CrossMark

ORIGINAL ARTICLE

Genetic diversity and population structure analysis of mandarin germplasm by nuclear, chloroplastic and mitochondrial markers

Andres Garcia-Lor[1] · François Luro[2] · Patrick Ollitrault[1,3] · Luis Navarro[1]

© Springer-Verlag Berlin Heidelberg 2015

Abstract The mandarin horticultural varietal group (a basic taxon of the cultivated citrus (*Citrus reticulata*)) is highly polymorphic. It includes also 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 (simple sequence repeat (SSR), indel, single nucleotide polymorphism (SNP)), chloroplastic and mitochondrial markers; (2) to evaluate genetic diversity and detect redundancies; (3) to determine the possible presence of citrus ancestral genome introgressions into the mandarin genome; and (4) to determine the genetic structure within the mandarin group. Fifty microsatellites (SSRs), 24 insertion-deletions (indels), 67 SNPs, 8 chloroplastic SSRs (cpDNA) and 4 mitochondrial (mtDNA) indel markers were analysed for 191 genotypes, including the 4 main citrus ancestral species (*C. reticulata*, *C. maxima*, *C. medica* and *C. micrantha*) and *Fortunella*. *C. maxima* was the main genome introgressed in the mandarin germplasm. Seven clusters were revealed by Structure analysis at the nuclear level (N) within the mandarin germplasm. At least four of these clusters showed a clear introgression from other ancestral species. Moreover, most of the mandarins appeared to be complex mixtures of these groups. The maternal indel analysis (mtDNA and cpDNA) revealed ten cytotypes in which mandarins were represented in seven of them. This work provides new insights into the organisation of the mandarin germplasm and its structure at the nuclear and cytoplasmic levels and will be useful to design more efficient breeding programmes and management of citrus germplasm collections.

Communicated by W.-W Guo

This article is part of the Topical Collection on *Population Structure*

✉ Luis Navarro
lnavarro.ivia@gmail.com

[1] Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada, Valencia, Spain

[2] UMR AGAP Institut National de la Recherche Agronomique (INRA Corse), 20230 San Giuliano, France

[3] UMR AGAP, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Station de Roujol, 97170 Petit-Bourg, Guadeloupe

Introduction

Citrus is the most important fruit crop in the world, with a production of over 128 million tons and a cultivated area of 9.2 million hectares (FAOSTAT 2013). Among the commercial citrus fruits, mandarins are the second most important group in the fresh fruit market worldwide. '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, 2013a). Moreover, a recent phylogenetic study (Garcia-Lor et al. 2013a) revealed a close relationship between the genus *Fortunella* and the mandarin group. Mandarin germplasm was classified as *Citrus 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 the 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. ('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* (satsumas), *C. reticulata* ('Ponkan', 'Dancy', clementines), *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 (Nicolosi et al. 2000; Barret and Rhodes 1976; 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 (Barkley et al. 2006; Koehler-Santos et al. 2003; Tapia Campos et al. 2005). 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 facultative apomictic group, including the recently published introgression of ancestral genomes, like *C. maxima*, in some mandarins (Wu et al. 2014; Curk et al. 2014, 2015). 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 included 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 (mainly Asia in mandarins), 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 the detection of misidentifications and redundancies (Krueger and Roose 2003).

To clarify to which genetic group we refer by using the term 'mandarin' in the text of this work, we have used the following nomenclature: *C. reticulata* for 'mandarin' as a true species (one of the four ancestors of the cultivated citrus); *C. reticulata* (Sw) for 'mandarin' according to the Swingle classification; *C. reticulata* (Tan) for one of the 17 'mandarin' species represented in this work according to the Tanaka classification; and 'mandarin-like' for genotypes that are phenotypically similar to mandarins.

The goals of this work were (1) to characterise the mandarin germplasm using nuclear (simple sequence repeats (SSRs), indels, single nucleotide polymorphisms (SNPs)), chloroplastic SSRs (cpDNA) and mitochondrial indel markers (mtDNA); (2) to evaluate its genetic diversity and detect redundancies; (3) to detect introgressions of other citrus ancestral taxa into the mandarin germplasm; and (4) to determine the genetic structure within the mandarin group.

In this study, we have observed the introgression of several ancestral genomes in the mandarin genotypes considered previously as pure mandarins and also that many mandarins appear to have a very complex genetic organisation (mixture of mandarin genomes).

