26	0.016	0.026	0.026	0.011	0.413	0.413		7	0.857	0.065	
	Pum	2	0.001	0.004	0.004	0.000	0.000	0.000		3	0.378	0.181	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	27	0.014	0.028	0.028	0.007	0.263	0.263		9	0.978	0.054	
	Pap	11	0.008	0.017	0.017	0.004	0.256	0.256		4	0.867	0.129	
	Mic	5	0.004	0.007	0.007	0.003	0.361	0.361		3	0.833	0.222	
	Ere	5	0.007	0.016	0.016	0.004	0.276	0.276		2	1.000	0.500	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	64	0.014	0.030	0.030	0.007	0.246	0.246	-0.956	25	0.938	0.014	0.522
	Whole Pop	65	0.014	0.029	0.029	0.008	0.274	0.274	-0.794	31	0.940	0.011	
	SD		0.001										
LCYB	Mand	11	0.002	0.006	0.006	0.000	0.062	0.062		2	0.143	0.119	
	Pum	6	0.003	0.007	0.007	0.002	0.298	0.298		3	0.600	0.131	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	2	0.000	0.001	0.001	0.000	0.281	0.281		2	0.200	0.154	
	Pap	14	0.007	0.026	0.026	0.002	0.083	0.083		6	1.000	0.096	
	Mic	5	0.003	0.005	0.005	0.002	0.430	0.430		3	0.833	0.222	
	Ere	5	0.005	0.019	0.019	0.001	0.071	0.071		2	1.000	0.500	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	37	0.009	0.028	0.028	0.004	0.141	0.141	0.253	20	0.898	0.019	0.723
	Whole Pop	37	0.009	0.028	0.028	0.004	0.134	0.134	0.433	24	0.898	0.018	
	SD		0.000										
NCED3	Mand	8	0.003	0.003	0.003	0.003	0.761	0.761		4	0.659	0.090	
	Pum	2	0.002	0.003	0.003	0.001	0.461	0.461		3	0.711	0.086	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	1	0.001	0.000	0.000	0.001	-	-		2	0.556	0.075	
	Pap	7	0.005	0.012	0.012	0.003	0.293	0.293		5	0.933	0.122	
	Mic	2	0.002	0.004	0.004	0.001	0.311	0.311		3	0.833	0.222	
	Ere	3	0.005	0.008	0.008	0.005	0.622	0.622		2	1.000	0.500	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	21	0.008	0.015	0.015	0.006		0.391	-0.096	16	0.922	0.012	0.715
	Whole Pop	21	0.008	0.015	0.015	0.006	0.387	0.387	0.231	21	0.929	0.010	
	SD		0.000										
AOC	Mand	9	0.007	0.021	0.029	0.001	0.039	0.054		4	0.736	0.075	
	Pum	1	0.001	0.002	0.003	0.000	0.000	0.000		2	0.467	0.132	
	Cit	1	0.000	0.000	0.000	0.000	-	-		2	0.200	0.154	
	For	5	0.002	0.005	0.007	0.001	0.155	0.210		5	0.800	0.100	
	Pap	1	0.001	0.003	0.004	0.000	0.000	0.000		2	0.533	0.172	
	Mic	9	0.007	0.013	0.017	0.005	0.312	0.423		3	0.833	0.222	
	Ere	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	35	0.010	0.024	0.032	0.004	0.133	0.181	0.851	16	0.909	0.018	0.785
	Whole Pop	37	0.010	0.024	0.032	0.004	0.116	0.158	-0.390	18	0.894	0.017	
	SD		0.001										
MRP4	Mand	2	0.001	0.001	0.000	0.000	-	0.309		3	0.385	0.149	

	Pum	1	0.001	0.002	0.004	0.000	0.000	0.000		2	0.467	0.132	
	Cit	1	0.000	0.000	0.000	0.001	-	-		2	0.356	0.159	
	For	3	0.001	0.003	0.000	0.001	-	0.281		4	0.800	0.089	
	Pap	7	0.003	0.004	0.004	0.002	-	0.485		4	0.867	0.129	
	Mic	6	0.004	0.000	0.000	0.006	-	-		3	0.833	0.222	
	Ere	6	0.008	0.000	0.000	0.012	-	-		2	1.000	0.500	
	Pon	1	0.001	0.002	0.005	0.000	0.000	0.000		2	0.600	0.129	
	AncTaxa	35	0.008	0.010	0.010	0.003	0.323	0.314	-0.685	19	0.925	0.016	0.769
	Whole Pop	40	0.007	0.010	0.009	0.002	0.286	0.239	-0.946	25	0.897	0.022	
	SD		0.000										
CCC1	Mand	12	0.005	0.019	0.019	0.001	0.033	0.033		8	0.890	0.060	
	Pum	2	0.001	0.000	0.000	0.001	-	-		3	0.378	0.181	
	Cit	1	0.000	0.001	0.001	0.000	0.000	0.000		2	0.200	0.154	
	For	9	0.004	0.018	0.018	0.000	0.018	0.018		6	0.778	0.137	
	Pap	21	0.012	0.042	0.042	0.004	0.091	0.091		4	0.867	0.129	
	Mic	1	0.001	0.004	0.004	0.000	0.000	0.000		2	0.667	0.204	
	Ere	6	0.008	0.018	0.018	0.005	0.288	0.288		2	1.000	0.500	
	Pon	1	0.001	0.003	0.003	0.000	0.000	0.000		2	0.533	0.172	
	AncTaxa	33	0.007	0.029	0.029	0.001	0.048	0.048	-1.144	25	0.930	0.017	0.630
	Whole Pop	39	0.008	0.028	0.028	0.002	0.059	0.059	-0.814	32	0.927	0.015	
	SD		0.001										
HKT1	Mand	8	0.013	0.018	0.070	0.006	0.081	0.318		8	0.912	0.049	
	Pum	8	0.018	0.024	0.019	0.000	0.000	0.000		4	0.778	0.091	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	5	0.009	0.012	0.028	0.004	0.145	0.346		4	0.822	0.072	
	Pap	1	0.002	0.004	0.000	0.000	-	0.000		2	0.536	0.123	
	Mic	5	0.011	0.014	0.039	0.006	0.149	0.409		3	0.833	0.222	
	Ere	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	17	0.013	0.018	0.038	0.006	0.158	0.332	-0.906	17	0.918	0.016	0.494
	Whole Pop	18	0.013	0.017	0.037	0.006	0.173	0.386	-0.832	19	0.894	0.018	
	SD		0.001										
LAPX	Mand	3	0.003	0.005	0.008	0.000	0.000	0.000		3	0.473	0.136	
	Pum	5	0.006	0.008	0.023	0.003	0.121	0.345		6	0.889	0.075	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	8	0.014	0.010	0.026	0.018	0.710	1.898		6	0.867	0.085	
	Pap	7	0.013	0.017	0.016	0.007	0.448	0.427		5	0.933	0.122	
	Mic	2	0.004	0.007	0.014	0.000	0.000	0.000		3	0.833	0.222	
	Ere	1	0.004	0.007	0.000	0.000	-	0.000		2	1.000	0.500	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	19	0.011	0.016	0.030	0.006	0.194	0.367	-0.827	19	0.906	0.020	0.588
	Whole Pop	19	0.011	0.015	0.028	0.005	0.190	0.352	-0.755	19	0.882	0.021	
	SD		0.001										
NADK	Mand	1	0.000	0.000	0.000	0.003	-	-		2	0.143	0.119	
	Pum	2	0.001	0.001	0.000	0.000	-	0.000		2	0.200	0.154	
	Cit	6	0.006	0.006	0.025	0.007	0.289	1.133		2	0.356	0.159	
	For	2	0.001	0.001	0.000	0.000	-	0.000		2	0.200	0.154	
	Pap	6	0.009	0.011	0.000	0.000	-	0.000		3	0.733	0.155	
	Mic	5	0.009	0.006	0.000	0.024	-	3.660		3	0.833	0.222	
	Ere	4	0.012	0.011	0.000	0.020	-	1.894		2	1.000	0.500	
	Pon	5	0.009	0.011	0.000	0.000	-	0.000		2	0.600	0.129	
	AncTaxa	28	0.010	0.011	0.004	0.009	2.043	0.856	-1.436	15	0.827	0.033	0.522
	Whole Pop	28	0.009	0.009	0.005	0.010	2.117	1.094	-1.458	15	0.789	0.029	
	SD		0.005										
PIP1	Mand	13	0.013	0.023	0.011	0.000	0.000	0.000		3	0.275	0.148	
	Pum	15	0.028	0.049	0.015	0.000	0.000	0.000		5	0.800	0.100	
	Cit	20	0.037	0.061	0.029	0.003	0.094	0.044		7	0.911	0.077	
	For	5	0.009	0.015	0.000	0.000	-	0.000		3	0.644	0.101	
	Pap	17	0.046	0.082	0.070	0.000	0.000	0.000		6	0.929	0.084	
	Mic	2	0.007	0.012	0.000	0.000	-	0.000		2	0.667	0.204	
	Ere	2	0.011	0.018	0.000	0.000	-	0.000		2	1.000	0.500	
	Pon	15	0.031	0.053	0.038	0.000	0.000	0.000		4	0.800	0.172	
	AncTaxa	26	0.037	0.065	0.029	0.000	0.015	0.007	-0.197	24	0.888	0.032	0.438
	Whole Pop	26	0.036	0.061	0.026	0.000	0.012	0.005	-0.261	27	0.842	0.035	
	SD		0.008										
SOS1	Mand	1	0.001	0.002	0.007	0.000	0.000	0.000		2	0.527	0.064	
	Pum	3	0.002	0.003	0.004	0.001	0.166	0.219		3	0.511	0.164	
	Cit	1	0.001	0.002	0.000	0.000	-	0.000		2	0.533	0.095	
	For	5	0.003	0.008	0.002	0.000	0.000	0.000		5	0.822	0.097	
	Pap	10	0.011	0.017	0.006	0.006	0.895	0.339		5	0.933	0.122	
	Mic	2	0.003	0.006	0.008	0.000	0.000	0.000		2	0.667	0.204	
	Ere	2	0.004	0.009	0.000	0.000	-	0.000		2	1.000	0.500	
	Pon	2	0.002	0.006	0.007	0.000	0.000	0.000		2	0.600	0.129	
	AncTaxa	35	0.014	0.025	0.028	0.006	0.206	0.234	-0.413	21	0.938	0.012	0.814
	Whole Pop	36	0.014	0.025	0.029	0.005	0.184	0.215	-0.294	24	0.917	0.014	

	SD		0.001										
TSC	Mand	2	0.003	0.004	0.000	0.000	-	0.000		3	0.670	0.007	
	Pum	1	0.002	0.000	0.000	0.005	-	-		2	0.533	0.009	
	Cit	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	For	6	0.005	0.007	0.020	0.000	0.000	0.000		4	0.644	0.023	
	Pap	7	0.010	0.011	0.010	0.008	0.795	0.673		6	1.000	0.096	
	Mic	3	0.005	0.007	0.018	0.000	0.000	0.000		3	0.833	0.049	
	Ere	4	0.012	0.018	0.037	0.000	0.000	0.000		2	1.000	0.250	
	Pon	0	0.000	0.000	0.000	0.000	-	-		1	0.000	0.000	
	AncTaxa	24	0.008	0.010	0.009	0.005	0.509	0.488	-1.601	17	0.907	0.016	0.542
	Whole Pop	24	0.007	0.009	0.007	0.004	0.583	0.472	-1.597	18	0.880	0.000	
	SD		0.001										

Supplementary Information 4. NJ tree with all the SNP markers in the whole population studied, ancestral *Citrus* species, relatives, secondary species and interspecific hybrids (1000 bootstraps performed). Branch support over 50% represented



ANNEX CHAPTER 2

Clymenia's phylogeny within the 'true citrus fruit trees'.

Clymenia polyandra (Tan.) Swing. is one of the six genera of the 'true citrus fruit trees' (*Citrus*, *Poncirus*, *Fortunella*, *Microcitrus*, *Eremocitrus* and *Clymenia*). The genus *Clymenia* is closely related to *Citrus*, but it is not well characterized. Chemotaxonomic work (Berhow *et al.*, 2000) suggests that *Clymenia* is closely allied with *Fortunella* and may be a hybrid between *Fortunella* and *Citrus*. *Clymenia* is considered a primitive genus in the 'true citrus fruit trees group' and may be a link between that group and the 'near citrus fruit trees' group (Krueger and Navarro, 2007).

It was not included in our previous study Garcia-Lor *et al.* (2013a) due to the unavailability of the vegetal material. After obtaining the DNA extract, we proceeded to the sequencing of the 18 genes involved in primary and secondary metabolite biosynthesis pathways that determine citrus fruit quality (sugars, acids, flavonoids and carotenoids) and nine putative genes involved in stress response as it was described in Garcia-Lor *et al.* (2013a).

From sequencing data, the available *Clymenia* accession appeared totally homozygous. It presented 60 specific SNP loci (not present in the other genera) in homozygosity, 31 in the coding regions and 29 in the non-coding regions.

Eight specific InDels were found in the *Clymenia* sequences, seven in the non-coding regions (2bp in ACO, 4bp in INVA, 1bp in DXS, 1bp in PSY, 18bp and 3bp in MPR4 and 1bp in PIP1A) and one (1bp) in the coding region of PIP1A gene fragment.

Among all models tested using Phylemon website (http://phylemon.bioinfo.cipf.es; Sánchez *et al*., 2011) for the phylogenetic analysis with SNP data, the model with the best fit was JC (with SH-like branch supports alone). This model takes into account the nucleotide substitution model JC (one substitution class; A = T = C = G). The phylogenetic relationships between *Citrus* species and their relatives inferred from maximum likelihood method using this model are represented in Figure 1. Branch support (BS) is represented in all branches. The different 'true citrus fruit trees' genotypes were rooted using *Severinia buxifolia* as outgroup. *Clymenia*, *Microcitrus* and *Eremocitrus* forms a clade with a high branch support (BS = 1).

Swingle (1967) described *Clymenia* as one of the six genus belonging to the 'true citrus fruit tress' group. It differs from all the species of the subgenus *Citrus* of the genus *Citrus* in some morphologic characters, like the leaves and the pulp-vesicles, but it is obviously related. Berhow et al. (2000) suggests that *Clymenia* is closely related with *Fortunella* and may be a hybrid between *Fortunella* and *Citrus*. In our analysis, *Clymenia* is placed in the same clade than *Microcitrus* and *Eremocitrus*, which are clearly differentiated from *Citrus* and *Fortunella* clusters. Moreover, the null amount of heterozygosity in the gene fragments analysed testifies that

Clymenia polyandra cannot be an interspecific or intergeneric hybrid. Our analysis is in agreement with Bayer et al. (2009) and Morton (2009), who observed *Clymenia* closely to *Microcitrus* and *Eremocitrus* in a phylogenetic study with cpDNA markers. Moreover, Swingle and Reece (1967) observed that *C. medica* is closely related to *Clymenia* and as our results show, the branch including *Clymenia*, *Microcitrus* and *Eremocitrus* is sister of the one formed by *C. maxima* and *C. medica*, confirming their probable close relationship.

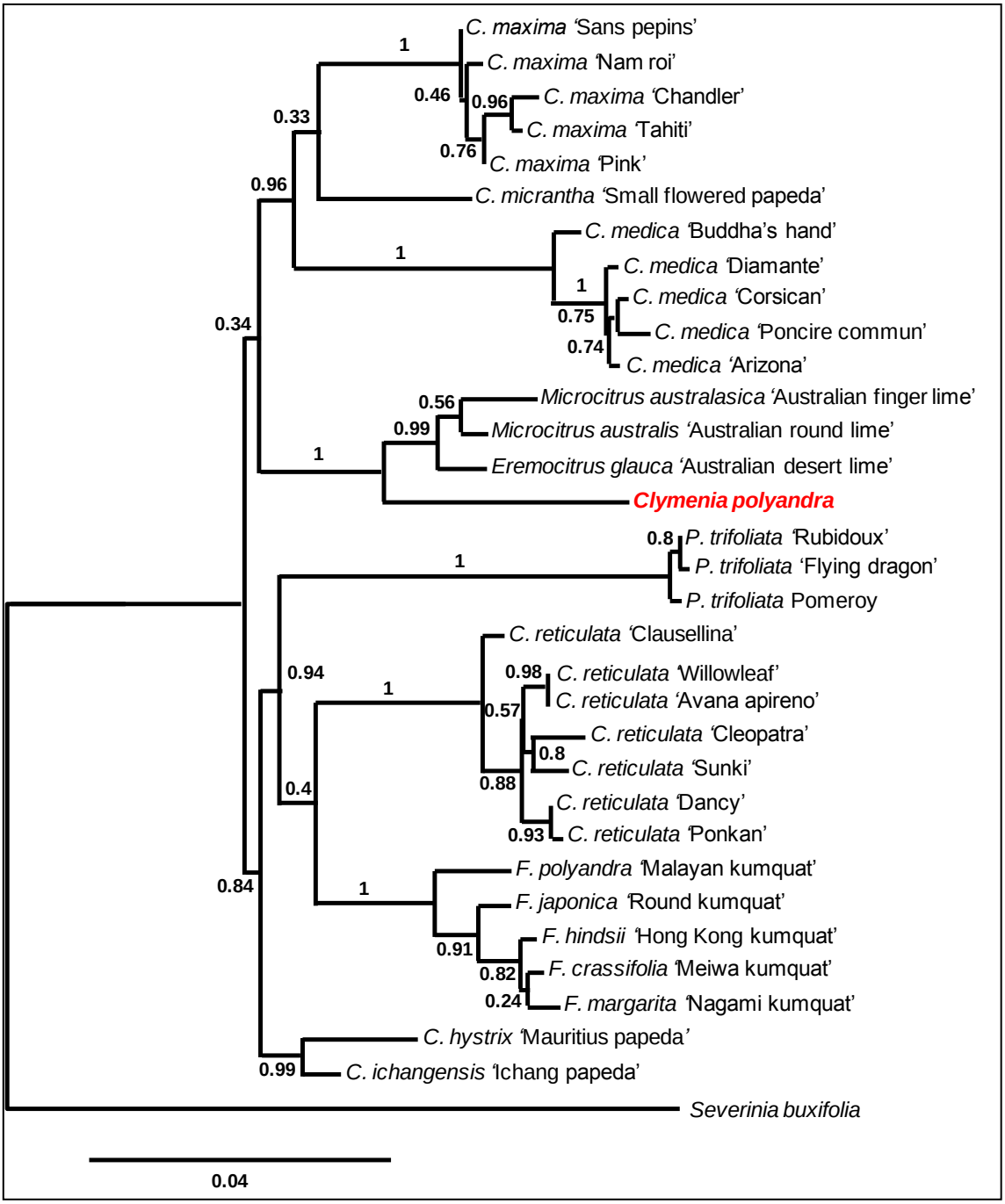


Figure 1. Phylogeny of the 'true citrus fruit trees' genera for the 27 genes sequenced.

CHAPTER 3

Citrus (Rutaceae) SNP markers based on Competitive Allele-Specific PCR; transferability across the Aurantioideae subfamily

Andres Garcia-Lor, Gema Ancillo, Luis Navarro, and Patrick Ollitrault

Applications in Plant Sciences (2013) 1: 1200406

Abstract

- *Premise of the study*

Single nucleotide polymorphism (SNP) markers based on Competitive Allele-Specific PCR (KASPar) were developed from sequences of three *Citrus* species. Their transferability was tested in 63 *Citrus* genotypes and 19 relative genera of the subfamily Aurantioideae to estimate the potential of SNP markers, selected from a limited intrageneric discovery panel, for ongoing broader diversity analysis at the intra- and intergeneric levels and systematic germplasm bank characterization.

- *Methods and Results*

Forty-two SNP markers were developed using KASPar technology. Forty-one were successfully genotyped in all of the *Citrus* germplasm, where intra- and interspecific polymorphisms were observed. The transferability and diversity decreased with increasing taxonomic distance.

- *Conclusions*

SNP markers based on the KASPar method developed from sequence data of a limited intrageneric discovery panel provide a valuable molecular resource for genetic diversity analysis of germplasm within a genus and should be useful for germplasm fingerprinting at a much broader diversity level.

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most frequent type of DNA sequence polymorphism. Their abundance and uniform distribution in genomes make them very powerful genetic markers. Several SNP genotyping methods have been developed. For low-to-medium throughput genotyping, the KBioscience Competitive Allele-Specific PCR genotyping system (KASPar; KBioscience Ltd., Hoddeston, United Kingdom) appears to be an interesting approach (Cuppen, 2007) that has been successfully applied in animals and plants (Nijman *et al*., 2008; Bauer *et al*., 2009; Cortés *et al*., 2011). For genetic diversity studies with SNP markers, it is very important to determine the representativeness of the discovery panel (Albrechtsen *et al.*, 2010). Ascertainment bias of the SNP markers affects the evaluation of genetic parameters, as was observed for the *Citrus* genus using SNP markers mined in a single Clementine cultivar (Ollitrault *et al*., 2012a). Recently, Garcia-Lor *et al*. (2013a) sequenced 27 amplified nuclear gene fragments for 45 genotypes of *Citrus*, which resulted in the identification of 1097 SNPs. Taking advantage of these previously obtained SNP data, the objective of this work was to implement a set of polymorphic SNP markers for systematic germplasm bank characterization within the *Citrus* genus and to investigate their transferability across the Aurantioideae [Engler] subfamily. More generally, the objective was to estimate the usefulness of SNP markers developed using KASPar technology, which were selected from a limited intrageneric discovery panel, for broader diversity analysis at the intra- and intergeneric levels.

METHODS AND RESULTS

The 42 SNP markers used in this study were selected from SNPs identified by Garcia-Lor *et al*. (2013a) in 27 nuclear genes. Most cultivated citrus (except for *C. aurantifolia* Christm.) Swingle) arose from interspecific hybridization of three ancestral taxa: *C. medica* L., *C. reticulata* Blanco, and *C. maxima* (Burm.) Merr. (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012a). Therefore, we selected SNPs between and within these three taxa (based on seven *C. reticulata*, five *C. maxima*, and five *C. medica* accessions). Primers were defined by KBioscience (http://www.kbioscience.co.uk/) from each SNP-locus flanking sequence (Appendix S1). Two allele-specific oligonucleotides and one common oligonucleotide were defined for each locus (Table 1). The KASPar system uses two Förster Resonance Energy Transfer (FRET) cassettes, where fluorometric dye is conjugated to the primer but quenched via resonance energy transfer. In this system, sample DNA is amplified in a thermal cycler using allele-specific primers, leading to the separation of fluorometric dye and quencher when the FRET cassette primer is hybridized with DNA (Cuppen, 2007). Normalized signals of each SNP allele (*x* and *y*) were provided by KBioscience services. Automatic allele calls provided by KlusterCaller software were visually checked with two-dimensional plot representations using SNPViewer software (KBioscience Ltd.).

Eighty-four accessions (Appendix 1) were genotyped for the 42 SNP markers. The sample set included representatives of the two tribes of the Aurantioideae (Clausenae and Citreae). In *Clausenae*, the subtribe Clauseniae was represented by four genotypes (three genera). Within the Citreae*,* three subtribes were represented: Triphasilinae (one genus was included), Balsamocitrinae (represented by six genera), and Citrinae (11 genera represented). For the Citrinae, we adopted the subdivision of this tribe into three groups (as proposed by Swingle and Reece, 1967), namely the primitive citrus fruit group (four accessions of four genera), the near citrus fruit group (three accessions of two genera), and the 'true citrus fruit trees' group (48 accessions of six genera). High-molecular-weight genomic DNA was extracted from leaf samples using a DNeasy Plant Mini Kit (Qiagen, Madrid, Spain) according to the manufacturer's instructions.

From the 42 SNP primers tested, only one did not produce polymorphisms. To check the accuracy of the allele call for the 41 other markers, we compared the KASPar genotyping data with Sanger sequencing data available for 35 accessions of the 'true citrus fruit trees' (Garcia-Lor *et al.*, 2013a). The conformity level was 95.41%, while 2.99% did not agree and 1.60% were missing data.

The allele number and the percentage of missing data are presented for each taxon (Table 2). The expected (H_e) and observed heterozygosity (H_o) were evaluated for *C. reticulata*, *C. maxima*, *C. medica,* the *Citrus* genus, and the 'true citrus fruit trees' excluding the *Citrus*

genus. Data analysis was conducted with Powermarker version 3.25 (Liu and Muse, 2005) and Darwin (Perrier and Jacquemoud-Collet, 2006) software.

The missing data rate was very low in *Citrus* (0.9%) and, generally, in the 'true citrus fruit trees' group (0.6%, excluding the *Citrus* genus). The missing data rate increased to 6.5% and 6.7% in the close citrus and primitive citrus groups of the Citrinae subtribe, respectively, reaching a level of 9.8% and 22.4% for the two other subtribes of the Citreae tribe, the Triphasilinae and the Balsamocitrinae, respectively. Missing data reached 26.8% in the Clauseniae tribe. These results indicate an increasing loss of transferability with increasing taxonomic distance. As expected due to the discovery panel, the *Citrus* genus was the most polymorphic (an average of two alleles per locus; H_e = 0.30 and H_o = 0.23), followed by the 'true citrus fruit trees' group excluding the *Citrus* genus (alleles per locus [*A*] = 1.32; H_e = 0.09 and H_o = 0.02). Diversity within and between the other taxa decreased considerably (data not shown). However, despite this important loss of polymorphism, all citrus relatives were differentiated when missing amplification was considered to represent null alleles, providing molecular fingerprinting for traceability in germplasm bank management.

Among the *Citrus* ancestral taxa, *C. reticulata* was the most polymorphic (*A* = 1.37; H_e = 0.11), followed by *C. medica* (*A* = 1.15; H_e = 0.04), and *C. maxima* (*A* = 1.10; H_e = 0.03). Considering as subpopulation the three species used in the discovery panel the F_{st} value was very high (0.842). The high level of differentiation between *C. reticulata*, *C. maxima*, and *C. medica* for this SNP panel was well illustrated by NJ analysis (Figure 1). The relative position of the accessions of secondary species (*C. aurantium* L., *C. aurantifolia*, *C. limon* (L.) Osbeck, *C. paradisi* Macf., and *C. sinensis* (L.) Osbeck) and hybrids (clementine, tangor, and tangelo) agrees with previous molecular studies (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012a; Garcia-Lor *et al.*, 2012a). Therefore, these markers should be useful as phylogenetic tracers of DNA fragments in secondary cultivated citrus species.

Table 1. Characteristics of 41 SNP primers used for genotyping of the Aurantioideae family

ID	Gene	SNP-specific primers	Common primer	GBA
EMA-M30	Malic enzyme (EMA)	AlleleX: GCCTATTCATATAATTTAGATGTCAGGAAA AlleleY: CCTATTCATATAATTTAGATGTCAGGAAG	GTTTAGCCCGCACTTTCTTTCTCTTT	JX630064
ACO-P353	Aconitase (ACO)	AlleleX: ATGTCTGCAGAGAAAACCAGTAAAATG AlleleY: CAATGTCTGCAGAGAAAACCAGTAAAATA	TCTCTGTTTTGAAGCTAATTCCCACTCAA	JX630065
ACO-C601	Aconitase (ACO)	AlleleX: ATAAAGGCTTATGAAAGAAAGTTTCAACTC AlleleY: CATAAAGGCTTATGAAAGAAAGTTTCAACTT	CTGAAGCTAATTTGCAGACATGGAACATT	JX630065
F3'H-P30	Flavonoid 3'-hydroxylase (F3'H)	AlleleX: CCCACTTGGCCTACGACGCT AlleleY: CCACTTGGCCTACGACGCC	CTCGGACCATAATCAGCAAAGACCAT	JX630066
F3'H-M309	Flavonoid 3'-hydroxylase (F3'H)	AlleleX: ACGTCATGAGCTCTACCACCATA AlleleY: CGTCATGAGCTCTACCACCATG	GACCAAAGGGACAGAATCTAATGAGTTTA	JX630066
F3'H-C341	Flavonoid 3'-hydroxylase (F3'H)	AlleleX: GAGCTCATGACGTCAGCTGGATT AlleleY: GAGCTCATGACGTCAGCTGGATA	GCAATCGAGGGTATAAAATCACCAATGTT	JX630066
PEPC-M316	Phosphoenolpyruvate carboxylase (PEPC)	AlleleX: TAAAGAGCAATGAATTTCTTCAAACCTAA AlleleY: AAAGAGCAATGAATTTCTTCAAACCTAG	GTGCATTTAAGAACTGAGAAGGCATAGAA	JX630067
PEPC-C328	Phosphoenolpyruvate carboxylase (PEPC)	AlleleX: TAAAGCTGACTTAAAGAGCAATGAATTC AlleleY: CTTAAAGCTGACTTAAAGAGCAATGAATTT	GAAGGCATAGAATATTCCAYTAGGTTTGAA	JX630067
SOS1-M50	Salt overly sensitive 1 (SOS1)	AlleleX: GGTTTAGTACTGAGTAAGTTACTTGC AlleleY: AAATGGTTTAGTACTGAGTAAGTTACTTGT	GGACTTTTTCAGGTTTTGCATGTTGTCAA	JX630068
CCC1-M85	Cation chloride cotransporter (CCC1)	AlleleX: CATTGTGGTTATGAGGTATCCAGAG AlleleY: AACATTGTGGTTATGAGGTATCCAGAA	CAGTAAGGTTTTCACGGCGCCATAT	JX630069
CCC1-P727	Cation chloride cotransporter (CCC1)	AlleleX: ATCAACCACCCAGCTTACTGCTAT AlleleY: CAACCACCCAGCTTACTGCTAC	GGCACATTCTCTACTAACAAATCCATGTA	JX630069
TRPA-M593	Vacuolar citrate/H+ symporter (TRPA)	AlleleX: AACGTGGCAGCAGCAGTGATG AlleleY: AACGTGGCAGCAGCAGTGATC	TCCCAGTGGCCACTGGCATCAT	JX630070
INVA-M437	Acid invertase (INVA)	AlleleX: GTTCAGCAGATCCTTCGCTGGAA AlleleY: CAGCAGATCCTTCGCTGGAG	ACAGCGGAGTCCAATGTGGAGTTTA	JX630071
INVA-P855	Acid invertase (INVA)	AlleleX: GGCACTGTCAATAGAATCCTCACAAT AlleleY: GCACTGTCAATAGAATCCTCACAAC	CCTGCAAATATACATACACAATGTTCCAAA	JX630071
MDH-MP69	Malate dehydrogenase (MDH)	AlleleX: AGGCCACTGAAACTCACAAGTGAT AlleleY: GGCCACTGAAACTCACAAGTGAG	CTGGTGTGAGGTTCAACTCCAAGAA	JX630072
MDH-M519	Malate dehydrogenase (MDH)	AlleleX: CAGCCTCAACCAAGGTCTTTACTATA AlleleY: AGCCTCAACCAAGGTCTTTACTATG	GATGACCTCTTCAACATCAACGCCAA	JX630072
ATMR-C372	MRP-like ABC transporter (ATMR)	AlleleX: GAATCATTATTGATGGAATCGACATTTCG AlleleY: AGAATCATTATTGATGGAATCGACATTTCA	ACCTTAGGTCATGAAGCCCCAACAA	JX630073
ATMR-M728	MRP-like ABC transporter (ATMR)	AlleleX: GTTTGATTTAATGGAAGTCATATGTATCTTTTT AlleleY: TGATTTAATGGAAGTCATATGTATCTTTTG	AAAGTTCAACATTTTGGCATGTTTTAGCTT	JX630073
CHS-P57	Chalcone synthase (CHS)	AlleleX: CAAGTATGGTAGTTTCAGAAGTGGTA AlleleY: CAAGTATGGTAGTTTCAGAAGTGGTT	AAAACAACCCTGGAAGCCGCGTTTT	JX630074
CHS-M183	Chalcone synthase (CHS)	AlleleX: GTTGGAGCTGACCCATTCCTG AlleleY: GTTGGAGCTGACCCATTCCTC	GTTAAGTTCCATGAAAGGAGAAGACTCTT	JX630074
CHI-M598	Chalcone isomerase (CHI)	AlleleX: CGTCACTTTCACGCCGTCCG AlleleY: CGTCACTTTCACGCCGTCCC	TGCGACTTTGTTGATCCTGGAGGTT	JX630075
PKF-C64	Phosphofructokinase (PKF)	AlleleX: ACTCCCTCTCCCTTCTGTTCTC AlleleY: CACTCCCTCTCCCTTCTGTTCTA	GGCCATCGACGATTTTGAAAGGGTT	JX630076
PKF-M186	Phosphofructokinase (PKF)	AlleleX: CGTCCGTAACATTACAGATTCAAGAT AlleleY: CGTCCGTAACATTACAGATTCAAGAC	CCGAACAGATTTGGAAACAATTTCGCAAT	JX630076

Table 1. Continued

NADK2-M285	NADH kinase (NADK2)	AlleleX: CATCTTCTCTTGGTGATACAAGAAAGAA AlleleY: ATCTTCTCTTGGTGATACAAGAAAGAG	AACTCATTTCTAGATCTGATGAGCAGGTT	JX630077
DFR-M240	Dihydroflavonol 4-reductase (DFR)	AlleleX: CCGAAGAGGGAAACTTTGATGAAG AlleleY: CCGAAGAGGGAAACTTTGATGAAC	GAAAAACTCCAGTGCAGCCTCGAAT	JX630078
LAPX-M238	Ascorbate peroxidase (LAPX)	AlleleX: GAATTGACCATGGTTTGTGTTTTATTTTC AlleleY: GAATTGACCATGGTTTGTGTTTTATTTTG	GGCAACAACTCCAGCCAACTTCAA	JX630079
PSY-M30	Phytoene synthase (PSY)	AlleleX: GTCCATTTGATATGCTTGATGCTGG AlleleY: GTCCATTTGATATGCTTGATGCTGC	CGACAGGAAATTTGGTTACTGTATCTGAT	JX630080
PSY-C461	Phytoene synthase (PSY)	AlleleX: CGCAGGCCTATTAAACTCTTGTCA AlleleY: CGCAGGCCTATTAAACTCTTGTCT	AAGTTCTGCATGCTACCCTTCTCAATATT	JX630080
AOC-M290	Ascorbate oxydase (AOC)	AlleleX: AAGGGGTGCATCTGAGCCAAAG AlleleY: AAAGGGGTGCATCTGAGCCAAAA	CTGCGTTGAAAACTAATGGTACTGTACTT	JX630081
AOC-C593	Ascorbate oxydase (AOC)	AlleleX: GCCATACCCATGGAATTCGGCT AlleleY: GCCATACCCATGGAATTCGGCA	GGGGTAACTGGAGGGCTCCATT	JX630081
DXS-C545	1-deoxyxylulose 5-phosphate synthase (DXS)	AlleleX: ACCAAATGCATCATGAACGCTTTCC AlleleY: ACCAAATGCATCATGAACGCTTTCG	GGGGCTTGCAGGATTCCCCAAA	JX630082
DXS-M618	1-deoxyxylulose 5-phosphate synthase (DXS)	AlleleX: GGTCTTGGTATGTACTTCG AlleleY: CTGCTGGTCTTGGTATGTACTTCA	CCTACAATTTCTCTAGATTGATGAAAGGAA	JX630082
FLS-P129	Flavonol synthase (FLS)	AlleleX: GGCTTCCGCGATGGAACGTA AlleleY: GGCTTCCGCGATGGAACGTG	CGATCTCGACGACCCCGTTCAA	JX630083
FLS-M400	Flavonol synthase (FLS)	AlleleX: CCGTCTTCTATCAACTACCGCTTT AlleleY: CGTCTTCTATCAACTACCGCTTC	TTCACCGGTAAGAAGGAGGGTTGTT	JX630083
LCY2-M379	Lycopene β-cyclase 2 (LCY2)	AlleleX: TGATGAGTTTGAAGACATAGGACTTG AlleleY: GTTGATGAGTTTGAAGACATAGGACTTA	CGGCCAAGTTTTGTCCAAACAGTCTA	JX566716
LCYB-M480	Lycopene β-cyclase (LCYB)	AlleleX: GAATAACCTTAATAACTTTAGCTTGGTGG AlleleY: GAATAACCTTAATAACTTTAGCTTGGTGA	GCTGCAAAAATGCATAACCAATGGTGTTA	JX630084
LCYB-P736	Lycopene β-cyclase (LCYB)	AlleleX: GATTCGCATCTGAACAACAATTCGG AlleleY: CGCATCTGAACAACAATTCGC	GAAAAGTAGGAATTTTGCTATTTGCCTCTT	JX630084
HYB-M62	β-Carotene hydroxylase (HYB)	AlleleX: AAAACAAAACATACGGTGAAAGAGTTGAT AlleleY: AACAAAACATACGGTGAAAGAGTTGAG	GGCTTCTTTAATGGCAAAAACCGAAGAAA	AF315289
HYB-C433	β-Carotene hydroxylase (HYB)	AlleleX: GAGCAAATGTGCCAAACATTTCAGC AlleleY: AGAGCAAATGTGCCAAACATTTCAGT	GTACAGGGTGGAGAGGTGCCTT	JX630087
TSC-C80	Tréhalose-6-phosphate synthase (TSC)	AlleleX: TCTTGACCACTTGGAAAATGTTCTTT AlleleY: CTTGACCACTTGGAAAATGTTCTTG	GCCTCTTTTGACAACAACAGGCTCAT	JX630084
NCED3-M535	9-cis-epoxy hydroxy carotenoid dyoxygenase 3 (NCED3)	AlleleX: GACACCTTGTTCTTGTCATAAATCACA AlleleY: ACACCTTGTTCTTGTCATAAATCACC	CAAGTGGTGTTCAAGTTGAATGAGATGAT	JX630086

(ID) SNP locus name; Gene (Genes amplified); SNP-specific primers (Allele X and Y); Common primer (Reverse primer); Allele X and Y (Alleles identified). GBA (GenBank accessions of the genes amplified). Genomic sequences of *C. reshni* included in GenBank are in Appendix S1 (see Supplemental Data with the online version of this article).

Table 2. Results of initial primer screening in different *Citrus* species and subtribes of the subfamily Aurantioideae.

	C. reticulata (*N*=12)			*C. maxima* (*N*=11)			*C. medica* (*N*=6)			*Citrus* (*N*=32)		True citrus* (*N*=16)		Balsamocitrinae (*N*=6)		Near Citrus (*N*=3)		Primitive Citrus (*N*=4)		Triphasilinae (*N*=1)		Clauseniae (*N*=4)		Aurantioideae (*N*=84)		
Marker	*A*	H_o	H_e	*A*	H_o	H_e	*A*	H_o	H_e	*A*	MD	*A*	MD	*A*	MD	*A*	H_o	*A*	MD	*A*	MD	*A*	MD	*A*	H_o	H_e	MD
EMA-M30	2	0.73	0.37	1	0.00	0.00	1	0.00	0.00	2	3.13	1	0.00	1	66.67	0	100.00	0	100.00	1	0.00	1	75.00	2	0.29	0.26	17.86
ACO-P353	1	0.00	0.00	2	0.55	0.37	1	0.00	0.00	2	0.00	1	0.00	2	33.33	1	0.00	2	0.00	1	0.00	1	25.00	2	0.16	0.27	3.57
F3H-M309	2	0.33	0.30	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	16.67	1	0.00	1	0.00	1	0.00	1	0.00	2	0.07	0.09	1.19
F3H-C341	1	0.00	0.00	1	0.00	0.00	2	0.17	0.37	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	0.04	0.10	1.19
PEPC-M316	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	3.13	1	0.00	1	16.67	1	0.00	1	25.00	1	0.00	1	50.00	2	0.11	0.29	5.95
SOS1-M50	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	75.00	2	0.12	0.35	3.57
CCC1-M85	2	0.67	0.37	1	0.00	0.00	1	0.00	0.00	2	0.00	2	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	50.00	2	0.24	0.31	2.38
TRPA-M593	2	0.58	0.33	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	25.00	2	0.20	0.37	1.19
INVA-M437	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	50.00	1	0.00	1	0.00	1	0.00	1	50.00	2	0.13	0.29	5.95
MDH-M519	2	0.42	0.33	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	16.67	1	0.00	1	25.00	1	0.00	1	50.00	2	0.18	0.24	4.76
ATMR-M728	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	3.13	2	6.25	1	83.33	1	33.33	1	0.00	1	0.00	0	100.00	2	0.15	0.37	14.29
CHS-P57	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	0.08	0.26	0.00
CHI-M598	2	0.09	0.08	1	0.00	0.00	2	0.17	0.14	2	3.13	2	6.25	2	33.33	1	0.00	2	0.00	1	0.00	1	50.00	2	0.22	0.37	7.14
PKF-M186	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	0	100.00	1	33.33	1	75.00	0	100.00	1	25.00	2	0.13	0.37	14.29
NADK2-M285	1	0.00	0.00	1	0.00	0.00	2	0.33	0.24	2	0.00	2	6.25	1	50.00	1	33.33	1	25.00	0	100.00	1	25.00	2	0.20	0.34	9.52
DFR-M240	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	3.13	2	0.00	1	83.33	2	33.33	2	0.00	0	100.00	1	0.00	2	0.17	0.37	10.71
LAPX-M238	2	0.50	0.35	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	75.00	2	0.22	0.26	3.57
PSY-M30	2	0.67	0.35	1	0.00	0.00	1	0.00	0.00	2	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	25.00	2	0.27	0.37	1.19
AOC-M290	2	0.45	0.29	1	0.00	0.00	1	0.00	0.00	2	6.25	1	0.00	2	16.67	2	0.00	2	0.00	1	0.00	2	0.00	2	0.28	0.30	3.57
DXS-M618	2	0.50	0.30	1	0.00	0.00	1	0.00	0.00	2	0.00	2	0.00	2	16.67	2	0.00	1	0.00	1	0.00	1	50.00	2	0.21	0.27	3.57
DXS-C545	1	0.00	0.00	1	0.00	0.00	2	0.33	0.35	2	0.00	1	0.00	1	16.67	1	0.00	1	0.00	1	0.00	1	0.00	2	0.07	0.13	1.19
FLS-P129	1	0.00	0.00	2	0.27	0.34	1	0.00	0.00	2	0.00	2	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	0.18	0.26	0.00
FLS-M400	2	0.50	0.30	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	2	16.67	1	0.00	1	0.00	1	0.00	2	0.00	2	0.18	0.27	1.19
LCY2-M379	2	0.67	0.35	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	0.27	0.26	1.19
LCYB-P736	1	0.00	0.00	2	0.18	0.37	1	0.00	0.00	2	0.00	1	0.00	1	66.67	1	0.00	1	0.00	1	0.00	1	25.00	2	0.06	0.16	5.95
LCYB-M480	2	0.33	0.24	1	0.00	0.00	1	0.00	0.00	2	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	25.00	2	0.23	0.34	1.19
HYB-M62	2	0.42	0.33	1	0.00	0.00	1	0.00	0.00	2	3.13	2	0.00	1	0.00	1	0.00	1	25.00	1	0.00	1	75.00	2	0.27	0.37	5.95
CCC1-P727	2	0.58	0.37	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	25.00	2	0.16	0.34	1.19
TSC-C80	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	3.13	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	0.07	0.16	1.19
ACO-C601	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	33.33	1	0.00	1	0.00	1	0.00	1	25.00	2	0.06	0.19	3.57
F3H-P30	1	0.00	0.00	2	0.10	0.09	1	0.00	0.00	2	6.25	1	0.00	1	0.00	1	0.00	1	0.00	0	100.00	1	0.00	2	0.10	0.24	4.76
NCED3-M535	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	16.67	1	0.00	1	0.00	1	0.00	1	25.00	2	0.13	0.29	2.38
INVA-P855	1	0.00	0.00	1	0.00	0.00	2	0.33	0.24	2	0.00	2	0.00	2	16.67	2	0.00	2	0.00	2	1.00	2	0.00	2	0.22	0.37	2.38
MDH-MP69	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	2	6.25	1	33.33	1	0.00	1	0.00	1	0.00	1	50.00	2	0.18	0.37	5.95
ATMR-C372	1	0.00	0.00	1	0.00	0.00	2	0.50	0.37	2	0.00	1	0.00	1	16.67	1	0.00	1	0.00	1	0.00	1	25.00	2	0.09	0.12	2.38
CHS-M183	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	66.67	1	0.00	1	0.00	1	0.00	1	25.00	2	0.06	0.31	5.95
PKF-C64	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	3.13	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	2	0.05	0.14	1.19
PSY-C461	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	33.33	1	0.00	1	0.00	1	0.00	1	25.00	2	0.06	0.17	3.57
AOC-C593	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	0.00	1	25.00	2	0.06	0.17	1.19
HYB-C433	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	2	0.00	1	33.33	1	0.00	1	0.00	1	0.00	2	0.10	0.19	1.19
PEPC-C328	1	0.00	0.00	1	0.00	0.00	1	0.00	0.00	2	0.00	1	0.00	1	16.67	1	0.00	2	0.00	1	0.00	1	0.00	2	0.08	0.18	1.19
Mean	1.37	0.18	0.11	1.10	0.03	0.03	1.15	0.04	0.04	2	0.91	1.32	0.61	1.22	22.36	1.07	6.50	1.12	6.71	0.93	9.78	1.07	26.83	2	0.15	0.26	4.15

N: sample size; *A*: number of alleles; H_o: observed heterozygosity; H_e: expected heterozygosity; MD: missing data (%); *True citrus excluding the *Citrus* genus.

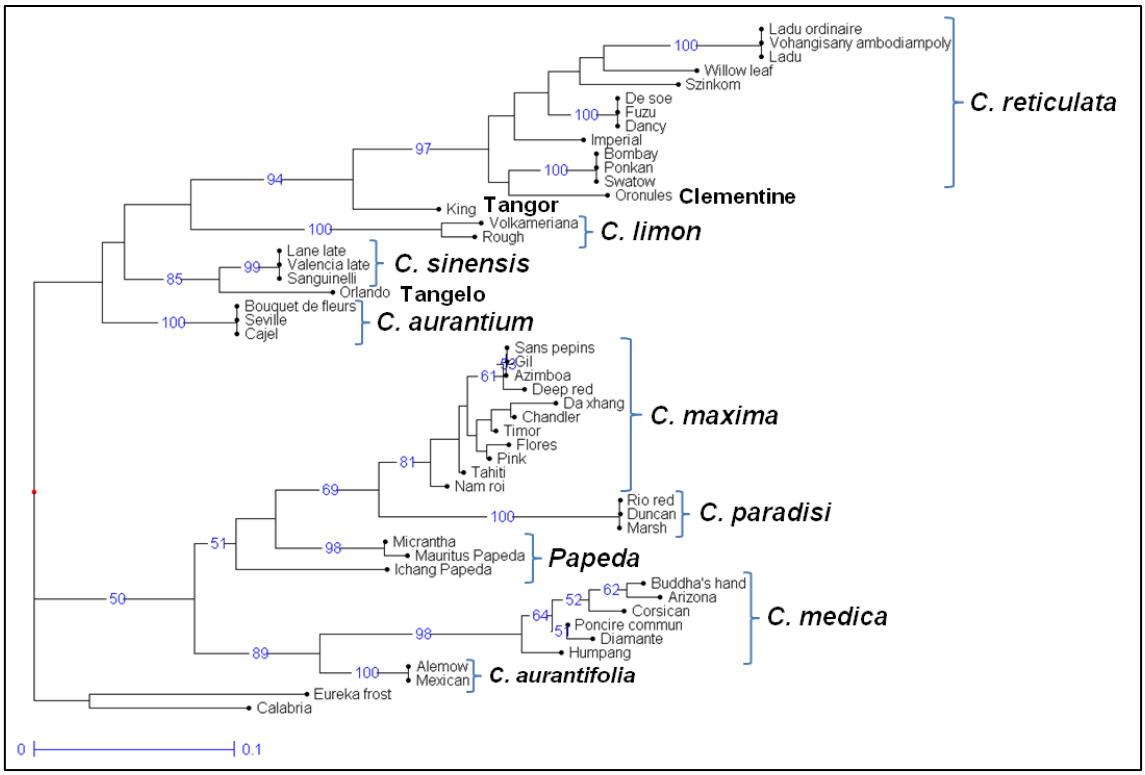


Figure 1. Neighbor-joining analysis based on simple matching dissimilarities from 41 SNP loci for 50 accessions belonging to the genus *Citrus*, including secondary species and hybrids. Numbers near nodes are bootstrap values based on 1000 resamplings (only values >50% are indicated)

CONCLUSIONS

Forty-one SNP markers were successfully developed from SNP loci mined by Sanger sequencing in a discovery panel including 17 genotypes of the three main cultivated *Citrus* ancestral taxa. The genotyping data displayed high conformity with previous sequencing data. Genotyping was highly successful within the *Citrus* genus, and the genetic organization displayed by this SNP marker panel was in agreement with previous studies. The frequency of missing data was higher for the citrus relatives and increased with taxonomic distances within the Aurantioideae subfamily, suggesting incomplete transferability. The polymorphism revealed within the relatives of the 'true citrus fruit trees' group remained relatively high but decreased strongly when considering the other citrus relatives. However, all citrus relative genotypes were differentiated. The markers that were developed appeared to be useful for phylogenic studies within the 'true citrus fruit trees'. Therefore, SNP markers based on the KASPar method developed from sequence data of a limited intra-generic discovery panel provide a valuable molecular resource for genetic diversity analysis of germplasm within a genus and should be useful for germplasm fingerprinting at a much broader diversity level.

APPENDIX CHAPTER 3

Appendix 1. Analyzed accessions

Species name, latin name or common name, accession number, ex-situ germplasm bank.

IVIA: Carretera Moncada, Naquera, km4.4, Apartado Oficial ,46113 Moncada (Valencia), Spain.

INRA/CIRAD: Station INRA 20230 San Giuliano, France.

1. Citreae

Balsamocitrinae: ***Aegle marmelos*** (L.) Corr., 345, IVIA; ***Aeglopsis chevalieri*** Swing., 308, IVIA; ***Afraegle paniculata*** (Schum.) Engl., 273, IVIA; ***Balsamocitrus*** dawei Stapf., 372, IVIA; ***Feroniella oblata*** Swing., 585, IVIA; ***Swinglea glutinosa*** (Blanco) Merr., 292, IVIA.

Citrinae

True citrus fruit:

Citrus**:** ***C. maxima***: Azimboa, 420, IVIA; Chandler, 207, IVIA; Da xanh, 589, IVIA; Deep red, 277, IVIA; Flores, 673, INRA/CIRAD; Gil, 321, IVIA; Nam roi, 590, IVIA; Pink, 275, IVIA; Sans Pepins, 710, INRA/CIRAD; Tahiti, 727, INRA/CIRAD; Timor, 707, INRA/CIRAD. ***C. medica***: Arizona, 169, IVIA; Buddha hand, 202, IVIA; Corsican, 567, IVIA; Diamante, 560, IVIA; Humpang, 722, INRA/CIRAD; Poncire Commun, 701, INRA/CIRAD. ***C. reticulata***: Bombay, 518, INRA/CIRAD; Dancy, 434, IVIA; De soe, 713, INRA/CIRAD; Imperial, 576, IVIA; Fuzhu, 571, IVIA; Ladu, 595, INRA/CIRAD; Ladu ordinaire, 590, INRA/CIRAD; Ponkan, 482, IVIA; Swatow, 175, INRA/CIRAD; Szinkom, 597, INRA/CIRAD; Vohangisany ambodiampoly, 437, SRA; Willow leaf, 154, IVIA. ***Papeda*****:** ***C. hystrix*** DC.: Combava, 178, IVIA; ***C. ichangensis*** Swing.: Papeda Ichang, 358, IVIA; ***C. micrantha*** Wester: Micrantha, IVIA.

Secondary species: ***C. aurantifolia***: Alemow, 288, IVIA; Calabria, 254, IVIA; Mexican, 164, IVIA. ***C. aurantium***: Bouquet de fleurs, 139, IVIA; Cajel, 108, IVIA; Seville, 117, IVIA. ***C. limon***: Eureka frost, 297, IVIA; Rough lemon, 333, IVIA; Volkamer lemon, 432, IVIA; ***C. paradisi***: Duncan, 274, IVIA; Marsh, 176, IVIA; Rio red, 289, IVIA. ***C. sinensis***: Lane late, 198, IVIA; Sanguinelli, 34, IVIA; Valencia late, 363, IVIA.

Hybrids: Clementine, Clemenules, 22, IVIA; Tangelo, Orlando, 101, IVIA; Tangor, King, 477, IVIA.

Clymenia polyandra (Tan.) Swing., 584, IVIA.

Eremocitrus: *E. Glauca*, 346, IVIA.

Fortunella: *F. crassifolia*, 280, IVIA; *F. hinsii*, 281, IVIA; *F. japonica*, 381, IVIA; *F. margarita*, 38, IVIA; *Fortunella* sp., 98, IVIA.

Microcitrus: *M. australasica*, 150, IVIA; *M. australis*, 313, IVIA; *M. australis* x *M. Australasica*, 378, IVIA; Australian Wild Lime, 314, IVIA; New Guinea Wild Lime, 315, IVIA.

***Poncirus trifoliata*:** Flying Dragon, 537, IVIA; Pomeroy, 374, IVIA; Rich 75, 236, IVIA; Rubidoux, 217, IVIA.

Near citrus fruit: ***Atalantia ceylanica*** (Arn.) Oliv., 172, IVIA; ***Atalantia citroides*** Pierre ex Guill., 284, IVIA; ***Citropsis gilletiana*** Swing. and M.Kell, 517, IVIA.

Primitive citrus fruit: ***Hesperethusa crenulata*** (Roxb.) Roem., 580, IVIA; ***Pleiospermium*** sp., 380, IVIA; ***Severinia buxifolia***, 147, IVIA; ***Severinia disticha*** (Blanco) Swing., 418, IVIA.

Triphasilinae: ***Triphasia trifolia*** (Burm. F.) P. Wils., 182, IVIA.

2. Clauseneae

Clauseniae: ***Clausena excavata*** Burm. f., 311, IVIA; ***Clausena lansium*** (Lour.) Skeels, 343, IVIA; ***Glycosmis pentaphylla*** (Retz.) Corrêa, 148, IVIA; ***Murraya koenigii*** (L.) Spreng., 377, IVIA.

Appendix S1

Genomic sequences of *C. reshni* included in GenBank corresponding to each SNP locus. SNPs genotyped for the considered loci are shown between brackets.

> Locus LCY2-M379. GenBank accession: JX566716 [organism=Citrus reshni] PCR product lycopene β-cyclase 2 genomic DNA.

GTTTCGCAAATAATTGATTCAACATCATCACATTCATTTTCGCTATTTCCATTAGGCCGCCAAAATGCAT
GTTCAAGAAAGGCGGATCATCATCATCATCACAGGATCCGGACAAGCAAGTTTGGTAACTTCCTAGAGTT
GACACCGGAGTCGGAACCTGAATTCTTAGTCTTTGATCTCCCCTGGTTTCATCCGTCCGATCGTATTCGA
TATGACGTGATCATCATTGGCACTGGACCTGCCGGCCTCCGTCTAGCTGAGCAAGTCTCATCGCGTCATG
GTATCAAGGTATGTTGTGTTGATCCTTCACCTCTTTCTACGTGGCCTAACAACTATGGAGTTTGGGTTGA
TGAGTTTGAAGACATAGGACTT**[A/G]**TAGACTGTTTGGACAAAACTTGGCCGATGACTTGTGTTTTTAT
TAATGATCACAAGACCAAGTATCTAGACAGGCCCTACGGTCGTGTTAGTAGAAATATTTTGAAGACAAAG
TTATTAGAGAATTGTGTTTTAAATGGCGTTAGGTTTCATAAGGCTAAAGTTTGGCATGTGAATCATCAGG
AGTTCGAGTCTTCGATTGTTTGTGATGATGGRAATGAGATTAAGGCTAGCTTGATTGTTGATGCTAGTGG
CTTTGCTAGTAGTTTTGTTGASTATGATAAGCCAAGAAACCATGGATACCAAATTGCTCATGGGATTTTA
GCTGAGGTTGAGAGTCACCCTTTTGATTTAGATAAA

> Locus EMA-M30. GenBank accession: JX630064 [organism=Citrus reshni] PCR product Malic enzyme (EMA) genomic DNA.

GTTTAGCCCGCACTTTCTTTCTCTTTCTG**[T/C]**TTCCTGACATCTAAATTATATGAATAGGCTTTTGTT
GTCAAATGGACTGAAATAATTAGGATGCAACAGAAATTAACTGCATRTTGCACCACCATTTAAGAACAGT
TTGTTACAATGTGAACAAGTCCACTGGAAAAATCCATTAACAAATATTTGAATTAGCCGTGAACGTAAGT
GTTCTCTTGGCAAACGTGTAAAATCRTTAGAGCTTGTTTACTTGGTGATTGATAAACTAGTTGTGTTTTA
TTCACCGGCAGCGGAAGCCTTGGCAAAACAAGTGACAGAAGAGAACTTTGAGAAGGGATTGATCTACCCA
CCATTTTCTAATATTAGAAAAATTTCAGCCAATATAGCTGCTAATGTTGCTGCTAAGGCATATGAACTA

> Locus ACO-P353. GenBank accession: JX630065 [organism=Citrus reshni] PCR product aconitase genomic DNA.

TTGCTTTTCCATGTGGTTGTATATTTACAAAATTAAGTTACATGGTCCTGTTTCAATGTCTGTAGTGGCC
TGCAAAAGTACTTGAACCAACAAGGTTTTCACATTGTTGGCTATGGCTGCACTACTTGTATTGGAAACTC
TGGAGATCTTGATGAATCAGTTGCTACTGCAATTACAGAAAATGGTAACTGTTAATTATCTTTGGTACCT
TTTAGAATCAGTTTAGACTGCATTACAAACTTTAGCATAAGTTTTAATCTGGTTACACATTAACTCTCAA
ACCATTTACATCATAATTTTTGTGGCAGTTTGTGATCCATCATTCTCTGTTTTGAAGCTAATTCCCACTC
AA**[C/T]**ATTTTACTGGTTTTCTCTGCAGACATTGTTGCAGCTGCTGTGCTTTCCGGTAATCGGAACTTT
GAAGGTCGTGTACATCCTTTGACAAGAGCTAACTATCTTGCATCTCCTCCATTAGTTGTTGCTTATGCCC
TTGCTGGCACAGTAAGTATATAACTTCTAGTCAAATATTCTTATAGAATTGTTGCTATCCTTATGATCTG
AAGCTAATTTGCAGACATGGAACATTATTATAATTTACAACTAGGAGTTGAAACTTTCTTTCATAAGCCT
TTATGCTAGTTACATGACATGCTTTTGAATCAACCAATGTCCATAATCCGTTAATTTTTTATTTTAAAA

> Locus ACO-C601. GenBank accession: JX630065 [organism=Citrus reshni] PCR product aconitase genomic DNA.

TTGCTTTTCCATGTGGTTGTATATTTACAAAATTAAGTTACATGGTCCTGTTTCAATGTCTGTAGTGGCC
TGCAAAAGTACTTGAACCAACAAGGTTTTCACATTGTTGGCTATGGCTGCACTACTTGTATTGGAAACTC
TGGAGATCTTGATGAATCAGTTGCTACTGCAATTACAGAAAATGGTAACTGTTAATTATCTTTGGTACCT
TTTAGAATCAGTTTAGACTGCATTACAAACTTTAGCATAAGTTTTAATCTGGTTACACATTAACTCTCAA
ACCATTTACATCATAATTTTTGTGGCAGTTTGTGATCCATCATTCTCTGTTTTGAAGCTAATTCCCACTC
AACATTTTACTGGTTTTCTCTGCAGACATTGTTGCAGCTGCTGTGCTTTCCGGTAATCGGAACTTTGAAG
GTCGTGTACATCCTTTGACAAGAGCTAACTATCTTGCATCTCCTCCATTAGTTGTTGCTTATGCCCTTGC
TGGCACAGTAAGTATATAACTTCTAGTCAAATATTCTTATAGAATTGTTGCTATCCTTATGATCTGAAGC
TAATTTGCAGACATGGAACATTATTATAATTTACAACTAG**[G/A]**AGTTGAAACTTTCTTTCATAAGCCT
TTATGCTAGTTACATGACATGCTTTTGAATCAACCAATGTCCATAATCCGTTAATTTTTTATTTTAAAA

> Locus F3'H-P30. GenBank accession: JX630066 [organism=Citrus reshni] PCR product flavonoid 3',5'-hydroxylase genomic DNA. F3'H-M309 and F3'H-C341.

CGCCGGTGCACCCACTTGGCCTACGACGC**[T/C]**CAAGACATGGTCTTTGCTGATTATGGTCCGAGGTGG
AAACTCTTAAGAAAGATAAGCAATCTGCACATGCTTGGTGGAAAAGCCCTATATGATTGGAGTAACGTGC
GTAACATTGAGCTAGGCCACATGCTTCGAGCCATTTGTGAGTCTAGCCAGCGAAACGAGCCCGTGGTGGT
CCCGGAGATGTTGACGTACGCCATGGCAAACATGATAGGTCAAGTCATACTAAGCCGTCGAGTTTTTGTG
ACCAAAGGGACAGAATCTAATGAGTTTAAGGACATGGTGGTAGAGCTCATGACGTCAGCTGGATTTTTCA
ACATTGGTGATTTTATACCCTCGATTGCTTGGTTGGATTTACAAGGGATCGAGCGTGGGATGAAGAAATT
ACATAACAGATTTGATGTCCTGTTAACAAAGATGATTGAAGAGCACATGGCTTCAACTWATGAACGTAAA
AGGAAGCCAGATTTTCTCGACATTGTCATGGCTAATAGAGAAAATTCTGATGGAGAAAGGCTCACCATCA
CCAACATCAAAGCACTTCTCTTGGTAATTTGTGTCTTCAAAACTTTACCCTTTTTTTTATCTCACTTTTG
TATTTATTATTACGCTCATGYATTTAAGGTTATCAAAGTTGCACTTACGAAATATTTATTTCATACctcg
tgcctcTTGGTCTTTCTTTGTGTTGTTCATTTTTGTACATATGTGAAATGCAGTATAATCTTGATAAATA
TATTATTCT

> Locus F3'H-M309. GenBank accession: JX630066 [organism=Citrus reshni] PCR product flavonoid 3',5'-hydroxylase genomic DNA.

CGCCGGTGCACCCACTTGGCCTACGACGCTCAAGACATGGTCTTTGCTGATTATGGTCCGAGGTGGAAAC
TCTTAAGAAAGATAAGCAATCTGCACATGCTTGGTGGAAAAGCCCTATATGATTGGAGTAACGTGCGTAA
CATTGAGCTAGGCCACATGCTTCGAGCCATTTGTGAGTCTAGCCAGCGAAACGAGCCCGTGGTGGTCCCG
GAGATGTTGACGTACGCCATGGCAAACATGATAGGTCAAGTCATACTAAGCCGTCGAGTTTTTGTGACCA
AAGGGACAGAATCTAATGAGTTTAAGGA**[T/C]**ATGGTGGTAGAGCTCATGACGTCAGCTGGATTTTTCA
ACATTGGTGATTTTATACCCTCGATTGCTTGGTTGGATTTACAAGGGATCGAGCGTGGGATGAAGAAATT
ACATAACAGATTTGATGTCCTGTTAACAAAGATGATTGAAGAGCACATGGCTTCAACTWATGAACGTAAA
AGGAAGCCAGATTTTCTCGACATTGTCATGGCTAATAGAGAAAATTCTGATGGAGAAAGGCTCACCATCA
CCAACATCAAAGCACTTCTCTTGGTAATTTGTGTCTTCAAAACTTTACCCTTTTTTTTATCTCACTTTTG
TATTTATTATTACGCTCATGYATTTAAGGTTATCAAAGTTGCACTTACGAAATATTTATTTCATACctcg
tgcctcTTGGTCTTTCTTTGTGTTGTTCATTTTTGTACATATGTGAAATGCAGTATAATCTTGATAAATA
TATTATTCT

> Locus F3'H-C341. GenBank accession: JX630066 [organism=Citrus reshni] PCR product flavonoid 3',5'-hydroxylase genomic DNA.

CGCCGGTGCACCCACTTGGCCTACGACGCTCAAGACATGGTCTTTGCTGATTATGGTCCGAGGTGGAAAC
TCTTAAGAAAGATAAGCAATCTGCACATGCTTGGTGGAAAAGCCCTATATGATTGGAGTAACGTGCGTAA
CATTGAGCTAGGCCACATGCTTCGAGCCATTTGTGAGTCTAGCCAGCGAAACGAGCCCGTGGTGGTCCCG
GAGATGTTGACGTACGCCATGGCAAACATGATAGGTCAAGTCATACTAAGCCGTCGAGTTTTTGTGACCA
AAGGGACAGAATCTAATGAGTTTAAGGACATGGTGGTAGAGCTCATGACGTCAGCTGGAT**[T/A]**TTTCA
ACATTGGTGATTTTATACCCTCGATTGCTTGGTTGGATTTACAAGGGATCGAGCGTGGGATGAAGAAATT
ACATAACAGATTTGATGTCCTGTTAACAAAGATGATTGAAGAGCACATGGCTTCAACTWATGAACGTAAA
AGGAAGCCAGATTTTCTCGACATTGTCATGGCTAATAGAGAAAATTCTGATGGAGAAAGGCTCACCATCA
CCAACATCAAAGCACTTCTCTTGGTAATTTGTGTCTTCAAAACTTTACCCTTTTTTTTATCTCACTTTTG
TATTTATTATTACGCTCATGYATTTAAGGTTATCAAAGTTGCACTTACGAAATATTTATTTCATACctcg
tgcctcTTGGTCTTTCTTTGTGTTGTTCATTTTTGTACATATGTGAAATGCAGTATAATCTTGATAAATA
TATTATTCT

> Locus PEPC-M316. GenBank accession: JX630067 [organism=Citrus reshni] PCR product phosphoenolpyruvate carboxylase genomic DNA.

AATTttAACTCTCCTTGTCAAGTTTGAACCATAAACTGCCAATTATTGATCTATTTGTGATGTTCCACAA
ATGACTTTTGTAGGAACACTAAACCTCTGAAGTCCATTCTACCTTTWAYRCCTTAATTAAAGAATTTTGA
AAATTCWTTTTGTCTAAATGATTTTGAACAATCGGCTAATGGTAGATATTGTACCAACTTTTTATATGTA
ATATGAAATTTTGGTTATTTATGTAGYCTTATTTATTGGAAAGTGCATTTAAGAACTGAGAAGGCATAGA
ATATTCCA**[T/C]**TAGGTTTGAAGAAATTCATTGCTCTTTAAGTCAGCTTTAAGTGAATATCCTTGTTAT
AAACTTTAGTGAGAGTGAATGCATTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTT
TCCCAYAACAGACTAGCTRAGCTTCAATTTTGATTTTCTTTTCTGAATGARTTTTGAAAATATTCGATAG
GACAATACTGAAATTTTTGCATTGTGGCTCTCACTTCTTATTTGATTTAATATTTAGAGAMAATTYMTTT
TTATTAATTTGATTTMTTTYTTCCTATAGTTCCTGGAGCCTCTAGARCTCTGTTACAGATCACTCTGTGC
TTGTGGTGATCGGCCAATAGCCGATGGAAGCCTTCTTGATT

> Locus PEPC-C328. GenBank accession: JX630067 [organism=Citrus reshni] PCR product phosphoenolpyruvate carboxylase genomic DNA.

AATTttAACTCTCCTTGTCAAGTTTGAACCATAAACTGCCAATTATTGATCTATTTGTGATGTTCCACAA
ATGACTTTTGTAGGAACACTAAACCTCTGAAGTCCATTCTACCTTTWAYRCCTTAATTAAAGAATTTTGA
AAATTCWTTTTGTCTAAATGATTTTGAACAATCGGCTAATGGTAGATATTGTACCAACTTTTTATATGTA
ATATGAAATTTTGGTTATTTATGTAGYCTTATTTATTGGAAAGTGCATTTAAGAACTGAGAAGGCATAGA
ATATTCCAYTAGGTTTGAAG**[G/A]**AATTCATTGCTCTTTAAGTCAGCTTTAAGTGAATATCCTTGTTAT
AAACTTTAGTGAGAGTGAATGCATTGGAGTCTCTCTTCCAGCAATTTGCTATTTTATATGAAGTTCTCTT
TCCCAYAACAGACTAGCTRAGCTTCAATTTTGATTTTCTTTTCTGAATGARTTTTGAAAATATTCGATAG
GACAATACTGAAATTTTTGCATTGTGGCTCTCACTTCTTATTTGATTTAATATTTAGAGAMAATTYMTTT
TTATTAATTTGATTTMTTTYTTCCTATAGTTCCTGGAGCCTCTAGARCTCTGTTACAGATCACTCTGTGC
TTGTGGTGATCGGCCAATAGCCGATGGAAGCCTTCTTGATT

> Locus SOS1-M50. GenBank accession: JX630068 [organism=Citrus reshni] PCR product Salt Overly Sensitive 1 genomic DNA.

TATGTTTACCCACTGGACTTTTTCAGGTTTTGCATGTTGTCAAAACCAG**[G/A]**CAAGTAACTTACTCAG
TACTAAACCATTTGATTGATTACATCCAAAATCTTGAGAAGGTTGGCTTGTTAGAAGAAAAGGAGATGCT
TCATCTTCATGATGCTGTCCAGGTATCTTTTTTTTGCATTGATCTTCATTCTATGCACTATACTTTATTT
AGTTCTTTGTACAATTTAGTATTATTTTCTTGCAGTCTGACTTGAAAAGGCTTCTAAGGAATCCTCCTTT
GGTGAAGTTTCCCAAAATAAGTGATTTGATTTGTGCCCATCCCTTGCTAAGGGAGCTTCCTCCCAGTGTC
CGTGAACCACTTGAACTTTCCACAAAAGAAATCATGAAACTCAGTGGCATGACACTGTACAGGGAGGGGT
CCAAGCCAAGTGGTATCTGGCTTATATCTAACGGTGTTGTTAAGGTAAATAATGTGTTTACATGGAAAAA
TTGTATCCT

> Locus CCC1-M85. GenBank accession: JX630069 [organism=Citrus reshni] PCR product cation-chloride cotransporter genomic DNA.

TATGTCAGAAGGCTTCCGTGGAATTGTCCAGACCATGGGTCTTGGTAATCTCAAGCCCAACATTGTGGTT
ATGAGGTATCCAGA**[G/A]**ATATGGCGCCGTGAAAACCTTACTGAAATCCCAGCCACCTTTGTTGGAATA
ATTAATGACTGTATTGTTGCTAACAAGGCYGTTGTTATTGTCAAGGGCCTTGATGAATGGCCCAATGAGT
ACCAAAGGCAATATGGTACAATCGATTTGTATTGGATTGTRAGAGACGGAGGTCTCATGCTCTTACTCTC
TCAGCTCCTGCTTACAAAGGAGAGCTTTGAAAGCTGTAAGATTCAAGTCTTCTGCATTGCTGAGGAGGAT
TCAGATGCAGCGGTGCTGAAGGCTGATGTAAAGAAGTTCCTATATGATCTTCGGATGCAGGCTGAAGTTA
TTGTTATATCTATGAAATCATGGGATGAGCAAACAGAGAATGGACCTCAACAAGATGAATCATTGGATGC
TTTTATTGCTGCTCAGCATCGGATTAAAAATTACCTGGCTGAAATGAAGGCTGAAGCTCAGAAATCAGGG
ACTCCGTTGATGGCTGATGGGAAGCCGGTGGTCGTGAATGAGCAACAGGTGGAGAAGTTTCTTTACACAA
CATTGAAGCTGAATTCGACAATACTGAGACACTCGAGAATGGCTGCAGTTGTGCTTGTTAGTCTACCGCC
GCCTCCGATCAACCACCCAGCTTACTGCTACATGGAATACATGGATTTGTTAGTAGAGAATGTGCC

> Locus CCC1-P727. GenBank accession: JX630069 [organism=Citrus reshni] PCR product cation-chloride cotransporter genomic DNA.

TATGTCAGAAGGCTTCCGTGGAATTGTCCAGACCATGGGTCTTGGTAATCTCAAGCCCAACATTGTGGTT
ATGAGGTATCCAGAGATATGGCGCCGTGAAAACCTTACTGAAATCCCAGCCACCTTTGTTGGAATAATTA
ATGACTGTATTGTTGCTAACAAGGCYGTTGTTATTGTCAAGGGCCTTGATGAATGGCCCAATGAGTACCA
AAGGCAATATGGTACAATCGATTTGTATTGGATTGTRAGAGACGGAGGTCTCATGCTCTTACTCTCTCAG
CTCCTGCTTACAAAGGAGAGCTTTGAAAGCTGTAAGATTCAAGTCTTCTGCATTGCTGAGGAGGATTCAG
ATGCAGCGGTGCTGAAGGCTGATGTAAAGAAGTTCCTATATGATCTTCGGATGCAGGCTGAAGTTATTGT
TATATCTATGAAATCATGGGATGAGCAAACAGAGAATGGACCTCAACAAGATGAATCATTGGATGCTTTT
ATTGCTGCTCAGCATCGGATTAAAAATTACCTGGCTGAAATGAAGGCTGAAGCTCAGAAATCAGGGACTC
CGTTGATGGCTGATGGGAAGCCGGTGGTCGTGAATGAGCAACAGGTGGAGAAGTTTCTTTACACAACATT
GAAGCTGAATTCGACAATACTGAGACACTCGAGAATGGCTGCAGTTGTGCTTGTTAGTCTACCGCCGCCT
CCGATCAACCACCCAGCTTACTGCTA**[T/C]**ATGGAATACATGGATTTGTTAGTAGAGAATGTGCC

> Locus TRPA-M593. GenBank accession: JX630070 [organism=Citrus reshni] PCR product vacuolar citrate/H+ symporter genomic DNA.

CACTTCACATTGAAACCATTTTCACACCAAACAATTTCTACATCTTCCTGGGACCTCTCCTGTGCGCTGT
TATATGTGTATGTGTGAAGCTCGATGGGCAGGCGACAAGCAGGAACATGTTGGGTATTCTTGCTTGGGTC
TTCGCTTGGTGGCTCACGGAGGCCGTACCCATGCCCATTACCTCTATGGCGCCTCTGTTTCTGTTCCCTC
TGTTTGGTATTTCTTCTGCTGATGCTGTTGCTCATTCTTACATGGATGATGTTATTGCCCTCGTTCTTGG
TAGCTTTATTCTTGCTCTCGCCGTTGAGCACTACAACATTCACAGAAGATTGGCCTTAAATGTAAGTTCC
CATAATGCATCATCATCATCATGTCATTAATCGTTACGATTTCTTTTTCAGAAAAATTATCAGTGACAAA
AGATGAATTAATTATGTATGGACAATCCTATACCATATAATATATATTAATAACTACAGATAACTATTCT
ATTCTGTGGAGAGCCAATGAATCCGCCCTTGCTGCTTCTTGGGATATGTGGCACGACAGCATTCGTGAGC
ATGTGGATGCATAACGTGGCAGCAGCAGTGAT**[C/G]**ATGATGCCAGTGGCCACTGGGATCTTACAGAAC
TTGCCAGAGGTTCATCTTCAATCAACCCTTGTTAGGAAGTATTGCAAAGCTGTGGTGCTCGGGGTCATCT
ACTCTGCAGCCGTAGGAGGGATGAGCACACTTACTGGAACAGGTGTTAATCTAATATTGGTCGGGATGTG
GAAGACCTATTTTCCAGAAGCAAACCCGT

> Locus INVA-M437. GenBank accession: JX630071 [organism=Citrus reshni] PCR product acid invertase genomic DNA.

ATTCATGGTTATTTATTTATAATTGAGCTCCCCTTTTGCTTAATATATTAAAGCAGTAACAACTTTGGGT
AATATGCTACAGGGCATTCCAAGGACAGTGGCGCTTGATACAAAAACTGGTAGTAATCTCCTYCAATGGC
CAGTGGAGGAAGTAGACAGTTTGCGATTGACCAGCAAAGAATTTAAAAAGATTGAGCTCAAGCCAGGGTC
AGTGATGCCGCTTGATGTTGGCTCAGCTACTCAGGTATGGAGATAGAGATACATTTATGCTTAATTAGTT
TGTCGATATCTCAATTTGAAAAGCACAAAGTAGGCAAATATAGCTTACATGGAAATGTTTGGGCAATGTG
AACAGCTGGACATAGTGGCCGAGTTTGAGCTAGACAAGGCGGCTTTAGAGAAAACAGCGGAGTCCAATGT
GGAGTTTAGCTGCAGTTCCAGC**[T/C]**AAGGATCTGCTGAACGCGGAGCATTAGGCCCCTTTGGCCTTCT
GGTTCTTGCAGATGACAGCCTAWCCGAGCAAACTCCAGTCTATTTCTACATTGCGAAAGGAAAGGATGGA
AGTCTCAAGACTTACTTCTGCACTGATCAATCAAGGTACCGTATTAATTACATGACTYGACTCTTGCATC
AAATTAAATCAARCCACGTGCAATGGTGTAATCCATTACTTAGCGCATTGTTAATTTCTTGTAGATCTTC
TGAGGCAAATGATGTCAATAAATCAAAATATGGTAGCTTTGTTCCAGTACTGGAAGGCGAGAAATTCTCA
ATGAGAGTATTGGTGAGCATATATCATGTTATTGTCCAAACGAACACATGTACATGTTGGCACTGTCAAT
AGAATCCTCACAATCAATTTGGAACATTGTGTATGTATATTTGCAGGTGGATCATTCGATAGTCGAA

> Locus INVA-P855. GenBank accession: JX630071 [organism=Citrus reshni] PCR product acid invertase genomic DNA.

ATTCATGGTTATTTATTTATAATTGAGCTCCCCTTTTGCTTAATATATTAAAGCAGTAACAACTTTGGGT
AATATGCTACAGGGCATTCCAAGGACAGTGGCGCTTGATACAAAAACTGGTAGTAATCTCCTYCAATGGC
CAGTGGAGGAAGTAGACAGTTTGCGATTGACCAGCAAAGAATTTAAAAAGATTGAGCTCAAGCCAGGGTC
AGTGATGCCGCTTGATGTTGGCTCAGCTACTCAGGTATGGAGATAGAGATACATTTATGCTTAATTAGTT
TGTCGATATCTCAATTTGAAAAGCACAAAGTAGGCAAATATAGCTTACATGGAAATGTTTGGGCAATGTG
AACAGCTGGACATAGTGGCCGAGTTTGAGCTAGACAAGGCGGCTTTAGAGAAAACAGCGGAGTCCAATGT
GGAGTTTAGCTGCAGTTCCAGCGAAGGATCTGCTGAACGCGGAGCATTAGGCCCCTTTGGCCTTCTGGTT
CTTGCAGATGACAGCCTAWCCGAGCAAACTCCAGTCTATTTCTACATTGCGAAAGGAAAGGATGGAAGTC
TCAAGACTTACTTCTGCACTGATCAATCAAGGTACCGTATTAATTACATGACTYGACTCTTGCATCAAAT
TAAATCAARCCACGTGCAATGGTGTAATCCATTACTTAGCGCATTGTTAATTTCTTGTAGATCTTCTGAG
GCAAATGATGTCAATAAATCAAAATATGGTAGCTTTGTTCCAGTACTGGAAGGCGAGAAATTCTCAATGA
GAGTATTGGTGAGCATATATCATGTTATTGTCCAAACGAACACATGTACATGTTGGCACTGTCAATAGAA
TCCTCACAA**[T/C]**CAATTTGGAACATTGTGTATGTATATTTGCAGGTGGATCATTCGATAGTCGAA

> Locus MDH-MP69. GenBank accession: JX630072 [organism=Citrus reshni] PCR product malate dehydrogenase genomic DNA.

GCCTTTGGCCCCAAGGCAGGCCAACTTCCACAGTCAAAACCCTCTGGTGTGAGGTTCAACTCCAAGAA**[A
/C]**TCACTTGTGAGTTTCAGTGGCCTCAAGGCAGTGACATCAGTTATCTGTGAATCAGATACCTCTTTCT
TGAACAAGGAGAGTTGTTCAGCTCTTCGAAGCACTTTTGCAAGAAAAGCCCAAAGTTCAGAGCAGAGGCC
TCAGAATGCCCTACAGCCTCAGGCTTCTTTTAAAGTAGCAGTTCTTGGAGCTGCTGGTGGAATAGGTCAA
CCCTTAGCACTTCTAATCAAGATGTCCCCACTAGTATCAGCCCTTCACCTCTATGATGTAATGAATGTCA
AGGGAGTTGCTGCTGACCTCAGTCACTGCAACACTCCCTCTCAAGTTCTGGATTTCACAGGACCTGAAGA
ATTAGCCAGTGCTTTGAAAGGGGTGAATGTCGTCGTCATACCTGCTGGAGTTCCAAGAAAGCCTGGGATG
ACCCGTGATGACCTCTTCAACATCAACGCCAATATAGTAAAGACCTTGGTTGAGGCTGTTGCTGATAACT
GCCCTGATGCCTTCATCCATATTATCAGCAATCCAGTTAATTCAACAGTGCCAATTGCTGCAGAAGTTCT

GAAGCAGAAGGGTGTTTATGATCCGAAGAAGCTTTTTGGTGTTACCACACTGGATGTCGTGAGAGCAAAC
ACCTTTGTTGCTCAAA

> Locus MDH-M519. GenBank accession: JX630072 [organism=Citrus reshni] PCR product malate dehydrogenase genomic DNA.

GCCTTTGGCCCCAAGGCAGGCCAACTTCCACAGTCAAAACCCTCTGGTGTGAGGTTCAACTCCAAGAAAT
CACTTGTGAGTTTCAGTGGCCTCAAGGCAGTGACATCAGTTATCTGTGAATCAGATACCTCTTTCTTGAA
CAAGGAGAGTTGTTCAGCTCTTCGAAGCACTTTTGCAAGAAAAGCCCAAAGTTCAGAGCAGAGGCCTCAG
AATGCCCTACAGCCTCAGGCTTCTTTTAAAGTAGCAGTTCTTGGAGCTGCTGGTGGAATAGGTCAACCCT
TAGCACTTCTAATCAAGATGTCCCCACTAGTATCAGCCCTTCACCTCTATGATGTAATGAATGTCAAGGG
AGTTGCTGCTGACCTCAGTCACTGCAACACTCCCTCTCAAGTTCTGGATTTCACAGGACCTGAAGAATTA
GCCAGTGCTTTGAAAGGGGTGAATGTCGTCGTCATACCTGCTGGAGTTCCAAGAAAGCCTGGGATGACCC
GTGATGACCTCTTCAACATCAACGCCAA**[T/C]**ATAGTAAAGACCTTGGTTGAGGCTGTTGCTGATAACT
GCCCTGATGCCTTCATCCATATTATCAGCAATCCAGTTAATTCAACAGTGCCAATTGCTGCAGAAGTTCT
GAAGCAGAAGGGTGTTTATGATCCGAAGAAGCTTTTTGGTGTTACCACACTGGATGTCGTGAGAGCAAAC
ACCTTTGTTGCTCAAA

> Locus ATMR-C372. GenBank accession: JX630073 [organism=Citrus reshni] PCR product MRP-like ABC transporter genomic DNA.

CAGGTAGCTGGCCTTAGATTATTACTTATTACAGTTTCTTGAAAACTGATATAAATATATTTGTTTGAGC
AGCCACGGATCATGTTCAGTACCAACTGTTAAATTACAACTGATCTAGGCCTCCTCAATTGATATTGCTT
GGGATAAATTACTGATTATTTACCATAACATGTTAAAAACTTTGGCAAACAGGTCAGATATCGCTCCAAC
ACTCCTCTGGTTCTCAAAGGTATTACACTCAGCATTCACGGGGGAGAGAAGATTGGTGTAGTTGGGCGTA
CAGGAAGTGGGAAGTCAACTTTAATTCAAGTTTTCTTTAGGCTGGTGGAGCCTTCAGGAGGGAGAATCAT
TATTGATGGAATCGACATTTC**[G/A]**TTGTTGGGGCTTCATGACCTAAGGTCTCGCTTTGGGATCATTCC
TCAAGAACCTGTCCTTTTTGAAGGAACTGTGAGAAGCAACATTGATCCAATTGGTCAGTATTCAGATGAA
GAAATCTGGAAGGTATGCCATTCCTTTTTTCTGATATGTGTCTCCTACATTTATGATCAAAGTTTGTGGG
TCTGTTTGCTGCATTAGCTAACTTATTATTATTTTTGTAGAGCCTCGAGCGATGTCAACTTAAAGATGTG
GTAGCTGCAAAGCCTGATAAACTCGATTCTTTAGGTAACTTCACTTCCTCCCTTTTCCTTGAATTTTCAG
TTTGATTTAATGGAAGTCATATGTATCTTTTTAGAAGCTAAAACATGCCAAAATGTTGAACTTTGTAGTG
GCTGATAG

> Locus ATMR-M728. GenBank accession: JX630073 [organism=Citrus reshni] PCR product MRP-like ABC transporter genomic DNA.

CAGGTAGCTGGCCTTAGATTATTACTTATTACAGTTTCTTGAAAACTGATATAAATATATTTGTTTGAGC
AGCCACGGATCATGTTCAGTACCAACTGTTAAATTACAACTGATCTAGGCCTCCTCAATTGATATTGCTT
GGGATAAATTACTGATTATTTACCATAACATGTTAAAAACTTTGGCAAACAGGTCAGATATCGCTCCAAC
ACTCCTCTGGTTCTCAAAGGTATTACACTCAGCATTCACGGGGGAGAGAAGATTGGTGTAGTTGGGCGTA
CAGGAAGTGGGAAGTCAACTTTAATTCAAGTTTTCTTTAGGCTGGTGGAGCCTTCAGGAGGGAGAATCAT
TATTGATGGAATCGACATTTCATTGTTGGGGCTTCATGACCTAAGGTCTCGCTTTGGGATCATTCCTCAA
GAACCTGTCCTTTTTGAAGGAACTGTGAGAAGCAACATTGATCCAATTGGTCAGTATTCAGATGAAGAAA
TCTGGAAGGTATGCCATTCCTTTTTTCTGATATGTGTCTCCTACATTTATGATCAAAGTTTGTGGGTCTG
TTTGCTGCATTAGCTAACTTATTATTATTTTTGTAGAGCCTCGAGCGATGTCAACTTAAAGATGTGGTAG
CTGCAAAGCCTGATAAACTCGATTCTTTAGGTAACTTCACTTCCTCCCTTTTCCTTGAATTTTCAGTTTG
ATTTAATGGAAGTCATATGTATCTTTT**[T/C]**AGAAGCTAAAACATGCCAAAATGTTGAACTTTGTAGTG
GCTGATAG

> Locus CHS-P57. GenBank accession: JX630074 [organism=Citrus reshni] PCR product chalcone synthase genomic DNA.

GGCCTCCGTGTTGCTAAAGACATAGCTGAAAACAACCCTGGAAGCCGCGTTTTGCT**[T/A]**ACCACTTCT
GAAACTACCATACTTGGGTTTCGCCCACCAAACAAGTCCCGCCCTTATGACCTTGTTGGGGCAGCTCTCT
TTGGTGATGGAGCTGCTGCTGTGATCGTTGGAGCTGACCCATTCCTGGATAAAGAGTCTTCTCCTTTCAT
GGAACTTAACTATGCAGTCCAACAATTCTTACCAGGGACACAGAATGTCATCGATGGGCGTCTTTCTGAA
GAGGGTATAAACTTCAAGCTTGGCAGGGACCTTCCTCAGAAGATTGAAGAAAATATTGAGGAGTTTTGCA
AGAAGCTCATGGCCAAAGCTGGTTTACAAGATTTCAATGATTTGTTCTGGGCAGTTCATCCTGGAGGACC
GGCAATTCTGAACCGACTGGAAAGCAATCTCAAGTTGAATAATCAGAAGCTTGAATGCAGCAGGAGGGCA
TTGATGGATTATGGGAATGTGAGCAGCAACACTATCTTTTATGTTATGGATTATATGAGGGAGGAGTTGA
AGAGGAAAGGAGATGAGG

> Locus CHS-M183. GenBank accession: JX630074 [organism=Citrus reshni] PCR product chalcone synthase genomic DNA.

GGCCTCCGTGTTGCTAAAGACATAGCTGAAAACAACCCTGGAAGCCGCGTTTTGCTTACCACTTCTGAAA
CTACCATACTTGGGTTTCGCCCACCAAACAAGTCCCGCCCTTATGACCTTGTTGGGGCAGCTCTCTTTGG
TGATGGAGCTGCTGCTGTGATCGTTGGAGCTGACCCATTCCT**[G/C]**GATAAAGAGTCTTCTCCTTTCAT
GGAACTTAACTATGCAGTCCAACAATTCTTACCAGGGACACAGAATGTCATCGATGGGCGTCTTTCTGAA
GAGGGTATAAACTTCAAGCTTGGCAGGGACCTTCCTCAGAAGATTGAAGAAAATATTGAGGAGTTTTGCA
AGAAGCTCATGGCCAAAGCTGGTTTACAAGATTTCAATGATTTGTTCTGGGCAGTTCATCCTGGAGGACC
GGCAATTCTGAACCGACTGGAAAGCAATCTCAAGTTGAATAATCAGAAGCTTGAATGCAGCAGGAGGGCA
TTGATGGATTATGGGAATGTGAGCAGCAACACTATCTTTTATGTTATGGATTATATGAGGGAGGAGTTGA
AGAGGAAAGGAGATGAGG

> Locus CHI-M598. GenBank accession: JX630075 [organism=Citrus reshni] PCR product chalcone isomerase genomic DNA.

TATATTATAATCAATTATTTTCCACATTAATTAACTAATAATAATTTTGAAGAATACTAAAGAGTTTATA
CATTTCTTTTTTCCTCTTGCGTTACGTGTAATGATAATAAATTAACAATACAGGTGCATTAAATATTTAA
ATTCACACTATCCGTATGGGAATCCTTTTCCGTCATAAACGCTGCTTAAAGAGTAGTGAACGTCAGTACT
WCACTCAAAATCTAAAACAGAATCCAACAGAAGAMACCAACGGCGAAAATCCGTTACCTGTGACGACGTC
TCTGAAGAACTCAACGGATTCCGTCAACTCCTCYGCAGTCTTCCCCTTCCATTTGCCGGCGAGTAACGGC
ACGGCGTYMTCCTCCAAGTACACTCCTATCGCCGTGAACTTCACGAACTTCCCTTCAATCTCCAATCCTC
TCTCCCCTGCCACGTCAGCGTCAAGAGCACCGAGACGTTAAAAACAAGTGAAATACAATGATGAAACAAM
AGTCAAATCATACAATCCGCCGGYGGTGCGCAGGRTACTAGCATACTACTAACCTGCGCCGCCGAGGAAA
TGCGACTTTGTTGATCCTGGAGGTTGCA**[C/G]**GGACGGCGTGAAAGTGACGTTCTCGACCTGCAGTTCG
GTGACGGACGGTGAGGG

> Locus PKF-C64. GenBank accession: JX630076 [organism=Citrus reshni] PCR product phosphofructokinase genomic DNA.

TTCAGTTTATAGCGAACTCCAAACGAGTCGAATCGATCACGCACTCCCTCTCCCTTCTGTTCT**[C/A]**AA
AAACCCTTTCAAAATCGTCGATGGCCCCGCTAGCTCCGCCGCCGGCAATCCAGGTCAGTTTCGTACTACC
ACTTCATCAATAAAACAATTTTCGTCCGTAACATTACAGATTCAAGATCTTTTCTTTTTGTCATATATAG
ATGAGATTGCGAAATTGTTTCCAAATCTGTTCGGGCAACCGTCCGCATTGTTGGTGCCGAACGGTGCTGA
CGCGGTGCGATCTGATGAGAAGTTGAAAATCGGCGTCGTCTTGTCTGGAGGTCAGGCGCCAGGTGGACAC
AATGTGATCTCTGGAATCTATGGTGAGTATAAATCTGAAAATGTAATATAAGCGTGATTTGTGTGAAAAT
TGGCCTTTTAAACGTGATTTGCTTATGTTTTGGTGGCAGATTACTTGCAGGATCGCGCGAAAGGGAGTGT
ACTGTATGGATTCAGAGGAGGTCCAGCTGGAATCATGAAGTGCAAATACGTTGAWCTAACTTCCGATTAT
ATTTATCCCTATAGAAACCAGGTATAACTTTGAGTATAATGTCAATGTTTTGAGTAATAAATAGTACATA
TTAATTAATCTTTGTAGAATTTAGACCATTTGCATATTAAATTTTGGCTTGACAAGTAAATGTAAAATGT
GCATGTTAAAGAAATGAAGTAAGCATCTAACCCCTTTG

> Locus PKF-M186. GenBank accession: JX630076 [organism=Citrus reshni] PCR product phosphofructokinase genomic DNA.

TTCAGTTTATAGCGAACTCCAAACGAGTCGAATCGATCACGCACTCCCTCTCCCTTCTGTTCTCAAAAAC
CCTTTCAAAATCGTCGATGGCCCCGCTAGCTCCGCCGCCGGCAATCCAGGTCAGTTTCGTACTACCACTT
CATCAATAAAACAATTTTCGTCCGTAACATTACAGATTCAAGA**[T/C]**CTTTTCTTTTTGTCATATATAG
ATGAGATTGCGAAATTGTTTCCAAATCTGTTCGGGCAACCGTCCGCATTGTTGGTGCCGAACGGTGCTGA
CGCGGTGCGATCTGATGAGAAGTTGAAAATCGGCGTCGTCTTGTCTGGAGGTCAGGCGCCAGGTGGACAC
AATGTGATCTCTGGAATCTATGGTGAGTATAAATCTGAAAATGTAATATAAGCGTGATTTGTGTGAAAAT
TGGCCTTTTAAACGTGATTTGCTTATGTTTTGGTGGCAGATTACTTGCAGGATCGCGCGAAAGGGAGTGT
ACTGTATGGATTCAGAGGAGGTCCAGCTGGAATCATGAAGTGCAAATACGTTGAWCTAACTTCCGATTAT
ATTTATCCCTATAGAAACCAGGTATAACTTTGAGTATAATGTCAATGTTTTGAGTAATAAATAGTACATA
TTAATTAATCTTTGTAGAATTTAGACCATTTGCATATTAAATTTTGGCTTGACAAGTAAATGTAAAATGT
GCATGTTAAAGAAATGAAGTAAGCATCTAACCCCTTTG

> Locus NADK2-M285. GenBank accession: JX630077 [organism=Citrus reshni] PCR product NADH kinase genomic DNA.

GTATTAGTGTTGAAAAAGCCTGGGCCAGCACTCATGGAAGAAGCTAAAGAGGTACCATGCAAAGTCTTTT
ATGTAATGTCAAAATAGTTTTTTGAATTTCACTTTGAAGCGATTCTTACATCTAAACAAATGTTTGTATT
AAGAAGATGCATATTATTGTGTTTCAGTTGCTCTACTTGATAATATGTCAACTAAACCTTCTACATTGCT
GATCTGTATTTCCATATCCACTCTATAAATATGTAGCTGCTATAACTCATTTCTAGATCTGATGAGCAGG
TTGC**[T/C]**TCTTTCTTGTATCACCAAGAGAAGATGAATATTCTTGTTGAGCCAGATGTGCAC

> Locus DFR-M240. GenBank accession: JX630078 [organism=Citrus reshni] PCR product dihydroflavonol 4-reductase genomic DNA.

GGCTATGCTGTTCGTGCTACTGTTCGCGATCCTGGTCTGGTTCATTTGCTGATCTTAATTAATTTTTGTT
AAACATTATCATAAATTTGCAAGTTCAACAGAATTTTAAAATGACTGTGGGCTATACGACAGATAACAAA
AAGAAAGTGAAACATTTGCTGGAGTTGCCGAAGGCAAGCACTCACCTGACTTTATGGAAAGCCGATTTAG
CCGAAGAGGGAAACTTTGATGAA**[G/C]**CGATTCGAGGCTGCACTGGAGTTTTTCATCTGGCCACGCCTA
TGGACTTTGAGTCCAAGGATCCTGAGGTATCGGTATCATCGTTACTCTTTAGTCTTTAGTTTCTTTTGAA
TAATACCAATAAATATTTATCCCGTTCATCGCAGATTTTTTTTTTTTTTTTTAATTTTGAA

> Locus LAPX-M238. GenBank accession: JX630079 [organism=Citrus reshni] PCR product ascorbate peroxidase genomic DNA.

TTTTGGGACGATCAGGCACCCAGATGAGCTTGCTCATGAGGCTAACAATGGTCTTGATATTGCTGTCAGG
CTCTTGGAGCCCATCAAGCAGCAGTTTCCTATCTTGTCCTACGCTGATTTCTATCAGGTAATTATTATTT
ATATCCAACTGTTGACTACAGAAAATGATTTGCTTTATGATCACTTTCTATGGATTACTTTGGATTGGTG
AATTGACCATGGTTTGTGTTTTATTTT**[C/G]**TTGAAGTTGGCTGGAGTTGTTGCCGTTGAAGTTACCGG
AGGGCA

> Locus PSY-M30. GenBank accession: JX630080 [organism=Citrus reshni] PCR product phytoene synthase genomic DNA.

GGGTCGTCCATTTGATATGCTTGATGCTG**[G/C]**ATTATCAGATACAGTAACCAAATTTCCTGTCGACAT
TCAGGTTAGACTATGTTTTCAAGATCAAATTAKATTTTAACAAAATGGTTGTTATAGTACTCTCTCTACT
CTCTTAAGTGTACTTGTATTAAATTAAAATAAGGAACAACTTCTGCTTTCTAATTGGTTTTTAAAACATT
AAGCCTTGATGCATAATGACAGACCTTATTTACATTTAATTGAGTCATRCCATTTTTGCATTTTCAATTT
ATCCAGGAGACCGAAGATGTGATGAGGTGATGCTACATGCTTACTAAGAACAATTCCGTTTCTCTAAATT
GCTCCATTATTTATTAGGACTCTTGAAGTTAACAGATAGCAATAGTGAATTTACTTCTCTGAAAAATTTA
CTTATCTGAAAACAAAGTTCTGCATGCTACCCTTCTCAATATTCAGACAAGAGTTTAATAGGCCTGCGAT
ATCTAAATAAAGGATGCAGTTTATGACTGAACCACCTCCCTGCAACGTTATCTTTTGTACCTTGATCTTT
CTTCAGAAAATGTTCTATTAAAAGTATTTCCAGTGGACCCTTAACCCAAT

> Locus PSY-C461. GenBank accession: JX630080 [organism=Citrus reshni] PCR product phytoene synthase genomic DNA.