Materials and methods

Analysed germplasm

One hundred ninety-one genotypes were studied to determine their nuclear diversity. Throughout the text, these genotypes will be referred to by identification number (ID), shown in Online Resource 1. Genotype classification was performed according to the Swingle and Reece (1967) and Tanaka (1954) systems. A summary of the genotypes used is shown in Table 1. Plant material for the analysis was collected from the germplasm collections of the *Instituto Valenciano de Investigaciones Agrarias* (IVIA, Valencia, Spain), mainly obtained from American and Mediterranean sources, and the *Station de Recherches Agronomiques* (CIRAD-INRA, Corsica, France), which include many genotypes of Asiatic origin (China, Japan, Philippines, India, Vietnam…). The databases of both collections are based on the Tanaka system of classification. These genotypes belong to the four ancestral species (26 *C. reticulata* (Sw, mandarins), *C. indica* and *C. tachibana*),

Table 1 Summary of genotypes employed in the study

Swingle system	Tanaka system	NG/S	NGSA
C. reticulata hybrid	*C. amblycarpa*	2	2
C. reticulata	*C. deliciosa*	11	1
C. reticulata	*C. daoxianensis*	1	0
C. reticulata	*C. depressa*	5	2
C. reticulata	*C. erythrosa*	2	1
C. indica	*C. indica*	1	1
C. hystrix	*C. hystrix*	1	1
C. reticulata	*C. kinokuni*	4	2
C. maxima	*C. maxima*	10	10
C. medica	*C. medica*	6	6
C. micrantha	*C. micrantha*	1	1
C. reticulata	*C. nobilis*	6	2
C. reticulata	*C. paratangerina*	2	2
C. reticulata	*C. reshni*	1	1
C. reticulata	*C. reticulata*	49	3
C. reticulata	*C. suavissima*	1	1
C. reticulata	*C. succosa*	1	1
C. reticulata	*C. suhuiensis*	7	2
C. reticulata	*C. sunki*	3	2
C. tachibana	*C. tachibana*	1	1
C. reticulata	*C. tangerina*	11	2
C. sinensis	*C. tankan*	1	0
C. reticulata	*C. temple*	2	0
C. reticulata	*C. unshiu*	8	2
Fortunella	*Fortunella*	4	4
C. reticulata	Hybrid mandarin	28	0
C. reticulata	Tangelo	4	0
C. reticulata	Tangor	12	0
?	Bintangor	1	0
C. reticulata	*C. clementina*	3	0
?	Unknown	2	0

Genotypes' classification based on Swingle and their classification within our databases based on the Tanaka system

NG/S number of genotypes per species, *NGSA* number of genotypes from each species included within an ancestral population

ten *C. maxima* (pummelos), six *C. medica* (citrons), two *Papeda* (*C. hystrix* D.C. and *C. micrantha*) and four *Fortunella* (kumquats: *F. crassifolia* Swing., *F. hindsii* (Champ.) Swing., *F. japonica* (Thunb.) Swing. and *F. margarita* (Lour.) Swing.). The 26 mandarin genotypes considered as *C. reticulata* by Swingle and Reece (1967) were considered by Tanaka (1977) as 15 species. The other genotypes (141 '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 (Poir.) Tenore was added as an out-group for neighbour-joining analysis.

For the maternal phylogeny, besides the ancestral species and interspecific hybrids detailed before, eight extra genotypes (secondary species) were analysed (two *C. sinensis* (L.) Osb., two *C. aurantium* L., two *C. paradisi* Macf., one *C. aurantifolia* (Christm.) Swing. and one *C. limon* Osb.).