GGGTCGTCCATTTGATATGCTTGATGCTGCATTATCAGATACAGTAACCAAATTTCCTGTCGACATTCAG
GTTAGACTATGTTTTCAAGATCAAATTAKATTTTAACAAAATGGTTGTTATAGTACTCTCTCTACTCTCT
TAAGTGTACTTGTATTAAATTAAAATAAGGAACAACTTCTGCTTTCTAATTGGTTTTTAAAACATTAAGC
CTTGATGCATAATGACAGACCTTATTTACATTTAATTGAGTCATRCCATTTTTGCATTTTCAATTTATCC
AGGAGACCGAAGATGTGATGAGGTGATGCTACATGCTTACTAAGAACAATTCCGTTTCTCTAAATTGCTC
CATTATTTATTAGGACTCTTGAAGTTAACAGATAGCAATAGTGAATTTACTTCTCTGAAAAATTTACTTA
TCTGAAAACAAAGTTCTGCATGCTACCCTTCTCAATATTC**[T/A]**GACAAGAGTTTAATAGGCCTGCGAT
ATCTAAATAAAGGATGCAGTTTATGACTGAACCACCTCCCTGCAACGTTATCTTTTGTACCTTGATCTTT
CTTCAGAAAATGTTCTATTAAAAGTATTTCCAGTGGACCCTTAACCCAAT

> Locus AOC-M290. GenBank accession: JX630081 [organism=Citrus reshni] PCR product ascorbate oxydase genomic DNA.

CTGACAAGATTCTTCCATGCCCACGTTGTAATAGTATGGACACTAAGTTCTGCTACTACAACAATTACAA
TGTAAACCAACCACGACACTTCTGCAAGAACTGCCAGAGATACTGGACAGCTGGTGGGACAATGCGTAAT
GTACCTGTAGGTGCTGGTCGTCGAAAGAACAAGAACTCAGCTTCTCACTACCGTCACATAACTGTCTCGG
AAGCCCTCCAAAACGTCCGAACTGATGTTCCGAATGGGGTCCACCATCCTGCGTTGAAAACTAATGGTAC
TGTACTTAC**[C/T]**TTTGGCTCAGATGCACCCCTTTGTGAATCAATGGCATCAGTTCTGAATATTGCTGA
TAAAACAATGAGGAATTGCACGAGAAATGGGTTTCATAAACCTGAGGAGTTGAGAATTCGACTTACTTAC

AGAGGTGGAGAAAATGGGGATAATTATGCACATGGATCTCCGGTGCCAGTTTCAAATTCAAAGGATGAGG
CAGGCAAAACTACTTCACAGGAGGCAGTTGTGCAGAATTGTCAAGGCTTCCCTCCTCATGTGGCTTGCTT
TCCTGGGGCTCCGTGGCCATACCCATGGAATTCGGCTCAATGGAGCCCTCCAGTTACCCCACCTGCGATC
CTTCCTCCAGGCTTCCCTATGCCATTCTACCCTCCAGCAGCTTACTGGG

> Locus AOC-C593. GenBank accession: JX630081 [organism=Citrus reshni]
PCR product ascorbate oxydase genomic DNA.

CTGACAAGATTCTTCCATGCCCACGTTGTAATAGTATGGACACTAAGTTCTGCTACTACAACAATTACAA
TGTAAACCAACCACGACACTTCTGCAAGAACTGCCAGAGATACTGGACAGCTGGTGGGACAATGCGTAAT
GTACCTGTAGGTGCTGGTCGTCGAAAGAACAAGAACTCAGCTTCTCACTACCGTCACATAACTGTCTCGG
AAGCCCTCCAAAACGTCCGAACTGATGTTCCGAATGGGGTCCACCATCCTGCGTTGAAAACTAATGGTAC
TGTACTTACTTTTGGCTCAGATGCACCCCTTTGTGAATCAATGGCATCAGTTCTGAATATTGCTGATAAA
ACAATGAGGAATTGCACGAGAAATGGGTTTCATAAACCTGAGGAGTTGAGAATTCGACTTACTTACAGAG
GTGGAGAAAATGGGGATAATTATGCACATGGATCTCCGGTGCCAGTTTCAAATTCAAAGGATGAGGCAGG
CAAAACTACTTCACAGGAGGCAGTTGTGCAGAATTGTCAAGGCTTCCCTCCTCATGTGGCTTGCTTTCCT
GGGGCTCCGTGGCCATACCCATGGAATTCGGC**[T/A]**CAATGGAGCCCTCCAGTTACCCCACCTGCGATC
CTTCCTCCAGGCTTCCCTATGCCATTCTACCCTCCAGCAGCTTACTGGG

> Locus DXS-C545. GenBank accession: JX630082 [organism=Citrus reshni]
PCR product 1-deoxyxylulose 5-phosphate synthase genomic DNA.

TTCATATGAAGAGTCTCTCTAAAGAGGTAAAARCGTGYGCGTCTGATTGATGGGAATCCTGTTCTTTCTT
GAGGATATTTCATTTGTTCATAACATAGTTCGTACAATTTTCAGGATCTTGAACAACTGGCAGCAGAGCT
TAGAGCAGATATTGTTAACAGTGTATCGAAGACAGGTGGGCATCTTAGTGCAAACTTAGGAGTGGTGGAG
CTAACACTTGCTTTGCATCGTGTTTTCAACACACCTGACGATAAAATTATATGGGATGTTGGCCATCAGG
TAATTAATTGAAGACACTTGTTAATTCGCTACTGCCCTGTCTCAAACGAATCATGGCTGAACAAATTAAA
GACCCAAACATATATCAGTGTTACTGAATGGCTGACCTTGAATCCTGCAGGCTTATGTACACAAAATTCT
GACTGGAAGAAGATCCAGAATGAACACCATGAGGAAGACTTCGGGGCTTGCAGGATTCCCCAAAAGAGA**[**
G/C]GAAAGCGTTCATGATGCATTTGGTGCAGGACATAGTTCCACAAGCATCTCTGCTGGTCTTGGTATG
TACTTCACTCTCTTAATATTTTCCTTTCATCAATCTAGAGAAATTGTAGGATGCAGAATACATAATTGAG
AATTTCCTAATCTAATCATATTTTTTAATAATAGGTATGGC

> Locus DXS-M618. GenBank accession: JX630082 [organism=Citrus reshni]
PCR product 1-deoxyxylulose 5-phosphate synthase genomic DNA.

TTCATATGAAGAGTCTCTCTAAAGAGGTAAAARCGTGYGCGTCTGATTGATGGGAATCCTGTTCTTTCTT
GAGGATATTTCATTTGTTCATAACATAGTTCGTACAATTTTCAGGATCTTGAACAACTGGCAGCAGAGCT
TAGAGCAGATATTGTTAACAGTGTATCGAAGACAGGTGGGCATCTTAGTGCAAACTTAGGAGTGGTGGAG
CTAACACTTGCTTTGCATCGTGTTTTCAACACACCTGACGATAAAATTATATGGGATGTTGGCCATCAGG
TAATTAATTGAAGACACTTGTTAATTCGCTACTGCCCTGTCTCAAACGAATCATGGCTGAACAAATTAAA
GACCCAAACATATATCAGTGTTACTGAATGGCTGACCTTGAATCCTGCAGGCTTATGTACACAAAATTCT
GACTGGAAGAAGATCCAGAATGAACACCATGAGGAAGACTTCGGGGCTTGCAGGATTCCCCAAAAGAGAG
GAAAGCGTTCATGATGCATTTGGTGCAGGACATAGTTCCACAAGCATCTCTGCTGGTCTTGGTATGTACT
TC**[G/A]**CTCTCTTAATATTTTCCTTTCATCAATCTAGAGAAATTGTAGGATGCAGAATACATAATTGAG
AATTTCCTAATCTAATCATATTTTTTAATAATAGGTATGGC

> Locus FLS-P129. GenBank accession: JX630083 [organism=Citrus reshni]
PCR product flavonol synthase genomic DNA.

TTCAAATGGCACAATTCCAGCAGAGTTCGTAAGACCCGAAAAAGAACAGCCAGCAAGCACAACGTACCAC
GGCCCCGCTCCTGAAATCCCCACGATCGATCTCGACGACCCCGTTCAAGACAGACTCG**[T/C]**ACGTTCC
ATCGCGGAAGCCAGCCGGGAGTGGGGGATTTTCCAGGTTACAAACCACGGGATACCTAGTGACCTCATCG
GTAAACTGCAAGCCGTCGGCAAAGAATTTTTTGAGCTCCCTCAGGAAGAGAAAGAAGTGTATTCTCGTCC
GGCTGATGCAAAAGACGTGCAAGGATACGGCACAAAGTTACAGAAAGAAGTCGAAGGAAAGAAATCTTGG
GTTGATCATCTCTTCCACAGGGTTTGGCCTCCGTCTTCTATCAACTACCGCTTCTGGCCCAACAACCCTC
CTTCTTACCGGTGAATGTTATGCATCTCTTATCTTTTTCAATTCTTTT

> Locus FLS-M400. GenBank accession: JX630083 [organism=Citrus reshni] PCR product flavonol synthase genomic DNA.

TTCAAATGGCACAATTCCAGCAGAGTTCGTAAGACCCGAAAAAGAACAGCCAGCAAGCACAACGTACCAC
GGCCCCGCTCCTGAAATCCCCACGATCGATCTCGACGACCCCGTTCAAGACAGACTCGTACGTTCCATCG
CGGAAGCCAGCCGGGAGTGGGGGATTTTCCAGGTTACAAACCACGGGATACCTAGTGACCTCATCGGTAA
ACTGCAAGCCGTCGGCAAAGAATTTTTTGAGCTCCCTCAGGAAGAGAAAGAAGTGTATTCTCGTCCGGCT
GATGCAAAAGACGTGCAAGGATACGGCACAAAGTTACAGAAAGAAGTCGAAGGAAAGAAATCTTGGGTTG
ATCATCTCTTCCACAGGGTTTGGCCTCCGTCTTCTATCAACTACCGCTT**[C/T]**TGGCCCAACAACCCTC
CTTCTTACCGGTGAATGTTATGCATCTCTTATCTTTTTCAATTCTTTT

> Locus LCYB-M480. GenBank accession: JX630084 [organism=Citrus reshni] PCR product lycopene β-cyclase genomic DNA.

AAGATTCAGAACCAGGAGCTTAGGTTTGGTCTCAAGAAGTCTCGTCAAAAGAGGAATATGAGTTGTTTCA
TTAAGGCTAGTAGTAGTGCTCTTTTGGAGCTAGTTCCTGAAACCAAGAAGGAAAATCTTGAATTTGAGCT
TCCCATGTATGACCCATCAAAGGGCCTTGTTGTAGACCTAGCAGTTGTCGGTGGTGGCCCAGCTGGGCTT
GCTGTTGCTCAGCAAGTTTCAGAGGCGGGGCTTTCGGTTTGCTCGATTGATCCATCTCCCAAATTGATTT
GGCCAAATAATTATGGTGTTTGGGTGGATGAATTTGAGGCCATGGATTTGCTTGATTGCCTTGATACTAC
TTGGTCTGGTGCTGTTGTGCACATTGATGATAATACAAAGAAGGATCTTGATAGACCTTATGGCAGAGTT
AATAGGAAGTTGCTGAAGTCGAAAATGCTGCAAAAATGCATAACCAATGGTGTTAAGTT**[G/C]**CACCAA
GCTAAAGTTATTAAGGTTATTCATGAAGAGTCCAAATCTTTGTTGATTTGCAATGATGGTGTGACAATTC
AGGCTGCCGTGGTTCTTGATGCTACGGGATTCTCTAGGTGTCTTGTGCAGTATGATAAACCCTATAATCC
AGGTTACCAAGTGGCATATGGAATACTAGCTGAGGTAGAAGAGCACCCGTTTGATTTAGACAAGATGGTT
TTCATGGATTGGAGAGATTCGCATCTGAACAACAATTCGGAGCTCAAAGAGGCAAATAGCAAAATTCCTA
CTTTTCTTTATGCCATGCCCTTTTCGTCAAACAGGATATTTCTTGAAGAGACTTCGCTAGTGGCGCGGCC
TGGAGTGCCAATGAAAGATATCCAGGAAAGAATGGTGGCTAGATTAAAGCACTTAGGCATAAAAGTTAGA
AGCATTGAAGAGGATGAGCATTGTGTCATTCCGAT

> Locus LCYB-P736. GenBank accession: JX630084 [organism=Citrus reshni] PCR product lycopene β-cyclase genomic DNA.

AAGATTCAGAACCAGGAGCTTAGGTTTGGTCTCAAGAAGTCTCGTCAAAAGAGGAATATGAGTTGTTTCA
TTAAGGCTAGTAGTAGTGCTCTTTTGGAGCTAGTTCCTGAAACCAAGAAGGAAAATCTTGAATTTGAGCT
TCCCATGTATGACCCATCAAAGGGCCTTGTTGTAGACCTAGCAGTTGTCGGTGGTGGCCCAGCTGGGCTT
GCTGTTGCTCAGCAAGTTTCAGAGGCGGGGCTTTCGGTTTGCTCGATTGATCCATCTCCCAAATTGATTT
GGCCAAATAATTATGGTGTTTGGGTGGATGAATTTGAGGCCATGGATTTGCTTGATTGCCTTGATACTAC
TTGGTCTGGTGCTGTTGTGCACATTGATGATAATACAAAGAAGGATCTTGATAGACCTTATGGCAGAGTT
AATAGGAAGTTGCTGAAGTCGAAAATGCTGCAAAAATGCATAACCAATGGTGTTAAGTTCCACCAAGCTA
AAGTTATTAAGGTTATTCATGAAGAGTCCAAATCTTTGTTGATTTGCAATGATGGTGTGACAATTCAGGC
TGCCGTGGTTCTTGATGCTACGGGATTCTCTAGGTGTCTTGTGCAGTATGATAAACCCTATAATCCAGGT
TACCAAGTGGCATATGGAATACTAGCTGAGGTAGAAGAGCACCCGTTTGATTTAGACAAGATGGTTTTCA
TGGATTGGAGAGATTCGCATCTGAACAACAATTCG**[G/C]**AGCTCAAAGAGGCAAATAGCAAAATTCCTA
CTTTTCTTTATGCCATGCCCTTTTCGTCAAACAGGATATTTCTTGAAGAGACTTCGCTAGTGGCGCGGCC
TGGAGTGCCAATGAAAGATATCCAGGAAAGAATGGTGGCTAGATTAAAGCACTTAGGCATAAAAGTTAGA
AGCATTGAAGAGGATGAGCATTGTGTCATTCCGAT

> Locus TSC-C80. GenBank accession: JX630085 [organism=Citrus reshni] PCR product trehalose-6-phosphate synthase genomic DNA.

GTGGCACCACCAGCACGCCGACCCTCATTTTGGCTCATGCCAGGCTAAAGAGCTTCTTGACCACTTGGAA
AATGTTCTT**[T/G]**CTAATGAGCCTGTTGTTGTCAAAAGAGGCCAACACATTGTTGAGGTCAAGCCACAG
GTATGTCAACATCAATCTTTTTAATACAGTTGTTACATAAACTTATATGATATGTTACGAAGAAACGGTG
TAATGCCCTTTGTTTTATATCTCATTGATATTGCTCGGATGATCATTCTAATGTGGTGAATTGGGTAGGG
AGTAAGCAAAGGCATTGTTGTAAAAAACTTGATTTCAACTATGCGAAGTAGGGGGAAGT

> Locus NCED3-M535. GenBank accession: JX630086 [organism=Citrus reshni] PCR product 9-cis-epoxy hydroxy carotenoid dyoxygenase 3 genomic DNA.

CCGTTTTGTTCAAGAACGTAGCTTAGGCCRCCCCGTATTCCCCAAAGCCATTGGCGAGCTTCACGGCCAC
ACGGGCATCGCAAGATTGCTTCTCTTCTACAGCAGAGCGCTCTTCGGTCTCGTTGACCCCAGCCACGGCA

CTGGCGTTGCCAACGCCGGCCTTGTTTACTTCAACAACCGTTTGTTGGCCATGTCTGAAGATGACTTGCC
TTATCACGTGCGCGTCACTCCATCCGGCGACCTCAAAACGGTCGGCCGTTTCKACTTCAGCGGCCAGCTC
AAGTCCACGATGATAGCTCATCCGAAAGTTGATCCCGTGACGGGTGATTTGTTTGCTCTGAGTTATGACG
TTGTCAAGAAGCCTTACTTGAAGTACTTTCGGTTTTCGCCCCAAGGGATCAAGTCTCCGGACGTTGAGAT
TCCCCTTGAGGAGCCTACAATGATGCATGATTTCGCAATCACTGAGAATTTTGTTGTGGTGCCTGACCAG
CAAGTGGTGTTCAAGTTGAATGAGATGATCCGAGGTGGCTCCCC**[T/G]**GTGATTTATGACAAGAACAAG
GTGT

> Locus HYB-C433. GenBank accession: JX630087 [organism=Citrus reshni]
PCR product β-Carotene hydroxylase genomic DNA.

GTGGCAAATGGAGGTACTTCAAACAAATCACACATGTCCTAATGTTATTGGTTGGTTRTATGAA
CAGAAAATTTCGCCCCTCTTTTGATGATGCTTACATGTTATGTATCCGTACAGGGTGGAGAGGTGCCTTT
A**[G/A]**CTGAAATGTTTGGCACATTTGCTCTCTCTGTTGGTGCTGCTGTAAGTTCAATCACCTTCTTCCT
TACAATGATTTGAAAACAAGACTAGAATTTTGGTTCTRATAGGAGCCGCGGTGGGGATGTTACAAACTTG
ATCGATCTTTAACATAAAAACTGTAAAAATGAGGGGCTTGTGTGAATTTTCAATGTGAAGGCCTTTTCTG
GCAAATTATATGATGATGATTCGCATTGGGTACC

CHAPTER 4

Genetic diversity and population-structure analysis of mandarin germplasm by nuclear (SSR, indel) and mitochondrial markers.

Andres Garcia-Lor, François Luro, Gema Ancillo, Luis Navarro, and Patrick Ollitrault

Submitted

Abstract

- *Background and Aims*

The mandarin horticultural varietal group is highly polymorphic. It is closely related with one of the basic taxa of the cultivated citrus (*Citrus reticulata*), but it also includes genotypes introgressed by other species. The precise contribution of ancestral species to the mandarin group is not known. The goals of this work were: 1) to characterise the mandarin germplasm using nuclear (SSR, indel) and mitochondrial markers; 2) to evaluate genetic diversity and detect redundancies; 3) to quantify the contributions of the citrus ancestral genomes to the mandarin germplasm; and 4) to determine the genetic structure within the mandarin group.

- *Methods*

Fifty microsatellite (SSR), 24 insertion-deletion (indel), and four mitochondrial (mtDNA) indel markers were analysed for 223 genotypes. The Structure software was applied to nuclear data to check and quantify potential interspecific introgressions in the mandarin germplasm and to determine the optimal number of clusters within it.

- *Key results*

The *C. maxima* and *Papeda* genomes were the main genomes introgressed in the *C. reticulata* background of the mandarin germplasm. By Structure analysis, seven clusters were revealed at the nuclear level (N) within the mandarin germplasm. Five of these clusters should be parental mandarin groups (N1–N5), and the other two included genotypes of known or supposed hybrid origin (N6 and N7). The contributions of these parental groups to the mandarin genotypes were estimated.

The mitochondrial indel analysis revealed four mitotypes in which mandarin and ‘mandarin-like’ genotypes were represented. Two cytoplasmic (C) groups clusterized pure mandarins (C1 and C2) while interspecific mandarin hybrids were found associated with the two other mitotypes (C3 and C4).

- *Conclusions*

This work provides new insights into the organisation of the mandarin germplasm and its structure at the nuclear and cytoplasmic level. These insights will be useful for better breeding and management of citrus germplasm collections.

INTRODUCTION

Citrus is the most important fruit crop in the world, with a production of 123,755,751 tons and a cultivated area of 8,643,502 ha (FAOSTAT, 2010). Among the commercial citrus fruits, mandarins are the second most important group in the fresh-fruit market worldwide. Spain is the second largest mandarin producer and the largest mandarin exporter in the world, with a total production of 1,708,200 tons in a cultivated area of 90,900 ha (FAOSTAT, 2010).

'Mandarin' is a common name given to most small, easy-peeling citrus fruits. This term includes interspecific hybrids, which make mandarins the most genetically and phenotypically polymorphic group of true *Citrus* (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012, 2013). Moreover, a recent phylogenetic study (Garcia-Lor *et al.*, 2013) revealed a close relationship between genus *Fortunella* and the mandarin group. Mandarin germplasm was classified as *C. reticulata* Blanco by Swingle and Reece (1967) and Mabberley (1997). On the contrary, Webber (1943) classified mandarin genotypes into four different groups: king, satsuma, mandarin, and tangerine. Tanaka (1954) divided mandarins into five groups that included 36 species, based on morphological differences in the tree, leaves, flowers, and fruits. Group 1 included *C. nobilis* Lour. (cultivars like 'King'), *C. unshiu* Marc. (satsumas) and *C. yatsushiro* Hort. ex Tanaka; group 2 included *C. keraji* Hort. ex Tanaka, *C. oto* Hort. ex Yuichiro and *C. toragayo* Hort. ex Yuichiro; group 3 included 14 species, including some of the most economically important varieties, such as *C. reticulata* ('Ponkan'), *C. deliciosa* Tenore ('Willowleaf' or 'Common mandarin'), *C. clementina* Hort. ex Tanaka (clementines) and *C. tangerina* Hort. ex Tanaka ('Dancy'); group 4 included *C. reshni* Hort. ex Tanaka ('Cleopatra'), *C. sunki* Hort. ex Tanaka ('Sunki') and *C. tachibana* (Mak.) Tanaka; and group 5 included the species *C. depressa* Hayata ('Shekwasha') and *C. lycopersicaeformis* (Lush.) Hort. ex Tanaka. Hodgson (1967) divided the mandarins into four species: *C. unshiu* (satsuma), *C. reticulata* ('Ponkan', 'Dancy', clementine), *C. deliciosa* ('Willowleaf') and *C. nobilis* ('King').

None of these citrus classification systems is perfect, but the Tanaka system seems better adapted to the horticultural features of each group, whereas the Swingle system simplifies it to the extreme. At present, *C. reticulata* (mandarin) is considered to be one of the four ancestral groups of the cultivated citrus (Barrett HC, 1976; Nicolosi *et al.*, 2000; Krueger and Navarro, 2007), along with *C. maxima* (Burm.) Merr. (pummelo), *C. medica* L. (citron) and *C. micrantha* Wester (papeda). The centre of diversification of *C. reticulata* is located in Asia, from Vietnam to Japan (Tanaka, 1954). This group is highly polymorphic, as revealed by molecular markers (Coletta Filho *et al.*, 1998; Ollitrault *et al.*, 2012a), chromosomal banding patterns (Yamamoto and Tominaga, 2003) and phenotypic characters, such as fruit pomology and the chemical variability of peel and leaf oils (Lota *et al.*, 2000; Fanciullino *et al.*, 2006), as well as tolerance to biotic and abiotic stresses. Several germplasm collections have been characterised by morphological characteristics and/or molecular markers (Koehler-Santos *et al.*, 2003; Tapia Campos *et al.*, 2005; Barkley *et al.*, 2006). This phenotypic and genetic variability reflects a long history of cultivation, in which many mutations and natural hybridisations have

given rise to the existing diversity within this mainly facultatively apomictic group. The intraspecific organisation of mandarins and the determinants of the group's phenotypic diversity remain poorly understood.

In addition to the taxonomic complexity of the mandarin group, the genotypes introduced in citrus germplasm collections are sometimes of doubtful origin. The origin of these genotypes can be from plant explorations in regions of natural genetic diversity, selection of new materials from hybridisations or mutations, or by exchange between germplasm collections (Krueger and Navarro, 2007). The assignation of a cultivar name and/or membership in a species can be done arbitrarily, with no molecular basis, leading to possible mistakes in assignation or duplication of material (Krueger and Navarro, 2007). For these reasons, molecular studies are important for detection of misidentifications and redundancies (Krueger and Roose, 2003).

In this work, we will use the term 'mandarin' in four different ways: 'mandarin' as a true species (*C. reticulata*; one of the four ancestors of the other cultivated genotypes); 'mandarin' according to the Swingle classification (*C. reticulata* [Sw]); 'mandarin' according to the Tanaka classification (17 species represented in this work, in which *C. reticulata* is included [Tan]); and 'mandarin-like' genotypes that are phenotypically similar to mandarins.

The goals of this work were: 1) to characterise the mandarin germplasm using nuclear (SSR, indel) and mitochondrial markers; 2) to evaluate genetic diversity and detect redundancies; 3) to quantify the contributions of the citrus ancestral genomes to the mandarin germplasm; and 4) to determine the structure within the mandarin group.

MATERIALS AND METHODS

Diversity analysis

Two-hundred-and-twenty-three genotypes were studied with regard to their nuclear diversity, using 50 SSR and 24 indel markers. Throughout the text, these genotypes will be referred to by identification number (ID), shown in Supplementary information 1. Genotype classification was performed according to the Swingle (Swingle and Reece, 1967) and Tanaka (Tanaka, 1954) systems. A summary of the genotypes used can be found in Table 1. Plant material for the analysis was collected from the germplasm collections of the Instituto Valenciano de Investigaciones Agrarias (IVIA, Valencia, Spain) and the Station de Recherches Agronomiques (CIRAD-INRA, Corsica, France). These genotypes belong to the four ancestral species (30 *C. reticulata* [Sw, mandarins], 11 *C. maxima* [pummelos], six *C. medica* [citrons], four *Papeda* [*C. ichangensis* Swingle, *C. histrix* D.C., *C. latipes* (Swingle) Tan. and *C. micrantha*], and four *Fortunella* [kumquats: *F. crassifolia* Swing., *F. hindsii* (Champ.) Swing., *F. japonica* (Thunb.) Swing., and *F. margarita* (Lour.) Swing.]. The 30 mandarin genotypes considered as *C. reticulata* by Swingle (1967) were considered by Tanaka (1977) as 17 species. The other genotypes (168 'mandarin-like' accessions, intra- and interspecific hybrids) were not assumed in any of the previously mentioned main taxa, in order to decipher their structure and determine whether their Tanaka classification in the germplasm-bank data was properly assigned in our databases. *Severinia buxifolia* was added as out-group for neighbour-joining analysis.

For the maternal-origin study, the same genotypes were analysed, including the ancestral species and interspecific hybrids.

Genotyping

Fifty SSR markers located along the nine linkage groups of the reference genetic map of clementine (Ollitrault *et al.*, 2012b) and 24 indel markers identified in a discovery panel representative of genus *Citrus* (Garcia-Lor *et al.*, 2012, 2013) were used (Supplementary information 2). To assess the maternal origin of the mandarin germplasm, four mitochondrial indel markers (*nad2*, *nad5*, *nad7*, and *rrn5/rrn18*; Froelicher *et al.*, 2011) were used.

Amplifications by polymerase chain reaction (PCR) and analyses with a capillary genetic fragment analyser (CEQ/GeXP Genetic Analysis System; Beckman Coulter, Fullerton, CA, USA) were performed as described in (Garcia-Lor *et al.*, 2012). The Genetic Analysis System software (GenomeLab GeXP, v. 10.0) was used for data collection and analysis.

Table 1. Summary of genotypes employed in the study, their classification based on Swingle, and their classification within our databases based on the Tanaka system.

Swingle system	Species name in databases based on the Tanaka system	NG/S	NGSA
C. reticulata hybrid	*C. amblycarpa*	2	2
C. reticulata	*C. deliciosa*	13	2
C. reticulata	*C. daoxianensis*	1	0
C. reticulata	*C. depressa*	5	2
C. reticulata	*C. erythrosa*	3	2
C. reticulata	*C. halimii*	1	0
C. indica	*C. indica*	1	1
C. hystrix	*C. hystrix*	1	1
C. ichangensis	*C. ichangensis*	1	1
C. ichangensis x *C. reticulata* var. *austera*	*C. junos*	1	0
C. limon	*C. karna*	1	0
C. reticulata	*C. kinokuni*	5	2
C. latipes	*C. latipes*	1	1
C. reticulata var. *austera*? x *Fortunella*?	*C. madurensis*	1	0
C. maxima	*C. maxima*	11	11
C. medica	*C. medica*	6	6
C. micrantha	*C. micrantha*	1	1
C. reticulata	*C. nobilis*	7	2
C. reticulata	*C. paratangerina*	2	2
C. reticulata	*C. reshni*	1	1
C. reticulata	*C. reticulata*	53	3
C. sinensis	*C. shunkokan*	1	0
C. reticulata	*C. suavissima*	1	1
C. reticulata	*C. succosa*	1	1
C. reticulata	*C. suhuiensis*	8	2
C. reticulata	*C. sunki*	4	2
C. tachibana	*C. tachibana*	1	1
C. reticulata	*C. tangerina*	11	2
C. sinensis	*C. tankan*	1	0
C. reticulata	*C. temple*	3	0
C. reticulata	*C. unshiu*	8	2
C. reticulata	Citrandarin	1	0
Fortunella	*Fortunella*	4	4
C. reticulata	Hybrid mandarin	30	0
C. reticulata	Tangelo	4	0
C. reticulata	Tangor	16	0
?	Bintangor	1	0
C. reticulata	*C. clementina*	3	0
?	Unknown	7	0

(NG/S) Number of genotypes per species; (NGSA) Number of genotypes from each species included within an ancestral population.

Data analysis

The allelic data obtained with the SSR, indel, and mtDNA markers was used to calculate a genetic dissimilarity matrix using the simple matching dissimilarity index (d_{i-j}) between pairs of accessions (units), with the Darwin5 software, version 5.0.159 (Perrier and Jacquemoud-Collet, 2006). Weighted neighbour-joining (NJ) analyses (Saitou and Nei, 1987) were computed with the same software to describe the population-diversity organisation, and robustness of branches was tested using 1000 bootstraps.

Population structure was inferred with the program Structure, v. 2.3.3 (http://cbsuapps.tc.cornell.edu/structure), which implements a model-based clustering method using genotype data (Pritchard *et al.*, 2000; Falush *et al.*, 2003). When there is a known population structure, it allows to calculate their contribution to genomes of genotypes of unknown origin. In cases of unknown population structure, the Structure program helps to

assign the optimal number of populations within the sample data set under study, based on the parameters of Evanno *et al.* (2005).

F-statistics were calculated with the program GENETIX, v. 4.03 (Belkhir *et al.*, 2002), based on the parameters of Wright (1969), and Weir and Cockerham (1984). Some other genetic population statistics were estimated from the allele data using the program PowerMarker, v. 3.25 (Liu and Muse, 2005).

RESULTS

SSR and indel analysis

Genetic-diversity statistics were calculated for each SSR and indel marker in the entire population (Supplementary information 2). Using the SSR markers, we detected 592 alleles. Allele numbers varied between five (TAA1) and 19 (mCrCIR02D04b). The average number of alleles and the H_e (expected heterozygosity) value per locus were 11.8 and 0.67, respectively. The whole population had an observed heterozygosity (H_o) of 0.61. F_w (Wright fixation index) values varied from −0.43 (CAC23) to 0.50 (MEST256). The average F_w value over all SSR loci was close to zero (0.05).

We detected a total of 80 alleles with the indel markers. Allele number per locus ranged from two (eight markers) to seven (IDHyb-1, IDDFR), with an average of 3.3. The values ranged from 0.02 (IDPEPC3) to 0.67 (IDDFR), with a median value of 0.20. F_w values varied from −0.50 (IDF'3H) to 0.94 (IDINVA1). The overall H_o and F_w value among all loci were 0.20 and 0.26, respectively.

Genetic population statistics within the whole population, i.e., all 'mandarin-like' genotypes of unknown or supposed hybrid origin and the 30 genotypes selected from all Tanaka species represented in our collections (Supplementary information 1), are summarised in Figure 1. Gene diversity (GD) and the H_o values were higher among SSR markers than indel markers, reflecting the higher maximum allele frequencies (MAF) of the latter. Comparing the whole population (AG), all 'mandarin-like' genotypes (AM), and mandarins from Tanaka species (MT), the mean allele number decreased at each step for SSR and indel markers (SSRs: AG = 11.84 > AM = 8.6 > MT = 6.76; indels: AG =3.33 > AM = 2.64 > MT = 2.02), and the GD was higher in AG than in AM or MT for both kinds of markers. For AG, Ho was slightly lower than He, leading to slightly positive Fw for SSR and indel markers. In AM and MT, Ho values were higher than He, providing negative Fw values for both kinds of markers.

Rare alleles

The 'mandarin-like' population included 25 genotypes with unique alleles (Supplementary information 3), ranging from one (12 genotypes) to 20 (*C. junos* Sieb. ex Tan.) unique alleles per genotype.

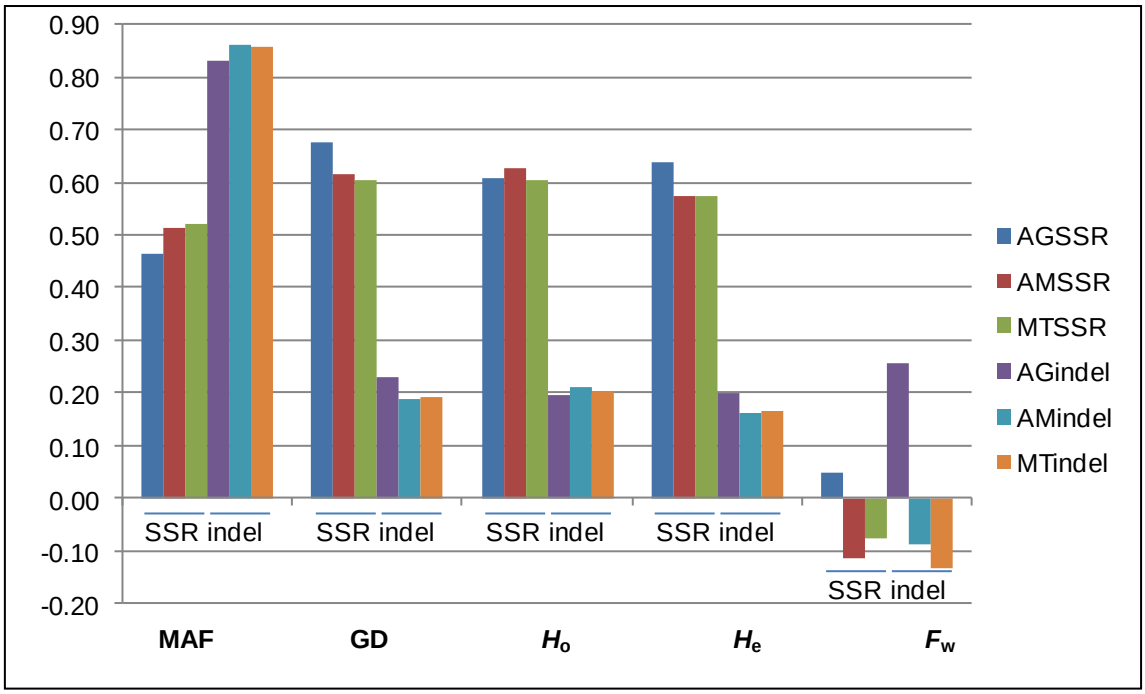


Figure 1. Genetic population statistics within the whole population, all 'mandarin-like' genotypes and Tanaka mandarin species. Comparison between SSR and indel markers.

(AGSRR) All genotypes analysed with SSR markers; (AMSSR) all mandarin-like genotypes analysed with SSR markers; (MTSSR) mandarins defined as true by Tanaka, analysed with SSR markers; (AGindel) all genotypes analysed with indel markers; (AMindel) all 'mandarin-like' genotypes analysed with indel markers; (MTindel) mandarins species defined by Tanaka, analysed with indel markers; (MAF) maximum allele frequency; (GD) gene diversity; (H_o) observed heterozygosity; (H_e) expected heterozygosity; (F_w) Wright's fixation index.

Classifications by NJ analysis

For the whole data set (SSR and indel markers), NJ analysis (Figure 2a) revealed a clear differentiation between the five main taxa studied, the four ancestral *Citrus* groups (papeda, citron, pummelo, and mandarin) and kumquat, with very high bootstrap support. The combination of both SSR and indel markers revealed high intraspecific diversity in the mandarin group, which was not well resolved (low bootstrap support in many branches; Figure 2b). From the whole data set, 35 genotypes were reduced to 14 multilocus genotypes (MLGs; Supplementary information 4). Some of these were mutations of the same genotype (for example 'Willowleaf' and 'Willowleaf seedless', 'Murcott' and 'Murcott seedless', some mutations of *C. unshiu*, and the mutations of *C. clementina*), others are duplications of the same genotype that are present in both collections ('Imperial Australia' [ID-98] and 'Imperial' [ID-121]), and others are possible redundant genotypes collected in different locations.

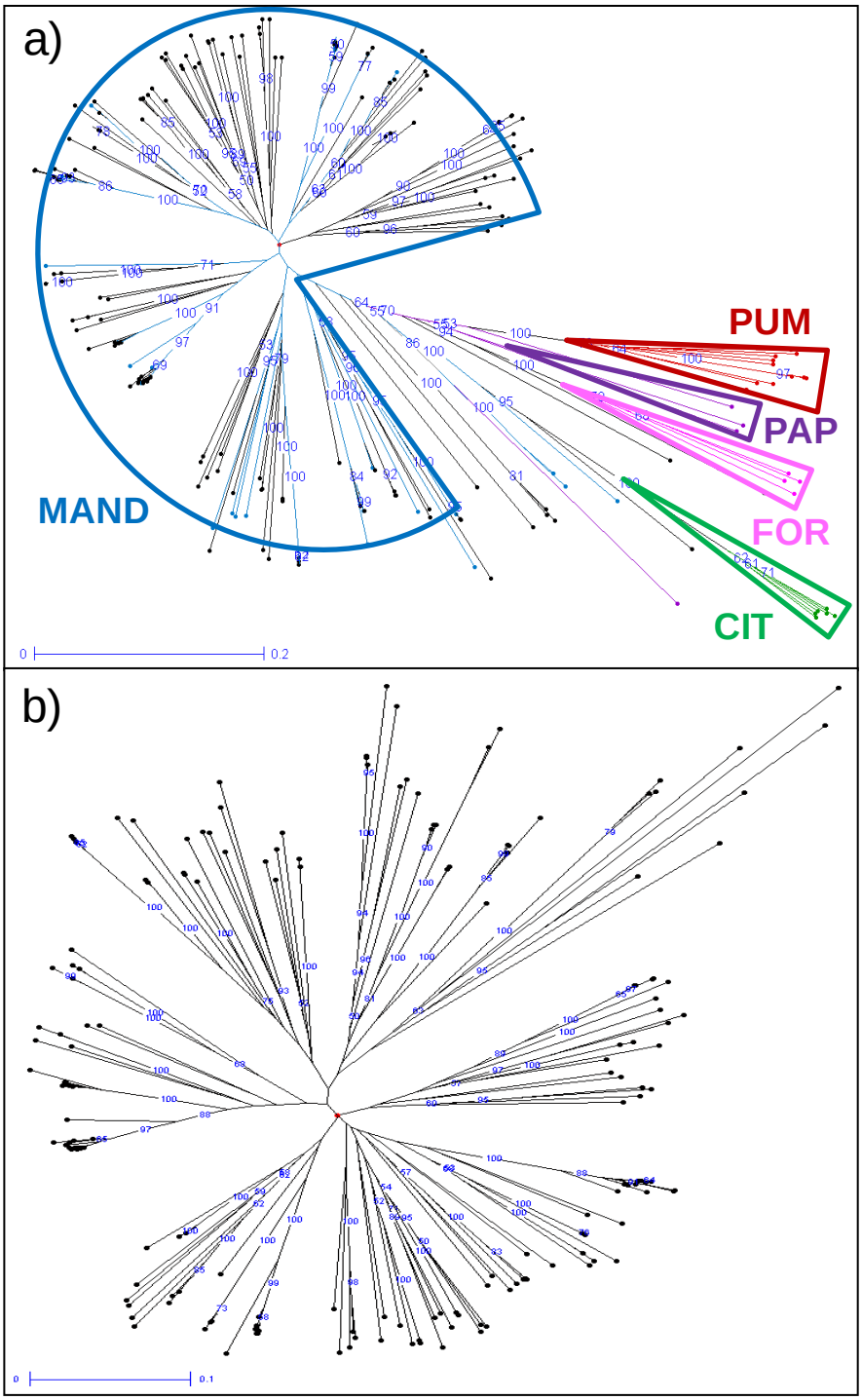


Figure 2. NJ analyses with 1000 bootstraps. Bootstrap values over 50 are represented. a) Entire data set (223 genotypes) representing the four ancestral *Citrus* species (*C. reticulata*, *C. maxima*, *C. medica*, *Papeda*) and *Fortunella*. b) 'Mandarin-like' genotypes (198, without the *C. maxima*, *C. medica*, *Papeda,* and *Fortunella* genotypes)

(MAND) Mandarin, (FOR) *Fortunella*, (CIT) Citron, (PAP) *Papeda*, (PUM) Pummelo

Contribution of the ancestral taxa to the mandarin group and modern hybrids; analysis with the Structure software

The indel and SSR data were analysed with the Structure software to assess the contribution to the mandarin germplasm of the four ancestral *Citrus* taxa (*C. reticulata* [Sw], *C. maxima*, *C. medica* and *Papeda*) and *Fortunella*, using an admixture model and the option of correlated allele frequencies between populations. The degree of admixture alpha was inferred from the data. The burn-in period was set to 500000 and MCMC (Markov Chain Monte Carlo) repetitions were set to 1000000; 10 runs of Structure with K = 5 (five populations assumed) were performed. These populations were as follows: mandarin [Sw] (30 samples, representing 17 Tanaka species), pummelo (11 samples), citron (six samples), papeda (four samples) and kumquat (four samples). The other samples analysed (168) were assumed to have been derived from these ancestral populations (Supplementary information 1). Assuming an admixture model between the four ancestral citrus species and *Fortunella* (Supplementary information 1, genotypes 1–55), the relative proportion of these genomes in the mandarin group and recent hybrids was inferred using Structure, v. 2.3.3 (Figure 3), with the complete data set (SSRs + indel).

Twenty of the 55 genotypes assumed to belong to one of the ancestral citrus populations, as well as *Fortunella*, appeared to contain a certain degree of contribution from other ancestors. This was particularly the case for genotypes considered as mandarin species by Tanaka. The two *C. amblycarpa* (Hassk.) Ochse (only differing by five SSR markers) had a very high contribution from the *Papeda* genome (~65%), with the remainder (~35%) from *C. reticulata*. *Citrus depressa* (ID-5) had contributions from *C. reticulata* (~65%) and *Papeda* (~35%). *Citrus erythrosa* Hort. ex Tan. cv 'San hu hong chu' (ID-8) had almost 10% introgression from *Papeda* and the remainder from *C. reticulata*. *Citrus indica* Tan. (ID-9) seems to have a tri-hybrid genome origin (41% each from *C. reticulata* and *C. medica*; 18% from *Papeda*). *Citrus kinonuni* cv 'Vietnam à peau fine' (ID-10) had almost 10% introgression from *Papeda*, with the remainder of the genome from *C. reticulata*. The two *C. nobilis*, cv 'Campeona' (ID-12) and cv 'Geleking' (ID-13), had introgressions of 10 and 23%, respectively, from *C. maxima*, with the remainders of their genomes from *C. reticulata*. *Citrus reshni* (ID-165) had introgression from the *Papeda* genome (11%), with the remainder from *C. reticulata*. *Citrus suavissima* Hort. ex Tan. cv 'Ougan' (ID-20) derived most of its genome from *C. reticulata* (~90%), with some introgression from *C. maxima* (~10%). *Citrus succosa* Hort. ex Tan. cv 'Ben di zao' (ID-21) had a contribution from *C. reticulata* (~90%), with the remainder of its genome from *C. maxima* (~7%). The two *C. sunki* (ID-24 and ID-25) had introgression from *Papeda* of 13 and 20%, respectively, with the remainders apparently from the *C. reticulata* genome. *Citrus tachibana* (ID-26) appeared to have equal contributions from the *C. reticulata* and *Papeda* genomes. The two *C. unshiu* (ID-29 and ID-30) had a small introgression from the *C. maxima* genome (8.80%). Two ancestral *Papeda*, *C. ichangensis* Swing. and *C. latipes* (Swing.) Tan., also had contributions from *C. maxima*, 6.30 and 34.60%, respectively.

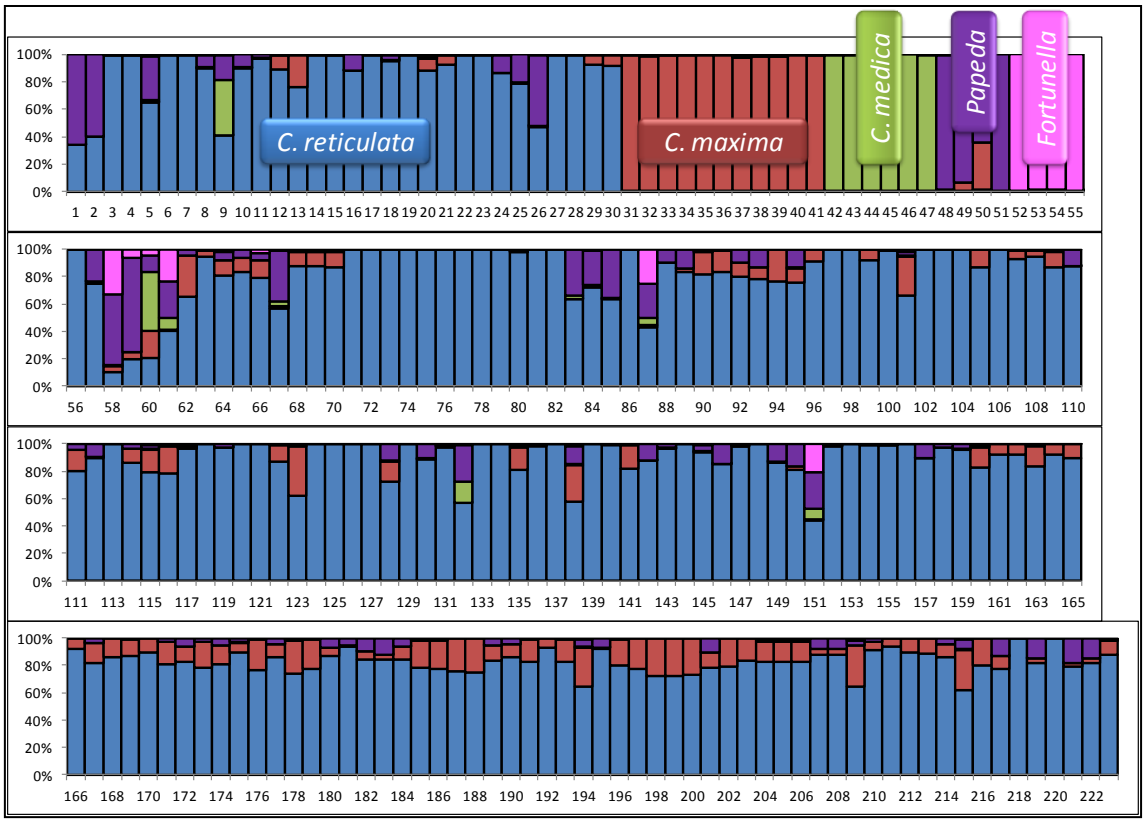


Figure 3. Structure analysis of 223 genotypes representing the four ancestral *Citrus* species (*C. reticulata*, *C. maxima*, *C. medica*, *Papeda*) and *Fortunella*. Dark blue, *C. reticulata* (1); brown, *C. maxima* (2); green, *C. medica* (3); purple, *Papeda* (4); pink, *Fortunella* (5). Genotypes 56–223 are genotypes without assigned populations.

The contribution of mandarin to the genomes of the 168 'mandarin-like' genotypes that were not included in any of the five pre-assumed populations (Supplementary information 1, genotypes ID-56/ID-223) was on average ~85.13%. Pummelo, papeda, kumquat and citron contributions were 8.00%, 5.14%, 1.03% and 0.70%, respectively (Figure 4). Contributions in individual genotypes lower than 2% were not considered for the calculations.

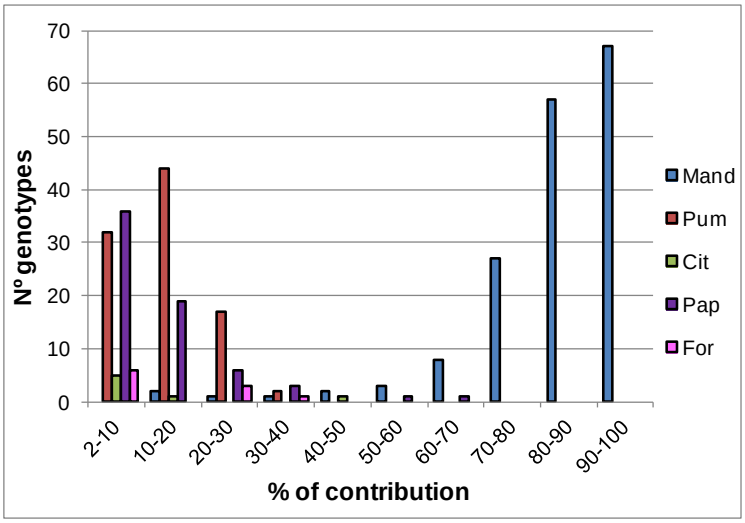


Figure 4. Contributions of the ancestral genomes (mandarin, pummelo, citron, papeda) and kumquat to the 'mandarin-like' genotypes under study.

(Mand) mandarin, (Pum) pummelo, (Cit) citron, (Pap) papeda, (For) kumquat.

In the whole data set, only the citrus ancestors (*C. maxima*, *C. medica* and *Papeda*) and *Fortunella* did not exhibit any contribution from the mandarin genome. The 168 genotypes analysed with no assumed population had at least a 5% contribution from *C. reticulata*. Eighty-four genotypes had a *C. maxima* contribution of at least 5%, and 95 genotypes had a *C. maxima* contribution of at least 2%. *Papeda* contributed at least 5% to 45 genotypes, and at least 2% to 66 genotypes. Only five genotypes exhibited a contribution from *C. medica* of at least 5%: *C. karna* Raf. (ID-60), *C. madurensis* Lour. (ID-61), *C. kinokuni* Hort. ex Tan. (ID-87), *C. reticulata* cv "Nicaragua" (ID-132), and *C. sunki* (ID-151). Two others had a contribution higher than 2%. Only five genotypes [*C. halimii* Stone (ID-58), *C. junos* (ID-59), *C. madurensis* (ID-61), *C. kinokuni* (ID-87) and *C. sunki* (ID-151)] had a contribution from *Fortunella* higher than 5%, and two others had a contribution higher than 2%.

Clementines had a contribution from *C. maxima* (~10%), with the remainder of the genome from *C. reticulata* [Sw]. The six additional satsumas are identical to the ones included in the assumed ancestral mandarin group.

Tangelos had an on average contribution of ~26% from *C. maxima* and ~74% from *C. reticulata* [Sw]. Tangors had lower contributions from *C. maxima* (14%) than the tangelos, and higher contributions from *C. reticulata* (Sw, 82%), as well as very small and perhaps insignificant contributions from *Papeda* (3%) and *Fortunella* (1%).

Some other genotypes, not directly related to mandarins, had mixed profiles: *Citrus daoxianensis* (ID-57) exhibited introgression from *Papeda* (~23%), with the remainder from *C. reticulata* [Sw]. *Citrus halimii* (ID-58) had a complex constitution: 51% *Papeda*, 33% *Fortunella*, 10% *C. reticulata* [Sw], 4% *C. maxima* and 1% *C. medica*. *Citrus junos* (ID-59) was 69% *Papeda*, 20% *C. reticulata* [Sw], 6% *Fortunella* and 5% *C. maxima*. *Citrus karna* (ID-60) also had a complex mixture of genomes: 43% *C. medica*, 21% *C. reticulata* [Sw], 20% *C. maxima*, 12% *Papeda* and 4% *Fortunella*. *Citrus madurensis* (ID-61) is another complex hybrid: 40% *C. reticulata* [Sw], 26% *Papeda*, 23.5% *Fortunella*, 8.5% *C. medica* and 2% *C. maxima*.

Inferring clusters in the mandarin population

The statistics used to select the correct *K* value were the ones followed by Evanno *et al.*, (2005): the mean likelihood, L(*K*); the mean difference between successive likelihood values of *K*, L'(*K*); the absolute value of this difference, L''(*K*) and Δ*K*, which is the mean of the absolute values of L''(*K*) divided by the standard deviation of L(*K*). The likelihood distribution L(*K*) and Δ*K* were the main values used to choose the optimal *K* value of the population. Three consecutive analyses were performed to obtain the correct number of groups within the mandarin germplasm.

The first Structure analysis was performed with the whole population (223 genotypes) with no population assignation. In this case, the optimal Δ*K* was 2: one population consisting of all mandarins and hybrids, and the other formed by the other parental representatives, *C.*

maxima, *C. medica*, *Papeda* and *Fortunella* (Supplementary information 5). From this analysis (K = 2), 175 accessions (Supplementary information 1) with a contribution above 95% from the mandarin population were selected to perform another Structure analysis without population assignation. The highest ΔK value was obtained for K = 6 (Supplementary information 6).

The third Structure analysis was done after removing all the known hybrids (clementines, tangelos, tangors and recent hybrids) from the 175 previous accessions. The genotypes 'Wallent' and tangor 'Gailang' were also removed. This analysis aimed to determine whether the groups observed previously were coherent. A sample set of 121 genotypes (Supplementary information 1) was used and, as before, the highest ΔK was observed for K = 6 (Supplementary information 7). The two main differences between this analysis and that with 175 genotypes (explained previously), are as follows: (1) in the analysis with 175 genotypes, the *C. nobilis* Tanaka species formed a group, while in the analysis with 121 genotypes, *C. nobilis* was not identified as a pure group; conversely, (2) the analysis with 121 genotypes identified a group formed by a mixture of Tanaka mandarin species (*C. reshni*, *C. kinokuni* and *C. reticulata*) not previously recognised. The 'Ampefy' genotype was identified as an independent parental group in both analyses. 'Ampefy', 'Wallent' and 'Gailang' exhibited a high degree of mutual similarity, shared a high percentage of *C. sinensis* (L.) Osb. molecular-marker data (85.33, 97.33, and 96%, respectively), and also exhibited high heterozygosity. Therefore, they are very probably interspecific hybrids, like sweet orange.

Structure analyses were compared with an NJ tree (Figure 5) to validate the clustering. *Citrus nobilis* (tangor 'King') and some of its hybrids appeared as a cluster in the NJ tree, as in the structure analysis with 175 genotypes. However, previous analyses (Coleta-Filho *et al.*, 1998; Nicolosi *et al.*, 2000; Garcia-Lor *et al.*, 2012) considered *C. nobilis* as a tangor; therefore, we have not considered it as a true mandarin group. Finally, seven groups of 'mandarin-like' genotypes were identified at the nuclear level. Five were parental groups of the mandarin germplasm (Nuclear groups N1–N5) and two groups were of interspecific origin: N6, consisting of 'Ampefy' (ID-101), 'Wallent' (ID-194) and 'Gailang' (ID-215); and N7, the tangor 'King' group, consisting of 'King' (ID-201), 'Rodeking' (ID-92), 'King' (ID-93) and 'Sanh' (ID-93). Genotypes included in each of the five parental groups are presented in Table 2 in relation to the Tanaka classification.

Furthermore, combining the results from the NJ and Structure analyses allowed us to determine the group to which the mandarin genotypes belong (Supplementary information 8).

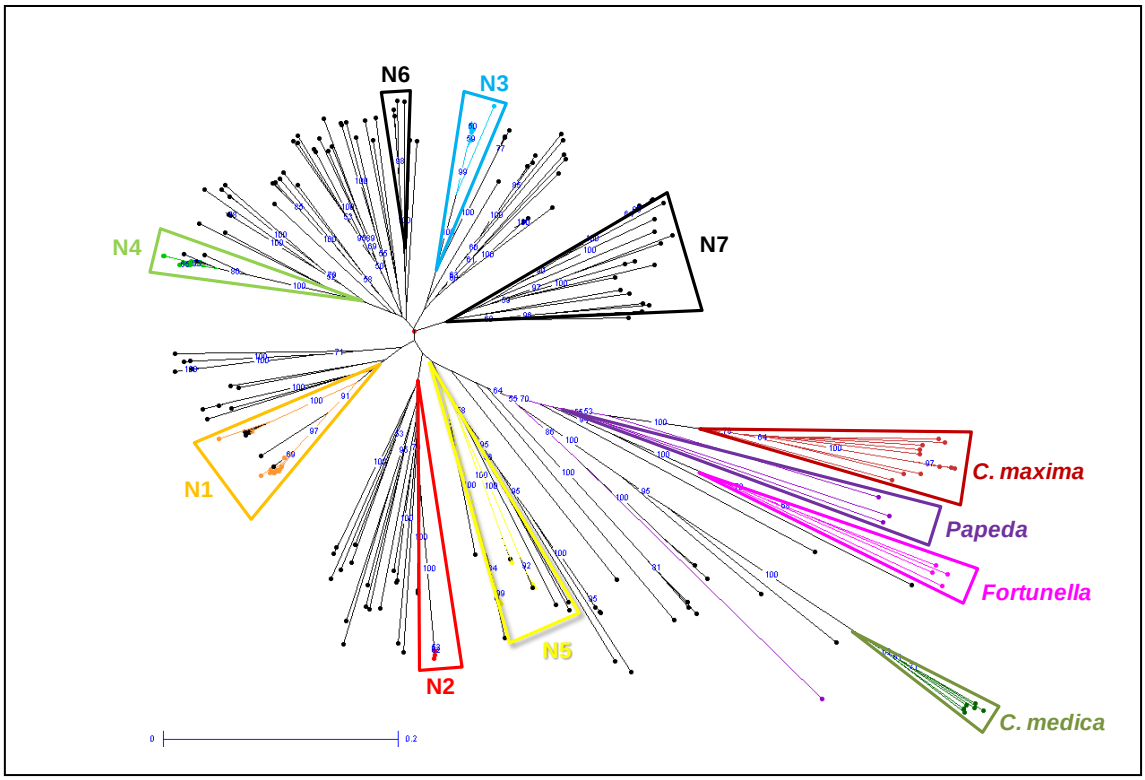


Figure 5. NJ tree analysis comparing the Structure populations found to the groups observed with the NJ tree. Bootstrap values over 50 are represented. One thousand resamplings were performed. Five mandarin groups and four parental populations (*C. maxima*, *C. medica*, *Papeda,* and *Fortunella*) were identified at the nuclear level. Orange, N1 (14/18 *C. reticulata* [Tan.]); red, N2 (all *C. unshiu*); light blue, N3 (7/9 *C. deliciosa*); light green, N4 (5/9 *C. tangerina*); yellow, N5 (*C. reshni* and two *C. reticulata* [Tan.]); brown, *C. maxima* (6); dark green, *C. medica* (7); purple, *Papeda* (8); pink, *Fortunella* (9). Interspecific hybrid groups: (N6) 'Ampefy', 'Gailang', and 'Wallent'; (N7) 'King' and hybrids.

(N) Nuclear group

Table 2. Parental mandarin groups identified at the nuclear level within the mandarin germplasm, based on the Structure and NJ tree analyses. Genotypes included in each group are compared with the Tanaka classification.

Group	Nº genotypes	Tanaka Species	%
N1	17	*C. reticulata*	73.91
	3	*C. suhuiensis*	13.04
	1	*C. tangerina*	4.35
	1	*C. erythrosa*	4.35
	1	Unknown	4.35
N2	7	*C. unshiu*	100.00
N3	11	*C. deliciosa*	100.00
N4	5	*C. tangerina*	45.45
	3	*C. reticulata*	27.27
	1	*C. deliciosa*	9.09
	1	*C. depressa*	9.09
	1	*C. paratangerina*	9.09
N5	4	*C. reticulata*	66.67
	1	*C. kinokuni*	16.67
	1	*C. reshni*	16.67

(N) Nuclear group.

Contribution of the various mandarin groups to the constitution of the other mandarin genomes

On the basis of the five mandarin groups identified as parental mandarins by the previous Structure analysis, a subsequent step was performed to quantify the contribution of these five groups to all genotypes of our 'mandarin-like' collection. We ran a new Structure analysis (Figure 6) with the whole collection, assigning as parental populations the five mandarin groups (N1–N5), the other *Citrus* ancestral populations (*C. maxima*, *C. medica*, *Papeda*) and *Fortunella*. The genotypes belonging to the other two groups identified as interspecific hybrids, N6 and N7, were removed in order to avoid biasing the contribution from the five parental mandarin groups. A list of genotypes included in this analysis (Structure 216, *K* = 9) is in Supplementary information 1.

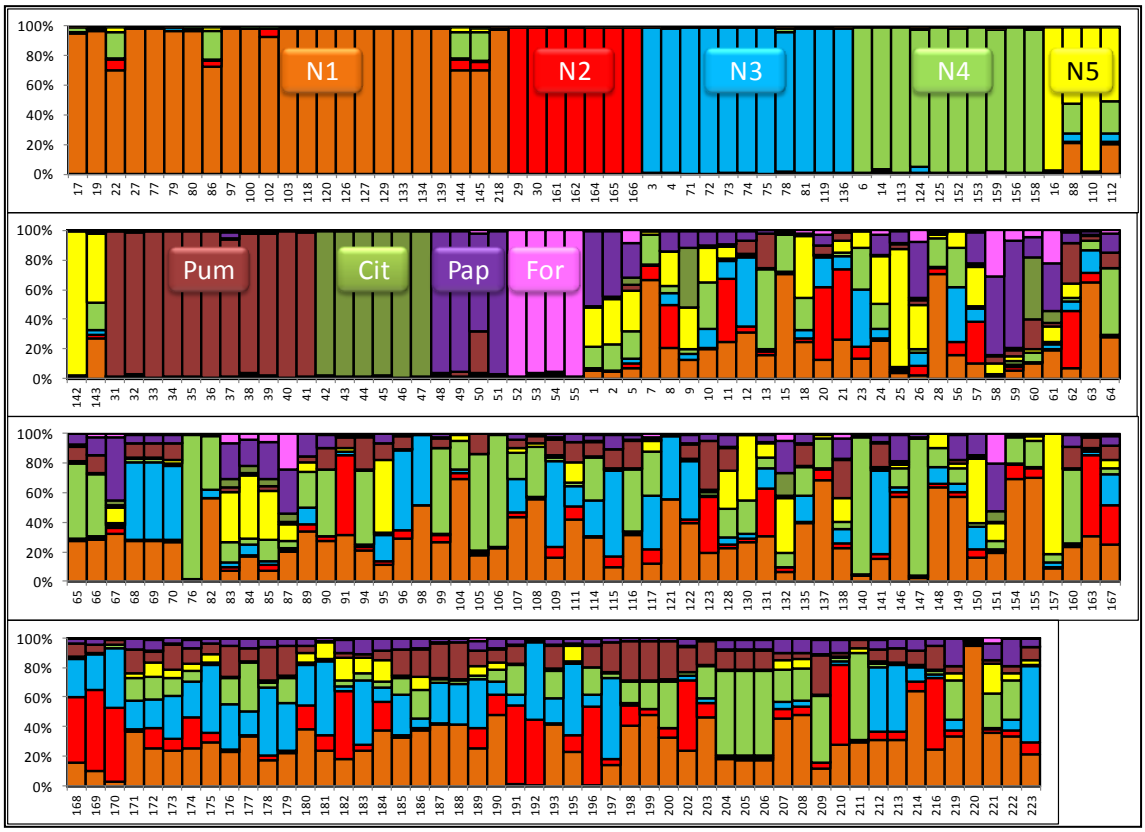


Figure 6. Structure analysis assuming nine populations (*K*=9). Five mandarin groups and four parental populations (*C. maxima*, *C. medica*, *Papeda,* and *Fortunella*). Orange, N1 (14/18 *C. reticulata* [Tan.]); red, N2 (all *C. unshiu*); light blue, N3 (7/9 *C. deliciosa*); light green, N4 (5/9 *C. tangerina*); yellow, N5 (*C. reshni* and two *C. reticulata* [Tan.]); brown, *C. maxima* (6); dark green, *C. medica* (7); purple, *Papeda* (8); pink, *Fortunella* (9). The remaining genotypes are without assigned populations.

(N) Nuclear group; (Pum) Pummelo; (Cit) Citron; (Pap) Papeda; (For) Fortunella.

Some genotypes of the five mandarin populations assigned exhibited an admixed genome structure (Figure 6): genotypes ID-22, ID-86, ID-102, ID-144 and ID-145 from N1; ID-78 from N3; ID-124 from N4; and ID-88, ID-112 and ID-143 from N5. These genotypes exhibited contributions greater than 5% from a non-mandarin genome. Therefore, they were removed from the parental mandarin group and excluded from the calculations of population statistics

(Table 3). H_o was higher than H_e for the five groups for both SSR and indel markers, leading to negative F_w values. The whole 'mandarin-like' population exhibited a similar pattern. The mandarin-like genotypes exhibited complex hybrid structures with contributions from more than two genomes.

The contributions of the five mandarin parental groups defined in this study into the other mandarins under study are summarised in Figure 7. The average contribution of N1 (18 genotypes, mainly *C. reticulata* [Tan.]) to the genotypes not included in any defined population was 28.25%; of N4 (nine genotypes, mainly C. *tangerina*), 16.69%; of N3 (10 *C. deliciosa* genotypes), 15.35%; of N2 (all *C. unshiu*), 10.44%; and of N5 (*C. reshni* (ID-16) and two *C. reticulata* (ID-110, ID-142), 8.88%. The rest of the genome contributions came from *Papeda* (8.85%), *C. maxima* (8.32%), *Fortunella* (1.85%) and *C. medica* (1.34%).

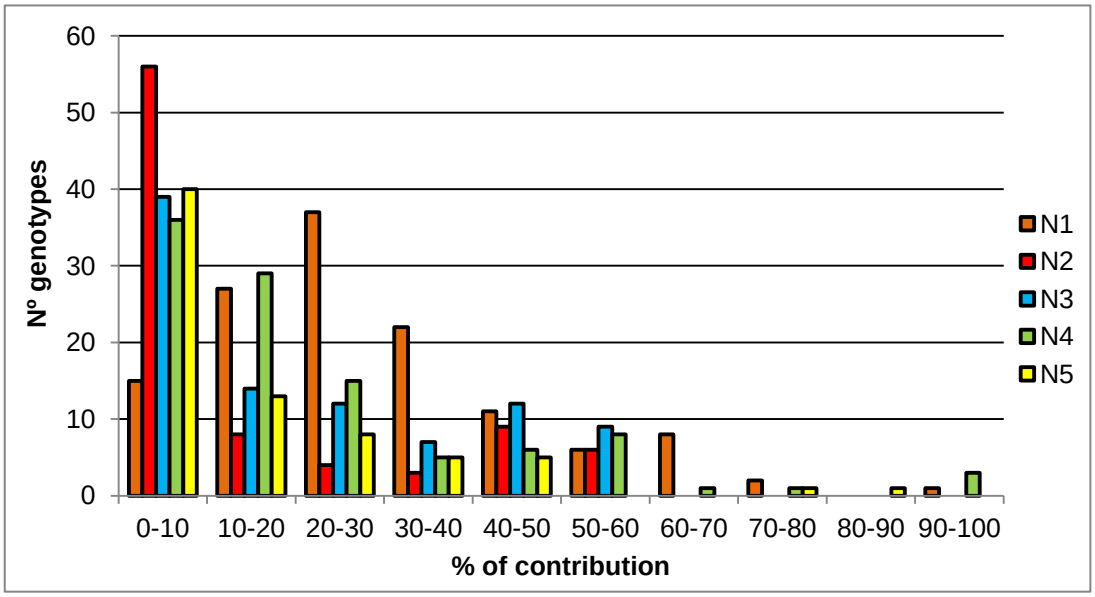


Figure 7. Contribution of the five parental mandarin groups (N1–N5) into the mandarin genome portion of each 'mandarin-like' genotype under study. Contributions lower than 2% were discarded.

To validate the contributions of the other basic taxa and *Fortunella*, the results of the Structure analysis presented in Figure 3 (223 genotypes, K = 5) and the present analysis (Figure 6; 216 genotypes, K = 9) were compared. High correlation coefficients between the two analyses were observed with regard to the contribution of the different ancestral taxa and *Fortunella* (*C. reticulata* [Sw], R^2 = 0.985; *C. maxima*, R^2 = 0.991; *C. medica*, R^2 = 0.999; *Papeda*, R^2 = 0.935; and *Fortunella*, R^2 = 0.994).

Data from hybrids with known parents were checked in order to validate the analysis of the parental-group contributions (from accessions ID-167/ID-193 and ID-195). Most of them agreed with their known origins; therefore, the origins of other genotypes can be accepted from this analysis. For example, the hybrid mandarin 'C-54-4-4' (ID-171 in Figure 6) had contributions from five different genomes, defined as populations 1, 3, 4, 6 and 7 of the present Structure analysis, which come from its supposed parents, clementines (ID-68, ID-69 and ID-70; genomes

from populations 1, 3 and 6) and tangor 'Murcott' (ID-207 and ID-208; genomes from populations 1, 4, 6 and 7).

Another example is the hybrid mandarin 'Simeto' (ID-192 in Figure 6), which was obtained from a cross between a *C. unshiu* and *C. deliciosa*. Our study confirms this cross (almost 50% each from *C. unshiu* and *C. deliciosa*). Tangelo 'Orlando' (ID-199 in Figure 6) is a cross between *C. paradisi* 'Duncan' and *C. tangerina* 'Dancy' (ID-28, ~70% from *C. reticulata* [Tan.] and ~20% from *C. tangerina*). The genomes contributing to 'Orlando' (ID-199) come from its supposed parents, *C. maxima* (26.9%), *C. reticulata* [Tan.; 48.1%] and *C. tangerina* (17.8%).

On the other hand, some examples of discrepancies between the Structure results and supposed parental origin can be explained by misidentified origin. The 'Fortune' mandarin (ID-178) was reported to come from a cross between a clementine and 'Dancy', made by Furr (1964). However, the structure analysis showed that 'Fortune' has a lower *C. reticulata* [Tan.] genome contribution (71.1%) than the parents (clementine [83.2%] and 'Dancy' [96.9%]), and a higher *C. maxima* contribution (22.4%) than clementine (10%) and 'Dancy' (0.7%). The false parental origin was confirmed by individual locus checking: in 16 out of 50 SSR markers and in one indel marker, 'Fortune' possesses a specific allele present in neither 'Dancy' nor clementine. Similar observations were made for 'Fremont' (ID-179; supposed hybrid between *C. clementina* and *C. reticulata* [Tan.] 'Ponkan' (ID-17) (Furr, 1964). Indeed, for 11 SSR markers, this hybrid possesses alleles that are not observed in its supposed parents. Moreover, 'Fremont' has a lower *C. reticulata* [Tan.] contribution to its genome (74.3%) than clementine (83.2%) and 'Ponkan' (99.2%), and a contribution from *C. maxima* (20.1%) higher than that in clementine (10%) and 'Ponkan' (0.2%).

Mitochondrial analysis

In the whole population, mitochondrial markers allowed discrimination of six mitotypes (Figure 8), previously described by (Froelicher *et al.*, 2011). One *Fortunella* genotype (*F. hindsii*; ID-53) was associated with the *Papeda* (*C. micrantha*) mitotype, and one *Papeda* (*C. latipes*; ID-50) had a *C. maxima* mitotype. In the mandarin group (194 genotypes), four mitotypes were distinguished: two of mandarins (C1 and C2), one identical to *C. maxima* (C3) and one identical to *C. micrantha* (*Papeda*, C4). The first mandarin mitotype (C1) included most of the genotypes studied. In the second mitotype (C2), 20 genotypes were present; 11 of them were acid mandarins (four *C. depressa* [ID-5, ID-83, ID-84, ID-85], three *C. sunki* [ID-24, ID-25, ID-150], *C. reshni* [ID-16], *C. daoxianensis* [ID-57], *C. indica* [ID-9] and 'Xien Khuang' [ID-142]), and nine were sweet genotypes (*C. tankan* Hay. [ID-63], *C. kinokuni* [ID-88], *C. tangerina* [ID-157] and six *C. reticulata* [Tan.]: 'Chiuka' [ID-110], 'Douhala' [ID-112], 'Lime sucrée' [ID-128], 'Macaque' [ID-130] and 'Sun chu sha' [ID-18, ID-143]. The *C. maxima* mitotype (C3) included the 'Ampefy' (ID-101), 'Suntara' (ID-138), 'Bendiguangju' (ID-163), 'Pet Yala' (ID-146), 'Yala' (ID-149), 'Kunembo' (ID-91), 'Ougan' (ID-20), and 'Kobayashi' (ID-123) mandarin cultivars, as well as the

tangor 'Dweet' (ID-203; *C. sinensis* × 'Dancy'). The *Papeda* mitotype (C4) included 'Nicaragua' (ID-132), 'Vietnam' (ID-87), and one *C. sunki* (ID-151).

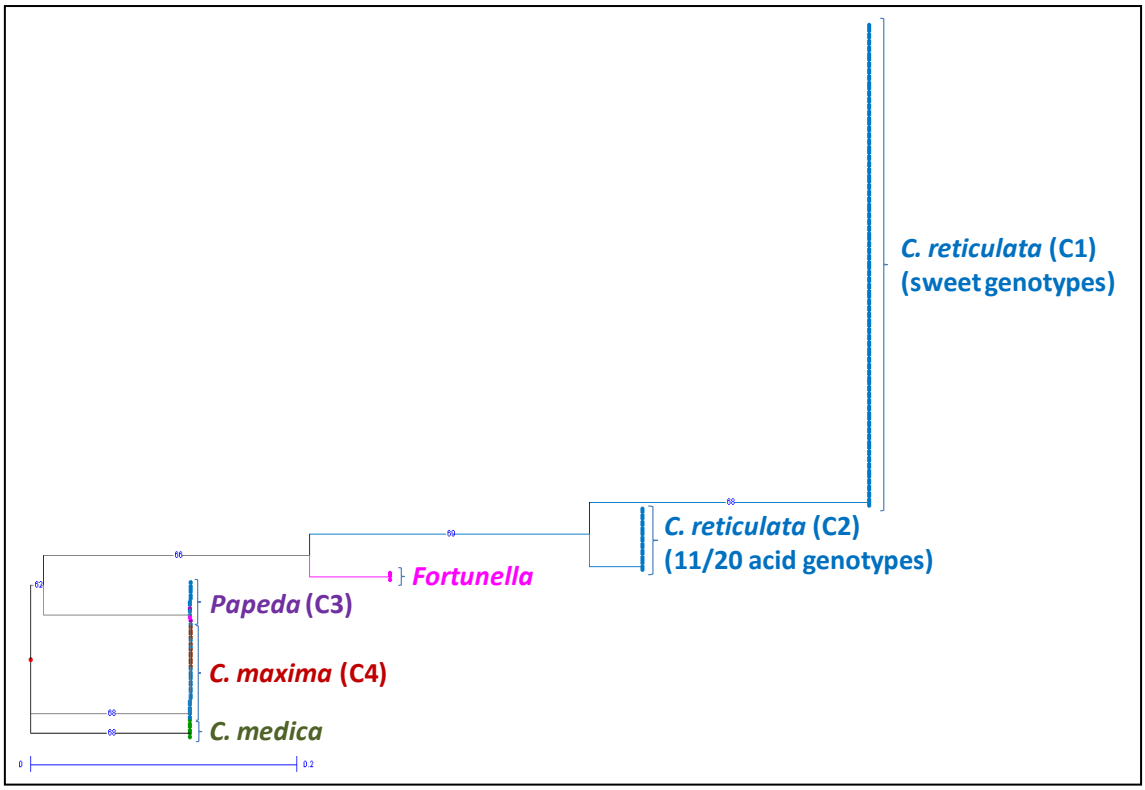


Figure 8. NJ tree of 223 varieties of *Citrus* with mitochondrial markers. Four mitotypes (C1-C4) were observed for the 'mandarin-like' genotypes: *C. reticulata* [Sw] (C1, C2), *Papeda* (C3) and *C. maxima* (C4).

DISCUSSION

Genetic structure of the studied population

SSR markers are more polymorphic than indel markers. The average numbers of alleles, gene diversity, and heterozygosity were all higher in SSR markers. The combination of both types of markers allowed differentiation of the mandarin group from the other ancestors and revealed diversity within the mandarin group (mainly from SSR markers), as reported by Garcia-Lor *et al.* (2012).

The clear differentiation of mandarins from *C. maxima*, *C. medica*, *C. micrantha* and *Fortunella* (Figure 2) has been described in several studies (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012; 2013). Moreover, as previously observed by (Federici *et al.*, 1998) and (Barkley *et al.*, 2006), the mandarin group was not well resolved (low bootstrap support in many branches), perhaps due to the large number of hybrids.

Several groups of accessions with identical MLGs were identified, such as 'Ellendale Leng' (ID-205) and 'Ellendale Taranco' (ID-206) (MLG10), or 'Willowleaf' (ID-3) and 'Willowleaf seedless' (ID-72, ID-73) (MLG12), which were produced by natural mutations and are probably distinguished only by point mutations. Therefore, the probability of distinguishing them by analysis of molecular markers such as SSRs or indels is very low. The groups MLG2, MLG3, MLG6, MLG11, MLG13 and MLG14 also contain such derivative mutants. On the other hand, the clusters MLG1, MLG5, MLG7, MLG8 and MLG9 include genotypes for which there is no clear prior information about their origin; therefore, they may represent either derivative mutants of this kind, or simply redundancies within the germplasm collections.

The overall F_w value among all loci and all genotypes was close to zero (0.12; 0.05 for SSR and 0.26 for indel), indicating that the observed heterozygosity was close to the expectation value under Hardy-Weinberg equilibrium, whereas higher structuration (positive F_w) was observed by Garcia-Lor *et al.* (2012). This may be due to the large proportion of mandarin hybrids within the population under study. The F_w values observed for all the mandarin-like genotypes and the representatives of the Tanaka mandarin species was close to zero. Therefore, it is a favourable situation for using the Structure software, which assumes that the populations are in Hardy-Weinberg equilibrium (Pritchard *et al.*, 2000).

Mitochondrial and nuclear data reveal interspecific hybridisation and introgression of ancestral genomes into mandarin varieties

Mitochondrial and chloroplastic markers have been previously used to reveal maternal phylogeny in *Citrus* (Green *et al.*, 1986; Yamamoto *et al.*, 1993; Bayer *et al.*, 2009; Morton, 2009). In our study, six mitotypes were found (pummelo, micrantha, citron, mandarin mitotype C1, mandarin mitotype C2 and *Fortunella*), all of them observed by (Froelicher Y *et al.*, 2011), who proposed a distinction between the acid mandarin and sweet mandarin mitotypes. The mandarin germplasm (194 genotypes) was represented in four of the six identified mitotypes;

two of them included mandarin and 'mandarin-like' genotypes (C1, C2), and two corresponded to other ancestral species (C3, C4). Our results, obtained with a large mandarin panel, show that the denomination of acid mandarin and sweet mandarin mitotypes proposed by Froelicher *et al.* (2011) is not apt: we found sweet mandarin genotypes that share the supposed acid mitotype (nine out of 20 sweet mandarins in the C2 mandarin mitotype). Some of the genotypes that does not fit with the hypothesis of Froelicher et al. (2011) may be result by hybridisations between the sweet and acid mandarin gene pools.

Three mandarins have a *Papeda* mitotype (C3), and seven have a *C. maxima* mitotype (C4). For example, 'Bendiguangju' mandarin (ID-163; *C. unshiu*, according to Tanaka classification) exhibited a pummelo rather than mandarin cytoplasm, as reported by Cheng *et al.* (2005) in a chloroplast DNA analysis and Froelicher *et al.* (2011) in a mitochondrial DNA analysis. At the nuclear level, however, we observed a close relationship between 'Bendiguangju' and satsumas, confirming the data of Nicolosi *et al.* (2000). The genotypes included in the *Papeda* and *C. maxima* mitotypes are interspecific hybrids, and not true mandarins, according to the Structure analysis; however, most of their genomes are derived from mandarins.

Among the mandarin species considered by Tanaka, we identified some interspecific hybrids, such as *C. amblycarpa,* which appears to be a cross between the papeda and mandarin gene pools with a maternal phylogeny from papeda, as already observed by Froelicher *et al.* (2011). These contributions from *C. reticulata* [Sw] and *Papeda* genomes were also observed in an SNP analysis (Ollitrault *et al.*, 2012a). By contrast, Federici *et al.* (1998) and Barkley *et al.* (2006) considered *C. amblycarpa* to be the result of a cross between *C. reticulata* and *C. aurantifolia*. The latter study observed contributions of three genomes: *C. reticulata* (~60%), *C. medica* (~25%) and *Papeda* (~15%). Our results show that *C. amblycarpa* genotypes had a high average heterozygosity, 51.33%, suggesting a potential origin from direct interspecific hybridisation.

Introgressions from other genomes were also found in other genotypes considered to be mandarin species by Tanaka: the *Papeda* genome (32.3%) is present in *C. depressa*, and the *C. maxima* (6.8%) genome is present in *C. succosa*. Similar genome contributions, albeit at different percentages, were found by Barkley *et al.* (2006). Those authors reported that the genome of *C. depressa* is shared between *C. reticulata* and *Papeda* in equal proportions, whereas we observed a higher contribution from *C. reticulata* (~65%) than Papeda (~35%).

Citrus indica clustered with the citron group at the nuclear level. It had contributions of 41% from citron and mandarin genomes and 18% from papeda, as well as a very high observed heterozygosity (61.33%), indicating that it was originated as an interspecific hybrid. *Citrus indica* is present in mandarin mitotype C2, whereas Nicolosi *et al.* (2000) clustered *C. indica* with the citron on the basis of cpDNA markers.

Citrus tachibana was considered to be a wild species of mandarin by Swingle and Reece (1967), and it was clustered with the mandarins by Nicolosi *et al.* (2000). Our results are

not in agreement with this theory, because *C. tachibana* clustered with the mandarin mitotype C2 and displays an equal contribution from the *C. reticulata* [Sw] and *Papeda* genomes at the nuclear level. The high H_o (54.67%) indicates that *C. tachibana* is an interspecific hybrid, and nuclear and mitochondrial data suggest that it is a direct hybrid between a mandarin of mitotype C2 as the maternal parent and a *Papeda.*

It is also remarkable that some other genotypes, like *C. unshiu* and *C. tankan*, have a small contribution from *C. maxima*, approximately 8 and 3%, respectively. This observation was also made by Nicolosi *et al.* (2000).

Citrus nobilis was considered as a species by Tanaka, but other authors (Coletta Filho *et al.*, 1998; Nicolosi *et al.*, 2000; Garcia-Lor *et al.*, 2012; 2013a) considered it as a tangor, with introgression from the *C. maxima* genome. Our results confirm this pummelo introgression in the various *C. nobilis* analysed (King: 10,9%; Campeona 10%).

The other tangors, tangelos, and clementines we analysed exhibited similar contributions from ancestral genomes to those reported by Garcia-Lor *et al.*, (2012), with higher introgression of pummelo in tangelos than in tangors. Some genotypes of unknown origin included in the study, and not related to the mandarin species defined by Tanaka, exhibited complex genomic structures.

Citrus junos appears to be a mixture of four genomes (papeda, mandarin, pummelo and kumquat). This observation is in contrast to the results obtained by Nicolosi *et al.* (2000), who clustered *C. junos* with the mandarins, of Mabberley (2004), who hypothesised that it was a cross between a *Papeda* and *C. maxima* and Tanaka (1954), who defined it as a relative of *C. ichangensis*. *Citrus junos* was considered to be a hybrid with a *Papeda* maternal phylogeny by Froelicher *et al.* (2011), a proposal that is confirmed by our results. This mixture of genomes leads to an observed heterozygosity of 46.67%.

Citrus halimii had a complex genomic constitution, with the main genome contributions from *Papeda* (51%), *Fortunella* (33%) and *C. reticulata* (10%). This result is in contrast with the results of Scora (1975) and Barkley *et al.* (2006), who considered *C. halimii* a hybrid of a citron and a kumquat based on morphological and phytochemical data and molecular data, respectively. Froelicher *et al.* (2011) found that *C. halimii* shared the *Papeda* mitotype, as we have seen in our work. However, its observed heterozygosity is low (21.33%), and the origin of this species remains unclear.

Citrus karna is still of unknown origin and has been proposed to be a natural hybrid (Swingle and Reece, 1967), as confirmed by its very high heterozygosity (66.3%). It appears to be a very complex admixture with five genome contributions: *Citrus medica* (43%) and *C. maxima* (20%) are the main contributors, and the cytoplasm is from *C. maxima*.

Our Structure analysis showed that many ‘mandarin-like’ genotypes are introgressed by other ancestral species, as reported by Barkley *et al.* (2006). In our work, the ancestor with the highest contribution to the mandarin germplasm was *C. maxima*, instead of the *Papeda* /

Fortunella group reported by Barkley *et al.* (2006). However, if we reduce the analysis to the 45 mandarin genotypes used in common between the two studies, we obtain similar results, with a contribution of 6% from the *Papeda* / *Fortunella* group in Barkley *et al.* (2006) and an 8% contribution from *Papeda* in our study. From the other two ancestral populations, *C. medica* and *Fortunella*, genome introgressions were identified in very few accessions.

It is also important to mention that recent whole-genome sequencing studies (Gmitter *et al.*, 2012; Shimizu *et al.*, 2012) have confirmed the introgression of ancestral genomes within some genotypes considered until now to be pure mandarins, such as 'Ponkan' or satsumas, which exhibit *C. maxima* genome introgression.

Organisation of the mandarin germplasm

The two main *Citrus* classification systems (Swingle and Reece, 1967; Tanaka, 1954) differ greatly in their treatments of the mandarins. The former system placed all mandarins in one species, *C. reticulata*, whereas the latter divided them into 36 species. Neither of the two systems is completely right, as discussed in many reports (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Barkley *et al.*, 2006). Different studies have tried to define groups within the mandarins. Coletta Filho *et al.* (1998) studied 35 accessions of mandarins and divided them into two main groups consisting of two and seven subgroups, which agreed partially with Tanaka's (Tanaka, 1954) and Webber´s (Webber HJ, 1943) taxonomic groups. Koehler-Santos *et al.* (2003) characterised 34 different genotypes from a Brazilian collection and described five groups, different from the ones found by Coletta Filho *et al.* (1998). Kacar *et al.* (2013) characterised 65 mandarin genotypes of the Tuzcu Citrus Variety Collection in Turkey, using 14 SSRs and 21 SRAP markers, resulting in two main groups: one including only tangelo 'Orlando', and the other including the rest (clementines, other tangelos, etc.).

In this work, a broad range of samples representing the mandarin germplasm (ancient cultivars from Asia, old and recent natural hybrids, and human-made hybrids) were analysed to clarify the structure of this highly diversified group. After three consecutive rounds of analyses with the Structure software in which the ancestral genotypes, interspecific hybrids, known recent hybrids, and hybrids detected with the programme were removed, five groups were defined as potential parental mandarins (N1–N5; Figure 6, Table 2). According to the analyses performed with the Structure software and NJ tree analysis, two more groups, including already known hybrids and their descendants, were identified as groups of the mandarin germplasm: N6, including 'Ampefy', 'Wallent' and 'Gailang'; and N7, the tangor 'King' group. The five parental mandarin groups exhibited higher allelic diversity for SSRs than for indel markers**.** The negative F_w values observed in these groups leads to fixation of heterozygosity within them, which may be due to apomixis and vegetative reproduction of citrus varieties. Significant differentiation between nuclear groups is confirmed by the F_{st} value (0.434). The global mandarin population has an F_w value close to 0, reflecting strong intergroup gene flow.

Four nuclear groups, N1, N2, N3, and N4, share the same mandarin mitotype (C1). Most of the genotypes sharing the other mandarin mitotype (C2) are also differentiated at the nuclear level, and 16 out of the 20 genotypes are clustered with mandarin nuclear group N5. Tanaka (1954) divided the acid mandarin genotypes in two groups, with *C. reshni*, *C. sunki* and *C. tachibana* in one group and *C. depressa* in another, which are joined in our analysis at both the nuclear and cytoplasmic levels.

Tanaka (1954) grouped the 36 mandarin species that he considered into five clusters. One cluster included *C. nobilis* and *C. unshiu*, which are separated in two different clusters in our study, N7 and N2, respectively, the first of which is of interspecific hybrid origin, mandarin × pummelo . The second cluster included species not analysed in our study. The third cluster had 14 species, including *C. clementina* (considered in our study as an hybrid and not a pure mandarin species), *C. reticulata* [Tan.], *C. deliciosa*, and *C. tangerina*, which appear in our work as different parental mandarin groups (N1, N3 and N4). The fourth Tanaka group was formed by *C. reshni*, *C. sunki*, and *C. tachibana*, and the fifth group included *C. depressa* and *C. lycopersicaeformis*. From these species, only *C. reshni* is included in a group in our analysis (N5). *Citrus sunki*, *C. tachibana*, and *C. depressa* seem to have resulted from *Papeda* introgression into a mandarin genome. Other Tanaka species, such as *C. erythrosa* and *C. suhuiensis*, seem to have originated from hybridisation between mandarin groups (*C. reticulata* [Tan.] and *C. tangerina* groups).

Hodgson (1967) divided the mandarins in four groups: *C. unshiu*, *C. reticulata* [Tan.] ('Ponkan', 'Dancy', clementine), *C. deliciosa* and *C. nobilis* ('King'). Only two groups are in agreement with our results, *C. unshiu* (N2) and *C. deliciosa* (N3). A third group, *C. nobilis* ('King'), is identified as a parental group in our analysis (N7), but is not a true mandarin group. The fourth group defined by Hodgson, *C. reticulata* [Tan.], included a known hybrid (*C. clementina*) and two genotypes separated between two groups in our analysis: 'Ponkan', within the *C. reticulata* [Tan.] group (N1), and 'Dancy', within the *C. tangerina* group (N4).

The contributions of the five parental mandarin groups defined in the mandarin germplasm, besides the contributions of the other ancestral taxa and *Fortunella*, were estimated for the entire 'mandarin-like' collection (Figure 6). This analysis revealed that the genomes of most 'mandarin-like' genotypes are complex admixtures of the five parental mandarin groups and even include contributions from the other ancestral populations.

Most of the hybrids with known origins displayed admixture coherent with the genomic structures of their supposed parents. Because most of these parents are themselves heterozygote admixed, the proportion of each genome in the hybrid variety is not inherited in an additive way (i.e., the sum of half shares of each parent), but instead depends on the recombination and segregation occurring in each parental gamete (Motohashi *et al.*, 1992; Coletta Filho *et al.*, 1998).

Some accessions' admixture structure did not agree with their supposed parents. In these cases, allele checking confirmed that the supposed parental origins were erroneous. Further analyses could provide more clues toward the identification of parents for these hybrids.

CONCLUSIONS

The mandarin horticultural varietal group is highly polymorphic. Many genotypes believed to be pure mandarins have introgressions from other basic taxa in their genomes. Moreover, some of them exhibited non-mandarin maternal phylogeny. Another characteristic of the mandarin group is that many genotypes originated from crosses between mandarins. Although this work has provided new insights into mandarin structuration, future sequencing of mandarin genotypes (single genes or whole genomes) will help to perform phylogenetic analyses and precisely determine the different genomic constitutions of this highly polymorphic group.

SUPPLEMENTARY INFORMATION CHAPTER 4

Supplementary information 1. Genotypes used in the study of the mandarin diversity, ordered by their appearance in Figure 1.

ID	Common name	Swingle system	Database based onTanaka system
1	*C. amblycarpa*	*C. reticulata* hybrid	*C. amblycarpa*
2	Nasnaran	*C. reticulata* hybrid	*C. amblycarpa*
3	Willowleaf	*C. reticulata*	*C. deliciosa*
4	de Chios	*C. reticulata*	*C. deliciosa*
5	*Citrus depressa*	*C. reticulata*	*C. depressa*
6	Vohangisany Ambodiampoly	*C. reticulata*	*C. depressa*
7	Fuzhu	*C. reticulata*	*C. erythrosa*
8	San hu hong chu	*C. reticulata*	*C. erythrosa*
9	Indian Wild Orange	*C. indica*	*C. indica*
10	Vietnam à peau fine	*C. reticulata*	*C. kinokuni*
11	Nan feng mi chu	*C. reticulata*	*C. kinokuni*
12	Campeona	*C. reticulata*	*C. nobilis*
13	Geleking	*C. reticulata*	*C. nobilis*
14	Ladu	*C. reticulata*	*C. paratangerina*
15	Ladu ordinaire	*C. reticulata*	*C. paratangerina*
16	Cleopatra	*C. reticulata*	*C. reshni*
17	Ponkan	*C. reticulata*	*C. reticulata*
18	Sun chu sha	*C. reticulata*	*C. reticulata*
19	Bombay	*C. reticulata*	*C. reticulata*
20	Ougan	*C. reticulata*	*C. suavissima*
21	Ben di zao	*C. reticulata*	*C. succosa*
22	de Soe	*C. reticulata*	*C. suhuiensis*
23	Szinkom	*C. reticulata*	*C. suhuiensis*
24	Sunki	*C. reticulata*	*C. sunki*
25	Sunki	*C. reticulata*	*C. sunki*
26	*C. tachibana*	*C. tachibana*	*C. tachibana*
27	Swatow	*C. reticulata*	*C. tangerina*
28	Dancy	*C. reticulata*	*C. tangerina*
29	Clausellina	*C. reticulata*	*C. unshiu*
30	Dobashi-Beni	*C. reticulata*	*C. unshiu*
31	Azimboa	*C. maxima*	*C. maxima*
32	Deep Red	*C. maxima*	*C. maxima*
33	Pink	*C. maxima*	*C. maxima*
34	Chandler	*C. maxima*	*C. maxima*
35	Gil	*C. maxima*	*C. maxima*
36	Da Xhang	*C. maxima*	*C. maxima*
37	Nam Roi	*C. maxima*	*C. maxima*
38	Flores	*C. maxima*	*C. maxima*
39	Timor	*C. maxima*	*C. maxima*
40	Sans Pepins	*C. maxima*	*C. maxima*
41	Tahiti	*C. maxima*	*C. maxima*
42	Arizona	*C. medica*	*C. medica*
43	Corsica	*C. medica*	*C. medica*
44	Buddha's hand	*C. medica*	*C. medica*
45	Diamante	*C. medica*	*C. medica*
46	Poncire Commun	*C. medica*	*C. medica*
47	Humpang	*C. medica*	*C. medica*
48	Mauritus Papeda	*C. hystrix*	*C. hystrix*
49	Ichang Papeda	*C. ichangensis*	*C. ichangensis*
50	Khasi Papeda	*C. latipes*	*C. latipes*
51	Micrantha	*C. micrantha*	*C. micrantha*
52	Meiwa Kumkuat	*Fortunella* hybrid	*F. crassifolia*
53	Hong Kong Kumkuat	*F. hindsii*	*F. hindsii*
54	Round Kumkuat	*F. japonica*	*F. japonica*
55	Nagami Kumkuat	*F. margarita*	*F. margarita*
56	Bintangor Sarawak	?	?
57	*Citrus daoxianensis*	*C. reticulata*	*C. daoxianensis*
58	*C. halimii*	*C. reticulata*	*C. halimii*
59	Yuzu	*C. ichangensis* x *C. reticulata* var. *austera*	*C. junos*
60	Karna	*C. limon*	*C. karna*
61	Calamondin	*C. reticulata* var. *austera*? x *Fortunella*?	*C. madurensis*
62	Shunkokan	?	*C. shunkokan*
63	Tankan SG	*C. sinensis*	*C. tankan*
64	Temple	*C. reticulata*	*C. temple*
65	Temple	*C. reticulata*	*C. temple*

66	Temple Sue Linda	*C. reticulata*	*C. temple*
67	Changsa	*C. reticulata*	*C. reticulata* x *P. trifoliata*
68	Clemenules	*C. reticulata*	*C. clementina*
69	Oronules	*C. reticulata*	*C. clementina*
70	Arrufatina	*C. reticulata*	*C. clementina*
71	Avana Apireno	*C. reticulata*	*C. deliciosa*
72	Willowleaf seedless	*C. reticulata*	*C. deliciosa*
73	Willowleaf seedless	*C. reticulata*	*C. deliciosa*
74	Salteñita	*C. reticulata*	*C. deliciosa*
75	Tardivo Di Ciaculli	*C. reticulata*	*C. deliciosa*
76	à peau lisse	*C. reticulata*	*C. deliciosa*
77	à peau rugueuse	*C. reticulata*	*C. reticulata*
78	Clemendor	*C. reticulata*	*C. deliciosa*
79	Empress	*C. reticulata*	*C. reticulata*
80	Late Emperor SG	*C. reticulata*	*C. reticulata*
81	Montenegrina	*C. reticulata*	*C. deliciosa*
82	Natal Tightskin	*C. reticulata*	*C. deliciosa*
83	Shekwasha	*C. reticulata*	*C. depressa*
84	Shekwasha	*C. reticulata*	*C. depressa*
85	Shekwasha	*C. reticulata*	*C. depressa*
86	Fuzhu	*C. reticulata*	*C. erythrosa*
87	Vietnam	*C. reticulata*	*C. kinokuni*
88	Vietnam	*C. reticulata*	*C. kinokuni*
89	Vietnam	*C. reticulata*	*C. kinokuni*
90	du Japon	*C. reticulata*	*C. nobilis*
91	Kunembo	*C. reticulata*	*C. nobilis*
92	Rodeking	*C. reticulata*	*C. nobilis*
93	King (Laï Vung)	*C. reticulata*	*C. nobilis*
94	Yellow King	*C. reticulata*	*C. nobilis*
95	Anana	*C. reticulata*	*C. reticulata*
96	Carvahal	*C. reticulata*	*C. reticulata*
97	Emperor	*C. reticulata*	*C. reticulata*
98	Imperial australia	*C. reticulata*	*C. reticulata*
99	Scarlet	*C. reticulata*	*C. reticulata*
100	Africa do Sul SG	*C. reticulata*	*C. reticulata*
101	Ampefy	*C. reticulata*	*C. reticulata*
102	Antillaise	*C. reticulata*	*C. reticulata*
103	Antsalaka Diego SG	*C. reticulata*	*C. reticulata*
104	Atumbua	*C. reticulata*	*C. reticulata*
105	Augustino	*C. reticulata*	*C. reticulata*
106	Batangas	*C. reticulata*	*C. reticulata*
107	Bower	*C. reticulata*	*C. reticulata*
108	Burgess	*C. reticulata*	*C. reticulata*
109	Capurro SG	*C. reticulata*	*C. reticulata*
110	Chiuka	*C. reticulata*	*C. reticulata*
111	Cravo	*C. reticulata*	*C. reticulata*
112	Douhala	*C. reticulata*	*C. reticulata*
113	East India SG	*C. reticulata*	*C. reticulata*
114	Enterprise	*C. reticulata*	*C. reticulata*
115	Federici	*C. reticulata*	*C. reticulata*
116	Fewtrell SG	*C. reticulata*	*C. reticulata*
117	Gayunan	*C. reticulata*	*C. reticulata*
118	Giant	*C. reticulata*	*C. reticulata*
119	Hall SG	*C. reticulata*	*C. deliciosa*
120	Hickson	*C. reticulata*	*C. reticulata*
121	Imperial	*C. reticulata*	*C. reticulata*
122	Improved	*C. reticulata*	*C. reticulata*
123	Kobayashi	*C. reticulata*	*C. reticulata*
124	Ladu x Szibat	*C. reticulata*	*C. reticulata*
125	Ladu x Szinking	*C. reticulata*	*C. reticulata*
126	Le Roux	*C. reticulata*	*C. reticulata*
127	Lebon SG	*C. reticulata*	*C. reticulata*
128	Lime sucrée	*C. reticulata*	*C. reticulata*
129	Lukan	*C. reticulata*	*C. reticulata*
130	Macaque	*C. reticulata*	*C. reticulata*
131	Nanfen Miguan	*C. reticulata*	*C. reticulata*
132	Nicaragua	*C. reticulata*	*C. reticulata*
133	Oneco	*C. reticulata*	*C. reticulata*
134	Pan American	*C. reticulata*	*C. reticulata*
135	Robinson	*C. reticulata*	*C. reticulata*
136	Small SG	*C. reticulata*	*C. deliciosa*
137	sud-est Martinique	*C. reticulata*	*C. reticulata*

138	Suntara	*C. reticulata*	*C. reticulata*
139	Tshello	*C. reticulata*	*C. reticulata*
140	Warnuco	*C. reticulata*	*C. reticulata*
141	Willowleaf x Blood	*C. reticulata*	*C. reticulata*
142	Xien Khuang	*C. reticulata*	*C. reticulata*
143	Sun Chu Sha	*C. reticulata*	*C. reticulata*
144	de Soe	*C. reticulata*	*C. suhuiensis*
145	de Soe	*C. reticulata*	*C. suhuiensis*
146	Pet Yala	*C. reticulata*	*C. suhuiensis*
147	Se hui gan	*C. reticulata*	*C. suhuiensis*
148	Szibat	*C. reticulata*	*C. suhuiensis*
149	Yala	*C. reticulata*	*C. suhuiensis*
150	Sunki	*C. reticulata*	*C. sunki*
151	Sunki	*C. reticulata*	*C. sunki*
152	Beauty of Glen Retreat	*C. reticulata*	*C. tangerina*
153	Brickaville	*C. reticulata*	*C. tangerina*
154	Da hong pao	*C. reticulata*	*C. tangerina*
155	Redskin	*C. reticulata*	*C. tangerina*
156	Sanguine Trabut	*C. reticulata*	*C. tangerina*
157	Sweet small	*C. reticulata*	*C. tangerina*
158	Zanzibar SG	*C. reticulata*	*C. tangerina*
159	Mandalina	*C. reticulata*	*C. tangerina*
160	Parson's special	*C. reticulata*	*C. tangerina*
161	Frost	*C. reticulata*	*C. unshiu*
162	Okitsu	*C. reticulata*	*C. unshiu*
163	Bendiguangju	*C. reticulata*	*C. unshiu*
164	Pucheng	*C. reticulata*	*C. unshiu*
165	Salzara	*C. reticulata*	*C. unshiu*
166	Kowano	*C. reticulata*	*C. unshiu*
167	A'-12	*C. reticulata*	*C. clementina* x (*C.unshiu* x *C.nobilis*)
168	Avasa 15	*C. reticulata*	*C. unshiu* x *C. clementina*
169	Avasa 16	*C. reticulata*	*C. clementina* x *C. deliciosa*
170	Avasa 17	*C. reticulata*	*C. unshiu* x *C. clementina*
171	C-54-4-4	*C. reticulata*	*C. clementina x* (*C.reticulata* x *C.sinensis*)
172	D-19	*C. reticulata*	*C. clementina* x *(C.unshiu* x *C.nobilis)*
173	Daisy	*C. reticulata*	(*C. clementina* x *C. tangerina*) x (*C. clementina* x *C.reticulata*)
174	E'-5	*C. reticulata*	*C. clementina* x (*C.unshiu x C.nobilis*)
175	Encore	*C. reticulata*	*C. nobilis* x *C. deliciosa*
176	Fairchild	*C. reticulata*	*C. clementina x* (*C. paradisi x C. tangerina*)
177	Fallglo	*C. reticulata*	(*C. clementina* x (*C. paradisi x C. tangerina*)) x *C. temple*
178	Fortune	*C. reticulata*	*C. clementina* x *C. tangerina*
179	Fremont	*C. reticulata*	*C. clementina* x *C. reticulata*
180	Gold Nugget	*C. reticulata*	(*C. clementina* x *C. nobilis*) x *(C. nobilis* x *C. tangerina)*
181	Honey	*C. reticulata*	*C. nobilis* x *C. deliciosa*
182	Kara	*C. reticulata*	*C. unshiu* x *C. nobilis*
183	Kinnow	*C. reticulata*	*C. nobilis* x *C. deliciosa*
184	N-27	*C. reticulata*	*C. clementina* x *(C.unshiu* x *C.nobilis)*
185	Nova	*C. reticulata*	*C. clementina* x *(C. paradisi* x *C. tangerina)*
186	Osceola	*C. reticulata*	*C. clementina* x *(C. paradisi* x *C. tangerina)*
187	Page	*C. reticulata*	(*C. paradisi* x *C. tangerina*) x *C. clementina*
188	Page	*C. reticulata*	(*C. paradisi* x *C. tangerina*) x *C. clementina*
189	Palazzelli	*C. reticulata*	*C. clementina x C. nobilis*
190	Pixie	*C. reticulata*	(*C. nobilis* x *C. tangerina*) x ?
191	Primosole	*C. reticulata*	*C. unshiu* x (*C. deliciosa* x ?)
192	Simeto	*C. reticulata*	*C. unshiu* x *C. deliciosa*
193	Sunburst	*C. reticulata*	((*C. clementina* x (*C. paradisi* x *C. tangerina*)) x ((*C. clementina* x (*C. paradisi* x *C. tangerina*))
194	Wallent	*C. reticulata*	?