Genotyping

Fifty SSR markers (Kijas et al. 1997; Froelicher et al. 2008; Luro et al. 2008; Aleza et al. 2011; Cuenca et al. 2011; Kamiri et al. 2011; Garcia-Lor et al. 2012, 2013a, b) located along the nine linkage groups of the reference genetic map of clementine (Ollitrault et al. 2012b), 24 indel markers identified in a discovery panel representative of the genus *Citrus* (Garcia-Lor et al. 2012, 2013a) and 67 SNP markers mined in 27 nuclear genes (Garcia-Lor et al. 2013a) and in clementine BAC-ends (Ollitrault et al. 2012a) were used (Online Resource 2). To assess the maternal origin of the mandarin germplasm, eight chloroplastic SSR markers (Cheng et al. 2005) (ccmp1, ccmp2, ccmp4, ccmp5, ccmp6, NTCP7, NTCP9 and NTCP28) and four mitochondrial indel markers (Froelicher et al. 2011) (*nad2*, *nad5*, *nad7*, *rrn5/rrn18*) were used. After the initial analysis, nad5 (no polymorphisms were found in our population) and three chloroplastic markers (NTCP28, ccmp1 and ccmp4) were discarded. In the latter, due to bad amplifications or due to a difference in polymorphisms of just one base that could be more confusing than clarifying.

For SSR and indel markers, 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.

SNPs were genotyped by competitive allele-specific PCR as described by Garcia-Lor et al. (2013b).

Data analysis

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

Population structure was inferred with the programme 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 programme 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 programme 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 programme PowerMarker, v. 3.25 (Liu and Muse 2005).

Results

Nuclear genetic diversity parameters

Genetic diversity statistics were calculated for each SSR, indel and SNP marker for the 191 genotypes analysed (Online Resource 3).

Using the SSR markers, we detected 529 alleles. Allele numbers varied between 4 (MEST107) and 18 (MEST192). The average number of alleles and the H_e (expected heterozygosity) value per locus were 10.6±0.508 and 0.63±0.021, respectively. The whole population had an observed heterozygosity (H_o) of 0.60±0.024. F_w (Wright fixation index) values varied from −0.40 (CAC23) to 0.48 (CimCrCIR01D11). The average F_w value over all SSR loci was close to 0 (0.04±0.024).

We detected a total of 74 alleles with the indel markers. Allele number per locus ranged from 2 (10 markers) to 7 (IDDFR), with an average of 3.1±0.182. The H_e values ranged from 0.01 (IDINVA2) to 0.66 (IDDFR), with a median value of 0.20. F_w values varied from −0.52 (IDF'3H) to 1.00 (IDINVA1). The overall H_o and F_w values among all loci were 0.19±0.033 and 0.29±0.067, respectively.

All SNP markers were biallelic (134 alleles identified). The whole population had an observed (H_o) and expected (H_e) heterozygosity of 0.27±0.023 and 0.26±0.016, respectively. The minimum H_o value was 0.01 (4 markers) and the maximum 0.52 (CCC1-M85). The overall F_w value among all loci was 0.12±0.059.

Genetic population statistics within the whole population (AG, [ID 1–191]), all 'mandarin-like' genotypes of unknown or supposed hybrid origin (AM, [ID 51–191]) and the 28 genotypes selected from all Tanaka species represented in our collections (MT, [ID 1–28]) (Online Resource 1) are summarised in Table 2. Gene diversity (GD) and the H_o values were higher among SSR markers than SNP and indel markers, reflecting the higher maximum allele frequencies (MAF) of the indel and SNP than SSR. 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=10.58> AM=6.84>MT=6.76; indels AG=3.08>AM=2.25>MT= 2.02) and for the SNPs (AG=2>AM=1.90>MT=1.79). The GD was higher in AG than in AM or MT for both kinds of markers. For AG, H_o was slightly lower than H_e, leading to slightly positive F_w for SSR, indel and SNP markers. In AM and MT, H_o values were higher than H_e, providing negative F_w values for both kinds of markers. In AM, 13 loci did not show polymorphism and in MT 14.

Rare alleles

The 'mandarin-like' population included 20 genotypes with unique alleles (Online Resource 4), ranging from 1 (11 genotypes) to 10 ('Nicaragua'; [ID 112]) unique alleles per genotype.

Classifications by NJ analysis

For the whole data set (SSR, indel and SNP markers), NJ analysis (Fig. 1) 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 the data from SSR, indel and SNP markers revealed high intraspecific diversity in the mandarin group, which was not well resolved (low bootstrap support in many branches). From the whole data set, 22 genotypes were reduced to 8 multilocus genotypes (MLGs; Online Resource 5). Some of these were originated by budsport mutations (different cultivars of *C. unshiu* or *C. clementina*), and others are possible redundant genotypes collected and named differently in different locations.