195	Wilking	*C. reticulata*	*C. nobilis* x *C. deliciosa*
196	Satsuma x Clementine	*C. reticulata*	?
197	Mapo	*C. reticulata*	*C.deliciosa* x *C.paradisi*
198	Minneola	*C. reticulata*	*C.paradisi* x *C.tangerina*
199	Orlando	*C. reticulata*	*C.paradisi* x *C.tangerina*
200	Seminole	*C. reticulata*	*C.paradisi* x *C.tangerina*
201	King	*C. reticulata*	*C. nobilis*
202	Kiyomi	*C. reticulata*	*C.unshiu* x *C.sinensis*
203	Dweet	*C. reticulata*	*C.tangerina* x *C.sinensis*
204	Ellendale	*C. reticulata*	*C.reticulata* x *C.sinensis*
205	Ellendale Leng	*C. reticulata*	*C.reticulata* x *C.sinensis*
206	Ellendale Taranco	*C. reticulata*	*C.reticulata* x *C.sinensis*
207	Murcott	*C. reticulata*	*C.reticulata* x *C.sinensis*
208	Murcott seedless	*C. reticulata*	*C.reticulata* x *C.sinensis*
209	Ortanique	*C. reticulata*	*C.reticulata* x *C.sinensis*
210	Umatilla	*C. reticulata*	*C.unshiu* x *C.sinensis*
211	Afourer	*C. reticulata*	(*C. reticulata* x *C. sinensis*) x ?
212	Bergamota	*C. reticulata*	?
213	Hybrida	*C. reticulata*	?
214	Neck	*C. reticulata*	?
215	Gailang	*C. reticulata*	?
216	Kiyomi	*C. reticulata*	*C. unshiu* x *C. sinensis*
217	(orange) Sanh	?	?
218	Bandipur (Népal)	?	?
219	Caibe	?	?
220	Importé de Chine marché Hanoï	?	?
221	Matieu (Laï Vung)	?	?
222	Paper (Qu'yt Giay)	?	?
223	S. E.	?	?

(ID) Identification number used in the whole article

Supplementary information 1 (Cont.).

ID	Germplasm bank code	PA	S 223 K=5	S 175 K=?	S 121 K=?	S 216 K=9
1	IVIA-478	Mandarin	1	0	0	-9
2	SRA-0100896	Mandarin	1	0	0	-9
3	IVIA-154	Mandarin	1	1	1	3
4	SRA-0100598	Mandarin	1	1	1	3
5	IVIA-238	Mandarin	1	0	0	-9
6	SRA-0100437	Mandarin	1	1	1	4
7	SRA-0100775	Mandarin	1	1	1	-9
8	SRA-0100769	Mandarin	1	1	1	-9
9	IVIA-550	Mandarin	1	0	0	-9
10	SRA-0100766	Mandarin	1	1	1	-9
11	SRA-0100839	Mandarin	1	1	1	-9
12	IVIA-193	Mandarin	1	1	1	-9
13	SRA-0100419	Mandarin	1	1	1	-9
14	SRA-0100595	Mandarin	1	1	1	4
15	SRA-0100590	Mandarin	1	1	1	-9
16	IVIA-385	Mandarin	1	1	1	5
17	IVIA-482	Mandarin	1	1	1	1
18	IVIA-483	Mandarin	1	1	1	-9
19	SRA-0100518	Mandarin	1	1	1	1
20	SRA-0100680	Mandarin	1	1	1	-9
21	SRA-0100582	Mandarin	1	1	1	-9
22	SRA-0100713	Mandarin	1	1	1	1
23	SRA-0100597	Mandarin	1	1	1	-9
24	IVIA-239	Mandarin	1	0	0	-9
25	SRA-0100971	Mandarin	1	0	0	-9
26	IVIA-237	Mandarin	1	0	0	-9
27	SRA-0100175	Mandarin	1	1	1	1
28	IVIA-434	Mandarin	1	1	1	-9
29	IVIA-019	Mandarin	1	1	1	2
30	SRA-0100681	Mandarin	1	1	1	2
31	IVIA-420	Pummelo	2	0	0	6
32	IVIA-277	Pummelo	2	0	0	6
33	IVIA-275	Pummelo	2	0	0	6
34	IVIA-207	Pummelo	2	0	0	6
35	IVIA-321	Pummelo	2	0	0	6
36	IVIA-589	Pummelo	2	0	0	6
37	IVIA-590	Pummelo	2	0	0	6
38	SRA-0100673	Pummelo	2	0	0	6
39	SRA-0100707	Pummelo	2	0	0	6
40	SRA-0100710	Pummelo	2	0	0	6
41	SRA-0100727	Pummelo	2	0	0	6
42	IVIA-169	Citron	3	0	0	7
43	IVIA-567	Citron	3	0	0	7
44	IVIA-202	Citron	3	0	0	7
45	IVIA-560	Citron	3	0	0	7
46	SRA-0100701	Citron	3	0	0	7
47	SRA-0100722	Citron	3	0	0	7
48	IVIA-178	Papeda	4	0	0	8
49	IVIA-358	Papeda	4	0	0	8
50	SRA-0100844	Papeda	4	0	0	8
51	IVIA-626	Papeda	4	0	0	8
52	IVIA-280	Fortunella	5	0	0	9
53	IVIA-281	Fortunella	5	0	0	9
54	IVIA-381	Fortunella	5	0	0	9
55	IVIA-038	Fortunella	5	0	0	9
56	SRA-0100683	NPA	-9	1	1	-9
57	IVIA-359	NPA	-9	0	0	-9
58	IVIA-278	NPA	-9	0	0	-9
59	IVIA-335	NPA	-9	0	0	-9
60	IVIA-242	NPA	-9	0	0	-9
61	IVIA-135	NPA	-9	0	0	-9
62	IVIA-241	NPA	-9	0	0	-9
63	SRA-0100524	NPA	-9	1	1	-9
64	IVIA-081	NPA	-9	1	1	-9
65	SRA-0100176	NPA	-9	1	1	-9
66	SRA-0100467	NPA	-9	1	1	-9
67	IVIA-452	NPA	-9	0	0	-9
68	IVIA-022	NPA	-9	1	0	-9
69	IVIA-132	NPA	-9	1	0	-9
70	IVIA-058	NPA	-9	1	0	-9
71	IVIA-189	NPA	-9	1	1	3
72	IVIA-340	NPA	-9	1	1	3
73	IVIA-383	NPA	-9	1	1	3
74	IVIA-361	NPA	-9	1	1	3

75	IVIA-186	NPA	-9	1	1	3
76	SRA-0100267	NPA	-9	1	1	4
77	SRA-0100277	NPA	-9	1	1	1
78	SRA-0100658	NPA	-9	1	1	3
79	SRA-0100416	NPA	-9	1	1	1
80	SRA-0100423	NPA	-9	1	1	1
81	SRA-0100553	NPA	-9	1	1	3
82	SRA-0100481	NPA	-9	1	1	-9
83	SRA-0100847	NPA	-9	0	0	-9
84	SRA-0100982	NPA	-9	0	0	-9
85	SRA-0100983	NPA	-9	0	0	-9
86	IVIA-571	NPA	-9	1	1	1
87	SRA-0100914	NPA	-9	0	0	-9
88	SRA-0100800	NPA	-9	1	1	5
89	SRA-0100764	NPA	-9	1	1	-9
90	SRA-0100279	NPA	-9	1	1	-9
91	SRA-0100326	NPA	-9	1	1	-9
92	SRA-0100431	NPA	-9	1	1	-
93	SRA-#22	NPA	-9	1	0	-
94	SRA-0100441	NPA	-9	1	1	-9
95	IVIA-390	NPA	-9	1	1	-9
96	IVIA-568	NPA	-9	1	1	-9
97	IVIA-394	NPA	-9	1	1	1
98	IVIA-576	NPA	-9	1	1	-9
99	IVIA-411	NPA	-9	1	1	-9
100	SRA-0100517	NPA	-9	1	1	1
101	SRA-0100495	NPA	-9	1	1	-
102	SRA-0100497	NPA	-9	1	1	1
103	SRA-0100527	NPA	-9	1	1	1
104	SRA-0100721	NPA	-9	1	1	-9
105	SRA-0100554	NPA	-9	1	1	-9
106	SRA-0100057	NPA	-9	1	1	-9
107	SRA-0100350	NPA	-9	1	1	-9
108	SRA-0100412	NPA	-9	1	1	-9
109	SRA-0100519	NPA	-9	1	1	-9
110	SRA-0100917	NPA	-9	1	1	5
111	SRA-0100434	NPA	-9	1	1	-9
112	SRA-0100767	NPA	-9	1	1	5
113	SRA-0100414	NPA	-9	1	1	4
114	SRA-0100521	NPA	-9	1	1	-9
115	SRA-0100417	NPA	-9	1	1	-9
116	SRA-0100418	NPA	-9	1	1	-9
117	SRA-0100600	NPA	-9	1	1	-9
118	SRA-0100420	NPA	-9	1	1	1
119	SRA-0100522	NPA	-9	1	1	3
120	SRA-0100523	NPA	-9	1	1	1
121	SRA-0100587	NPA	-9	1	1	-9
122	SRA-0100421	NPA	-9	1	1	-9
123	SRA-0100782	NPA	-9	0	0	-9
124	SRA-0100589	NPA	-9	1	1	4
125	SRA-0100588	NPA	-9	1	1	4
126	SRA-0100496	NPA	-9	1	1	1
127	SRA-0100425	NPA	-9	1	1	1
128	SRA-0100424	NPA	-9	1	1	-9
129	SRA-0100654	NPA	-9	1	1	1
130	SRA-0100426	NPA	-9	1	1	-9
131	SRA-0100700	NPA	-9	1	1	-9
132	SRA-0100693	NPA	-9	0	0	-9
133	SRA-0100429	NPA	-9	1	1	1
134	SRA-0100706	NPA	-9	1	1	1
135	SRA-0100139	NPA	-9	1	1	-9
136	SRA-0100526	NPA	-9	1	1	3
137	SRA-0100435	NPA	-9	1	1	-9
138	SRA-0110251	NPA	-9	0	0	-9
139	SRA-0100723	NPA	-9	1	1	1
140	SRA-0100439	NPA	-9	1	1	-9
141	SRA-0100440	NPA	-9	1	1	-9
142	SRA-0100868	NPA	-9	1	1	5
143	SRA-0100786	NPA	-9	1	1	5
144	SRA-0100653	NPA	-9	1	1	1
145	SRA-0100735	NPA	-9	1	1	1
146	SRA-0100694	NPA	-9	1	1	-9
147	SRA-0100586	NPA	-9	1	1	-9
148	SRA-0100596	NPA	-9	1	1	-9
149	SRA-0100655	NPA	-9	1	1	-9
150	SRA-0100705	NPA	-9	0	0	-9

151	SRA-0100970	NPA	-9	0	0	-9
152	SRA-0100261	NPA	-9	1	1	4
153	SRA-0100266	NPA	-9	1	1	4
154	SRA-0100591	NPA	-9	1	1	-9
155	SRA-0100428	NPA	-9	1	1	-9
156	SRA-0100264	NPA	-9	1	1	4
157	SRA-0100826	NPA	-9	1	1	-9
158	SRA-0100442	NPA	-9	1	1	4
159	SRA-GA1145	NPA	-9	1	1	4
160	IVIA-168	NPA	-9	1	1	-9
161	IVIA-175	NPA	-9	1	1	2
162	IVIA-195	NPA	-9	1	1	2
163	SRA-0100578	NPA	-9	1	1	2
164	SRA-0100657	NPA	-9	1	1	2
165	SRA-0100341	NPA	-9	1	1	2
166	SRA-0100167	NPA	-9	1	1	2
167	IVIA-424	NPA	-9	1	0	-9
168	IVIA-439	NPA	-9	1	0	-9
169	IVIA-440	NPA	-9	1	0	-9
170	IVIA-438	NPA	-9	1	0	-9
171	IVIA-453	NPA	-9	1	0	-9
172	IVIA-447	NPA	-9	1	0	-9
173	IVIA-362	NPA	-9	1	0	-9
174	IVIA-421	NPA	-9	1	0	-9
175	IVIA-155	NPA	-9	1	0	-9
176	IVIA-083	NPA	-9	1	0	-9
177	IVIA-466	NPA	-9	1	0	-9
178	IVIA-080	NPA	-9	1	0	-9
179	IVIA-082	NPA	-9	1	0	-9
180	IVIA-523	NPA	-9	1	0	-9
181	IVIA-209	NPA	-9	1	0	-9
182	IVIA-218	NPA	-9	1	0	-9
183	IVIA-033	NPA	-9	1	0	-9
184	IVIA-423	NPA	-9	1	0	-9
185	IVIA-074	NPA	-9	1	0	-9
186	IVIA-573	NPA	-9	1	0	-9
187	IVIA-079	NPA	-9	1	0	-9
188	IVIA-429	NPA	-9	1	0	-9
189	IVIA-188	NPA	-9	1	0	-9
190	IVIA-210	NPA	-9	1	0	-9
191	IVIA-414	NPA	-9	1	0	-9
192	IVIA-413	NPA	-9	1	0	-9
193	IVIA-200	NPA	-9	1	0	-9
194	IVIA-404	NPA	-9	1	0	-
195	IVIA-028	NPA	-9	1	0	-9
196	SRA-0100791	NPA	-9	1	0	-9
197	IVIA-190	NPA	-9	1	0	-9
198	IVIA-084	NPA	-9	1	0	-9
199	IVIA-101	NPA	-9	1	0	-9
200	IVIA-348	NPA	-9	1	0	-9
201	IVIA-477	NPA	-9	1	0	-
202	IVIA-405	NPA	-9	1	0	-9
203	IVIA-165	NPA	-9	1	0	-9
204	IVIA-194	NPA	-9	1	0	-9
205	IVIA-353	NPA	-9	1	0	-9
206	IVIA-575	NPA	-9	1	0	-9
207	IVIA-196	NPA	-9	1	0	-9
208	IVIA-371	NPA	-9	1	0	-9
209	IVIA-276	NPA	-9	1	0	-9
210	IVIA-100	NPA	-9	1	0	-9
211	SRA-0100741	NPA	-9	1	0	-9
212	SRA-0100164	NPA	-9	1	0	-9
213	SRA-0100714	NPA	-9	1	0	-9
214	SRA-0100674	NPA	-9	1	0	-9
215	SRA-0100575	NPA	-9	1	0	-
216	SRA-0100704	NPA	-9	1	0	-9
217	SRA-#45	NPA	-9	1	1	-
218	SRA-#NEPAL2	NPA	-9	1	1	1
219	SRA-#11	NPA	-9	1	1	-9
220	SRA-#27	NPA	-9	1	1	-9
221	SRA-#18	NPA	-9	1	1	-9
222	SRA-#8	NPA	-9	1	1	-9
223	SRA-0100433	NPA	-9	1	1	-9

(ID) Identification number used in the whole article; (PA) Population assigned in Structure software (S); (NPA) Not population assigned

Supplementary information 2. Statistical summary of the diversity for the SSR and indel markers employed in the genotyping of the whole dataset.

Marker	Type	MAF	*A*	GD	H_o	H_e	F_w
mCrCIR02D09	SSR	0.42	18	0.77	0.69	0.74	0.08
TAA41	SSR	0.31	18	0.82	0.70	0.80	0.13
mCrCIR06B07	SSR	0.53	8	0.62	0.47	0.57	0.18
mCrCIR01C07	SSR	0.38	16	0.76	0.75	0.73	-0.03
mCrCIR05A05	SSR	0.31	16	0.79	0.57	0.77	0.25
mCrCIR04H06	SSR	0.52	8	0.66	0.59	0.61	0.04
MEST46	SSR	0.44	12	0.70	0.68	0.66	-0.02
CAC15	SSR	0.77	6	0.37	0.42	0.33	-0.26
mCrCIR03C08	SSR	0.36	15	0.79	0.68	0.77	0.11
CAC23	SSR	0.57	7	0.53	0.64	0.44	-0.43
MEST256	SSR	0.55	14	0.64	0.30	0.61	0.50
mCrCIR02G12	SSR	0.41	12	0.77	0.73	0.74	0.02
MEST131	SSR	0.45	6	0.64	0.61	0.57	-0.08
MEST121	SSR	0.54	11	0.59	0.61	0.51	-0.18
MEST1	SSR	0.49	13	0.65	0.65	0.60	-0.08
MEST431	SSR	0.68	11	0.50	0.42	0.47	0.10
TAA15	SSR	0.27	13	0.83	0.77	0.81	0.04
mCrCIR03D12a	SSR	0.27	18	0.82	0.81	0.79	-0.02
mCrCIR02D04b	SSR	0.27	19	0.81	0.81	0.79	-0.03
mCrCIR03G05	SSR	0.35	10	0.79	0.62	0.76	0.18
mCrCIR07D06	SSR	0.39	11	0.74	0.70	0.71	0.01
MEST15	SSR	0.31	7	0.77	0.81	0.73	-0.11
MEST104	SSR	0.45	8	0.71	0.73	0.68	-0.08
mCrCIR01F08a	SSR	0.80	7	0.35	0.24	0.34	0.30
MEST88	SSR	0.34	8	0.73	0.57	0.68	0.16
mCrCIR05A04	SSR	0.64	8	0.50	0.39	0.43	0.09
mCrCIR07E12	SSR	0.57	14	0.64	0.61	0.61	0.00
MEST115	SSR	0.73	6	0.42	0.39	0.38	-0.03
mCrCIR06A12	SSR	0.65	8	0.54	0.49	0.50	0.03
MEST56	SSR	0.37	16	0.78	0.69	0.75	0.08
mCrCIR04H12	SSR	0.72	11	0.47	0.35	0.45	0.23
MEST192	SSR	0.37	18	0.81	0.74	0.80	0.07
mCrCIR02F12	SSR	0.42	11	0.72	0.64	0.68	0.05
mCrCIR01D11	SSR	0.41	9	0.70	0.33	0.65	0.49
MEST488	SSR	0.32	14	0.80	0.80	0.78	-0.03
mCrCIR01E02	SSR	0.42	12	0.73	0.64	0.69	0.07
mCrCIR01C06	SSR	0.28	18	0.80	0.81	0.78	-0.04
TAA1	SSR	0.66	5	0.48	0.47	0.41	-0.14
mCrCIR02A09	SSR	0.48	12	0.70	0.64	0.67	0.04
mCrCIR02G02	SSR	0.31	16	0.79	0.82	0.76	-0.08
mCrCIR02F07	SSR	0.76	14	0.41	0.24	0.40	0.39
mCrCIR07B05	SSR	0.34	15	0.79	0.51	0.76	0.33
mCrCIR01F04a	SSR	0.24	15	0.85	0.82	0.83	0.02
MEST86	SSR	0.40	9	0.69	0.65	0.63	-0.02
MEST107	SSR	0.76	6	0.38	0.30	0.34	0.10
mCrCIR03B07	SSR	0.59	13	0.61	0.49	0.58	0.15
mCrCIR07C07	SSR	0.45	9	0.73	0.59	0.70	0.15
mCrCIR07C09	SSR	0.55	10	0.64	0.68	0.60	-0.13
mCrCIR07F11	SSR	0.37	16	0.80	0.84	0.77	-0.09
mCrCIR02B07	SSR	0.31	15	0.77	0.79	0.73	-0.09
IDCHI	indel	0.97	4	0.06	0.03	0.06	0.56
IDHyb-2	indel	0.89	4	0.20	0.06	0.19	0.68
IDLCY2	indel	0.78	3	0.35	0.36	0.31	-0.17
IDTRPA	indel	0.79	2	0.33	0.39	0.28	-0.39
IDHyb-1	indel	0.92	7	0.15	0.10	0.14	0.32
IDPSY	indel	0.97	2	0.07	0.01	0.06	0.79
IDEMA	indel	0.84	5	0.28	0.24	0.26	0.07
IDDXS	indel	0.95	4	0.09	0.03	0.09	0.70
IDPEPC1	indel	0.91	2	0.17	0.09	0.15	0.38
IDPEPC2	indel	0.96	3	0.07	0.02	0.07	0.73
IDCAX	indel	0.53	5	0.60	0.50	0.53	0.06
IDAtGRC	indel	0.96	2	0.09	0.03	0.08	0.67
IDAVP	indel	0.98	2	0.04	0.02	0.04	0.42
IDDFR	indel	0.35	7	0.72	0.70	0.67	-0.03
IDINVA2	indel	0.99	3	0.01	0.01	0.01	-0.01
IDDFR2	indel	0.99	4	0.03	0.03	0.03	-0.02
IDPEPC3	indel	0.99	2	0.02	0.02	0.02	-0.02
IDINVA1	indel	0.96	3	0.07	0.00	0.07	0.94
IDFLS1	indel	0.66	4	0.47	0.53	0.38	-0.40
IDCHI2	indel	0.98	2	0.04	0.00	0.04	0.88
IDFLS2	indel	0.97	3	0.07	0.01	0.06	0.79

IDPKF	indel	0.49	3	0.62	0.53	0.55	0.04
IDF3'H	indel	0.64	2	0.46	0.53	0.35	-0.50
IDPSY2	indel	0.50	2	0.50	0.50	0.37	-0.33
Mean	SSR	0.47	12	0.67	0.61	0.64	0.05
Mean	indel	0.84	3	0.23	0.20	0.20	0.26

(MAF) Maximum allele frequency; (*A*) Allele number; (GD) Gene diversity; (H_o) Observed heterozygosity; (H_e) Expected heterozygosity; (F_w) Wright fixation index.

Supplementary information 3. Unique alleles present in some genotypes of the entire population.

Common name	Latin name Tanaka system	Germplasm bank code	UA
Cleopatra	*C. reshni*	IVIA-385	2
C. daoxianensis	*C. daoxianensis*	IVIA-359	1
Yuzu	*C. junos*	IVIA-335	20
Changsa	Citrandarin	IVIA-452	13
Osceola	*C. clementina* x (*C. paradisi* x *C. tangerina*)	IVIA-573	1
Pixie	(*C. nobilis* x *C. tangerina*) x ?	IVIA-210	1
Wilking	*C. nobilis* x *C. deliciosa*	IVIA-028	1
Nasnaran	*C. amblycarpa*	SRA-0100896	3
San hu hong chu	*C. erythrosa*	SRA-0100769	2
Vietnam	*C. kinokuni*	SRA-0100914	1
Rodeking	*C. nobilis*	SRA-0100431	1
Fewtrell SG	*C. reticulata*	SRA-0100418	1
Imperial	*C. reticulata*	SRA-0100587	1
Kobayashi	*C. reticulata*	SRA-0100782	2
Nicaragua	*C. reticulata*	SRA-0100693	10
Robinson	*C. reticulata*	SRA-0100139	1
Suntara	*C. reticulata*	SRA-0110251	6
Ougan	*C. suavissima*	SRA-0100680	1
de Soe	*C. suhuiensis*	SRA-0100735	2
Sunki	*C. sunki*	SRA-0100705	1
Sunki	*C. sunki*	SRA-0100970	2
Sunki	*C. sunki*	SRA-0100971	1
C. amblycarpa	*C. amblycarpa*	IVIA-478	8
Shunkokan	*C. shunkokan*	IVIA-241	4
C. tachibana	*C. tachibana*	IVIA-237	7

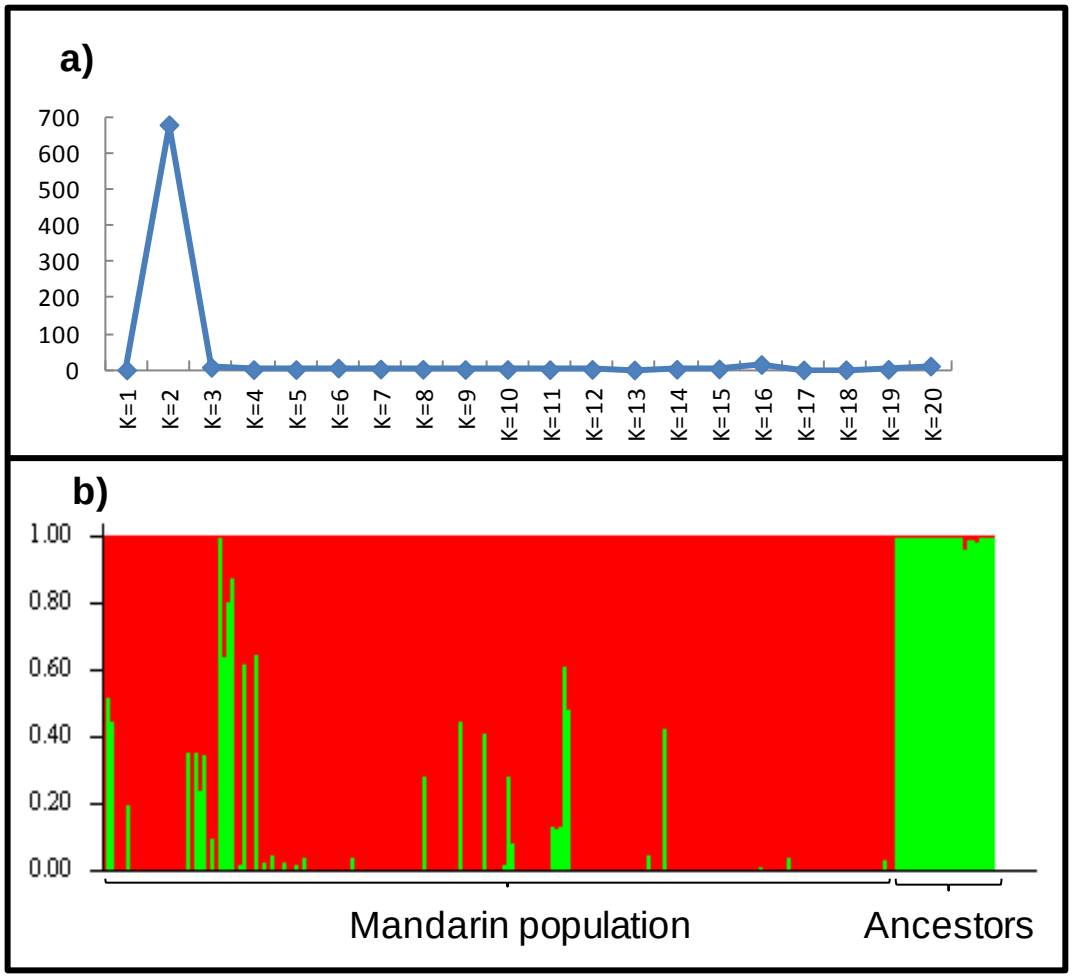
(UA) unique alleles

Supplementary information 4. Genotypes not distinguished with molecular markers.

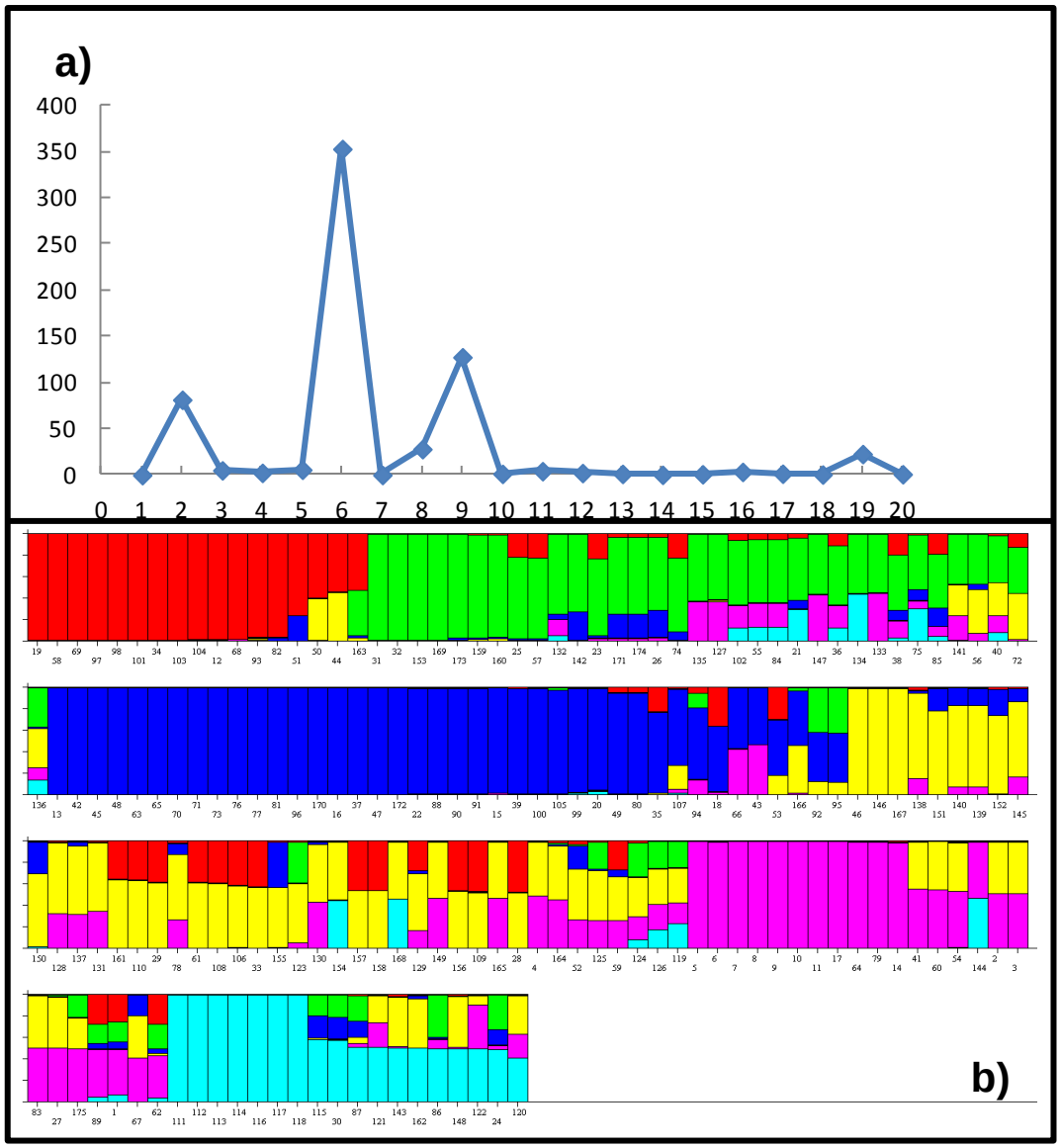
N	Common name	Swingle system	Database based on Tanaka system	Germplasm bank code	MLG
1	Xien Khuang	*C. reticulata*	*C. reticulata*	SRA-0100868	1
2	Chiuka	*C. reticulata*	*C. reticulata*	SRA-0100917	1
3	Clausellina	*C. reticulata*	*C. unshiu*	IVIA-019	2
4	Frost	*C. reticulata*	*C. unshiu*	IVIA-175	2
5	Okitsu	*C. reticulata*	*C. unshiu*	IVIA-195	2
6	Pucheng	*C. reticulata*	*C. unshiu*	SRA-0100657	2
7	Kowano	*C. reticulata*	*C. unshiu*	SRA-0100167	3
8	Dobashi-Beni	*C. reticulata*	*C. unshiu*	SRA-0100681	3
9	Lukan	*C. reticulata*	*C. reticulata*	SRA-0100654	4
10	Hickson	*C. reticulata*	*C. reticulata*	SRA-0100523	4
11	Pan American	*C. reticulata*	*C. reticulata*	SRA-0100706	4
12	à peau rugueuse	*C. reticulata*	*C. deliciosa*	SRA-0100277	4
13	Emperor	*C. reticulata*	*C. reticulata*	IVIA-394	4
14	Lebon SG	*C. reticulata*	*C. reticulata*	SRA-0100425	5
15	Le Roux	*C. reticulata*	*C. reticulata*	SRA-0100496	5
16	Antsalaka Diego SG	*C. reticulata*	*C. reticulata*	SRA-0100527	5
17	de Soe	*C. reticulata*	*C. suhuiensis*	SRA-0100653	6
18	de Soe	*C. reticulata*	*C. suhuiensis*	SRA-0100713	6
19	Caibe	?	Unknown	SRA-#11	7
20	Paper (Qu'yt Giay)	?	Unknown	SRA-#8	7
21	Sanguine Trabut	*C. reticulata*	*C. tangerina*	SRA-0100264	8
22	Ladu x Szinking	*C. reticulata*	*C. reticulata*	SRA-0100588	8
23	Vohangisany Ambodiampoly	*C. reticulata*	*C. depressa*	SRA-0100437	9
24	East India SG	*C. reticulata*	*C. reticulata*	SRA-0100414	9
25	Brickaville	*C. reticulata*	*C. tangerina*	SRA-0100266	9
26	Ellendale Leng	*C. reticulata*	*C. reticulata* x *C. sinensis*	IVIA-353	10
27	Ellendale Taranco	*C. reticulata*	*C. reticulata* x *C. sinensis ?*	IVIA-575	10
28	Tardivo Di Ciaculli	*C. reticulata*	*C. deliciosa*	IVIA-186	11
29	Avana Apireno	*C. reticulata*	*C. deliciosa*	IVIA-189	11
30	Willow leaf seedless	*C. reticulata*	*C. deliciosa*	IVIA-340	12
31	Willow leaf	*C. reticulata*	*C. deliciosa*	IVIA-154	12
32	Clemenules	*C. reticulata*	*C. clementina*	IVIA-022	13
33	Oronules	*C. reticulata*	*C. clementina*	IVIA-132	13
34	Murcott	*C. reticulata*	*C. reticulata* x *C. sinensis*	IVIA-196	14
35	Murcott seedless	*C. reticulata*	*C. reticulata* x *C. sinensis*	IVIA-371	14

(N) Number; (MLG) Multilocus genotypes: genotypes with the same number are not possible to be distinguished between them

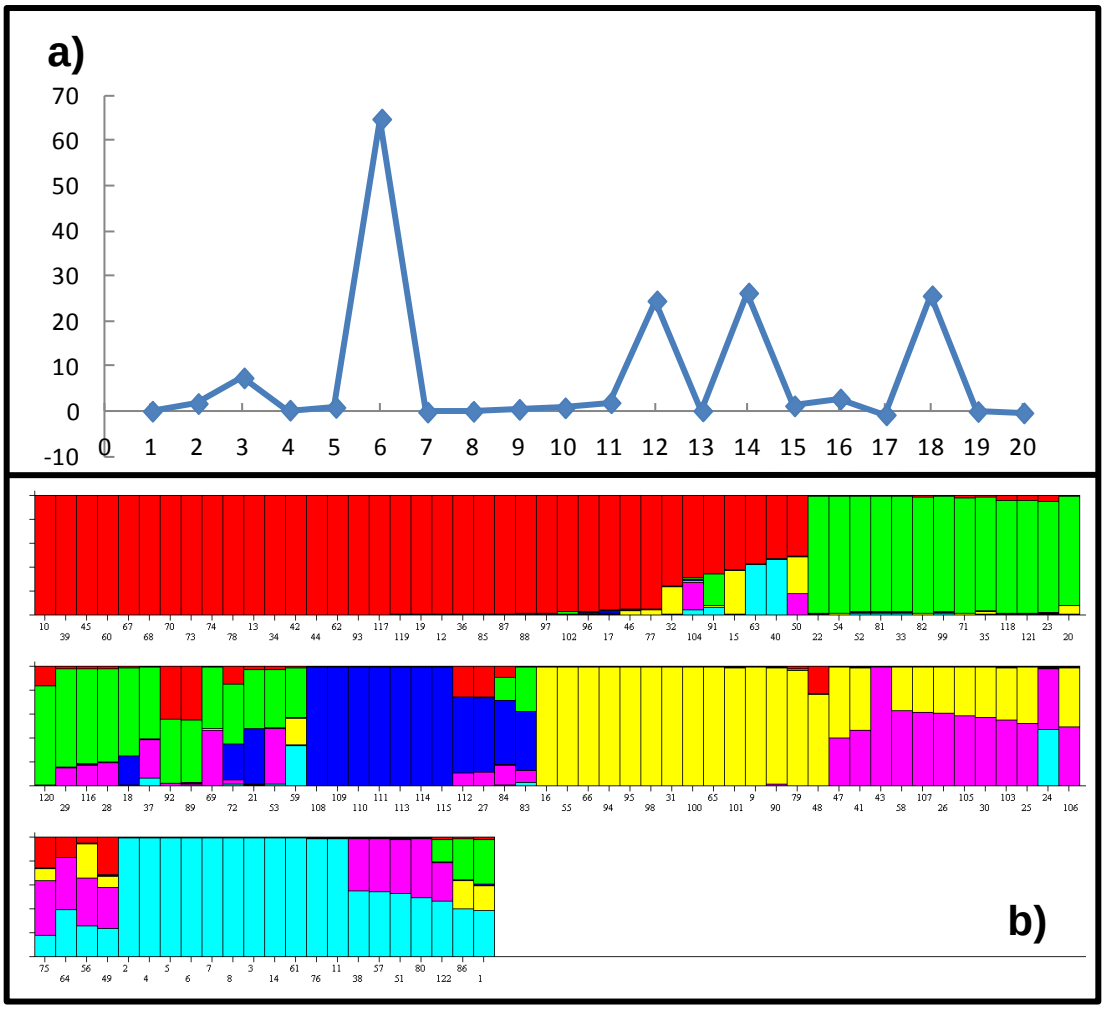
Supplementary information 5. Structure analysis 223 genotypes without assuming populations. **a)** Optimal Δ*K* value (number of populations within the population studied); **b)** Populations observed and its contributions to the rest of the genotypes. Population one (mandarins), population two (Ancestral populations: *C. maxima*, *C. medica*, *Papeda*, *Fortunella*).



Supplementary information 6. Structure analysis 175 genotypes (> 95% of mandarin genome) without assuming populations. **a)** Optimal Δ*K* value (number of populations within the population studied); **b)** Populations observed (*K* = 6) and its contributions to the rest of the genotypes. Population colors: Red, mainly *C. tangerina*; Green, tangor 'king' and hybrids; Dark blue, mainly *C. reticulata* [Tan.]; Yellow, 'Ampefy', 'Wallent' and 'Gailang'; Pink, *C. deliciosa*; Light blue, *C. unshiu*.



Supplementary information 7. Structure analysis 121 genotypes (mandarins without hybrids) without assuming populations. **a)** Optimal Δ*K* value (number of populations within the population studied); **b)** Populations observed (*K* = 6) and their contribution to the rest of the genotypes. Population colors: Red, mainly *C. reticulata* [Tan.]; Green, mixture of Tanaka's mandarin species; Dark blue, *C. unshiu*; Yellow, mainly *C. tangerina*; Pink, 'Ampefy'; Light blue, *C. deliciosa*.



Supplementary information 8. Assignation of the mandarin genotypes into a nuclear group from the parental mandarin groups identified in this work (N1 – N5). (?) Genotype not possible to assign to any cluster.

Unit	Common name	Latin name	Germplasm bank code	N
1	Bintangor Sarawak	?	SRA-0100683	3
2	*Citrus daoxianensis*	*C. daoxianensis*	IVIA-359	4
3	à peau lisse	*C. deliciosa*	SRA-0100267	2
4	à peau rugueuse	*C. reticulata*	SRA-0100277	1
5	Clemendor	*C. deliciosa*	SRA-0100658	3
6	de Chios	*C. deliciosa*	SRA-0100598	3
7	Empress	*C. reticulata*	SRA-0100416	1
8	Late Emperor Sg	*C. reticulata*	SRA-0100423	1
9	Montenegrina	*C. deliciosa*	SRA-0100553	3
10	Natal Tightskin	*C. deliciosa*	SRA-0100481	1
11	Avana Apireno	*C. deliciosa*	IVIA-189	3
12	Willowleaf	*C. deliciosa*	IVIA-154	3
13	Willowleaf seedless	*C. deliciosa*	IVIA-340	3
14	Willowleaf seedless	*C. deliciosa*	IVIA-383	3
15	Salteñita	*C. deliciosa*	IVIA-361	3
16	Tardivo Di Ciaculli	*C. deliciosa*	IVIA-186	3
17	Shekwasha	*C. depressa*	SRA-0100847	5
18	Shekwasha	*C. depressa*	SRA-0100982	5
19	Shekwasha	*C. depressa*	SRA-0100983	5
20	*Citrus Depressa*	*C. depressa*	IVIA-238	5
21	Vohangisany Ambodiampoly	*C. depressa*	SRA-0100437	2
22	Fuzhu	*C. erythrosa*	SRA-0100775	?
23	San hu hong chu	*C. erythrosa*	SRA-0100769	4
24	Fuzhu	*C. erythrosa*	IVIA-571	?
25	Nan feng mi chu	*C. kinokuni*	SRA-0100839	4
26	Vietnam	*C. kinokuni*	SRA-0100914	4
27	Vietnam	*C. kinokuni*	SRA-0100800	5
28	Vietnam	*C. kinokuni*	SRA-0100764	?
29	Vietnam à peau fine	*C. kinokuni*	SRA-0100766	5
30	Du Japon	*C. nobilis*	SRA-0100279	2
31	Geleking	*C. nobilis*	SRA-0100419	2
32	King (Laï Vung)	*C. nobilis*	SRA-#22	?
33	Kunembo	*C. nobilis*	SRA-0100326	4
34	Rodeking	*C. nobilis*	SRA-0100431	?
35	Yellow King	*C. nobilis*	SRA-0100441	2
36	Campeona	*C. nobilis*	IVIA-193	3
37	Ladu	*C. paratangerina*	SRA-0100595	2
38	Ladu Ordinaire	*C. paratangerina*	SRA-0100590	1
39	Cleopatra	*C. reshni*	IVIA-385	5
40	Africa Do Sul SG	*C. reticulata*	SRA-0100517	1
41	Ampefy	*C. reticulata*	SRA-0100495	2
42	Antillaise	*C. reticulata*	SRA-0100497	1
43	Antsalaka Diego Sg	*C. reticulata*	SRA-0100527	1
44	Atumbua	*C. reticulata*	SRA-0100721	?
45	Augustino	*C. reticulata*	SRA-0100554	2
46	Batangas	*C. reticulata*	SRA-0100057	2
47	Bombay	*C. reticulata*	SRA-0100518	1
48	Bower	*C. reticulata*	SRA-0100350	?
49	Burgess	*C. reticulata*	SRA-0100412	1
50	Capurro Sg	*C. reticulata*	SRA-0100519	2
51	Chiuka	*C. reticulata*	SRA-0100917	5
52	Cravo	*C. reticulata*	SRA-0100434	?
53	Douhala	*C. reticulata*	SRA-0100767	5
54	East India Sg	*C. reticulata*	SRA-0100414	2
55	Enterprise	*C. reticulata*	SRA-0100521	?
56	Federici	*C. reticulata*	SRA-0100417	3
57	Fewtrell Sg	*C. reticulata*	SRA-0100418	2
58	Gayunan	*C. reticulata*	SRA-0100600	3
59	Giant	*C. reticulata*	SRA-0100420	1
60	Hall Sg	*C. deliciosa*	SRA-0100522	3
61	Hickson	*C. reticulata*	SRA-0100523	1
62	Imperial	*C. reticulata*	SRA-0100587	1
63	Improved	*C. reticulata*	SRA-0100421	3
64	Kobayashi	*C. reticulata*	SRA-0100782	4
65	Ladu x Szibat	*C. reticulata*	SRA-0100589	2
66	Ladu x Szinking	*C. reticulata*	SRA-0100588	2
67	Le Roux	*C. reticulata*	SRA-0100496	1

68	Lebon Sg	*C. reticulata*	SRA-0100425	1
69	Lime Sucrée	*C. reticulata*	SRA-0100424	2
70	Lukan	*C. reticulata*	SRA-0100654	1
71	Macaque	*C. reticulata*	SRA-0100426	5
72	Nanfen Miguan	*C. reticulata*	SRA-0100700	4
73	Nicaragua	*C. reticulata*	SRA-0100693	5
74	Oneco	*C. reticulata*	SRA-0100429	1
75	Pan American	*C. reticulata*	SRA-0100706	1
76	Robinson	*C. reticulata*	SRA-0100139	?
77	Small Sg	*C. deliciosa*	SRA-0100526	3
78	sud-est Martinique	*C. reticulata*	SRA-0100435	1
79	Sun chu sha	*C. reticulata*	SRA-0100786	5
80	Suntara	*C. reticulata*	SRA-0110251	4
81	Tshello	*C. reticulata*	SRA-0100723	1
82	Warnuco	*C. reticulata*	SRA-0100439	2
83	Willowleaf x Blood	*C. reticulata*	SRA-0100440	3
84	Xien Khuang	*C. reticulata*	SRA-0100868	5
85	Anana	*C. reticulata*	IVIA-390	5
86	Carvahal	*C. reticulata*	IVIA-568	3
87	Emperor	*C. reticulata*	IVIA-394	1
88	Imperial Australia	*C. reticulata*	IVIA-576	1
89	Ponkan	*C. reticulata*	IVIA-482	1
90	Scarlet	*C. reticulata*	IVIA-411	2
91	Sun chu sha	*C. reticulata*	IVIA-483	5
92	Ougan	*C. suavissima*	SRA-0100680	4
93	Ben Di Zao	*C. succosa*	SRA-0100582	4
94	de Soe	*C. suhuiensis*	SRA-0100713	?
95	de Soe	*C. suhuiensis*	SRA-0100653	?
96	de Soe	*C. suhuiensis*	SRA-0100735	?
97	Pet Yala	*C. suhuiensis*	SRA-0100694	?
98	Se Hui Gan	*C. suhuiensis*	SRA-0100586	2
99	Szibat	*C. suhuiensis*	SRA-0100596	1
100	Szinkom	*C. suhuiensis*	SRA-0100597	3
101	Yala	*C. suhuiensis*	SRA-0100655	?
102	Sunki	*C. sunki*	SRA-0100971	5
103	Sunki	*C. sunki*	SRA-0100705	5
104	Sunki	*C. sunki*	SRA-0100970	4
105	Sunki	*C. sunki*	IVIA-239	5
106	*C. tachibana*	*C. tachibana*	IVIA-237	5
107	Brickaville	*C. tangerina*	SRA-0100266	2
108	Da Hong Pao	*C. tangerina*	SRA-0100591	?
109	Mandalina	*C. tangerina*	SRA-GA1145	2
110	Redskin	*C. tangerina*	SRA-0100428	?
111	Sanguine Trabut	*C. tangerina*	SRA-0100264	2
112	Swatow	*C. tangerina*	SRA-0100175	1
113	Sweet Small	*C. tangerina*	SRA-0100826	5
114	Zanzibar Sg	*C. tangerina*	SRA-0100442	2
115	Beauty of Glen Retreat	*C. tangerina*	SRA-0100261	2
116	Dancy	*C. tangerina*	IVIA-434	1
117	Parson's Special	*C. tangerina*	IVIA-168	2
118	Tankan Sg	*C. tankan*	SRA-0100524	1
119	Temple	*C. temple*	SRA-0100176	2
120	Temple Sue Linda	*C. temple*	SRA-0100467	2
121	Temple	*C. temple*	IVIA-81	2
122	Clausellina	*C. unshiu*	IVIA-19	4
123	Dobashi-Beni	*C. unshiu*	SRA-0100681	4
124	Frost	*C. unshiu*	IVIA-175	4
125	Okitsu	*C. unshiu*	IVIA-195	4
126	Bendiguangju	*C. unshiu*	SRA-0100578	4
127	Kowano	*C. unshiu*	SRA-0100167	4
128	Pucheng	*C. unshiu*	SRA-0100657	4
129	Salzara	*C. unshiu*	SRA-0100341	4
130	(Orange) Sanh	Unknown	SRA-#45	?
131	Bandipur (Népal)	Unknown	SRA-#NEPAL2	1
132	Caibe	Unknown	SRA-#11	?
133	Importé De Chine Marché Hanoï	Unknown	SRA-#27	1
134	Matieu (Laï Vung)	Unknown	SRA-#18	?
135	Paper (Qu'yt Giay)	Unknown	SRA-#8	?
136	S. E.	Unknown	SRA-0100433	3

(N) Nuclear group.

DISCUSSION

Citrus is one of the most important fruit crops in the world due to its economic importance. Despite the fact that its diversity (Krueger and Navarro, 2007) and origin have been widely studied (Webber *et al.*, 1967; Calabrese, 1992) the taxonomy, diversity and phylogeny of *Citrus* remain controversial. This is due to the large degree of morphological diversity found within this group, the sexual compatibility between the species and the apomixis of many genotypes (Scora, 1975).

In this PhD thesis a broad diversity of germplasm within the *Citrus* genus and *Citrus* relatives from the Aurantioideae subfamily has been studied in order to clarify their organization and phylogeny using different kind of molecular markers and different genotyping platforms.

1. New set of complementary markers have been developed.

Many different kinds of markers have been used to study the citrus diversity, from morphological characteristics (Barret and Rhodes, 1976; Ollitrault *et al.*, 2003), quantification of primary (Luro *et al.*, 2011) and secondary metabolites (Fanciullino *et al.*, 2006a), to molecular markers, isoenzymes (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003), RFLP (Federici *et al.*, 1998), RAPD, SCAR (Nicolosi *et al.*, 2000), AFLP (Liang *et al.*, 2007) and SSR (Luro *et al.*, 2001; Barkley *et al.*, 2006).

Several works have been recently published with the aim of developing diagnostic markers of the inter-specific differentiation in citrus. In the framework of this thesis, Garcia-Lor *et al.* (2012a) released for the first time in citrus insertion-deletion (indel) markers, and Garcia-Lor *et al.* (2013a) identified SNP markers mined in a large diversity panel, while (Ollitrault *et al.*, 2012a) analysed the value of SNPs mined in a single genotype of clementine. These recent papers, and the other previously cited, agree that most of the important commercial citrus species (secondary species) can be considered a mosaic of large DNA fragments of three ancestral species (*C. medica* L. –citrons-, *C. maxima* (Burm.) Merr. –pummelos- and *C. reticulata* Blanco –mandarins-) that resulted from a few inter-specific recombination events (*Curk et al.*, 2012). It is also accepted that *C. micrantha*, a member of the *Papeda* subgenus, is a potential parent of some limes (*C. aurantifolia* (L.) Christm.).

Indel markers developed in this thesis (Garcia-Lor *et al.*, 2012a; 2013a) seemed to be better phylogenetic markers than SSRs, as they are less polymorphic (low allele number) but display a higher organisation of genetic diversity at the interspecific level (F_{st} value higher than SSR). On the other hand, SSR markers showed a higher level of polymorphism and a better differentiation between varieties at intraspecific level.

Indels are more common in non-coding regions than in coding regions, as has been shown in other species like Brassica (Park *et al.*, 2010), melon (Morales *et al.*, 2004), or maize (Ching *et al.*, 2002). Indels play an important role in sequence divergence between closely related DNA sequences in animals, plants, insects and bacteria (Bapteste, 2002; Väli *et al.*, 2008; Vasemägi *et al.*, 2010). In humans, it has been suggested that indels are a major source

of gene defects. When they occur in coding regions they probably have functional roles and are considered to be a significant source of evolutionary change in eukaryotic and bacterial evolution (Britten *et al.*, 2003). They can also be included in genetic linkage maps, as it is the case of clementine (Ollitrault et al., 2012b).

The high level of SSR markers polymorphisms is due the high evolution rate of the number of repeats (Weber and Wong, 1993; Jarne and Lagoda, 1996), that can vary depending on the number of repeats or base composition (Bachtrog *et al.*, 2000). However, due to this important rate of variation, homoplasy should be relatively frequent, as Barkley *et al.* (2009) demonstrated in citrus and this limits the value of SSRs as phylogenetic markers.

Considering indel and SSR characteristics discussed before, these markers are complementary in diversity studies. Therefore, we have combined both kind of markers for the quantification of the exact contribution of ancestral genomes to the secondary species and some modern hybrids (Garcia-Lor *et al.*, 2012a) and for the study of the organization of the mandarin germplasm diversity (Garcia-Lor *et al.*, submitted) coming from two germplasm collections: IVIA Citrus Germplasm Bank of pathogen-free plants (Navarro *et al*., 2002), and the collection at the *Station de Recherches Agronomiques* (INRA/CIRAD).

From the 1097 SNPs identified by Garcia-Lor *et al.* (2013a) in a study based on Sanger sequencing of gene fragments in a broad discovery panel, forty-one of the mined SNP loci selected from a limited intra-generic discovery panel (*C. reticulata*, *C. maxima* and *C. medica*) were converted into efficient markers based on Competitive Allele-Specific PCR to perform a genotyping study through the KASPar genotyping system (KBiosciences) (Garcia-Lor et al., 2013b). The aim was to test their transferability across the Aurantioideae subfamily (Swingle and Reece, 1967). This genotyping method lost efficiency as the genetic distance was increasing. Within the *Citrus* genus, the secondary species and hybrids the missing data level was very low. It increased slightly in the close citrus and primitive citrus groups of the *Citrinae* subtribe, reaching higher levels for the two other subtribes of the *Citreae* tribe, the *Triphasilinae* and the *Balsamocitrinae*. The highest missing data level was found in the *Clauseniae* tribe. The conformity level between KASPar genotyping and Sanger sequencing was 95.41% (2.99% did not agree and 1.60% were missing data).

Moreover, 53 SNP loci where successfully integrated in a GoldenGate array and used for genetic diversity analysis (Ollitrault *et al*., 2012a) and genetic mapping (Ollitrault *et al*., 2012b). The level of conformity in Ollitrault *et al.* (2012a) was 99.2% with Sanger sequencing, confirming that this technique it is still a good method for SNP discovery.

These SNP markers will be important for the management of citrus germplasm collections and marker/trait association studies.

2. New insights have been obtained on the phylogeny of ancestral taxa.

For a biologically complex crop like citrus, the information obtained from nuclear gene sequences is more useful than the information from maternally-inherited chloroplast or mitochondrial sequences (Ramadugu *et al*., 2011; Puritz *et al*., 2012) due to the possibility of gene flow between sexually compatible species and the fact that the species belong to the same area of diversification. We have performed a study based on Sanger sequencing of gene fragments in a broad discovery panel (Garcia-Lor *et al.*, 2013a; annex chapter 2) to clarify the phylogenetic relationships between 'true citrus fruit trees' of the subtribe Citrinae (*Fortunella*, *Eremocitrus*, *Poncirus*, *Microcitrus*, *Clymenia* and *Citrus*). The starting dataset employed in this study was selected in order to avoid the ascertainment bias associated with a low genetic basis of a small discovery panel (Rosenblum and Novembre, 2007; Albrechtsen *et al.*, 2010; Ollitrault *et al.*, 2012a).

Nuclear phylogenetic analysis revealed that all 'true citrus fruit trees' species constitute a monophyletic clade, as it was previously shown (de Araújo *et al*., 2003; Bayer *et al*., 2009). The latter added two more species to this group, *Oxanthera* and *Feroniella* (not present in our study). An important observation was that *C. reticulata* and *Fortunella* are joined in a cluster that is differentiated from the clade that includes the three other basic taxa of cultivated citrus (*C. maxima*, *C. medica* and *C. micrantha*). These results confirm the taxonomic subdivision between the subgenera Metacitrus (East Asiatic floral zone) and Archicitrus (Indo-Malayan floral zone) and the geographical distribution of species divided by the 'Tanaka line' (Tanaka, 1954). Interestingly, some phenotypic traits (like the carotenoid content) differentiate these two clades. On one hand, *Fortunella*, *Poncirus* and *C. reticulata* are facultative apomictic species with high carotenoid contents, and on the other hand, *C. maxima* and *C. medica* are monoembryonic non-apomictic species, which have strong limitations in the carotenoid pathway. The apomixis might have been transferred to the secondary species via *C. reticulata* genome (Garcia-Lor *et al.*, 2013a). The speciation between *Fortunella*, *Poncirus* and *C. reticulata*, that share the same geographic distribution, might be explained by their different flowering periods (very precocious in *Poncirus* and late in *Fortunella*). However, gene flow probably occurred by accidental out-of-time flowering. Despite sharing the Indo-Malayan floral zone (Tanaka, 1954), *C. maxima* and *C. medica* were geographically separated, with a more intertropical specialisation for *C. maxima*.

The genus *Clymenia* (Annex chapter 2) is placed in the same clade than *Microcitrus* and *Eremocitrus*, which are clearly differentiated from *Citrus* and *Fortunella* clusters. Moreover, the null amount of heterozygosity in the gene fragments analysed indicates that *Clymenia* cannot be an interspecific or intergeneric hybrid. Our analysis is in agreement with Bayer et al. (2009) and Morton (2009), who observed *Clymenia* closely to *Microcitrus* and *Eremocitrus* in a phylogenetic study with cpDNA markers. From morphological data Swingle and Reece (1967) proposed that *C. medica* was closely related to *Clymenia.* In our results, the branch including *Clymenia*, *Microcitrus* and *Eremocitrus* is sister of the one formed by *C. maxima* and *C. medica*, confirming their probable relationship.

The higher level of non-synonymous to silent SNP rates per site (π_{nonsyn}/π_{sil}) found in citrus species than in other species, like white spruce (Pavy *et al.*, 2006) or *Arabidopsis thaliana* (Zhang *et al*., 2002), may be due to a lower purifying selection pressure in the 'true citrus fruit trees'. This can probably be attributed to the wide diversity encompassed by 'true citrus fruit trees' and the high genetic and phenotypic differentiation between the different taxa that have experienced allopatric evolution under highly differentiated environmental conditions.

Despite the fact that some genes exhibit selective pressure, the genetic organization found in citrus with SNP data is similar to previous SSR studies (Ollitrault *et al.*, 2010; Garcia-Lor *et al.*, 2013a), which suggests that the diversity existing in both kind of markers comes from similar types of evolution. For this reason, a neutral evolution pattern can be assumed in most of the SNP markers identified.

3. The origin of the secondary commercial species has been assessed.

As mentioned before, secondary species (*C. sinensis* –sweet orange-, *C. aurantium* –sour orange-, *C. paradisi* –grapefruit-, *C. limon* –lemon- and *C. aurantifolia* –lime-) and many recent hybrids come from interspecific hybridisations between the basic *Citrus* taxa (*C. maxima*, *C. reticulata*, *C. medica* and *C. micrantha*). Within these secondary species, we do not found intercultivar polymorphism at intraspecific level for *C. sinensis*, *C. aurantium* and *C. paradisi* (SSR, indel, mtDNA and SNP markers), whereas these species are highly heterozygous (SSRs, indels, SNPs). The same observation was made for clementine cultivars. Our results agree with previous molecular studies (Barkley *et al.*, 2006; Luro *et al.*, 2008) and confirm that most of the inter-varietal polymorphisms within these secondary species and in clementines and satsumas, arose from punctual mutation or movement of transposable elements (Breto *et al.*, 2001). Therefore, the quantification of the ancestral genomes contributions and the mosaic genome structure inferred from one or two genotypes can be extended to other cultivars of the same secondary species.

An important result of this research concerns to the origin of sweet orange (*C. sinensis)*. Roose *et al*., (2009) and (Garcia-Lor *et al.*, 2012a) showed that sweet orange posses almost a 75% of *C. reticulata* genome and a 25% from *C. maxima*, which indicated that a backcross 1 (BC1) [(*C. maxima* x *C. reticulata*) x *C. reticulata*] should be the most probable origin of *C. sinensis*. This BC1 theory was also proposed by Xu *et al.* (2013) from whole genome sequencing data. This theory differed with the hypothesis of Nicolosi *et al.* (2000) and Barkley *et al.* (2006) who proposed that sweet orange arose from a direct hybridization between *C. maxima* and *C. reticulata*. We have shown (Garcia-Lor *et al*., 2012a) that the two previous hypotheses were not in agreement with the genomic organisation of sweet orange. Indeed, based in multilocus SNP analysis we demonstrated the presence of nuclear genomic fragments in phylogenetic homozygosity inherited from *C. maxima* or *C. reticulata*. This result leads to state that the two parents of sweet orange were of interspecific origin (Garcia-Lor *et al.*, 2013a).

This conclusion was further confirmed by the whole genome sequencing analysis performed by the International Citrus Genome Consortium (Gmitter *et al.*, 2012).

It was proposed that grapefruit (*C. paradisi*) arose from a natural cross between *C. maxima* and *C. sinensis* (de Moraes *et al.,* 2007; Ollitrault *et al.,* 2012a). From SSR and indel data we have estimated (Garcia-Lor *et al.*, 2012a) that the genomic contributions of *C. reticulata* and *C. maxima* were respectively around 60% and 40%. This was confirmed with SNP data by (Garcia-Lor *et al.*, 2013a).

Citrus aurantium (sour orange) is thought to come from a natural hybridisation between *C. maxima* and *C. reticulata* (Nicolosi *et al.*, 2000; Uzun *et al.*, 2009). Our research (Garcia-Lor *et al.*, 2013a) showed a contribution from *C. reticulata* and *C. maxima* of 50%. Interestingly we also found that the genotype mandarin 'Suntara' share a lot of rare alleles with *C. aurantium* and could be either a parent or a hybrid from *C*. *aurantium* (Garcia-Lor *et al.*, 2012a).

This work has confirmed the hypothesis proposed by Nicolosi *et al.*, (2000) for the origin of *C. limon* (lemon), that resulted from a direct cross between *C. medica* and *C*. *aurantium*, as we have observed a tri-hybrid genome constitution (*C. medica*, *C. maxima* and *C. reticulata*) (Garcia-Lor *et al.*, 2012a, 2013a). *Citrus aurantifolia* (Mexican lime) was proposed by (Nicolosi *et al.*, 2000) to be a hybrid between *C. medica* and a *Papeda*. Most of our data fits with this theory, but in some SNP loci the *C. micrantha* (*Papeda*) used did not agree with this hypothesis.

Clementine is thought to have arisen from a cross between 'Willowleaf' mandarin and *C. sinensis* (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012a, b). Parental contributions observed in our work were not exactly the same, but the two studies agree with the previous hypothesis (Garcia-Lor *et al.*, 2012a, 2013a).

4. The genetic organisation of the mandarin germplasm was revealed.

An important focus of our research was the diversity of the mandarin-like genotypes, which are an increasing component of the citrus fresh fruit market (second most important group worldwide, FAOSTAT, 2010). The mandarin horticultural varietal group is highly polymorphic (Moore, 2001) and it is highly related with one of the basic taxa of the cultivated citrus (*C. reticulata*). It also includes genotypes introgressed by other species, like tangors (hybrids between *C. reticulata* and *C. sinensis*) and tangelos (hybrids between *C. reticulata* and *C. paradisi*). The precise contribution of the ancestral species to the mandarin group was not known.

Several botanical classifications have been proposed for mandarins. For Swingle and Reece (1967) all mandarins are included in *C. reticulata*, Webber (1943) divided the mandarins in four groups ('King', satsuma, mandarin and tangerine), Hodgson (1967) classified the mandarins in four species [*C. unshiu* (satsumas), *C. reticulata* ('Ponkan', 'Dancy', clementines), *C. deliciosa* ('Willowleaf') and *C. nobilis* ('King')], while Tanaka (1961) considered 36 mandarin species included in five groups. In addition to the taxonomic complexity of this citrus group, the

incorrect passport information of some genotypes and the redundancy present in citrus germplasm collections are extra problems (Krueger and Roose, 2003).

From molecular data, it is well documented that the mandarin group (*C. reticulata*) is clearly differentiated from the other *Citrus* species, *C. maxima, C. medica* and *C. micrantha* (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012a). Some works have tried to clarify the organization of the mandarin group (Coletta Filho *et al.*, 1998; Koehler-Santos *et al.*, 2003; Yamamoto and Tominaga, 2003; Tapia Campos *et al.*, 2005), but they are not conclusive. Recently, Froelicher *et al.* (2011) divided the mandarins in two groups, acid and sweet, based on mitochondrial indel markers. In this PhD thesis, joining the information coming from 50 SSRs and 25 indel markers dispersed throughout the genome, the introgression of other genomes (*C. maxima*, *C. medica*, *Papeda* or *Fortunella*) was quantified in a broad representation of the mandarin like germplasm (198 genotypes). The genome with the higher contribution is *C. maxima*, followed by *Papeda*, *C. medica* and *Fortunella* in a few genotypes. Similar contributions were observed by Barkley *et al.* (2006) in some genotypes. Our analysis clearly shows that some mandarins considered by Tanaka as species are not true mandarins, since they are hybrids between different ancestral taxa. This is the case of *C. amblycarpa*, *C. depressa*, *C. tachibana*, *C. succosa* (*C. reticulata* and *Papeda* genomes) and *C. indica* that has *C. reticulata* and *C. medica* genomic contributions. These results indicate that the Tanaka classification is not accurate and should be revised.

We have analysed the mandarin germplasm organization with two approaches, the Structure software (it uses a model-based clustering method using genotype data) and Neighbour Joining analysis (based in the simple matching dissimilarity index ($d_{i\text{-}j}$) between pairs of accessions). Both analyses come to the agreement that five groups can be defined to be the parental mandarins at nuclear level. Four are related with some Tanaka species [*C. reticulata* (N1), *C. unshiu* (N2), *C. deliciosa* (N3), *C. tangerina* (N4), while the last group includes different mandarin types (N5; acid mandarins, small fruit mandarins)]. Two more clusters including genotypes with clearly identified interspecific introgressions and their descendants were identified within the 'mandarin-like' germplasm, 'Ampefy', 'Wallent' and 'Gailang' group (N6, it shares a high percentage of more than 90%, of allelic similarity with the sweet oranges) and the tangor 'King' group (N7), which is parent of many hybrids.

Considering the five mandarin parental groups defined in this thesis, the contribution of these groups to the constitution of the other mandarin genomes was studied with the software Structure.

Most of the hybrids with known origin had a coherent genome structure when compared with their parents. In some cases they do not display totally additive contributions from their ancestors, which is logical considering the heterozygosity of their parents and the different reconstruction of the genomes through the mating process (Motohashi *et al*., 1992; Coletta Filho *et al.*, 1998). In other cases, our data contradict previous information in some genotypes, as it happened with 'Fortune' and 'Fremont' hybrids, which were supposed to come from a cross

between *C. clementina* and *C. tangerina* 'Dancy' and *C. clementina* x *C. reticulata* 'Ponkan', respectively, made by Furr (1964).

Mitochondrial markers are very useful to analyse the maternal phylogeny in citrus (Green *et al.*, 1986; Yamamoto *et al.*, 1993). In the mandarin germplasm studied in this work, four mitotypes were found. Two of them (C1, C2) were identified respectively by (Froelicher *et al.*, 2011) as sweet and acid mandarin mitotypes. However, in our study with more mandarin genotypes, C2 included acid but also sweet genotypes. The acid mandarins included in the C2 mytotype belong to two groups of acid genotypes identified by Tanaka (1954), *C. reshni*, *C. sunki* and *C. tachibana* in one group and *C. depressa* in another one. The mandarin mitochondrial mitotype group (C1) identified as 'sweet mandarin mitotype' by Froelicher *et al.* (2011) is divided in four groups with nuclear markers, *C. reticulata* (N1), *C. unshiu* (N2), *C. deliciosa* (N3), and *C. tangerina* (N4). The other two mitotypes observed correspond to the *Papeda* (C3; three genotypes) and the *C. maxima* mitotypes (C4; seven genotypes). The nuclear genetic structure of these ten last genotypes, sharing the *C. maxima* and *Papeda* mitotypes, were clear interspecific admixture, not true mandarins.

5. The genetic organization precludes association genetic studies based on linkage disequilibrium at the *Citrus* level but suggest potential application at mandarin germplasm level.

The data obtained with the three kinds of markers used for the diversity and phylogenetic studies (Indel, SSR and SNPs), revealed that the *Citrus* gene pool is highly structured in direct relation with the ancestral taxa differentiation. The deficit of heterozygous genotypes observed in the whole sample indicates a strong population subdivision (Hartl and Clark, 1997) and, therefore, a low gene flow between *C. medica*, *C. reticulata* and *C. maxima*. The differentiation between these sexually compatible taxa can be explained by the origin in three geographic zones and by an initial allopatric evolution. *Citrus maxima* originated in the Malay Archipelago and Indonesia, *C. medica* evolved in North-eastern India and the nearby region of Burma and China and *C. reticulata* diversification occurred over a region including Vietnam, Southern China and Japan (Webber *et al.*, 1967; Scora, 1975). This allopatric evolution resulted in a global genotypic and phenotypic divergence due to different selective pressures (found in some of the genes studied), mutation and genetic drift. Later on, human activity facilitated migration and hybridization among the differentiated gene pools of the basic taxa. However, the partial apomixis observed in most of the secondary species, which probably arose from the *C. reticulata* germplasm, has strongly limited the interspecific gene flow.

This evolution of *Citrus* resulted in a high and generalised Linkage Disequilibrium (LD) revealed in this PhD thesis (chapter 1; Garcia-Lor *et al.*, 2012a). This structure precludes association genetic studies at the genus level without developing additional recombinant populations from interspecific hybrids.

The decay of LD with increasing genetic distance, found in the mandarin group (our unpublished data), a less structured population, was lower than in the *Citrus* population. These results suggest that a LD-based association studies at the species level could be affordable. Anyway, the development of additional intraspecific hybrids, such as BC1 or F2, between mandarins or the generation of hybrids between a mandarin and interspecific species, would improve the success of genetic association studies by decreasing the LD between distant loci and limiting the risk of false associations between a marker and a phenotype, even though the marker is not physically linked to the locus responsible for the phenotypic variation.

6. Evolutionary patterns of different genes should be related with phenotypic polymorphisms.

Some genes of the different biosynthetic pathways studied presented interesting evolutionary patterns. Carotenoids are involved in different processes, like photosynthesis, fruit color, and precursors of vitamin A and have antioxidant capacity (Demmig-Adams *et al.*, 1996; Lee, 2002; Rao and Rao, 2007). Therefore, changes in the sequences of genes involved in their biosynthesis may have important consequences in these processes. Our analysis showed that the phytoene synthase (PSY), the first gene in the pathway, presented some amino acid changes considering the eight taxa studied, but not within individual taxon. Moreover, data indicates that it has undergone positive selection. The lycopene β-cyclase (LCYB) is an important enzyme for the conversion of lycopene into β-carotenoids (Fanciullino *et al.*, 2006b; Alquézar *et al.*, 2009). In our study, some amino acid changes were found different between *C. maxima* and *C. reticulata* that might be associated with the limitation in the conversion of lycopene in *C. maxima* (Fanciullino *et al.*, 2007). The β-carotene hydroxylase (HYB) is a very important enzyme involved in the catalyzation of β-carotene into β-cryptoxanthin and zeaxanthin (Fanciullino *et al.*, 2006b). We found strong differences between the three main citrus ancestors. *Citrus reticulata* continue the pathway and accumulate the products, however, *C. maxima* stops at this level and *C. medica* only convert β-carotene into β-cryptoxanthin.

Flavonoids are nother important compounds in fruit quality, which can give color to the leaves and flowers, are involved in the auxin transport, attract pollinators and are also antioxidant (Kaur and Kapoor, 2001; Winkel-Shirley, 2001). The enzyme chalcone isomerase (CHI) controls the second step of the flavonoid biosynthesis. In our work it appeared to be under positive selection at the inter-specific level due to differences between taxa. A second gene, the flavonoid 3'-hydroxylase (F3'H), showed positive selection only in *C. reticulata*. It has been shown to be important in flavonoid biosynthesis in *Arabidopsis* (Schoenbohm *et al.*, 2000) and grapevine (Bogs *et al.*, 2006). Therefore, understanding F3'H and CHI regulation and allelic functionality could be important for the analysis of molecular determinants of flavonoid composition in citrus fruits.

Within the genes studied in the acid biosynthesis, only the malic enzyme (EMA), which is involved in the last steps of the citric acid cycle, showed positive selection at interspecific

level, therefore it could be related with the different acid content existing between *Citrus* taxa (Penniston *et al.*, 2008).

Another important characteristic in the citrus fruit quality is the sugar content, which increases along the maturation process (Albertini *et al.*, 2006), but none of the genes studied presented positive selection and they are highly conserved.

For the genes related to plant stress response, the NADH kinase (NADK2) displayed a non-synonymous/synonymous ratio greater than 1. This enzyme plays an important role in the phosphorylation of NAD (H) and have been shown to change the sensitivity to abiotic stress in *A. thaliana* (Chai *et al.*, 2005).

Future prospects

This work has released new information about the genetic relationships of taxa in the *Citrus* genus and relative species that will help in the breeding of new, high-quality citrus cultivars and the conservation of the existing material.

Sanger sequencing of nuclear genes has provided information on the mosaic structure of secondary species and recent hybrids (Garcia-Lor *et al*., 2013a). Parallel sequencing of individual DNA molecules (454 Roche pyrosequencing) will allow to define multilocus haplotypes of heterozygous genotypes and to perform a deeper phylogenetic assignement of DNA fragments of the main cultivated species (Curk *et al*., 2012).

For all of the genes discussed in this report displaying amino acid variability of corresponding proteins (probably subjected to selection), it would be interesting to complete their full sequence (including promoter sequencing) and to perform allelic functional studies to decipher the molecular basis of the phenotypic variability in the species that were examined. It would also be interesting to obtain polymorphism information of partial or full sequence of the genes, missing in this work, from the biosynthesis pathway analysed.

Another contribution of this work has been the proper characterization of the two citrus germplasm collections analysed with molecular markers which will help in their management, the determination of which accessions must be preserved or removed in order not to lose diversity, and also which should be introduced to cover lack of diversity. A database including all the data generated in this work is being implemented and will help citrus breeders and geneticists in their research.

Gene banks are founded with the aim to conserve the genetic diversity of crop species, but large germplasm collections lead to management problems (space, maintenance cost, etc.) (van Hintum *et al.*, 2000). In this context, the concept of core collections was proposed to reduce the size of large germplasm collections (10-15% of the initial collection) and keep the maximum variability (at least the 80%), leading to a better use of the genetic resources present in the germplasms (Frankel and Brown, 1984; Pessoa-Filho *et al.*, 2010). In the near future, it will be afford for the first time in *Citrus* the establishment of a core collection, specifically in *C. reticulata*, which is the second most important group in the fresh fruit market worldwide.

CONCLUSIONS

According to the results obtained in this PhD thesis the following general conclusions can be established:

1) The development of nuclear indel markers, for the first time in citrus, has allowed us to demonstrate its usefulness for diversity and phylogenetic studies in the genus *Citrus*. They can become an important source of genetic markers with easy and inexpensive genotyping.
2) The comparison between indel, SNP and SSR markers shows their application as complementary molecular markers. Indel and SNP markers appear to be better phylogenetic markers for tracing the contributions of the ancestral species to the secondary species and modern cultivars and the SSR markers are more useful for intraspecific diversity analysis.
3) The contribution of each basic taxa (*C. reticulata*, *C. maxima*, *C. medica* and *C. micrantha*) to the genomes of secondary species and modern cultivars has been quantified, and the their origins are in agreement with those previously proposed in the case of sour orange, grapefruit, lemon and lime.
4) Regarding the sweet orange, it seems to have a different origin to what was previously proposed. The first study with indel and SSR markers suggested that *C. sinensis* could not be a direct cross between a mandarin (*C. reticulata*) and a pummelo (*C. maxima*), as it was previously believed. It could be the result of a backcross 1 (BC1) [(*C. maxima x C. reticulata*) x *C. reticulata*]. The study with SNP markers lead to the conclusion that the two parents of sweet orange were of interspecific origin due to the presence of nuclear genomic fragments in phylogenetic homozygosity inherited from *C. maxima* and *C. reticulata*.
5) No intercultivar polymorphisms had been previously found at intraspecific level for *C. sinensis*, *C. aurantium* and *C. paradisi*. Even in the attempt to find polymorphisms by gene seguencing (Garcia-Lor *et al.*, 2013a) and SNP genotyping (Garcia-Lor *et al.*, 2013b) in several secondary species genotypes, no polymorphisms were detected which confirms that most of the inter-varietal polymorphisms within these secondary species arise from punctual mutation or movement of transposable elements.
6) The initial differentiation between the basic species and a limited number of interspecific meiosis generated a genetic organisation of the citrus gene pool with high and generalised linkage disequilibrium. This structure does not allow association genetic studies at the genus level without the development of additional recombinant populations from interspecific hybrids. However, it could be possible to perform association genetic studies at intraspecific level in a less structured pool such as *C. reticulata*, where most of the breeding programs are been developed in several countries.

7) The core subset of markers identified can differentiate between accessions and study their origin. It could be useful for quick and inexpensive genotyping at interspecific and intraspecific level in existing or new accessions in germplasm banks.
8) Some of the SNP loci mined in this study have been converted into efficient markers to perform high throughput genotyping studies in Illumina GoldenGate Array and have been used for diversity analysis and genetic mapping. They will be useful for the management of citrus germplasm collections and marker/trait association studies.
9) Generally neutral evolution has been observed in the 27 genes studied (carotenoid, flavonoid, acid biosynthesis pathways and some related to plant response to stresses), but for a few genes [phytoene synthase (PSY), lycopene β-cyclase (LCYB), β-carotene hydroxylase (HYB), chalcone isomerase (CHI), flavonoid 3'-hydroxylase (F3'H), malic enzyme (EMA), NADH kinase 2 (NADK2)] positive selection was observed within or between the species studied, suggesting that these genes may play a key role in phenotypic differentiation. These seven genes are therefore major candidates for future studies, including complete gene sequencing and functional analysis of different alleles to analyse the molecular basis of the phenotypic differentiation of corresponding traits.
10) The nuclear phylogeny of *Citrus* and its sexually compatible relatives showed coherence with the geographic distribution and differentiation proposed by the 'Tanaka line' (Tanaka, 1954). *Citrus reticulata* and *Fortunella* share the same area of diversification, where the subgenus Metacitrus predominates (East-Asiatic floral zone), and appeared to be closely related. The cluster that joins *C. medica* and *C. maxima* is in agreement with the area of distribution where the subgenus Archicitrus predominates (Indo-Malayan floral zone).
11) The present study already allowed us to assign a phylogenetic inheritance of the genes that were examined for most of the genotypes of interspecific origin under study. With the next release of the pseudo chromosome sequence assembly of the reference haploid clementine genome (Gmitter *et al*., 2012), the assignation of the phylogenetic origin of these 27 genes will contribute to the deciphering of the interspecific mosaic genome structure of the secondary species.
12) Sanger sequencing it is still an important resource for many kinds of studies and it has a high level of accuracy as it has been shown in the works of Ollitrault *et al*. (2012a), where 99.2% of the SNPs in common were in agreement with the GoldenGate genotyping, and Garcia-Lor *et al*. (2013b), where the level of conformity with KASpar genotyping was 95.41%, while 2.99% did not agree and 1.60% was missing data.

13) SNP genotyping based on Competitive Allele-Specific PCR (KASpar) appears to be an interesting approach for low-to-medium throughput genotyping. The SNP markers developed from sequence data of a limited intra-generic discovery panel (three ancestral species, *C. medica*, *C. reticulata*, and *C. maxima*), provide a valuable molecular resource for genetic diversity analysis of germplasm within a genus and should be useful for germplasm fingerprinting at a much broader diversity level.
14) The transferability of these SNP markers to the genera of the subfamily Aurantioideae was not complete. The frequency of missing data was higher for the citrus relatives than within the *Citrus* genus and increased with taxonomic distances within the Aurantioideae subfamily.
15) The genotypes from the germplasm collections of the *Instituto Valenciano de Investigaciones Agrarias* (IVIA) and the *Station de Recherches Agronomiques* (CIRAD-INRA) have been well characterized through different kinds of molecular markers. This study has detected some redundancies and has improved the management of the citrus collections existing in their orchards. This data has updated the databases existing in both research centers.
16) Regarding the mandarin horticultural varietal group, it has been shown that it is a highly polymorphic group and that many genotypes, believed to be pure mandarins, have shown introgression in their genomes from *C. maxima*, *C. medica*, *Papeda* and *Fortunella*, even though some of them presented non-mandarin maternal origin.
17) A new organization of the mandarin germplasm has been defined in this study, showing that many genotypes have originated from the cross between mandarins, besides the genotypes that presented other ancestral genome contributions.
18) Although new insights in the mandarin germplasm structure have been released in this work, there is still a lot of work to do to clarify more precisely their phylogeny. Future sequencing of mandarin genotypes (single genes or whole genomes) will help to perform phylogenetic analysis and decipher the exact genome constitution of this highly polymorphic group.
19) In the near future, by using the entire citrus genome as a reference and resequencing data from the main secondary species, the resulting estimations of the relative levels of within and between taxa differentiation will be useful for deciphering the interspecific mosaic structure of the citrus secondary cultivated species and modern cultivars.

LITERATURE CITED

Abdurakhmonov IY, Abdukarimov A. 2008. Application of Association Mapping to Understanding the Genetic Diversity of Plant Germplasm Resources. *International Journal of Plant Genomics* ID 574927, doi:10.1155/2008/574927.

Abkenar AA, Isshiki S, Tashiro Y. 2004. Phylogenetic relationships in the "true citrus fruit trees" revealed by PCR-RFLP analysis of cpDNA. *Scientia Horticulturae* **102**: 233-242.

Agrama HA, Yan WG. 2009. Association mapping of straighthead disorder induced by arsenic in *Oryza sativa*. *Plant Breeding* **128**: 551-558.

Ahmad R, Struss D, Southwick S. 2003. Development and characterization of microsatellite in citrus. *Journal American. Society Horticultural Science* **128**: 584-590.

Albertini MV, Carcouet E, Pailly O, Gambotti C, Luro F, Berti L. 2006. Changes in organic acids and sugars during early stages of development of acidic and acidless citrus fruit. *Journal of Agricultural and Food Chemistry* **54**: 8335-8339.

Albrechtsen A, Nielsen FC, Nielsen R. 2010. Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular biology and evolution* **27**: 2534-2547.

Aleza P, Juarez J, Hernandez M, Pina JA, Ollitrault P, Navarro L. 2009. Recovery and characterization of a Citrus clementina Hort. ex Tan. 'Clemenules' haploid plant selected to establish the reference whole Citrus genome sequence. *BMC plant biology* **9**: 110.

Aleza P, Cuenca J, Juárez J, Pina JA, Navarro L. 2010. 'Garbí' Mandarin: a new late-maturing triploid hybrid. *HortScience* **45**: 139-141.

Aleza P, Froelicher Y, Schwarz S, Hernández M, Juárez J, Morillon R, Navarro L, Ollitrault P. 2011. Tetraploidization events by chromosome doubling of nucellar cells are frequent in apomictic citrus and are dependent on genotype and environment. *Annals of Botany* **108**: 37-50.

Alquézar B, Zacarias L, Rodrigo MJ. 2009. Molecular and functional characterization of a novel chromoplast-specific lycopene beta-cyclase from Citrus and its relation to lycopene accumulation. *Journal of experimental botany* **60**: 1783-1797.

Amar MH, Biswas MK, Zhang Z, Guo W. 2011. Exploitation of SSR, SRAP and CAPS-SNP markers for genetic diversity of Citrus germplasm collection. *Scientia Horticulturae* **128**: 220-227.

de Araújo EF, de Queiroz LP, Machado MA. 2003. What is Citrus? Taxonomic implications from a study of cp-DNA evolution in the tribe Citreae (Rutaceae subfamily Aurantioideae). *Organisms Diversity & Evolution* **3**: 55-62.

Bachtrog D, Agis M, Imhof M, Schlötterer C. 2000. Microsatellite Variability Differs Between Dinucleotide Repeat Motifs—Evidence from Drosophila melanogaster. *Molecular biology and evolution* **17**: 1277-1285.

Bapteste EP, Hervé. 2002. The Potential Value of Indels as Phylogenetic Markers: Position of Trichomonads as a Case Study. *Molecular Biology and Evolution* **19**: 972-977

Barkley NA, Roose ML, Krueger RR, Federici CT. 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *TAG Theoretical and Applied Genetics* **112**: 1519-1531.

Barkley NA, Krueger RR, Federici CT, Roose ML. 2009. What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. *Plant Systematics and Evolution* **282**: 71-86.

Barrett HC and Rhodes AM. 1976. A numerical taxonomic study ofaffinity relationships in cultivated Citrus and its close relatives. *Systematic Botany* **1**: 105-136.

Bastianel M, Schwarz SF, Coleta Filho HD, Lin LL, Machado M, Koller OC. 1998. Identification of zygotic and nucellar tangerine seedlings (Citrus spp.) using RAPD. *Genetics and Molecular Biology* **21**.

Bauer F, Elbers CC, Adan RAH, Loos RJF. 2009. Obesity genes identified in genome-wide association studies are associated with adiposity measures and potentially with nutrient-specific food preference. *American Journal of Clinical Nutrition* **90**: 951-959.

Bausher M, Shatters R, Chaparro J, Dang P, Hunter W, Niedz R. 2003. An expressed sequence tag (EST) set from Citrus sinensis L. Osbeck whole seedlings and the implications of further perennial source investigations. *Plant Science* **165**: 415-422.

Bayer RJ, Mabberley DJ, Morton C, et al. 2009. A molecular phylogeny of the orange subfamily(Rutaceae: Aurantioideae) using nine cpDNA sequences. *American Journal of Botany* **96**: 668-685.

Behrouz Golein, Talaie A, Zamani Z, Ebadi A, Behjatnia A. 2005. Assessment of genetic variability in some Iranian sweet oranges (*Citrus sinensis* [L.] Osbeck) and mandarins (*Citrus reticulata* Blanco) using SSR markers. *International Journal of Agriculture and Biology* **7**: 167-170.

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F. 2002. *GENETIX v. 4.03, Logiciel sous Windows pour la génétique des populations*. Laboratoire Génome et Population, Université de Montpellier 2, Montpellier, France. http://www.univ-montp2.fr.

Berhow MA, Hasegawa S, Kwan K, Bennett RD. 2000. Limonoids and the chemotaxonomy of Citrus and the Rutaceae family. In: Berhow, M.A., Hasegawa, S. and Manners, G.D., ed. *Citrus Limonoids: Functional Chemicals in Agriculture and Food.* American Chemical Society, Washington DC, 212-228.

Bernet GP, Gorris MT, Carbonell EA, Cambra M, Asins MJ. 2008. Citrus tristeza virus resistance in a core collection of sour orange based on a diversity study of three germplasm collections using QTL-linked markers. *Plant Breeding* **127**: 398-406

Berrin J, Pierrugues O, Brutesco C, et al. 2005. Stress induces the expression of AtNADK-1, a gene encoding a NAD(H) kinase in Arabidopsis thaliana. *Molecular Genetics and Genomics* **273**: 10-19.

Biswas M, Xu Q, Deng X. 2010. Utility of RAPD, ISSR, IRAP and REMAP markers for the genetic analysis of *Citrus* spp. *Sci. Horticult.* **124**: 254-261.

Bogs J, Ebadi A, McDavid D, Robinson SP. 2006. Identification of the Flavonoid Hydroxylases from Grapevine and Their Regulation during Fruit Development. *Plant Physiology* **140**: 279-291.

Botstein D, Risch N. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics* **33**: 228-237.

Bouvier F, Backhaus RA, Camara B. 1998. Induction and control of chromoplast-specific carotenoid genes by oxidative stress. *Journal of Biological Chemistry* **273**: 30651-30659.

Bové JM. 2006. Huanglongbing: A destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology* **88**: 7-37.

Bramley PM. 2002. Regulation of carotenoid formation during tomato fruit ripening and development. **53**: 2107-2113.

Breseghello F, Sorrells M. 2006. Association Analysis as a Strategy for Improvement of Quantitative Traits in Plants *Crop. Sci.* **46**.

Bretó MP, Ruiz C, Pina JA, Asins MJ. 2001. The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. *Molecular phylogenetics and evolution* **21**: 285-293.

Britten RJ, Rowen L, Williams J, Cameron RA. 2003. Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4661-4665.

Brookes AJ. 1999. The essence of SNPs. *Gene* **234**: 177-186.

Brown AHD. 1989. Core collections: a practical approach to genetic resources management. *Genome, 31, 818–824* .

Buckler E, Thornsberry J. 2002. Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* **5**: 107-111.

Cacciola S, Lio GMS. 2008. Management Of Citrus Diseases Caused By Phytophthora Spp. In: A Ciancio, KG Mukerji, eds. Springer Netherlands, 61-84.

Cai Q, Guy CL, Moore GA. 1994. Extension of the linkage map in Citrus using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. *Theoretical and Applied Genetics* **89**: 606-614.

Calabrese F. 1992. The history of citrus in the Mediterranean countries and Europe. *International Society of Citriculture* **1**: 35-38.

Calabrese F. 1998. La favolosa storia degli agrumi. *L'EPOS Società Editrice.* Palermo, Italia: 172.

Cameroon J, Frost H. 1968. Genetic, breeding and nucellar embryony. In: Reuther, W.; Batchelor, L.D.; Webber, H.J. (eds) *The Citrus Industry. V2.* University of California, Berkeley, U.S.A.

Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Francks CD, Kresovich S. 2008. Community resources and strategies for association mapping in sorghum. *Crop Science* **48**: 30-40.

Chai MF, Chen QJ, An R, Chen YM, Chen J, Wang XC. 2005. NADK2, an Arabidopsis chloroplastic NAD kinase, plays a vital role in both chlorophyll synthesis and chloroplast protection. *Plant Molecular Biology* **59**: 553-564.

Chai MF, Wei PC, Chen QJ, et al. 2006. NADK3, a novel cytoplasmic source of NADPH, is required under conditions of oxidative stress and modulates abscisic acid responses in Arabidopsis. *The Plant Journal* **47**: 665-674.

Chen CX, Bowman KD, Choi YA, Dang PM, Rao MN, Huang S, Soneji JR, McCollum TG, Gmitter FG Jr. 2008. EST-SSR genetic maps for *Citrus sinensis* and *Poncirus trifoliata*. *Tree Genetics and Genomes* **4**: 1-10.

Cheng Y, De Vicente MC, Meng H, Guo W, Tao N, Deng X. 2005. A set of primers for analyzing chloroplast DNA diversity in Citrus and related genera. *Tree physiology* **25**: 661-672.

Chin H, Roberts E. 1980. Recalcitrant Crop Seeds. *Tropical Press SDN. BHD,* Kuala Lumpur.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* **3**: 19.

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome research* **15**: 1496-1502.

Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD, Lewis LC. 2009. Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity* **100**: 556-564.

Coletta Filho HD, Machado MA, Targon MLPN, Moreira MCPQDG, Pompeu J Jr. 1998. Analysis of the genetic diversity among mandarins (*Citrus* spp.) using RAPD markers. *Euphytica* **102**: 133-139.

Copeland L, Turner JF. 1987. The regulation of glycolysis and the pentose phosphate pathway. In A Marcus, ed, *The Biochemistry of Plants: A Comprehensive Treatise*, Davies DD. Academic Press, New York **11**: 107-125.

Corazza-Nunes M, Machado MA, Nunes WMC, Cristofani M, Targon MLPN. 2002. Assessment of genetic variability in grapefruits (*Citrus paradisi* Macf.) and pummelos (*C. maxima* (Burm.) Merr.) using RAPD and SSR markers. *Euphytica* **126**: 169-176.

Cortés A, Chavarro M, Blair M. 2011. SNP marker diversity in common bean (Phaseolus vulgaris L.). *Theoretical and Applied Genetics* **123**: 827-845.

Cuenca J, Aleza P, Juarez J, Pina JA, Navarro L. 2010. 'Safor' mandarin: a new citrus mid-late triploid hybrid. *HortScience* **45**: 977-980.

Cuenca J, Froelicher Y, Aleza P, Juarez J, Navarro L, Ollitrault P. 2011. Multilocus half-tetrad analysis and centromere mapping in citrus: evidence of SDR mechanism for 2n megagametophyte production and partial chiasma interference in mandarin cv 'Fortune'. *Heredity* **107**: 462-470.

Cuenca J, Aleza P, Iborra E, Vicent A, Ollitrault P, Navarro L. 2012. Location of a chromosome region linked to Alternaria Brown Spot resistance from the evaluation of triploid mandarin populations. *XII International Citrus Congress- Valencia, Spain. S02O14*: 42.

Cunningham FX, J., Pogson B, ZaiRen S, McDonald KA, DellaPenna D, Gantt E. 1996. Functional analysis of the β and ε lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell* **8**: 1613-1626.

Cuppen E. 2007. Genotyping by Allele-Specific Amplification (KASPar). *Cold Spring Harbor Protocols* **9**: 172-173.

Curk F, Ancillo G, Garcia-Lor A, Luro F, Navarro L, Ollitrault P. 2012. Multilocus haplotyping by parallel sequencing to decipher the interspecific mosaic genome structure of cultivated citrus. *XII International Citrus Congress- Valencia, Spain. S01O06*: 28.

Das A. 2003. Citrus canker: a review. *Journal of Applied Horticulture* **5**: 52-60.

Davies FS, Albrigo LG. 1994. Cítricos. Ed Acribia, Zaragoza, España.

Del Caro A, Piga A, Vacca V, Agabbio M. 2004. Changes of flavonoids, vitamin C and antioxidant capacity in minimally processed citrus segments and juices during storage. *Food Chemistry* **84**: 99-105.

Delseny M, Han B, Hsing Yl. 2010. High throughput DNA sequencing: the new sequencing revolution. *Plant Science* **179**: 407-422.

Demmig-Adams B, Gilmore A, Adams W. 1996. In vivo functions of carotenoids in higher plants. *FASEB Journal* **10**: 403-412.

Deng Z, Gentile A, Nicolosi E, Continella G, Tribulato E. 1996. Parentage determination of some citrus hybrids by molecular markers. *Proc Int Soc Citricul* **2**: 849-854.

Dereeper A, Guignon V, Blanc G, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research* **36**: W465-W469.

Deu M, Glaszmann JC. 2004. Linkage disequilibrium in sorghum. In Proceedings of the Plant and Animal Genome XII Conference, San Diego, 10– 14 January, Town & Country Convention Center, San Diego, CA, W10

Dhuique-Mayer C, Caris-Veyrat C, Ollitrault P, Curk F, Amiot MJ. 2005. Varietal and interspecific influence on micronutrient contents in citrus from the Mediterranean area. *Journal of Agricultural and Food Chemistry* **53**: 2140-2145.