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

The SSR, indel and SNP 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 500, 000, and MCMC (Markov Chain Monte Carlo) repetitions were set to 1,000,000; 10 runs of Structure with $K=5$ (5 populations assumed) were performed. These populations were as follows: mandarin (Sw), *C. indica* and *C. tachibana* (28 samples, representing 17 Tanaka species), pummelo (10 samples), citron (6 samples), papeda (2 samples) and kumquat (4 samples). The other samples analysed (141) were assumed to have been derived from these ancestral populations (Online

Table 2 Genetic population statistics within the whole population, all 'mandarin-like' genotypes and Tanaka mandarin species

Groups	N	MAF		A		GD		H_o		H_e		F_w	
AG SSR	191	0.471	±0.022	10.58	±0.51	0.666	±0.020	0.603	±0.024	0.628	±0.021	0.039	±0.024
AG indel		0.833	±0.028	3.08	±0.18	0.227	±0.032	0.192	±0.033	0.199	±0.027	0.291	±0.067
AG SNP		0.758	±0.020	2.00	–	0.326	±0.022	0.268	±0.023	0.261	±0.016	0.118	±0.059
AM SSR	166	0.541	±0.026	6.84	±0.42	0.583	±0.025	0.594	±0.028	0.538	±0.026	−0.121	±0.032
AM indel		0.855	±0.029	2.25	±0.19	0.190	±0.036	0.210	±0.040	0.160	±0.030	−0.167	±0.044
AM SNP		0.798	±0.023	1.90	±0.06	0.276	±0.027	0.291	±0.030	0.223	±0.021	−0.249	±0.036
MT SSR	28	0.521	±0.029	6.76	±0.35	0.606	±0.029	0.606	±0.031	0.575	±0.028	−0.076	±0.032
MT indel		0.858	±0.028	2.02	±0.18	0.192	±0.035	0.202	±0.037	0.164	±0.029	−0.132	±0.057
MT SNP		0.835	±0.036	1.79	±0.11	0.230	±0.042	0.216	±0.043	0.188	±0.032	−0.078	±0.072

AG all genotypes of the population, *AM* all 'mandarin-like' genotypes, *MT* genotypes selected from all Tanaka species represented in our collections, *N* number of genotypes included in each group, *MAF* maximum allele frequency, *A* number of alleles, *GD* gene diversity, H_o heterozygosity observed, H_e heterozygosity expected, F_w Wright fixation index

Resource 1). Assuming an admixture model between the four ancestral citrus species and *Fortunella* (Online Resource 1, genotypes 1–50), the relative proportion of these genomes in