Ding S, Zhang X, Bao Z, Ling M. 1984. A new species of Poncirus from China. *Acta Botanica Yunnanica* **6**: 292-293.

Dong J, Qing-liang Y, Fu-sheng W, Li C. 2010. The mining of citrus EST-SNP and its application in cultivar discrimination. *Agricultural Sciences in China* **9**: 179-190.

Duran-Vila N. 1995. Cryopreservation of germplasm of citrus. In: Anonymous *Biotechnology in agriculture and forestry, cryopreservation of plant germplasm I.* Springer, Berlin, 70-96.

Duran-Vila N, Pina J, Ballester J, Juarez, J, Roistacher CN, Rivera-Bustamante R, Semancik, JS. 1988. The citrus exocortis disease: a complex of viroid RNAs. *Proceedings of the International Organization of Citrus Virologists* **10**: 152-164.

Ellwood SR, D'Souza NK, Kamphuis LG, Burgess TI, Nair RM, Oliver RP. 2006. SSR analysis of the Medicago truncatula SARDI core collection reveals substantial diversity and unusual genotype dispersal throughout the Mediterranean basin. *TAG Theoretical and Applied Genetics* **112**: 977-983.

Engelmann F. 1997. *In vitro* conservation methods. In: Ford-Lloyd, B.V., Newburry, J.H. and Callow, J.A. (eds), ed. *Biotechnology and Plant Genetic Resources: Conservation and Use.* CABI: Wallingford, UK, 119.

Erickson L. 1968. The general physiology of citrus. In: W Reuther, LD Batchelor, H Webber, eds, *The Citrus Industry, Vol II.* University of California Press, Berkley.

Erlund I. 2004. Review of the flavonoids quercetin, hesperetin, and naringenin. Dietary sources, bioactivities, bioavailability, and epidemiology. *Nutrition Research* **24**: 851-874.

Escribano P, Viruel MA, Hormaza JI. 2008. Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (Annona cherimola, Annonaceae), an underutilised subtropical fruit tree species. *Annals of Applied Biology* **153**: 25-32.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular ecology* **14**: 2611-2620.

Fajardo DS, Bonte DR, Jarret RL. 2002. Identifying and selecting for genetic diversity in Papua New Guinea sweetpotato Ipomoea batatas (L.) Lam. germplasm collected as botanical seed. *Genetic Resources and Crop Evolution* **49**: 463-470.

Falush D, Stephens M, Pritchard JK. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **164**: 1567-1587.

Fanciullino AL, Dhuique-Mayer C, Luro F, Casanova J, Morillon R, Ollitrault P. 2006a. Carotenoid diversity in cultivated citrus is highly influenced by genetic factors. *Journal of Agricultural and Food Chemistry* **54**: 4397-4406.

Fanciullino AL, Tomi F, Luro F, Desjobert JM, Casanova J. 2006b. Chemical variability of peel and leaf oils of mandarins. *Flavour and Fragrance Journal* **21**: 359-367.

Fanciullino AL, Dhuique-Mayer C, Luro F, Morillon R, Ollitrault P. 2007. Carotenoid biosynthetic pathway in the *Citrus* genus: number of copies and phylogenetic diversity of seven genes. *Journal of Agricultural and Food Chemistry* **55**: 7405-7417.

Fanciullino AL, Cercos M, Dhique-Mayer, Froelicher Y, Talón M, Ollitrault P, Morillon R. 2008. Changes in carotenoid content and biosynthetic gene expression in juice sacs of four orange varieties (*Citrus sinensis*) differing in flesh fruit color. *Journal of Agricultural and Food Chemistry* **56**: 3628-3638.

Fang D, Roose ML, Krueger RR, Federici CT. 1997. Fingerprinting trifoliate orange germ plasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. *Theoretical and Applied Genetics* **95**: 211-219.

Fang D, Krueger RR, Roose ML. 1998. Phylogenetic relationships among selected *Citrus* germplasm accessions revealed by inter-simple sequence repeat (ISSR) markers. *Journal of the American Society for Horticultural Science* **123**: 612-617.

Fantz PR. 1988. Nomenclature of the Meiwa and Changshou kumquats, intrageneric hybrids of Fortunella. *HortScience* **23**: 249-250.

Food and Agriculture Organization of the United Nations, FAOSTAT database. Available at http://faostat.fao.org/

Federici CT, Fang DQ, Scora RW, Roose ML. 1998. Phylogenetic relationships within the genus *Citrus* (Rutaceae) and related genera as revealed by RFLP and RAPD analysis. *Theoretical and Applied Genetics* **96**: 812-822.

Fideghelli C, Sansavini S. 2002. The fruit industry and the role of research in the Mediterranean Basin. *Acta Horticulturae (ISHS)* **582**: 69-76.

Flint-Garcia SA, Thornsberry JM, Buckler ES. 2003. Structure of linkage disequilibrium in plants. *Annual Review Plant Biology* **54**: 357-374.

Forment J, Gadea J, Huerta L, et al. 2005. Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. *Plant Molecular Biology* **57**: 375-391.

Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet JM, Cheynier V, This P. 2009. Quantitative Genetic Bases of Anthocyanin Variation in Grape (*Vitis vinifera* L. ssp. sativa) Berry: A Quantitative Trait Locus to Quantitative Trait Nucleotide Integrated Study. *Genetics* **183**: 1127-1139.

Frankel O, Brown A. 1984. Current plant genetic resources - a critical appraisal. In: VL Chopra, B Joshi, R Sharma, H Bansal, eds. *In Genetics: New Frontiers.* Oxford and IBH Publishing: New Delhi, India., 1-13.

Froelicher Y, Dambier D, Bassene JB, Costantino G, Lotfy S, Didout C, Beaumont V, Brottier P, Risterucci AM, Luro F, Ollitrault P. 2008. Characterization of microsatellite markers in mandarin orange (*Citrus reticulata* Blanco). *Molecular ecology resources* **8**: 119-122.

Froelicher Y, Mouhaya W, Bassene JB, Costantino G, Kamiri M, Luro F, Morillon R, Ollitrault P. 2011. New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. *Tree Genetics and Genomes* **7**: 49-61.

Frydman A, Weisshaus O, Bar-Peled M, Huhman DV, Sumner LW, Marin FR, Lewinsohn E, Fluhr R, Gressel J, Eyal Y. 2004. Citrus fruit bitter flavors: isolation and functional characterization of the gene Cm1,2RhaT encoding a 1,2 rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of citrus. *The Plant Journal* **40**: 88-100.

Fu B, Chen M, Zou M, Long M, He S. 2010. The rapid generation of chimerical genes expanding protein diversity in zebrafish. *BMC Genomics* **11**: 657.

Furr J. 1964. New tangerines for the desert. *Calif. Citrog.* **49**: 266-276.

Garcia-Lor A, Luro F, Navarro L, Ollitrault P. 2012a. Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. *Molecular Genetics and Genomics* **287**: 77-94.

Garcia-Lor A, Garcia-Martinez J, Perez-Amador M. 2012b. Identification of ovule and seed genes from Citrus clementina. *Tree Genetics & Genomes* **8**: 227-235.

Garcia-Lor A, Curk F, Snoussi-Trifa H, Morillon R, Ancillo G, Luro F, Navarro L, Ollitrault P. 2013a. A nuclear phylogenetic analysis; SNPs, indels and SSRs deliver new insights into the relationships in the "true citrus fruit trees" group (Citrinae, Rutaceae) and the origin of cultivated species. *Annals of botany* **111**: 1-19.

Garcia-Lor A, Ancillo G, Navarro L, Ollitrault P. 2013b. *Citrus* (Rutaceae) SNP markers based on Competitive Allele-Specific PCR; transferability across the Aurantioideae subfamily. *Applications in Plant Sciences* **1**: doi:10.3732/apps.1200406.

Garris AJ, McCouch SR, Kresovich S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics* **165**: 759-769.

Garvin M, Saitoh K, Gharret A. 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**: 915-934.

Gattuso G, Barreca D, Gargiulli C, Leuzzi U, Caristi C. 2007. Flavonoid composition of Citrus juices. *Molecules* **12**: 1641-1673.

Gaydou EM, Bianchini JP, Randriamiharisoa RP. 1987. Orange and mandarin peel oils differentiation using polymethoxylated flavone composition. *Journal of Agricultural and Food Chemistry* **35**: 525-529.

Genebank Standards. 1994. Food and Agriculture Organization of the United Nations, Rome, International Plant Genetic Resources Institute, Rome.

Germanà C, Sardo V. 1988. Correlations among some physiological and climatic parameters in orange trees. *International Citrus Congress (6th: 1988 : Tel Aviv, Israel)*: Margraf **1**: 525-534.

Ghislain M, DaPeng Z, Fajardo D, Huamán Z, Hijmans RJ. 1999. Marker-assisted sampling of the cultivated Andean potato Solanum phureja collection using RAPD markers. *Genetic Resources and Crop Evolution* **46**: 547-555.

Gilmore AM, Yamamoto HY. 1993. Linear models relating xanthophylls and lumen acidity to non-photochemical fluorescence quenching. Evidence that antheraxanthin explains zeaxanthin-independent quenching. *Photosynthesis Research* **35**: 67-78.

Gmitter FG Jr., Chen CX, Rao MN, Soneji JR. 2007. Citrus fruits. In: Kole KR (Ed.) *Fruits and Nuts, Genome Mapping and Molecular Breeding in Plants*. Springer, Heidelberg **4**: 265-279.

Gmitter FG Jr., Ollitrault P, Machado M, Reforgiato-Recupero G, Talon M, Roose ML, Navarro L, Wu G, Jaillon O, Morgante M, Rokhsar DS. 2012. Genome sequence analysis and comparisons reveal ancestral hybridization and admixture events in the origins of some citrus cultivars. *XII International Citrus Congress- Valencia, Spain. S03O01*: 61.

Golein B, Talaie A, Zamani Z, Ebadi A, Behjatnia A. 2005. Assessment of genetic variability in some Iranian sweet oranges (*Citrus sinensis* [L.] Osbeck) and mandarins (*Citrus reticulata* Blanco) using SSR markers. *International Journal of Agriculture and Biology* **7**: 167-170.

Gómez-Cadenas A, Tadeo FR, Talon M, Primo-Millo E. 1996. Leaf abscission induced by ethylene in water-stressed intact seedlings of Cleopatra mandarin requires previous abscisic acid accumulation in roots. *Plant Physiology* **112**: 401-408.

González-Arnao M, Juárez J, Ortega C, Navarro L, Duran-Vila N. 2003. Cryopreservation of ovules and somatic embryos of citrus using the encapsulation-dehydration technique. *Cryo Letters* **24**: 85-94.

Goodwin TW. 1980. The biochemistry of the carotenoids. *2nd edn. Vol.1. London: Chapman and Hall*.

Grauke LJ, Thompson TE, Marquard RD. 1995. Evaluation of pecan (*Carya illinoinensis* (Wangenh.) K. Koch) germplasm collections and designation of a core subset. *HortScience* **30**: 950-954.

Green RM, Vardi A, Galun E. 1986. The plastome of Citrus. Physical map, variation among Citrus cultivars and species and comparison with related genera. *Theoretical and Applied Genetics* **72**: 170-177.

Grenier C, Bramel-Cox P, Noirot M, Rao KEP, Hamon P. 2000. Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures. A. Using morpho-agronomical and passport data. *Theoretical and Applied Genetics* **101**: 190-196.

Groppo M, Pirani J, Salatino M, Blanco S, Kallunki J. 2008. Phylogeny of Rutaceae based on two noncoding regions from cpDNA. *American Journal of Botany* **95**: 985-1005.

Gross M. 2004. Flavonoids and cardiovascular disease. *Pharmaceutical Biology* **42**: 21-35.

Gulsen O, Roose ML. 2001a. Chloroplast and nuclear genome analysis of the parentage of lemons. *Journal of the American Society for Horticultural Science* **126**: 210-215.

Gulsen O, Roose ML. 2001b. Determination of genetic diversity and phylogenetic relations to citrus ancestors in lemons by DNA markers. *Bahce* **30**: 53-63.

Gulsen O, Roose ML. 2001c. Lemons: diversity and relationships with selected *Citrus* genotypes as measured with nuclear genome markers. *Journal of the American Society for Horticultural Science* **126**: 309-317.

Gulsen O, Uzun A, Canan I, Seday U, Canihos E. 2010. A new citrus linkage map based on SRAP, SSR, ISSR, POGP, RGA and RAPD markers. *Euphytica* **173**: 265-278.

Gupta PK, Rustgi S, Kulwal PL. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology* **57**: 461-485.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.

Hao CY, Zhang XY, Wang LF, Dong YS, Shang XW, Jia JZ. 2006. Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Molecular Breeding* **17**: 69-77.

Hartl DL, Clark AG. 1997. *Principles of population genetics*. Sinauer Associates Incorporated: Sunderland, US.

Hayashi K, Yoshida H, Ashikawa I. 2006. Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *TAG Theoretical and Applied Genetics* **113**: 251-260.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, Nielsen EE. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular ecology resources* **11 Suppl 1**: 123-136.

Herrero R, Asíns MJ, Carbonell EA, Navarro L. 1996. Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. *Theoretical and Applied Genetics* **92**: 599-609.

Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N. 2006. Multilocus Patterns of Nucleotide Diversity, Linkage Disequilibrium and Demographic History of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* **174**: 2095-2105.

Hirschberg J. 2001. Carotenoid biosynthesis in flowering plants. *Current opinion in plant biology* **4**: 210-218.

Hodgson RW. 1967. Horticultural varieties of *Citrus*. In: Reuther W, Webber HJ, Batchelor LD. eds. *The Citrus industry.* Berkeley, CA: University of California Press, 431-589.

Hu J, Zhu J, Xu H. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theoretical and Applied Genetics* **101**: 264-268.

Hyoungshin P, Kreunen SS, Cuttriss AJ, DellaPenna D, Pogson BJ. 2002. Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. *Plant Cell* **14**: 321-332.

HyunPyo K, KunHo S, HyeunWook C, SamSik K. 2004. Anti-inflammatory plant flavonoids and cellular action mechanisms. *Journal of Pharmacological Sciences* **96**: 229-245.

Ingvarsson PK. 2005. Nucleotide Polymorphism and Linkage Disequilibrium Within and Among Natural Populations of European Aspen (*Populus tremula* L., Salicaceae). *Genetics* **169**: 945-953.

IPGRI. 1999. Descriptors for Citrus. International Plant Genetic Resources Institute, Rome, Italy.

Isaacson T, Ronen G, Zamir D, Hirschberg J. 2002. Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants. *Plant Cell* **14**: 333-342.

Jaeger B, Goldbach H, Sommer K. 2000. Release from lime induced iron chlorosis by CULTAN in fruit trees and its characterisation by analysis. *Acta Horticulturae* **531**: 107-113.

Jarne P, Lagoda PJL. 1996. Microsatellites, from molecules to populations and back *Trends in ecology & evolution* **11**: 424-429.

Johnson RC, Hodgkin T. 1999. *Core collections for today and tomorrow*. International Plant Genetic Resources Institute (IPGRI): Rome, Italy.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**: 94.

Jung Y, Kwon H, Kang S, Kang J, Kim S. 2005. Investigation of the phylogenetic relationships within the genus *Citrus* (Rutaceae) and related species in Korea using plastid trnL-trnF sequences. *Scientia Horticulturae* **104**: 179-188.

Kacar Y, Uzun A, Polat I, Yesiloglu T, Yilmaz B, Gulsen O, Tuzcu O, Kamiloglu M, Kurt S, Seday U. 2013. Molecular characterization and genetic diversity analysis of mandarin genotypes by SSR and SRAP markers. *Journal of Food Agriculture & Environment* **11**: 516-521.

Kamiri M, Stift M, Srairi I, Costantino G, Moussadik AE, Hmyene A, Bakry F, Ollitrault P, Froelicher Y. 2011. Evidence for non-disomic inheritance in a Citrus interspecific tetraploid somatic hybrid between C. reticulata and C. limon using SSR markers and cytogenetic analysis. *Plant Cell Reports* **30**: 1415-1425.

Kato M, Ikoma Y, Matsumoto H, Sugiura M, Hyodo H, Yano M. 2004. Accumulation of Carotenoids and Expression of Carotenoid Biosynthetic Genes during Maturation in Citrus Fruit. *Plant Physiology* **134**: 824-837.

Kato M, Matsumoto H, Ikoma Y, Okuda H, Yano M. 2006. The role of carotenoid cleavage dioxygenases in the regulation of carotenoid profiles during maturation in citrus fruit. *Journal of experimental botany* **57**: 2153-2164.

Kaur C, Kapoor HC. 2001. Antioxidants in fruits and vegetables - the millennium's health. *International Journal of Food Science & Technology* **36**: 703-725.

Kawaii S, Tomono Y, Katase E, Ogawa K, Yano M. 1999. HL-60 differentiating activity and flavonoid content of the readily extractable fraction prepared from citrus juices. *J Agric Food Chem.* **47**: 128-135.

Kay J, Weitzman P. 1987. *Krebs' citric acid cycle: half a century and still turning*. London: Biochemical Society.

Kijas JMH, Fowler JCS, Thomas MR. 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within Citrus and related species. *Genome* **38**: 349-355.

Kijas JMH, Thomas MR, Fowler JCS, Roose ML. 1997. Integration of trinucleotide microsatellites into a linkage map of *Citrus*. *Theoretical and Applied Genetics* **94**: 701-706.

Kim I, Ko K, Kim C, Chung W. 2001. Isolation and characterization of cDNAs encoding β-carotene hydroxylase in Citrus. *Plant Science* **161**: 1005-1010.

Koca U, Berhow MA, Febres VJ, Champ KI, Carrillo-Mendoza O, Moore GA. 2009. Decreasing unpalatable flavonoid components in Citrus: the effect of transformation construct. *Physiologia Plantarum* **137**: 101-114.

Koehler-Santos P, Dornelles ALC, Freitas LB. 2003. Characterization of mandarin citrus germplasm from Southern Brazil by morphological and molecular analyses. *Pesquisa Agropecuária Brasileira* **38**: 797-806.

Kolkman J, Berry S, Leon A, et al. 2007. Single Nucleotide Polymorphisms and Linkage Disequilibrium in Sunflower. *Genetics* **177**: 457-468.

Kotzé J. 1981. Epidemiology and control of citrus black spot in South Africa. *Plant Disease* **65**: 945-950.

Krueger RR, Navarro L. 2007. Citrus germplasm resources. In: Khan IA, ed. *Citrus Genetics, Breeding and Biotechnology.* Wallingford, UK: CAB International, 45-140.

Krueger RR, Roose ML. 2003. Use of molecular markers in the management of citrus germplasm resources. *Journal of the American Society for Horticultural Science* **128**: 827-837.

Kubo T, Hohjo I, Hiratsuka S. 2001. Sucrose accumulation and its related enzyme activities in the juice sacs of satsuma mandarin fruit from trees with different crop loads. *Scientia Horticulturae* **91**: 215-225.

Külheim C, SuatHui Y, Maintz J, Foley WJ, Moran GF. 2009. Comparative SNP diversity among four Eucalyptus species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* **10**: 452.

Lee HS. 2002. Characterization of major anthocyanins and the color of red-fleshed budd blood orange (*Citrus sinensis*). *Journal of Agricultural and Food Chemistry* **50**: 1243-1246.

Lee RF. 2004. Certification Programs for Citrus. In: Naqvi SAMH (ed.) *Diseases of Fruits and Vegetables*, Springer Netherlands **1**: 291-305.

Li W, Liu G, He S. 1992. Leaf isozymes of mandarin. *Proc Int Soc Citricult* **1**: 217-220.

Li Y, Haseneyer G, Schon C, Ankerst D, Korzun V, Wilde P, Bauer E. 2011. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biology* **11**: 6.

Liang G, Xiong G, Guo Q, He Q, Li X. 2007. AFLP analysis and the taxonomy of Citrus. *Acta Horticulturae* **760**: 137-142.

Lijavetzky D, Cabezas J, Ibáñez A, Rodríguez V, Martínez-Zapater J. 2007. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* **8**: 424.

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

Liu YZ, Deng XX. 2007. Citrus breeding and genetics in China. *The Asian and Australasian Journal of Plant Science and Biotechnology* **1**: 23-28.

Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, Stamatakis AP, Linder CR. 2012. SATé-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. *Systematic Biology* **61**: 90-106.

Lo Piero AR, Puglisi I, Rapisarda P, Petrone G. 2005. Anthocyanins Accumulation and Related Gene Expression in Red Orange Fruit Induced by Low Temperature Storage. *Journal of Agricultural and Food Chemistry* **53**: 9083-9088.

Lota ML, Serra DR, Tomi F, Casanova J. 2000. Chemical variability of peel and leaf essential oils of mandarins from Citrus reticulata Blanco. *Biochemical systematics and ecology* **28**: 61-78.

Lotfy S, Luro F, Carreel F, Froelicher Y, Rist D, Ollitrault P. 2003. Application of cleaved amplified polymorphic sequence method for analysis of cytoplasmic genome among *Aurantioideae* intergeneric somatic hybrids. *Journal of the American Society for Horticultural Science* **128**: 225-230.

Luro F, Laigret F, Bove JM, Ollitrault P. 1995. DNA amplified fingerprinting, a useful tool for determination of genetic origin and diversity analysis in *Citrus*. *HortScience* **30**: 1063-1067.

Luro F, Rist D, Ollitrault P. 2001. Evaluation of genetic relationships in *Citrus* genus by means of sequence tagged microsatellites. *Acta Horticulturae* **546**: 237-242.

Luro F, Maddy F, Jacquemond C, Froelicher Y, Morillon R, Rist D, Ollitrault P. 2004. Identification and evaluation of diplogyny in clementine (*Citrus clementina*) for use in breeding. *Acta Horticulturae* **663**: 841-847.

Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R. 2008. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC genomics* **9**: 287.

Luro F, Gatto J, Costantino G, Pailly O. 2011. Analysis of genetic diversity in Citrus. *Plant Genetic Resources* **9**: 218.

Mabberley DJ. 1997. A classification for edible *Citrus* (Rutaceae). *Telopea* **7**: 167-172.

Mabberley DJ. 2004. *Citrus* (Rutaceae): a review of recent advances in etymology, systematics and medical applications. *Blumea* **49**: 481-498.

Malosetti M, Abadie T. 2001. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. *Genetic Resources and Crop Evolution* **48**: 381-390.

Marita JM, Rodriguez JM, Nienhuis J. 2000. Development of an algorithm identifying maximally diverse core collections. *Genetic Resources and Crop Evolution* **47**: 515-526.

Martinez-Godoy M, Mauri N, Juarez J, Marques MC, Santiago J, Forment J, Gadea J. 2008. A genome-wide 20 K citrus microarray for gene expression analysis. *BMC Genomics* **9**: 318.

McIntosh CA, Latchinian L, Mansell RL. 1990. Flavanone-specific 7-O-glucosyltransferase activity in Citrus paradisi seedlings: purification and characterization. *Archives of Biochemistry and Biophysics* **282**: 50-57.

McKhann HI, Camilleri C, Bérard A, Bataillon T, David JL, Reboud X, Le Corre V, Caloustian C, Gut IG, Brunel D. 2004. Nested core collections maximizing genetic diversity in Arabidopsis thaliana. *Plant Journal* **38**: 193-202.

Middleton E, Kandaswami C. 1992. Effects of flavonoids on immune and inflammatory cell functions. *Biochem Pharmacol* **43**: 1167-1179.

Mills R, Luttig C, Larkins C, Beauchamp A, Tsui C, Pittard WS, Devine SE. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome research* **16**: 1182-1190.

Moore GA. 2001. Oranges and lemons: clues to the taxonomy of Citrus from molecular markers. *Trends in Genetics* **17**: 536-540.

de Moraes A, dos Santos Soares Filho W, Guerra M. 2007. Karyotype diversity and the origin of grapefruit. *Chromosome Research* **1**: 115-121.

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

Moreno P, Ambrós S, Albiach-martí MR, Guerri J, Peña L. 2008. Citrus tristeza virus: a pathogen that changed the course of the citrus industry. *Molecular Plant Pathology* **9**: 251-268.

Morton CM. 2009. Phylogenetic relationships of the Aurantioideae (Rutaceae) based on the nuclear ribosomal DNAITS region and three noncoding chloroplast DNA regions, atpB-rbcL spacer, rps16, and trnL-trnF. *Organisms,Diversity&Evolution* **9**: 52-68.

Motohashi T, Matsuyama T, Akihama T. 1992. DNA fingerprinting in Citrus cultivars. *Proc Int Soc Citriculture* : 221-224.

Mouly PP, Arzouyan CR, Gaydou EM, Estienne JM. 1994. Differentiation of citrus juices by factorial discriminant analysis using liquid chromatography of flavanone glycosides. *Journal of Agricultural and Food Chemistry* **42**: 70-79.

Navarro L, Roistacher C, Murashige T. 1975. Improvement of Shoot-Tip Grafting Invitro for Virus-Free Citrus. *Journal of the American Society for Horticultural Science* **100**: 471-479.

Navarro L, Ballester J, Juárez J, Pina J, Arregui J, Bono R. 1981. Development of a program for desease-free citrus budwood in Spain. *Proc. Int. Soc. Citriculture* **1**: 452-456.

Navarro L, Pina J, Juarez J, Ballester-Olmos JF, Arregui JM, Ortega C, Navarro A, Duran-Vila N, Guerri J, Moreno P, Cambra M, Zaragoza S. 2002. The Citrus Variety Improvement Program in Spain in the period 1975–2001. In: N Duran-Vila, R Milne, J da Graça, eds. *Proceedings of the 15th Conference of the International Organization of Citrus Virologists*. Riverside, California, 306-316.

Navarro L. 2005. Necesidades y problemáticas de la mejora sanitaria y genética de los cítricos en España. *Phytoma.* **170**: 2-5.

Navarro L, Aleza P, Juárez J. 2006a. Mejora de la calidad de los cítricos. In: G. Llácer, M.J. Díez, J.M. Carrillo y M.L. Badenes (eds), *Mejora Genética de la calidad de las plantas.* Sociedad Española de Ciencias Hortícolas. Sociedad Española de Genética. Universidad Politécnica de Valencia, España. 579-596.

Navarro L, Juárez J, Aleza P, et al. 2006b. Selección de nuevos mandarinos triploides. *Agraria Comunitat Valenciana. 2ª época, año 2* **7**: 23-26.

Nicolosi E, Deng ZN, Gentile A, Malfa S, Continella G, Tribulato E. 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics* **100**: 1155-1166.

Nicolosi E. 2007. Origin and taxonomy. In: Khan I (ed) *Citrus genetics, breeding and biotechnology*. CAB International: Wallington, 19-43.

Nijman I, Kuipers S, Verheul M, Guryev V, Cuppen E. 2008. A genome-wide SNP panel for mapping and association studies in the rat. *BMC Genomics* **9**: 95.

Nogata Y, Sakamoto K, Shiratsuchi H, Ishii T, Yano M, Ohta H. 2006. Flavonoid composition of fruit tissues of citrus species. *Bioscience, Biotechnology and Biochemistry* **70**: 178-192.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature genetics* **30**: 190-193.

Novelli VM, Takita MA, Machado MA. 2004. Identification and analysis of single nucleotide polymorphisms (SNPs) in citrus. *Euphytica* **138**: 227-237.

Novelli VM, Cristofani M, Souza AA, Machado MA. 2006. Development and characterization of polymorphic microsatellite markers for the sweet orange (*Citrus sinensis* L. Osbeck). *Genetics and Molecular Biology* **29**: 90-96.

Olivares-Fuster Ó, Hernández-Garrido M, Guerri J, Navarro L. 2007. Plant somatic hybrid cytoplasmic DNA characterization by single-strand conformation polymorphism. *Tree physiology* **27**: 785-792.

de Oliveira RP, Cristofani M, Machado MA. 2002. Genetic mapping for citrus variegated chlorosis resistance. *Laranja* **23**: 247-261.

de Oliveira RP, Aguilar-Vildoso CI, Cristofani M, Machado MA. 2004. Skewed RAPD markers in linkage maps of Citrus. *Genetics and Molecular Biology* **27**: 437-441.

Ollitrault F, Terol J, Pina JA, Navarro L, Talon M, Ollitrault P. 2010. Development of SSR markers from *Citrus clementina* (Rutaceae) BAC end sequences and interspecific transferability in Citrus. *American Journal of Botany* **97**: e124-9.

Ollitrault F, Terol J, Alonso Martin A, Pina JA, Navarro L, Talon M, Ollitrault P. 2012. Development of InDel markers from *Citrus clementina* (Rutaceae) BAC-end sequences and interspecific transferability in *Citrus*. *American Journal of Botany*. **99**: 268-273.

Ollitrault P, and Luro F. 2001. Citrus. *In Tropical Plant Breeding.* **55**.

Ollitrault P, Jacquemond C, Dubois C, Luro F. 2003. Citrus. In: Hannon P, Seguin M, Perrier X, Glaszmann JC (eds). *Genetic diversity of cultivated tropical plants.* Montpellier/Enfield, NH: CIRAD/Science Publishers, Inc., 193-217.

Ollitrault P, Navarro L. 2012. Citrus. In: M Badenes, D Byrne, eds. *Fruit Breeding*. Springer New York Dordrecht Heildelberg London, 623-662.

Ollitrault P, Terol J, Garcia-Lor A, Bérard A, Chauveau A, Froelicher Y, Belzile C, Morillon R, Navarro L, Brunel D, Talon, M. 2012a. SNP mining in *C. clementina* BAC end sequences; transferability in the *Citrus* genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. *BMC Genomics* **13**: 13.

Ollitrault P, Terol J, Chen C, Federici CT, Lotfy S, Hippolyte I, Ollitrault F, Bérard A, Chauveau A, Costantino G, Kacar Y, Mu L, Cuenca J, Garcia-Lor A, Froelicher Y, Aleza P, Boland A, Billot C, Navarro L, Luro F, Roose ML, Gmitter FG, Talon M, Brune D. 2012b. A reference genetic map of C. clementina hort. ex Tan.; citrus evolution inferences from comparative mapping. *BMC Genomics* **13**: 593.

Olson JA. 1989. Provitamin A function of carotenoids: the conversion of β-carotene into vitamin A. *Journal of Nutrition* **119**: 105-108.

Omura M, Ueda T, Kita M, Komatsu A, Takanokura Y, Shimada T, Endo-Inagaki T, Nesumi H, Yoshida T. 2000. EST Mapping of Citrus. *Proceedings of the International Society of Citriculture IX Congress.* Orlando, FL, USA., 71-74.

Pang XM, Hu CG, Deng XX. 2003. Phylogenetic relationships among Citrus and its relatives as revealed by SSR markers. *Acta Genet Sin* **30**: 81-87.

Pang XM, Hu CG, Deng XX. 2007. Phylogenetic relationships within *Citrus* and its related genera as inferred from AFLP markers. *Genetic Resources and Crop Evolution* **54**: 429-436.

Park S, Yu HJ, Mum JH, Lee SC. 2010. Genome-wide discovery of DNA polymorphism in Brassica rapa. *Mol Genet Genomics* **283**: 135-145.

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

Peakall R, Smouse P. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288-295.

Peeters JP, Martinelli JA. 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. *Theoretical and Applied Genetics* **78**: 42-48.

Penjor T, Anai T, Nagano Y, Matsumoto R, Yamamoto M. 2010. Phylogenetic relationships of Citrus and its relatives based on rbcL gene sequences. *Tree Genetics and Genomes* **6**: 931-939.

Penniston K, Nakada S, Holmes R, Assimos D. 2008. Quantitative Assessment of Citric Acid in Lemon Juice, Lime Juice, and Commercially-Available Fruit Juice Products. *Journal of Endourology* **3**: 567-570.

Perrier X, Jacquemoud-Collet J. 2006. DARwin software. *http://darwin.cirad.fr/darwin **5.0.158.***

Perrotta G, Graniti A. 1988. Phomatracheiphila (Petri) Kanchaveli & Ghikashvili. In: Smith IM, Duarez J, Lelliott RA, Phillips DH, Archer SA (eds). *In: European Handbook of the Plant Disease.* Blackwell Scientific Publications, Oxford, 396-398.

Pessoa-Filho M, Rangel P, Ferreira M. 2010. Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. *BMC Plant Biology* **10**: 127.

Pino Del Carpio D, Basnet R, De Vos R, Maliepaard C, Paulo M, Bonnema G. 2011. Comparative Methods for Association Studies: A Case Study on Metabolite Variation in a Brassica rapa Core Collection. *PLoS ONE* **6**: e19624. doi:10.1371/journal.pone.0019624.

Plaxton WC. 1996. The organization and *regulation of plant glycolisis.*. *Annual reviews of plant physiology and plant molecular biology* **47**: 185-214.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Praloran J. 1977. Los agrios. Técnicas agrícolas y producciones tropicales. *Editorial Blume, Barcelona.*

Price A,L Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**: 904-909.

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

Puritz J, Addison J, Toonen R. 2012. Next-Generation Phylogeography: A Targeted Approach for Multilocus Sequencing of Non-Model Organisms. *PLoS ONE* **7**: e34241. doi:10.1371/journal.pone.0034241.

Quang N, Ikeda S, Harada K. 2008. Nucleotide variation in *Quercus crispula* Blume. *Heredity* **101**: 166-174.

Rafalski JA. 2010. Association genetics in crop improvement. *Current opinion in plant biology* **13**: 174-180.

Rafalski A, Morgante M. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* **20**: 103-111.

Ramadugu C, Keremane L, Lee R, Roose M. 2011. Single nucleotide polymorphisms in Citrus and members of Aurantioideae. *Plant & Animal Genomes XIX Conference, San Diego, CA.* W147.

Raman H, Raman R, Wood R, Martin P. 2006. Repetitive indel markers within the ALMT1 gene conditioning aluminium tolerance in wheat (*Triticum aestivum* L.). *Molecular Breeding* **18**: 171-183.

Rao AV, Rao LG. 2007. Carotenoids and human health. *Pharmacological Research* **55**: 207-216.

Rao MN, Soneji JR, Chen CX, Huang S, Gmitter FG, J. 2008. Characterization of zygotic and nucellar seedlings from sour orange-like citrus rootstock candidates using RAPD and EST-SSR markers. *Tree Genetics and Genomes* **4**: 113-124.

Reich DE, Goldstein DB. 2001. Detecting association in a case-control study while correcting for population stratification. *Genetic epidemiology* **20**: 4-16.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 11479-11484.

Reuther W. 1973. Climate and citrus behavior. pp. 281-337. En; Reuther W., L.D. Batchelor y H.J. Webber (eds.). *Citrus industry. Vol. 3.* University of California, Div. Agr. Sci., California.

Reuther W, Ríos-Castaño D. 1969. Comparison of growth, maduration and composition of citrus fruit in subtropical California and tropical Colombia. *Proc. First Intl. Citrus Symp.* **3**: 277-300.

Riju A, Chandrasekar A, Arunachalam V. 2007. Mining for single nucleotide polymorphisms and insertions/deletions in expressed sequence tag libraries of oil palm. *Bioinformation* **2**: 128-131.

Roberts E. 1973. Predicting the storage life of seeds. *Seed Science Technology* **1**: 499-514.

Rodrigo MJ, Marcos JF, Zacarías L. 2004. Biochemical and Molecular Analysis of Carotenoid Biosynthesis in Flavedo of Orange (*Citrus sinensis* L.) during Fruit Development and Maturation. *Journal of Agricultural and Food Chemistry* **52**: 6724-6731.

Roose M, Federici C, Mu L, Kwok K, Vu C. 2009. Map-based ancestry of sweet orange and other citrus variety groups. In: Gentile A, Tribulato E. (eds.) *Second International Citrus Biotechnology Symposium. Catania, Italy,* 28.

Rosenblum EB, Novembre J. 2007. Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *The Journal of heredity* **98**: 331-336.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. In: S In: Krawetz, Se Misener, eds. *Bioinformatics methods and protocols: methods in molecular biology.*, Humana Press edn: Totowa, NJ., 365-386.

Russo G, Recupero S, Puglisi A, Recupero G. 2004. New triploid citrus hybrids by Italian genetic improvement. *Rivista di Frutticoltura e di Ortofloricoltura.* **66**: 14-18.

Sachidanandam R, Weissman D, Schmidt S, et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928-933.

Sadka A, Dahan E, Or E, Roose ML, Marsh KB, Cohen L. 2001. Comparative analysis of mitochondrial citrate synthase gene structure, transcript level and enzymatic activity in acidless and acid-containing *Citrus* varieties. *Australian Journal of Plant Physiology* **28**: 383-390.

Saitou N, Nei M. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology Evolution* **4**: 406-425.

Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, de María A, Capella-Gutíerrez S, Huerta-Cepas J, Gabaldón T, Dopazo J, Dopazo H. 2011. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research* **39**: 1-5.

Sandmann G. 2001. Carotenoid biosynthesis and biotechnological application. *Archives of Biochemistry and Biophysics* **385**: 4-12.

Sanz ML, Villamiel M, Martínez-Castro I. 2004. Inositols and carbohydrates in different fresh fruit juices. *Food Chemistry* **87**: 325-328.

Scarano M, Tusa N, Abbate L, Lucretti S, Nardi L, Ferrante S. 2003. Flow cytometry, SSR and modified AFLP markers for the identification of zygotic plantlets in backcrosses between 'Femminello' lemon cybrids (2n and 4n) and a diploid clone of 'Femminello' lemon (*Citrus limon* L. Burm. F.) tolerant to mal secco disease. *Plant Science* **164**: 1009-1017.

Schijlen EGWM, Vos CHR, Tunen AJ, Bovy AG. 2004. Modification of flavonoid biosynthesis in crop plants. *Phytochemistry* **65**: 2631-2648.

Schoenbohm C, Martens S, Eder C, Forkmann G, Weisshaar B. 2000. Identification of the Arabidopsis thaliana flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biological Chemistry.* **381**: 749-753.

Scora RW. 1975. On the history and origin of citrus. *Bulletin of the Torrey Botanical Club* **102**: 369-375.

Scora RW, Kumamoto J, Soost RK,Nauer EM. 1982. Contribution to the origin of the grapefruit *Citrus paradisi* (Rutaceae). *Systematic Botany* **7**: 170-177.

Scora RW. 1988. Biochemistry, taxonomy and evolution of modern cultivated Citrus. In: Goren R, Mendel K. (eds.) *Proceedings of the 6th International Citrus Congress.*, Balaban Publishers; Weikersheim, Germany: Margraf Scientific Books edn: Philadelphia/Rehovot, 277-289.

Scott K, McIntyre C, Playford J. 2000. Molecular analyses suggest a need for a signifi cant rearrangment of Rutaceae subfamilies and a minor reassessment of species relationships within *Flindersia. Plant Systematics and Evolution* **223**: 15-27.

Shimada T, Fujii H, Endo T, Yazaki J, Kishimoto N, Shimbo K, Kikuchi S, Omura M. 2005. Toward comprehensive expression profiling by microarray analysis in citrus: monitoring the expression profiles of 2213 genes during fruit development. *Plant Science* **168**: 1383-1385.

Shimada T, Nakano R, Shulaev V, Sadka A, Blumwald E. 2006. Vacuolar citrate/H^+ symporter of citrus juice cells. *Planta* **224:** 472-480.

Shimizu T, Fujii H, Nishikawa F, Shimada T, Kotoda N, Yano K, Endo T. 2009. Data mining of citrus sequence data sets to develop microarrays for expression and genomic analysis. *Plant & Animal Genome XVII Conference, San Diego, California,* 414.

Shimizu T, Fujii H, Kotoda N, Yano K, Endo T. 2011. Data mining of citrus expression sequence data sets and application for functional genomic study. *Acta Horticulturae* **892**: 29-36.

Shimizu T, Yoshiuka T, Nagasaki H, Kaminuma E, Toyoda A, Fujiyama A, Nakamura Y. 2012. Whole genome sequencing and mapping analysis for idenfiying polymorphism among 11 citrus varieties. *XII International Citrus Congress- Valencia, Spain. S03O03,* 62.

Simsek O, Kacar YA, Yesiloglu T, Ollitrault P. 2011. Determination by SSCP markers of the allelic diversity of candidate genes for tolerance to iron chlorosis in citrus germplasm. : 85-91.

Soost R, Cameroon J. 1969. Tree and fruit characters of *Citrus* triploid from tetraploid by diploid crosses. *Hilgardia* **20**: 569-579.

Soost R, Cameron J. 1975. Citrus. In: Janick, J. and Moore, J.N. (eds) *Advances in Fruit Breeding.* Purdue University Press, West Lafayette, Indiana, 507–540.

Storey R, Walker RR. 1999. Citrus and salinity. *Scientia Horticulturae* **78**: 39-81.

Swingle W. 1939. Clymenia and Burkillanthus, new genera also three new species of Pleiospermium (Rutaceae - Aurantioideae). *J. Arnold Arbor.* **20**: 250-263.

Swingle W. 1943. The botany of Citrus and its wild relatives in the orange subfamily. In: Webber HJ, Batchelor LD (eds) *The citrus industry. History, world distribution, botany and wild relatives.* Berkeley CA: University of California **1**: 129-474.

Swingle W, Reece P. 1967. The botany of Citrus and its wild relatives. In: Reuther W, Webber HJ, Batchelor LD (eds) *The citrus industry. The botany of Citrus and its wild relatives.* Berkeley CA: University of California **1**: 190-430.

Tadeo F, Cercós M, Colmenero-Flores J, et al. 2008. Molecular Physiology of Development and Quality of Citrus. In: Anonymous *Advances in Botanical Research*. Academic Press, 147-223.

Tajima F. 1989a. DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. *Genetics* **123**: 229-240.

Tajima F. 1989b. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Talon M, Gmitter Jr. FG. 2008. Citrus Genomics. *International Journal of Plant Genomics.* Article ID 528361, 17 pages, doi:10.1155/2008/528361.

Tanaka T. 1954. Species problem in Citrus (Revisio Aurantiacearum IX). *Japanese Society for Promotion of Science, Tokyo, Japan*.

Tanaka T. 1961. Citologia: *semi-centennial commemoration papers on citrus studies.* Citologia Supporting Foundation: Osaka, Japan, 114.

Tanaka T. 1977. Fundamental discussion of *Citrus* classification. *Study in Citrologia, Osaka* **14**: 1-6.

Tanksley SD, McCouch SR. 1998. Seed banks and molecular maps: unlocking genetic potential from the wild. *Crop Genetic Resources* : 49-52.

Tapia Campos E, Gutiérrez Espinosa MA, Warburton ML, Santacruz Varela A, Villegas Monter Á. 2005. Characterization of mandarin (*Citrus* spp.) using morphological and AFLP markers. *Interciencia* **30**: 687-693.

Terol J, Conesa A, Colmenero JM, Cercos M, Tadeo F, Agustí J, Alós E, Andres F, Soler G, Brumos J, Iglesias DJ, Götz S, Legaz F, Argout X, Courtois B, Ollitrault P, Dossat C, Wincker P, Morillon R, Talon M. 2007. Analysis of 13000 unique Citrus clusters associated with fruit quality, production and salinity tolerance. *BMC genomics* **8**: 31.

Terol J, Naranjo MA, Ollitrault P, Talon M. 2008. Development of genomic resources for Citrus clementina: characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. *BMC genomics* **9**: 423.

Terol J, Carbonell J, Alonso R, et al. 2012. Sequencing of 150 citrus varieties: linking genotypes to phenotypes. *XII International Citrus Congress-Valencia, Spain. S03O02*: 61.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796-813.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nature genetics* **28**: 286-289.

Tokunaga T, Nii M, Tsumura T, Yamao M. 2005. Production of triploids and breeding seedless cultivar 'Tokushima 3X No.1' from tetraploid x diploid crosses in sudachi (Citrus sudachi Shirai). *Horticultural Research* **4**: 11-15.

Torres AM, Soost RK, Diedenhofen U. 1978. Leaf isozymes as genetic markers in Citrus. *American Journal of Botany* **65**: 869-881.

van Treuren R, Tchoudinova I, Soest LJM, Hintum TJL. 2006. Marker-assisted acquisition and core collection formation: a case study in barley using AFLPs and pedigree data. *Genetic Resources and Crop Evolution* **53**: 43-52.

Tucker GA. 1993. Introduction. In : Biochemistry of fruit ripening, Seymour G, Taylor J, Tucker G (eds.) London: Chapman and Hall: 1-37.

Tucker GA. 2003. Nutritional enhancement of plants. *Current opinion in biotechnology* **14**: 221-225.

Tudela D, Primo-Millo E. 1992. 1-Aminocyclopropane-1-carboxylic acid transported from roots to shoots promotes leaf abscission in Cleopatra mandarin (Citrus reshni Hort. ex Tan.) seedlings rehydrated after water stress. *Plant Physiology* **100**: 131-137.

Turner W, Waller J, Vanderbeld B, Snedden W. 2004. Cloning and characterization of two NAD kinases from Arabidopsis: identification of a calmodulin binding isoform. *Plant Physiology* **135**: 1243-1255.

Turner W, Waller J, Snedden W. 2005. Identification, molecular cloning and functional characterization of a novel NADK kinase from Arabidopsis thaliana (thale cress). *Biochemical Journal* **385**: 217-223.

Tuskan GA, DiFazio S, Jansson S, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596-1604.

Uzun A, Yesiloglu T, Aka-Kacar Y, Tuzcu O, Gulsen O. 2009. Genetic diversity and relationships within *Citrus* and related genera based on sequence related amplified polymorphism markers (SRAPs). *Scientia Horticulturae* **121**: 306-312.

Väli Ü, Brandström M, Johansson M, Ellegren H. 2008. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genetics* **9**: 8.

van Hintum TJL, Brown AHD, Spillane C, Hodgkin T. 2000. Core collections of plant genetic resources. *IPGRI Technical Bulletin No. 3*. Rome, Italy: International Plant Genetic Resources Institute.

Varshney R, Graner A, Sorrells M. 2005. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48-55.

Vasemägi A, Gross R, Palm D, Paaver T, Primmer CR. 2010. Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genomics* **11**: 156.

Vicent A, Badal J, Asensi MJ, Sanz N, Armengol J, García-Jiménez J. 2004. Laboratory Evaluation of citrus cultivars susceptibility and influence of fruit size on Fortune mandarin to infection by *Alternaria alternata* pv. citri. *European Journal of Plant Pathology* **110**: 245-251.

Vicent A, Armengol J, Sales R, García-Jiménez J, Alfaro-Lassala F. 2000. First Report of Alternaria Brown Spot of Citrus in Spain. *Plant Disease* **84**: 1044-1044.

Viruel M. 2010. Mejora genética y recursos fitogenéticos: "Nuevos avances en la conservación y utilización de los recursos fitogenéticos". *Editores: JM Carrillo, MJ Díez, M Pérez de la Vega, F Nuez. Edita: Ministerio de Medio Ambiente y Medio Rural y Marino. ISBN: 978-84-491-1014-6.* Pp: 319-349.

Volk G, Bonnart R, Shepherd A, Krueger RR, Lee R. 2012. Cryopreservation of citrus for long-term conservation. *XII International Citrus Congress, Valencia-Spain. S01O02*: 27.

Wang ML, Barkley NA, Yu JK, Dean RE, Newman ML, Sorrells ME, Pederson GA. 2005. Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. *Plant Genetic Resources* **3**: 45.

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *American Journal of Human Genetics* **68**: 738-752.

Webber HJ. 1943. Cultivated varieties of citrus. In: Webber HJ, Batchelor DL (eds.) *The citrus industry. History, world distribution, botany and varieties*, Berkeley CA: University of California, **1**: 475-668.

Webber HJ, Reuther W, Lawton H. 1967. History and development of the citrus industry. In: Reuther W, Webber HJ, Batchelor LD (eds) *The citrus industry. The botany of Citrus and its wild relatives.* Berkeley CA: University of California, **1**: 1–39.

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Human molecular genetics* **2**: 1123-1128.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Williams T, Roose M. 2004. 'TDE2' Mandarin hybrid (Shasta Gold Mandarin), 'TDE3' Mandarin hybrid (Tahoe Gold Mandarin) and 'TDE4' MAndarin hybrid (Yosemite Gold Mandarin): Three New Mid and Late-Season Triploid Seedless Mandarin Hybrids from California. *Proceedings International Society Citriculture* **1**: 394-398.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* **126**: 485-493.

Wright S. 1969. Evolution and the genetics of populations. *The theory of gene frequencies.* The University of Chicago Press: Chicago, vol2.

Wright S. 1978. Evolution and the genetics of populations. *A treatise in four volumes. Variability within and among natural populations*. The University of Chicago Press: Chicago, vol 4.

Xing C, Schumacher F, Xing G, Lu Q, Wang T, Elston R. 2005. Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. *BMC Genetics* **6**: S29.

Xu Q, Chen L, Ruan X, et al. 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nature genetics* **45**: 59-66.

Yakushiji H, Morinaga K, Nonami H. 1998. Sugar accumulation and partitioning in Satsuma mandarin tree tissues and fruit in response to drought stress. *Journal of the American Society for Horticultural Science* **123**: 719-726.

Yamamoto M, Kobayashi S, Nakamura Y, Yamada Y. 1993. Phylogenic relationships of *Citrus* revealed by diversity of cytoplasmic genomes. In: Hayashi T, Omura M, Scott NS (eds.). *Techniques on gene diagnosis and breeding in fruit trees.* Research Station, Okitsu, Japan, 39-46.

Yamamoto M, Tominaga S. 2003. High chromosomal variability of mandarins (Citrus spp.) revealed by CMA banding. *Euphytica* **129**: 267-274.

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **17**: 2859-2872.

Yan WG, Yong L, Agrama HA, Dagang L, Fangyuan G, Xianjun L, Guangjun R. 2009. Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). *Molecular Breeding* **24**: 277-292.