the mandarin group and recent hybrids was inferred using Structure, v. 2.3.3 (Fig. 2), with the complete data set (SSR+indel+SNP). Contributions lower than 5 % were not
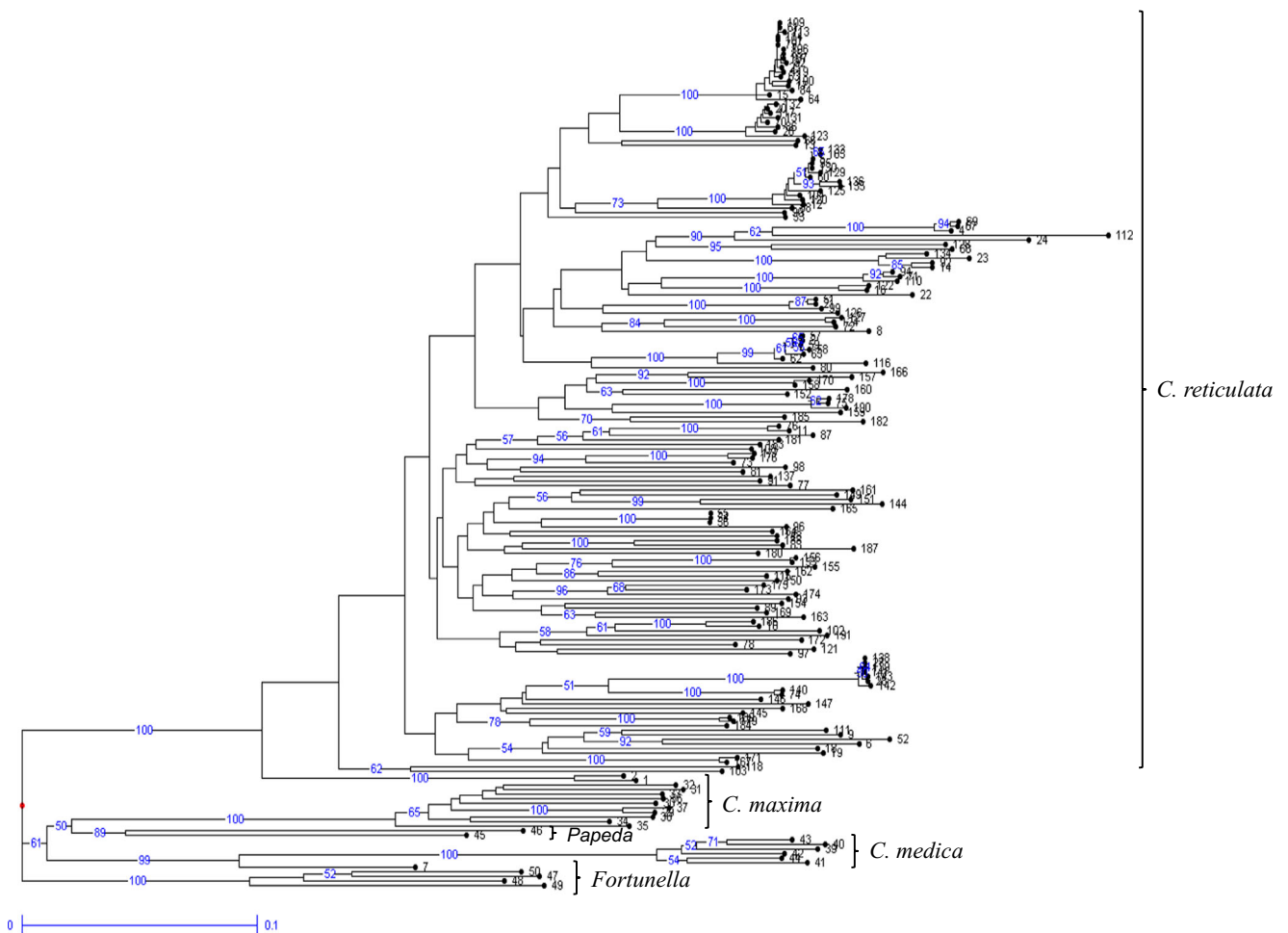

Fig. 1 NJ analyses with the SSR, indel and SNP data for the entire data set and all the 'mandarin-like' genotypes (1000 bootstraps). *Numbers* represented has correspondence with the ID numbers in Online Resource 1. Bootstrap values over 50 are represented. Entire data set (191 genotypes) representing the four ancestral *Citrus* species (*C. reticulata, C. maxima, C. medica, Papeda*) and *Fortunella*

considered (since they may be artefacts or due to homoplasy in the SSR markers, Barkley et al. (2009)).

Thirteen of the 50 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 (~55 %), with the remainder (~45 %) from *C. reticulata*. *Citrus deliciosa* [ID 3] had a 4 % contribution from other ancestors, *C. depressa* [ID 4] had introgression from *Fortunella* (~15 %). *C. indica* Tan. [ID 7] has a tri-hybrid genome origin according to the Structure analysis (25 % from *C. reticulata*, 45 % from *C. medica* and 30 % from *Papeda*). The two *C. nobilis* [ID 10 and ID 11] had around 15 % introgression from *C. maxima*. *Citrus suavissima* Hort. ex Tan. [ID 18] and *C. succosa* Hort. ex Tan. [ID 19] presented introgression from *C. maxima* (~7 %). *Citrus tachibana* [ID 24] appeared to have introgression from *Fortunella* (~19 %). The two *C. unshiu* [ID 27 and ID 28] had introgression from the *C. maxima* genome (~14 %). One ancestral *Papeda*, *C. hystrix* DC., had 6 % contribution from *Fortunella*.

The contribution of mandarin to the genomes of the 141 'mandarin-like' genotypes that were not included in any of the five pre-assumed populations (Online Resource 1, genotypes ID 51/ID 191) was on average ~87 %. Pummelo contributed on average 10 %, and papeda, kumquat and citron contributions were lower than 5 % (Fig. 3). Contributions in individual genotypes lower than 5 % were not considered for the calculations.

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 141 genotypes analysed with no assumed population had at least 60 % contribution from *C. reticulata*. Seventy-seven genotypes had a *C. maxima* contribution of at least 5 %, with a maximum of 50 %. *Papeda* contributed 5–10 % to 11 genotypes, *C. medica* contributed 10–20 % to only one genotype and *Fortunella* contributed 5–20 % to five genotypes.

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
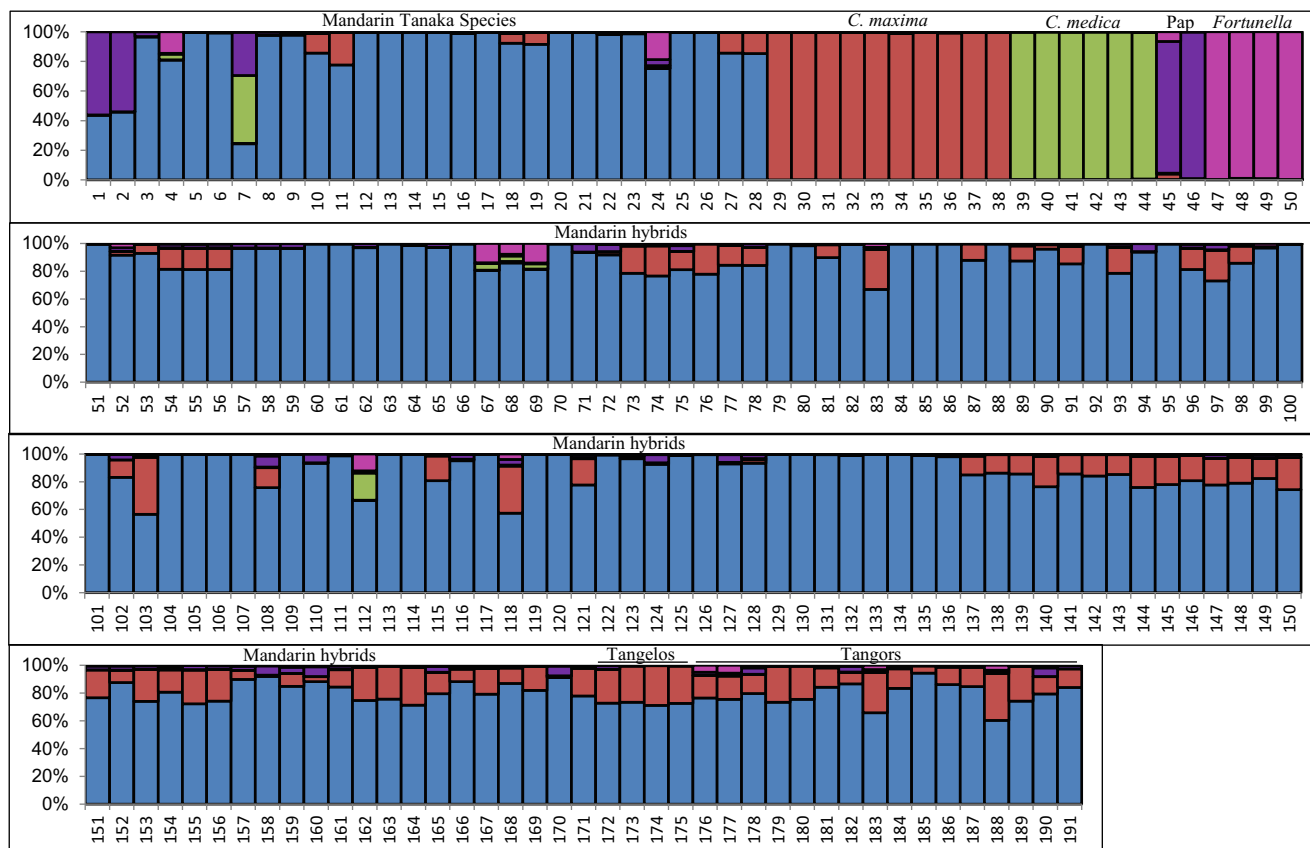