Yang Y, YueZhi P, Xun G, MouTian F. 2010. Genetic variation in the endangered Rutaceae species Citrus hongheensis based on ISSR fingerprinting. *Genetic Resources and Crop Evolution* **57**: 1239-1248.

Yonekura-Sakakibara K, Saito K. 2006. Review: genetically modified plants for the promotion of human health. *Biotechnology Letters* **28**: 1983-1991.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203.

Yu JM, Buckler ES. 2006. Genetic association mapping and genome organization of maize. *Current opinion in biotechnology* **17**: 155-160.

Zaragoza S. 2007. Aproximación a la historia de los cítricos. Origen, dispersión y evolución de su uso y cultivo. *Tesis doctoral. Universidad Politécnica de Valencia. Departamento de Producción Vegetal, Valencia, España*.

Zhang Y, Zhang Q, Yang Y, Luo Z. 2009. Development of japanese persimmon core collection by genetic distance sampling based on SSR markers. *Agric Sci China* **8**: 276-284.

Zhang L, Vision TJ, Gaut BS. 2002. Patterns of Nucleotide Substitution Among Simultaneously Duplicated Gene Pairs in Arabidopsis thaliana. *Molecular biology and evolution* **19**: 1464-1473.

Zhao KY, Aranzana MJ, Kim S, Lister C, Shindo C, Tang CL, Toomajian C, Zheng HG, Dean C, Marjoram P, Nordborg M. 2007. An Arabidopsis example of association mapping in structured samples. *PLoS Genetics* **3**: e4.

Zhu C, Gore M, Buckler ES, Yu J. 2008. Status and prospects of association mapping in plants. *Plant Genome* **1**: 5-20.

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. 2003. Single-Nucleotide Polymorphisms in Soybean. *Genetics* **163**: 1123-1134.

ANNEX

OTHER WORKS PERFORMED ALONG WITH THE PhD THESIS

Beside the work included in this PhD thesis, I contributed to the research of the IVIA/CIRAD laboratory in citrus genetics. I have been therefore associated to three papers published in peered reviewed articles.

Some of the indels and SNP markers developed during this thesis have been employed for the establishment of the reference genetic map of *C. clementina* (Ollitrault *et al*., 2012b), which constitutes a good framework for further marker-trait association studies, and helped the chromosome assembly of the reference whole genome citrus sequence (Gmitter *et al*., 2012).

Several of these SNP markers were also included on a GoldenGate array platform in addition to more than 600 SNPs mined in clementine BAC-end sequences to genotype 54 accession covering the main *Citrus* species and 52 inter-specific hybrids between pummelo and clementine (Ollitrault *et al*., 2012a). The SNP data confirmed the important stratification of the gene pools around *C. maxima*, *C. medica* and *C. reticulata* as well as previous hypothesis on the origin of secondary species. The implemented SNP marker set will be very useful for comparative genetic mapping in *Citrus* and genetic association in *C. reticulata*.

Besides the previous two works, I was involved in the characterization of the diversity of Tunisian citrus rootstocks (Snoussi *et al.*, 2012). Two hundred and one local accessions belonging to four facultative apomictic species (*C. aurantium*, sour orange; *C. sinensis*, orange; *C. limon*, lemon; and *C. aurantifolia*, lime) were collected and genotyped using 20 nuclear SSR markers and four indel mitochondrial markers. Multi-locus genotypes (MLGs) were compared to references from French and Spanish collections. The Tunisian citrus rootstock genetic diversity is predominantly due to high heterozygosity and differentiation between the four varietal groups. The phenotypic diversity within the varietal groups has resulted from multiple introductions, somatic mutations and rare sexual recombination events. Finally, this diversity study enabled the identification of a core sample of accessions for further physiological and agronomical evaluations. These core accessions will be integrated into citrus rootstock breeding programs for the Mediterranean Basin.

On the other hand, several works have been presented as poster and oral communications in different congresses as first author and some other as collaborations with other colleagues:

Articles in international journals

- **Ollitrault P, Terol J, Garcia-Lor A, Bérard A, Chauveau A, Froelicher Y, Belzile C, Morillon R, Navarro L, Brunel D, Talon, M. 2012a.** SNP mining in *C. clementina* BAC end sequences; transferability in the *Citrus* genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. *BMC Genomics* **13**: 13.
- **Ollitrault P, Terol J, Chen C, Federici CT, Lotfy S, Hippolyte I, Ollitrault F, Bérard A, Chauveau A, Costantino G, Kacar Y, Mu L, Cuenca J, Garcia-Lor A, Froelicher**

Y, Aleza P, Boland A, Billot C, Navarro L, Luro F, Roose ML, Gmitter FG, Talon M, Brune D. 2012b. A reference genetic map of C. clementina hort. ex Tan.; citrus evolution inferences from comparative mapping. *BMC Genomics* **13**: 593.

- **Snoussi H, Duval MF, Garcia-Lor A, Belfalah Z, Froelicher Y, Risterucci AM, Perrier X, Jacquemoud-Collet JP, Navarro L, Harrabi M, Ollitrault P. 2012.** Assessment of the genetic diversity of the Tunisian citrus rootstock germplasm. *BMC Genetics* **13**: 16.

Congress oral communications

- **Garcia-Lor A, Luro F, Ancillo G, Ollitrault P, Navarro L.** Genetic diversity and population structure of the mandarin germplasm revealed by nuclear and mitochondrial markers analysis. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S01O03, Pag. 27.
- **Curk F, Ancillo G, Garcia-Lor A, Luro F, Navarro L, Ollitrault P.** Multilocus haplotyping by parallel sequencing to decipher the interspecific mosaic genome structure of cultivated citrus. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S01O06, Pag. 28.
- **Ollitrault P, Terol J, Chen C, Federici CT, Lotfy S, Hippolyte I, Ollitrault F, Bérard A, Chauveau A, Cuenca J, Costantino G, Kacar Y, Mu L, Garcia-Lor A, Froelicher Y, Aleza P, Boland A, Billot C, Navarro L, Luro F, Roose ML, Gmitter Jr. FG, Talón M, Brunel D.** A reference genetic map of Citrus clementina; citrus evolution inferences from comparative mapping. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S03O05, Pag. 63.

Congress posters

- **Garcia-Lor A, Luro F, Navarro L, Ollitrault P.** Analysis of genetic diversity and population structure of *Citrus* Germplasm using nuclear (SSRs, indels) and mitochondrial markers. 2nd International Symposium on Genomics of Plant Genetic Resources. Bologna (Italy). 24-27 April 2010. P3.31.
- **Garcia-Lor A, Luro F, Navarro L, Ollitrault P.** Análisis de la diversidad genética y de la estructura poblacional del Germoplasma de mandarino mediante marcadores moleculares nucleares (SSRs, indels) y mitocondriales. V Congreso de Mejora Genética de Plantas. Madrid, 7-9 Julio 2010. Pag. 253-254.
- **Garcia-Lor A, Curk F, Luro F, Navarro L, Ollitrault P.** Nuclear and maternal phylogeny within Citrus and four related genera based on nuclear gene sequence SNPs and mitochondrial indels. Plant Genome Evolution. Amsterdam, The Netherlands, 4-6 September 2011. P2.23.
- **Garcia-Lor A, Curk F, Snoussi H, Morillon R, Ancillo G, Luro F, Navarro L, Ollitrault P.** Nuclear phylogeny of Citrus and four related genera. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S01P08, Pag. 31.
- **Curk F, Garcia-Lor A, Snoussi H, Froelicher Y, Ancillo G, Navarro L, Ollitrault P.** New insights on limes and lemons origin from targeted nuclear gene sequencing and cytoplasmic markers genotyping. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S01P09, Pag. 32.
- **Ollitrault P, Garcia-Lor A, Terol J, Curk F, Ollitrault F, Talon M, Navarro L.** Comparative values of SSRs, SNPs and indels for citrus genetic diversity analysis. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S02P05, Pag. 44.
- **Curk F, Ancillo G, Garcia-Lor A, Luro F, Ollitrault P, Navarro L.** Multilocus SNPs analysis allows phylogenetic assignation of DNA fragments to decipher the interspecific mosaic genome structure of cultivated *Citrus*. Plant Genome Evolution. Amsterdam, The Netherlands, 4-6 September 2011. P2.21.

- **Snoussi H, Duval MF, Garcia-Lor A, Perrier X, Jacquemoud-Collet JC, Navarro L, Ollitrault P.** Analysis of genetic diversity in Tunisian citrus rootstocks. XII International Citrus Congress- Valencia, Spain. November 18th-23rd 2012. S01P11, Pag. 33.
- **Terol J, Chen C, Federici CT, Lotfy S, Hippolyte I, Ollitrault F, Bérard A, Chauveau A, Costantino G, Kacar Y, Mu L, Cuenca J, Garcia-Lor A, Froelicher Y, Aleza P, Boland A, Billot C, Navarro L, Luro F, Roose ML, Gmitter FG, Talon M, Brune D.** A Reference Linkage Map of *C. clementina* Based on SNPs, SSRs and Indels. Plant & Animal Genomes XIX Conference. January 15-19, 2011. Town & Country Convention Center San Diego, USA. P.477.

Articles in international journals

Ollitrault et al. BMC Genomics 2012, 13:593
http://www.biomedcentral.com/1471-2164/13/593

BMC Genomics

RESEARCH ARTICLE Open Access

A reference genetic map of *C. clementina* hort. ex Tan.; citrus evolution inferences from comparative mapping

Patrick Ollitrault[1,2*], Javier Terol[3], Chunxian Chen[4], Claire T Federici[5], Samia Lotfy[1,6], Isabelle Hippolyte[1], Frédérique Ollitrault[2], Aurélie Bérard[7], Aurélie Chauveau[7], Jose Cuenca[2], Gilles Costantino[8], Yildiz Kacar[9], Lisa Mu[5], Andres Garcia-Lor[2], Yann Froelicher[1], Pablo Aleza[2], Anne Boland[10], Claire Billot[1], Luis Navarro[2], François Luro[8], Mikeal L Roose[5], Frederick G Gmitter[4], Manuel Talon[3] and Dominique Brunel[7]

Abstract

Background: Most modern citrus cultivars have an interspecific origin. As a foundational step towards deciphering the interspecific genome structures, a reference whole genome sequence was produced by the International Citrus Genome Consortium from a haploid derived from Clementine mandarin. The availability of a saturated genetic map of Clementine was identified as an essential prerequisite to assist the whole genome sequence assembly. Clementine is believed to be a 'Mediterranean' mandarin × sweet orange hybrid, and sweet orange likely arose from interspecific hybridizations between mandarin and pummelo gene pools. The primary goals of the present study were to establish a Clementine reference map using codominant markers, and to perform comparative mapping of pummelo, sweet orange, and Clementine.

Results: Five parental genetic maps were established from three segregating populations, which were genotyped with Single Nucleotide Polymorphism (SNP), Simple Sequence Repeats (SSR) and Insertion-Deletion (Indel) markers. An initial medium density reference map (961 markers for 1084.1 cM) of the Clementine was established by combining male and female Clementine segregation data. This Clementine map was compared with two pummelo maps and a sweet orange map. The linear order of markers was highly conserved in the different species. However, significant differences in map size were observed, which suggests a variation in the recombination rates. Skewed segregations were much higher in the male than female Clementine mapping data. The mapping data confirmed that Clementine arose from hybridization between 'Mediterranean' mandarin and sweet orange. The results identified nine recombination break points for the sweet orange gamete that contributed to the Clementine genome.

Conclusions: A reference genetic map of citrus, used to facilitate the chromosome assembly of the first citrus reference genome sequence, was established. The high conservation of marker order observed at the interspecific level should allow reasonable inferences of most citrus genome sequences by mapping next-generation sequencing (NGS) data in the reference genome sequence. The genome of the haploid Clementine used to establish the citrus reference genome sequence appears to have been inherited primarily from the 'Mediterranean' mandarin. The high frequency of skewed allelic segregations in the male Clementine data underline the probable extent of deviation from Mendelian segregation for characters controlled by heterozygous loci in male parents.

Keywords: *C. clementina*, *C. sinensis*, *C. maxima*, SSRs, SNPs, Indels, Genetic maps

* Correspondence: patrick.ollitrault@cirad.fr
[1]CIRAD, UMR AGAP, F-34398 Montpellier, France
[2]IVIA, Centro Proteccion Vegetal y Biotechnologia, Ctra. Moncada-Náquera Km 4.5, 46113 Moncada, Valencia, Spain
Full list of author information is available at the end of the article

Ollitrault et al. BMC Genomics 2012, 13:13
http://www.biomedcentral.com/1471-2164/13/13

BMC Genomics

RESEARCH ARTICLE Open Access

SNP mining in *C. clementina* BAC end sequences; transferability in the *Citrus* genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping

Patrick Ollitrault[1,2*], Javier Terol[3], Andres Garcia-Lor[2], Aurélie Bérard[4], Aurélie Chauveau[4], Yann Froelicher[1], Caroline Belzile[2], Raphaël Morillon[1,3], Luis Navarro[2], Dominique Brunel[4] and Manuel Talon[3]

Abstract

Background: With the increasing availability of EST databases and whole genome sequences, SNPs have become the most abundant and powerful polymorphic markers. However, SNP chip data generally suffers from ascertainment biases caused by the SNP discovery and selection process in which a small number of individuals are used as discovery panels. The ongoing International Citrus Genome Consortium sequencing project of the highly heterozygous Clementine and sweet orange genomes will soon result in the release of several hundred thousand SNPs. The primary goals of this study were: (i) to estimate the transferability within the genus *Citrus* of SNPs discovered from Clementine BACend sequencing (BES), (ii) to estimate bias associated with the very narrow discovery panel, and (iii) to evaluate the usefulness of the Clementine-derived SNP markers for diversity analysis and comparative mapping studies between the different cultivated *Citrus* species.

Results: Fifty-four accessions covering the main *Citrus* species and 52 interspecific hybrids between pummelo and Clementine were genotyped on a GoldenGate array platform using 1,457 SNPs mined from Clementine BES and 37 SNPs identified between and within *C. maxima*, *C. medica*, *C. reticulata* and *C. micrantha*. Consistent results were obtained from 622 SNP loci. Of these markers, 116 displayed incomplete transferability primarily in *C. medica*, *C. maxima* and wild *Citrus* species. The two primary biases associated with the SNP mining in Clementine were an overestimation of the *C. reticulata* diversity and an underestimation of the interspecific differentiation. However, the genetic stratification of the gene pool was high, with very frequent significant linkage disequilibrium. Furthermore, the shared intraspecific polymorphism and accession heterozygosity were generally enough to perform interspecific comparative genetic mapping.

Conclusions: A set of 622 SNP markers providing consistent results was selected. Of the markers mined from Clementine, 80.5% were successfully transferred to the whole *Citrus* gene pool. Despite the ascertainment biases in relation to the Clementine origin, the SNP data confirm the important stratification of the gene pools around *C. maxima*, *C. medica* and *C. reticulata* as well as previous hypothesis on the origin of secondary species. The implemented SNP marker set will be very useful for comparative genetic mapping in *Citrus* and genetic association in *C. reticulata*.

* Correspondence: patrick.ollitrault@cirad.fr
[1]CIRAD, UMR AGAP, Avenue Agropolis, TA A-108/02, 34398 Montpellier, Cedex 5, France
Full list of author information is available at the end of the article

Snoussi et al. BMC Genetics 2012, 13:16
http://www.biomedcentral.com/1471-2156/13/16

BMC Genetics

RESEARCH ARTICLE Open Access

Assessment of the genetic diversity of the Tunisian citrus rootstock germplasm

Hager Snoussi[1*], Marie-France Duval[2], Andres Garcia-Lor[3], Zina Belfalah[1], Yann Froelicher[2], Ange-Marie Risterucci[2], Xavier Perrier[2], Jean-Pierre Jacquemoud-Collet[2], Luis Navarro[3], Moncef Harrabi[4] and Patrick Ollitrault[2,3*]

Abstract

Background: Citrus represents a substantial income for farmers in the Mediterranean Basin. However, the Mediterranean citrus industry faces increasing biotic and abiotic constraints. Therefore the breeding and selection of new rootstocks are now of the utmost importance. In Tunisia, in addition to sour orange, the most widespread traditional rootstock of the Mediterranean area, other citrus rootstocks and well adapted to local environmental conditions, are traditionally used and should be important genetic resources for breeding. To characterize the diversity of Tunisian citrus rootstocks, two hundred and one local accessions belonging to four facultative apomictic species (*C. aurantium*, sour orange; *C. sinensis*, orange; *C. limon*, lemon; and *C. aurantifolia*, lime) were collected and genotyped using 20 nuclear SSR markers and four indel mitochondrial markers. Multi-locus genotypes (MLGs) were compared to references from French and Spanish collections.

Results: The differentiation of the four varietal groups was well-marked. The groups displayed a relatively high allelic diversity, primarily due to very high heterozygosity. Sixteen distinct MLGs were identified. Ten of these were noted in sour oranges. However, the majority of the analysed sour orange accessions corresponded with only two MLGs, differentiated by a single allele, likely due to a mutation. The most frequent MLG is shared with the reference sour oranges. No polymorphism was found within the sweet orange group. Two MLGs, differentiated by a single locus, were noted in lemon. The predominant MLG was shared with the reference lemons. Limes were represented by three genotypes. Two corresponded to the 'Mexican lime' and 'limonette de Marrakech' references. The MLG of 'Chiiri' lime was unique.

Conclusions: The Tunisian citrus rootstock genetic diversity is predominantly due to high heterozygosity and differentiation between the four varietal groups. The phenotypic diversity within the varietal groups has resulted from multiple introductions, somatic mutations and rare sexual recombination events. Finally, this diversity study enabled the identification of a core sample of accessions for further physiological and agronomical evaluations. These core accessions will be integrated into citrus rootstock breeding programs for the Mediterranean Basin.

Background

Worldwide production of citrus in 2009 reached greater than 120 million tons [1], making citrus the leading cultivated tree crop in the world. Oranges represent the majority of citrus production (54% in 2009) with over 67 million tons in 2009 [1]. The other significant cultivated citrus are mandarins, lemons and grapefruits. The citrus production of the Mediterranean Basin is second only to Brazil. Cultivars are vegetatively propagated by bud-grafting onto seedling rootstocks. This ensures tree uniformity, early tree production, and tolerance to pathogens including *Phytophthora* sp., parasitic nematodes and viruses. Rootstocks are also significant in the adaptation of the tree to several abiotic constraints affecting the Mediterranean citriculture. These include water resource scarcity and soil salinity. Citrus rootstocks are generally apomictic and seed-propagated. Therefore, both scions and rootstocks are clonally propagated. Sour orange (C. *aurantium* L.), one of the most important citrus rootstocks in the world, is still the

* Correspondence: hagersnoussi@gmail.com; patrick.ollitrault@cirad.fr
[1]Horticultural Laboratory, Tunisian National Agronomic Research Institute (INRAT), Rue Hedi Karray, 2049 Ariana, Tunisia
[2]Department BIOS, TGU AGAP, International Center for of Agricultural Research for Development (CIRAD), Avenue Agropolis, TA A-75/02, F-34398 Montpellier Cedex 5, France
Full list of author information is available at the end of the article

Congress oral communications

S01O03

Genetic diversity and population structure of the mandarin germplasm revealed by nuclear and mitochondrial markers analysis

Garcia-Lor A.[1], Luro F. [2] , Ancillo G. [1] , Ollitrault P. [3] , and Navarro L. [1]

[1]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Proteccion Vegetal Y Biotecnología, Spain; [2] Institut National de la Recherche Agronomique (INRA), Unité de recherche GEQA, France; and [3] Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BIOS, France. angarcia@ivia.es

Mandarins (*C. reticulata*) are considered as one of the four main species involved in the origin of cultivated citrus. However, the classification of the mandarin germplasm is still controversial and numerous cases of introgression from other species are known or suspected in this germplasm. The main objective of this work was to analyze the genetic diversity structure of mandarin germplasm and its relationship with the other citrus species. Fifty microsatellite (SSR) markers, 25 Insertion-Deletion (InDel) nuclear markers and four mitochondrial InDel markers were genotyped for 223 accessions. 'Structure' software was applied on nuclear data to check and quantify potential interspecific introgressions in the mandarin germplasm, mainly the pummelo and papeda genomes. Within the mandarin germplasm without identified introgression, seven clusters were revealed by 'Structure' analysis. Five of them should be true basic mandarin groups and the other two include genotypes of known or supposed hybrid origin. The contributions of these seven groups to the mandarin genotypes were estimated. The mitochondrial InDel analysis revealed eight mitotypes, in which the mandarin germplasm was represented in four of them. In this work, new insights in the organization of mandarin germplasm and its structure have been found, and different mandarin core collections were determined. This will allow a better management and use of citrus germplasm collections and to perform genetic association studies.

S01O06

Multilocus haplotyping by parallel sequencing to decipher the interspecific mosaic genome structure of cultivated citrus

Curk F.[1], Ancillo G.[2], Garcia-Lor A.[2], Luro F.1, Navarro L.[2], and Ollitrault P.[3]

[1]INRA, UR1103 Génétique et Ecophysiologie de la Qualité des Agrumes (INRA UR GEQA), France; [2]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; and [3]Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD), BIOS, France. curk@corse.inra.fr

Recent studies support the theory that four basic taxa (*Citrus medica*, *Citrus maxima*, *Citrus reticulata* and *Citrus micrantha*) have generated all cultivated Citrus species. It is supposed that the genomes of most of the actual citrus cultivars are interspecific mosaics of large DNA fragments issued from a limited number of interspecific meiotic events. In the present work, we analyzed how haplotypic multilocus study of closely linked SNPs allows phylogenetic assignment of DNA fragments for the main cultivated species. We have developed a new method based on universal primers to prepare the amplicons to be analyzed by 454 technology (Roche). It was applied for direct multilocus haplotyping of 12 gene fragments of 48 *Citrus* genotypes. Moreover, Sanger sequencing was performed on a subset of these amplicons (seven gene fragments of 24 citrus genotypes) to validate the 454 results. Consensus haplotype sequences were successfully identified from 454 sequencing. Sanger and 454 results were mostly identical. *C. reticulata* was the most polymorphic basic taxa. The average differentiation between the basic taxa was about 20 SNPs/kb. These polymorphisms were enough for unambiguous multilocus differentiation of the basic species and assignment of phylogenetic origin for each haplotype of the secondary species. Multilocus haplotyping by parallel sequencing will be a powerful tool to decipher the interspecific mosaic genome structure of cultivated citrus.

S03O05

A reference genetic map of *Citrus clementina*; citrus evolution inferences from comparative mapping

Ollitrault P.[1], Terol J.[2], Chen C. [3] , Federici C.T. [4] , Lotfy S. [5] , Hippolyte I. [1] , Ollitrault F. [6] , Bérard A. [7] , Chauveau A. [7] , Cuenca J. [6] , Costantino G. [8] , Kacar Y. [9] , Mu L. [4] , Garcia-Lor A.[6] , Froelicher Y. [1] , Aleza P. [6] , Boland A. [10] , Billot C. [1] , Navarro L. [6] , Luro F. [8] , Roose M.L. [3] , Gmitter Jr. F.G. [3] , Talón M. [2] , and Brunel D. [7]

[1] Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BIOS, France; [2] Instituto Valenciano de Investigacion Agrarias (IVIA), Genomic Center, Spain; [3] Citrus Research and Education Center, University of Florida (CREC), USA; [4] University of California, Riverside (UCR), Botany and Plant science, USA; [5] Institut National de la Recherche Agronomique (INRA), Morroco; [6] Instituto Valenciano de Investigacion Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; [7] Institut National de la Recherche Agronomique (INRA), UR EPGV, France; [8] Institut National de la Recherche Agronomique (INRA), UR GEQA, France; [9] Faculty of Agriculture, University of Çukurova, Department of Horticulture, Turkey; and [10] Centre de l'Energie Atomique (CEA), DSV/Institut de Génomique, France. patrick.ollitrault@cirad.fr

The availability of a saturated genetic map of clementine was identified by the ICGC as an essential prerequisite to assist the assembly of the reference whole genome sequence based on a 'Clemenules' clementine derived haploid. The primary goals of the present study were to establish a clementine reference map, and to perform comparative mapping with pummelo and sweet orange. Five parental genetic maps were established with SNPs, SSRs and InDels. A medium density reference map (961 markers for 1084.1 cM) of clementine was established and used by the ICGC to facilitate the chromosome assembly of the haploid genome sequence. Comparative mapping with pummelo and sweet orange revealed that the linear order of markers was highly conserved. The map should allow reasonable inferences of most citrus genomes by mapping next-generation sequencing data against the haploid reference genome sequence. Significant differences in map size were observed between species, suggesting variations in the recombination rates. Skewed segregations were frequent and higher in the male than female clementine. The mapping data confirmed that clementine arose from hybridization between 'Mediterranean' mandarin and sweet orange and identified nine recombination break points for the sweet orange gamete that contributed to the clementine genome. Moreover it appears that the genome of the haploid clementine used to establish the citrus reference genome sequence has been inherited primarily from the 'Mediterranean' mandarin.

Congress posters

Analysis of genetic diversity and population structure of the *Citrus* Germplasm using nuclear markers (SSRs, INDELs) and mitochondrial markers.

A. Garcia-Lor[1], F. Luro[2], L. Navarro[1] and P. Ollitrault[1,3]

[1] Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada (Valencia), Spain
[2] Institut National de Recherche Agronomique (INRA), San Giuliano, France
[3] Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France

Previous molecular markers studies (ISSR, RAPD, SCAR, AFLP and SSR) have shown that most of the genetic diversity of cultivated *Citrus* (except *C. aurantifolia*) comes from the recombination between three main species: *C. medica* (citron), *C. reticulata* (mandarin) and *C. maxima* (pummelo). However the precise contribution of these basic species to the genome constitution of secondary species (*C. sinensis, C. limon, C. aurantium, C. paradisi*) and recent hybrids is not known.

In this study, 58 nuclear markers and 4 mitochondrial markers were used to investigate the genetic diversity among 106 *Citrus* accessions, representing the three main ancestors groups, secondary species and several hybrids from the 20th century breeding programs. For the nuclear analysis, 50 simple sequence repeats (SSRs) developed from genomic libraries and ESTs databases were used. Moreover, 10 Insertion-Deletion (INDEL) markers were developed from genomic sequences of some primary and secondary metabolites determining the citrus fruit quality (sugars, acids, flavonoids and carotenoids. All the SSR markers and one INDEL are included in a consensus genetic map of Clementine x Chandler and are distributed along all the linkage groups, representing positively the global diversity of *Citrus*.

Genetic diversity statistics were calculated for each SSR and INDEL marker, within the entire population and within and between the different specified *Citrus* groups. The organizations of the genetic diversity among all the accessions were determined by constructing neighbor-joining trees for the different sets of primers.

INDEL markers are less polymorphic than SSRs, display a higher structuration of genetic diversity and appear as better phylogenetic markers to trace the contribution of the three ancestral species.

Population structure was studied using the Structure software, version 2.2.3, (http://cbsuapps.tc.cornell.edu/structure) which implements a model-based clustering method for inferring population structure using genotype data. The relative proportion of ancestral taxa genomes in the secondary species and recent hybrids was assigned.

Mitochondrial markers revealed the maternal phylogeny of citrus germplasm accessions in agreement with previous studies with chloroplastic markers.

This analysis allowed a better understanding of the organization of genetic diversity among citrus cultivars, opening the way for a better management of citrus germplasm bank and breeding programs.

Análisis de la diversidad genética y de la estructura poblacional del Germoplasma de mandarino mediante marcadores moleculares nucleares (SSRs, INDELs) y mitocondriales.

A. Garcia-Lor[1], F. Luro[3], L. Navarro[1] y P. Ollitrault[1-2]

[1]Centro de Protección Vegetal y Biotecnología, IVIA. Ctra. Moncada-Náquera km 4.5 46113 Moncada, Valencia.

[2]Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UPR 75, Avenue Agropolis - TA A-75/02–34398, Montpellier, cedex5, France.

[3] Unité de recherche GEQA, Institut National de la de Recherche Agronomique (INRA), San Giuliano, France.

Palabras clave: cítricos, variabilidad, microsatélite, marcadores filogenéticos, *C. reticulata*, inserción, delección.

INTRODUCCIÓN

Estudios previos con marcadores moleculares (ISSR, RAPD, SCAR, AFLP y SSR) han mostrado que la mayor parte de la diversidad genética de los cítricos cultivados (excepto *C. aurantifolia*) procede de la recombinación entre tres especies principales: *C. medica* L. (cidro), *C. reticulata* Blanco (mandarino) y *C. maxima* L. Osbeck (zamboa) (Swingle and Reece, 1967; Tanaka, 1977). Sin embargo, la contribución precisa de estas especies al grupo mandarino no es conocida. Por ello, en este trabajo se han empleado además de marcadores microsatélites (SSR), marcadores de inserción-delección (INDEL). Estos últimos son menos polimórficos que los marcadores microsatélites, presentan una mayor organización de la diversidad genética y parecen ser mejores marcadores filogenéticos para determinar la contribución de las especies ancestrales a la colección de mandarinos. El origen materno del citoplasma de las variedades estudiadas ha sido analizado mediante marcadores de tipo mitocondrial, siendo éste concordante con estudios previos realizados con marcadores cloroplásticos.

MATERIALES Y MÉTODOS

El material vegetal empleado consta de 84 variedades del banco de germoplasma del IVIA (formado mayoritariamente por variedades de origen americano y europeo, de aparición relativamente reciente) y 124 variedades del banco de germoplasma de Córcega (formado mayoritariamente por variedades ancestrales de origen asiático).

Se emplearon 50 marcadores moleculares nucleares SRRs (Simple Sequence Repeat) (Kijas et al., 1995; Luro et al., 2008, Froelicher et al., 2008) que están distribuidos a lo largo de todo el genoma, según el mapa genético consenso de Clementino x Chandler (obtenido por Patrick Ollitrault y colaboradores), lo cual hace que los resultados que se han obtenido representen la variabilidad genética global. Además se utilizaron 8 marcadores INDEL, desarrollados a partir de secuencias genómicas de genes implicados en la biosíntesis de metabolitos primarios y secundarios que determinan la calidad de los cítricos (flavonoides, azúcares, acidez y carotenos), como son: Chalcona isomerasa (CHI), Enzima málico (EMA), Fosfoenolpiruvato carboxilasa (PEPC), Transportador vacuolar citrato/H^+ (TRPA), Deoxixilulosa 5-fosfato sintasa (DXS), β-Caroteno hidroxilasa (Hy-b) y Fitoeno sintasa (PSY).

Para genotipar los SSRs e INDEL se empleó el Analizador Genético Automático CEQ^{TM}8000 de Beckman Coulter. Los resultados fueron analizados mediante diversas herramientas de análisis genético: DARwin (http://darwin.cirad.fr/darwin) para hacer análisis de grupo, GENEPOP 4.0 (http://genepop.curtin.edu.au/index.html) para determinar parámetros de genética poblacional y Structure version 2.2.3, (http://cbsuapps.tc.cornell.edu/structure) para representar la organización genética de la población estudiada, definiendo grupos y observando la contribución relativa de cada uno de ellos a los distintos genotipos de la población.

RESULTADOS Y DISCUSIÓN

Como se muestra en los datos obtenidos para la población de mandarino (Tabla 1), para los marcadores SSR la diversidad genética es elevada, así como el número medio de alelos por *locus* en comparación con lo observado con los marcadores INDEL. Esto es debido principalmente a su mayor polimorfismo (PIC). Los marcadores microsatélites parecen ser mejores para la diferenciación intraespecífica y los marcadores INDEL para la diferenciación interespecífica. Existen muchas variedades clasificadas como *C. reticulata* Blanco (según Tanaka) que se encuentran dispersas en la población, lo cual indica que podrían estar sujetas a una diferenciación mayor, incluso asignar variedades a otras especies. Con el programa Structure se diferencian entre 8 y 10 grupos dentro del germoplasma de mandarino, así como la proporción relativa de estos y de las especies ancestrales en cada variedad, siendo acorde con los resultados obtenidos con DARwin. Los valores muy bajos del coeficiente de endogamia (FIS) confirman la existencia de mezcla genética frecuente entre los diferentes grupos de mandarino.

Los datos de marcadores mitocondriales han permitido diferenciar tres orígenes maternales al nivel de los mandarinos (dos mitotipos de mandarino y uno de zamboa), confirmando la introgresión de zamboa en algunas variedades de mandarino. La alta variabilidad genética observada en el grupo mandarino y los distintos parámetros genéticos analizados, muestran una situación favorable para realizar estudios de genética de asociación entre caracteres genotípicos y fenotípicos. Además, se pretende establecer una colección base que represente la variabilidad global del grupo mandarino.

Referencias

Froelicher, Y. et al. 2008. Characterization of microsatellite markers in mandarin orange (*Citrus reticulata* Blanco). Molecular ecology resources 8: 119-122.

Kijas, J.M.H. et al. 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within Citrus and related species. Genome 38: 349-355

Luro et al. 2008. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. BMC Genomics 9: 287.

Swingle, WT. and Reece, PC. 1967. The botany of Citrus and its wild relatives. In: Reuter W, Webber HJ, Batchelor LD (eds) The Citrus Industry 1: 190-430.

Tanaka, T. 1977. Fundamental discussion of Citrus classification. Stud Citrologia 14: 1-6.

Tabla 1. Estadísticas población grupo mandarino.

	INDEL	SSRs
Nº medio de alelos/*locus*	3.38	8.22
Heterocigosidad esperada	0.14	0.61
Heterocigosidad observada	0.15	0.62
PIC(Polimorphic Information Content)	0.14	0.71
FIS (Coeficiente de endogamia)	-0.09	-0.01

[P2.23]

Nuclear and maternal phylogeny within Citrus and four related genera based on nuclear genes sequence SNPs and mitochondrial InDels

A. Garcia-Lor*[1], F. Curk[1,2], F. Luro[2], L. Navarro[1], P. Ollitrault[1,3]
[1]IVIA, Spain, [2]INRA, France, [3]CIRAD, France

Despite considerable morphological differentiation *Citrus*, *Fortunella*, *Poncirus*, *Microcitrus* and *Eremocitrus* genera are sexually compatible. Species of these genera are mainly diploid (2n=18). If the origin of cultivated *Citrus* from four basic taxa (*C. maxima, C. medica, C. reticulata* and *C. micrantha*) is now well documented, their phylogenetic relationships with *Citrus* wild species and related genera is still unclear. In the present work we analyse their nuclear and maternal phylogeny by using respectively SNPs on gene sequences and mitochondrial InDels.

A total of 7.15 kb were amplified by PCR from 11 genes (Table1) and sequenced (Sanger) for 33 genotypes. The varietal sample was composed of 7*C. reticulata*, 5 *C. maxima*, 5 *C. medica*, 4 papeda, 5 *Fortunella*, 3 *Poncirus*, 2*Microcitrus*, 1 *Eremocitrus. Severinia buxifolia* was used as outgroup. SNPs were mined using BioEdit and SeqMan softwares and phylogenetic analysis done in "http://phylemon.bioinfo.cipf.es" with different approaches (Phylip (v. 3.68), PhymlBest AIC Tree (v. 1.02b), PhyML (v. 3.00). For maternal phylogeny, 4 InDel markers developed by (Froelicher et al. 2011) have been used.

The average frequency per Kb of SNPs and InDels were respectively 59.88 and1.33 in coding region and 110.99 and 16.31 in non-coding ones. A total of 506SNP and 23 InDels were identified (Table1). Within *Citrus*, the papeda group was the most polymorphic species, with 185 polymorphisms, followed by *C. reticulata* (125), *C. maxima* (48), *and C. medica* (27).

A new mitotype was observed for *Microcitrus australasica* while two different mitotypes were identified for *Fortunella*. Nuclear and mitochondrial phylogenetic analysis reveal that *C. reticulata* and*Fortunella* form a consistent clade clearly differentiated from the clade includingthe other basic taxa of cultivated citrus (*C. maxima, C. medica* and *C. micrantha*).

Inclusion of more genes sequences is undergoing and will improve the resolution of the phylogenetic analysis.

Table 1. Statistics in the population studied.

Gene	CS	CDS	NCS	SCF	SNCF	ICF	INCF
Chalcone isomerase	652	206	446	53.40	170.40	0	17.94
Chalcone synthase	565	565	0	35.40	-	0	-
Flavonol Synthase	473	419	54	90.69	111.11	0	55.56
Flavonoid 3'-hydroxylase	613	569	44	70.30	45.45	0	0
Enzyme malique	428	128	300	54.69	86.67	7.81	13.33
Vacuolar citrate/H+ symporter	795	657	138	60.88	115.94	0	7.25
Malate dehydrogenase	712	712	0	39.33	-	0	-
Acid invertase	673	409	264	85.57	136.36	0	3.79
Lycopene β-cyclase	738	738	0	88.08	-	6.78	-
Lycopene β-cyclase	941	941	0	39.32	-	0	-
9-cis-epoxy hydroxy carotenoid dyoxygenase	560	560	0	41.07	-	0	-

(CS) Cleaned sequence (bp); (CDS) Coding sequence (bp); (NCS) Non-coding sequence (bp); (SCF) SNP frequency in coding region, x/Kb; (SNCF) SNP frequency in non-coding region; (ICF) InDel frequency in coding region, x/kb; (INCF) InDel frequency in non-coding region

S01P08

Nuclear phylogeny of Citrus and four related genera

Garcia-Lor A. [1], Curk F.[2], Snoussi H.[3], Morillon R.[4], Ancillo G.[1], Luro F.[2], Navarro L.[1], and Ollitrault P.[4]

[1]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; [2]INRA, UR1103 Génétique et Ecophysiologie de la Qualité des Agrumes (INRA UR GEQA), DGAP, France; [3]Institut National de la Recherche Agronomique de Tunisie (INRAT), Laboratoire d'Horticulture, Tunisie; and [4]Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BIOS, France. angarcia@ivia.es

Despite considerable differences in morphology, the genera representing "true citrus fruit trees" are sexually compatible, but their phylogenetic relationships remain unclear. Most of the important commercial species of *Citrus* are believed to be of interspecific origin. By studying SNP and InDel polymorphisms of 27 nuclear genes on 45 genotypes of Citrus and related taxa, the average molecular differentiation between species was estimated, and the phylogenetic relationship between "true citrus fruit trees" was clarified. A total of 16238 bp of DNA was sequenced for each genotype, and 1097 SNPs and 50 InDels were identified. Nuclear phylogenetic analysis revealed that *Citrus reticulata* and *Fortunella* form a clade clearly differentiated from the other two basic taxa of cultivated citrus (*Citrus maxima*, *Citrus medica*). A few genes displayed positive selection patterns within or between species, but most of them displayed neutral patterns. The phylogenetic inheritance patterns of the analysed genes were inferred for commercial *Citrus* species. The SNPs and InDels identified are potentially very useful for the analysis of interspecific genetic structures. The nuclear phylogeny of *Citrus* and its sexually compatible relatives was consistent with their geographic origin. The positive selection observed for a few genes will orient further work to analyze the molecular basis of the variability of the associated traits. This study presents new insights into the origin of *Citrus sinensis*.

S01P09

New insights on limes and lemons origin from targeted nuclear gene sequencing and cytoplasmic markers genotyping

Curk F.[1], Garcia-Lor A.[2], Snoussi H.[3], Froelicher Y.[4], Ancillo G.[2], Navarro L.[2], and Ollitrault P.[4]

[1]INRA, UR1103 Génétique et Ecophysiologie de la Qualité des Agrumes (INRA UR GEQA), DGAP, France; [2]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; [3]INRAT (Institut National de la Recherche Agronomique de Tunisie), Tunisia; and [4]Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD), BIOS, France. curk@corse.inra.fr

It is believed that Citrus medica, Citrus maxima, Citrus reticulata and Citrus micrantha have generated all cultivated Citrus species. Depending on the classification, lemons and limes are classified either into two species, Citrus limon and Citrus aurantifolia (Swingle and Reece) or into more than 30 (Tanaka). In order to study the molecular phylogeny of this Citrus group, we analyzed 20 targeted sequenced nuclear genes and used 3 mitochondrial and 3 chloroplastic markers for 21 lemons and limes compared with representatives of the 4 basic taxa. We observed 3 main groups, each one derived from direct interspecific hybridizations: (1) the Mexican lime group (C. aurantifolia), including Citrus macrophylla, arising from hybridization between papeda (C. micrantha) and citron (C. medica); (2) the yellow lemon group (C. limon) that are hybrids between sour orange (Citrus aurantium, which is believed to be a hybrid betweenC. Maxima and C. reticulata) and citron; and (3) a rootstock lemon/lime group (Rough lemon and Rangpur lime) that are hybrids between the acid small mandarin group and citron. We also identified different probable backcrosses and genotypes with more complex origins. None of the analyzed limes and lemons shared the C. medica cytoplasm, while this taxon is the common nuclear contributor of all limes and lemons. Limes and lemons appear to be a very complex citrus varietal group with the contribution of the 4 basic taxa. Neither the Swingle and Reece nor the Tanaka classifications fit with the genetic evidence.

S02P05

Comparative values of SSRs, SNPs and InDels for citrus genetic diversity analysis

Ollitrault P.[1], Garcia-Lor A.[2], Terol J.3, Curk F.[4], Ollitrault F.[2], Talon M.[3], and Navarro L.[2]

[1]Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BIOS, France; [2]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; [3]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Genómica, Spain; and [4]Institut National de la Recherche Agronomique (INRA), GAP, France. patrick.ollitrault @cirad.fr

SSRs have long been considered as almost ideal markers for genetic diversity analysis. With the increasing availability of sequencing data, SNPs and InDels become major classes of codominant markers with genome wide coverage. We have analyzed the respective values of SSRs, InDels, and SNPs for intra and interspecific *Citrus* genetic diversity analysis. Moreover, we have compared the diversity structure revealed by markers mined in a single heterozygous genotype (the clementine) and markers mined in a large interspecific survey. A random set of 25 markers was selected for each marker class to genotype 48 citrus accessions. SSRs were the most polymorphic markers at the intraspecific level allowing complete varietal differentiation within basic taxa (*Citrus reticulata*, *Citrus maxima*, *Citrus medica*). However, SSRs gave the lowest values for interspecific differentiation, followed by SNPs and InDels, that displayed low intraspecific variability but high interspecific differentiation. A clear effect of the discovery panel was observed for SNPs and InDels. The ascertainment biases associated with the clementine heterozygosity mining resulted mainly in an over estimation of within C. reticulata diversity and an underestimation of the interspecific differentiation. Therefore SSRs are very useful for intraspecific structure analysis while SNPs and InDels mined in large discovery panel will be more powerful to decipher the interspecific mosaic structure of secondary cultivated species.

[P2.21]

Multilocus snps analysis allows phylogenetic assignation of DNA fragments to decipher the interspecific mosaic genome structure of cultivated *Citrus*

F. Curk*[1,2], G. Ancillo[2], A. Garcia-Lor[2], F. Luro[1], P. Ollitrault3,2, L. Navarro[2]
[1]*INRA, France,* [2]*IVIA, Spain,* [3]*CIRAD, France*

All current studies seem to support the theory that four basic taxa (*C. medica*, *C. maxima*, *C. reticulata* and *C. micrantha*) have generated all cultivated *Citrus* species. It is supposed that the genomes of most of the modern *Citrus* cultivars, vegetatively propagated, are interspecific mosaic of large DNA fragments issued from a limited number of inter-specific meiosis. In the present work we analyse how multilocus study of closely linked SNPs allows a phylogenetic assignation of DNA fragments of the main cultivated species.

Genomic fragments of 25 genes dispersed in the different chromosomes covering more than 12,5 Kb were amplified by PCR and sequenced (Sanger) for 24 accessions representative of 10 species. Moreover we checked the potential of parallel pyrosequencing (454 Roche) for direct multilocus haplotyping of heterozygous genotypes. Amplified fragments from 7 genes in 8 genotypes were obtained by using an original new method based on universal primers. *C. clementina* (Clementine) was used as model for secondary species. Citrus reticulata was the most polymorph basic taxa with an average of 4.2 SNPs/kb. The average differentiation between the basic taxa was about 20 SNPs/kb. For each amplified gene fragment, this polymorphism was enough for unambiguous multilocus differentiation of the basic species and assignation of a phylogenetic origin for the secondary species. A preliminary reconstitution of phylogenetic structure of chromosome 3 is proposed for sweet orange, sour orange, grapefruit, lemon and lime. Consensus haplotype sequences were successfully obtained from 454 sequencing with genotype sequence in total agreement with Sanger control. Each haplotype sequence of Clementine was univocally assigned to one of the haplotype clusters of the basic taxa. Phylogenetic origin of specific DNA fragments can be assigned from multilocus analysis of closely linked SNPs. Multilocus haplotyping by parallel sequencing of individual DNA molecule will be a very powerful tool to decipher the interspecific mosaic genome structure of cultivated citrus.

Keywords: snp, mosaic sturcutre genome, phylogeny, Citrus

S01P11

Analysis of genetic diversity in Tunisian citrus rootstocks

Snoussi H.[1], Duval M.F.[2], Garcia-Lor A.[3], Perrier X.[2], Jacquemoud-Collet J.C.[2], Navarro L.[3], and Ollitrault P.[2]

[1]Tunisian National Agronomic Research Institute (INRAT), Horticultural Laboratory, Tunisia; [2]International Center for of Agricultural Research for Development (CIRAD), Department BIOS. TGU. AGAP, France; and [3]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain. hagersnoussi@gmail.com

Breeding and selection of new citrus rootstocks are nowadays of the utmost importance in the Mediterranean Basin because the citrus industry faces increasing biotic and abiotic constraints. In Tunisia, citrus contributes significantly to the national economy, and its extension is favored by natural conditions and economic considerations. Sour orange, the most widespread traditional rootstock of the Mediterranean area, is also the main one in Tunisia. In addition to sour orange, other citrus rootstocks well adapted to local environmental conditions are traditionally used and should be important genetic resources for breeding. Prior to initiation of any breeding program, the exploration of Tunisian citrus rootstock diversity was a priority. Two hundred and one local accessions belonging to four facultative apomictic species (*Citrus aurantium*, sour orange; *Citrus sinensis*, sweet orange; *Citrus limon*, lemon; and *Citrus aurantifolia*, lime) were collected and genotyped using 20 nuclear SSR markers and four InDel mitochondrial markers. Sixteen distinct Multi-locus genotypes (MLGs) were identified and compared to references from French and Spanish collections. The differentiation of the four varietal groups was well-marked. Each group displayed a relatively high allelic diversity, primarily due to very high heterozygosity. The Tunisian citrus rootstock genetic diversity is predominantly due to high heterozygosity and differentiation between the four varietal groups. The phenotypic diversity within the varietal groups has resulted from multiple introductions, somatic mutations and rare sexual recombination events. This diversity study enabled the identification of a core sample of accessions for further physiological and agronomic evaluations. These core accessions will be integrated into citrus rootstock breeding programs for the Mediterranean Basin.

Reference SNPs, SSRs and InDels *C. clementina* linkage map

Patrick Ollitrault[1,2], Javier Terol[3], Chunxian Chen[4] Claire T. Federici[5],Samia Lotfy [6,1], Isabelle Hippolyte[1], Frédérique Ollitrault[2], Aurélie Bérard[7], Gilles Costantino[8], Lisa Mu[5], Yildiz Kacar[9], Jose Cuenca[2], Andres Garcia[2], Yann Froelicher[1], Anne Boland[10], Claire Billot[11], Luis Navarro[2], François Luro[9], Mikeal L. Roose[5], Frederick G. Gmitter[4], Manuel Talon[3] and Dominique Brunel[7]

1 CIRAD, UPR 75, Avenue Agropolis, TA A-75/02, 34398 Montpellier, Cedex 5, France

2 Centro de Protection Vegetal y Biotechnologia, IVIA, Apartado Oficial ,46113 Moncada (Valencia), Spain

3 Centro de genomica, IVIA, Apartado Oficial , 46113 Moncada (Valencia), Spain

4 Citrus Research and Education Center, University of Florida, Lake Alfred, FL 33850, USA

5 Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

6 Institut National de la Recherche Agronomique, BP 293, 14 000 Kénitra, Morocco

7 INRA, UR EPGV, 2 rue Gaston Cremieux, 91057, Evry, France

8 INRA, UR GEQA San Giuliano 20230 San Nicolao, France

9 Department of Horticulture, Faculty of Agriculture, University of Çukurova, 01330, Adana, Turkey

10 CNG, 2 rue Gaston Cremieux, 91057, Evry, France

11 CIRAD, UMR DAP, Avenue Agropolis - TA A-96 / 03, 34398 Montpellier Cedex 5, France

Corresponding author: patrick.ollitrault@cirad.fr

An haploid *C. clementina* was chosen by the International Citrus Genomic Consortium (ICGC) to establish the reference whole Citrus genome sequence. The implementation of a dense clementine linkage map was part of the objectives of this global collaborative project. Two inter-specific populations between *C. clementina* and *C. maxima* were used for this purpose. 156 hybrids of Nules Clementine x Pink pummelo and 200 hybrids of Chandler pummelo x Nules clementine were genotyped with 1003 markers. 306 were SSRs markers (66 from genomic bank, 207 from ESTs and 33 from clementine BACEnd sequences –BES-), 34 were Indels markers mined from BES and 663 SNPs mined from Clementine BES or identified by candidate gene sequencing. 901 markers were successfully mapped in the 9 clementine linkage groups. Important segregation distortion were observed for clementine when used as male parent while it followed Mendelian segregation for most markers when used as female parent. However marker order was mostly conserved between the male and female maps; thus, data of the two populations were joined to establish the reference Clementine genetic map. Total Clementine linkage map size is 1250 cM with linkage groups from 105 cM until 210 cM. This map is strongly anchored on a large diploid clementine BAC library resource. It is a powerful tool for Citrus genetic and supports the alignment of the haploid Clementine whole genome sequence in the framework of the collaborative project of the ICGC.