Fig. 2 Structure analysis of 191 genotypes representing the four ancestral *Citrus* species (*C. reticulata*, *C. maxima*, *C. medica*, *Papeda*) and *Fortunella*. Dark blue, *C. reticulata* (Sw.) (1); dark red, *C. maxima* (2); green, *C. medica* (3); purple, *Papeda* (4); pink, *Fortunella* (5). Genotypes 51–191 are genotypes without assigned populations. *Numbers'* correspondence in Online Resource 1
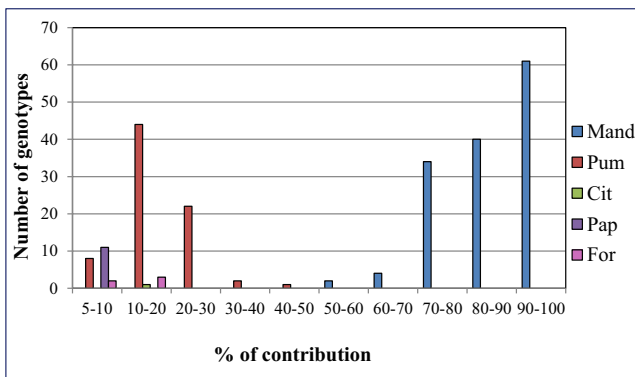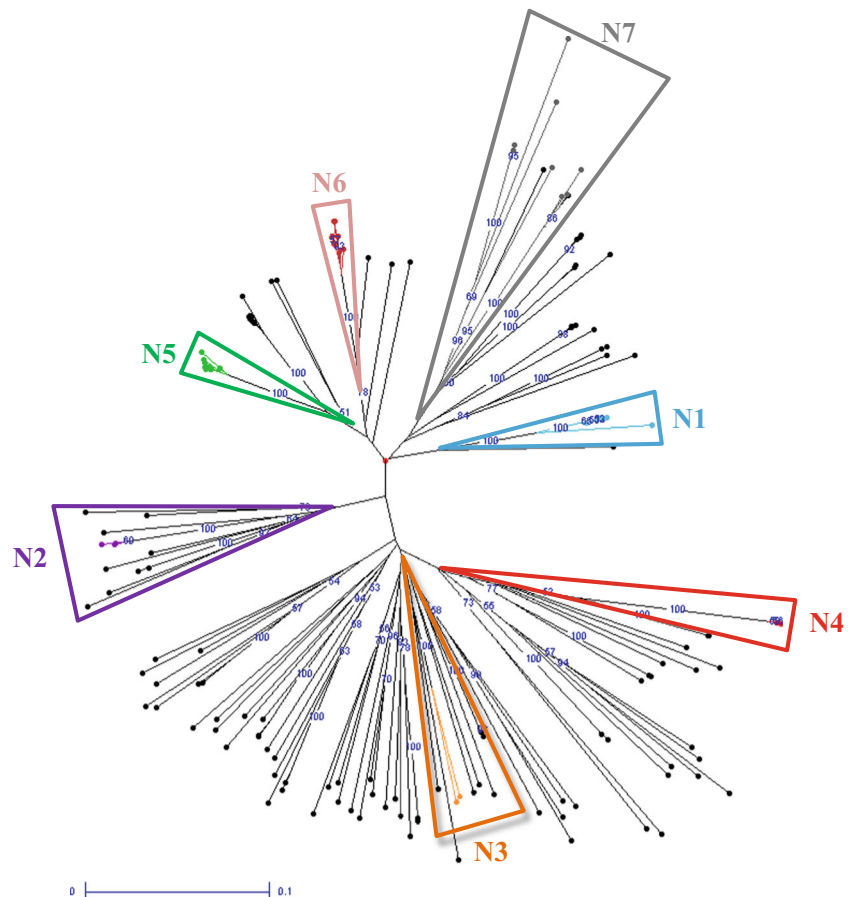
Fig. 3 Contributions of the ancestral genomes (mandarin, pummelo, citron, papeda) and kumquat to the 'mandarin-like' genotypes under study. (*Mand*, *blue*) mandarin, (*Pum*, *dark red*) pummelo, (*Cit*, *green*) citron, (*Pap*, *purple*) papeda, (*For*, *pink*) kumquat

ΔK, 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 ΔK were the main values used to choose the optimal *K* value of the population. One analysis was performed to obtain the correct number of groups within the mandarin germplasm.

One hundred sixty-six genotypes were selected from the first Structure analysis (191 genotypes, *K*=5, Fig. 2), discarding the genotypes of the four ancestral taxa and three of the mandarin genotypes considered by Tanaka as species, but that showed strong introgression from other ancestors (the two *C. amblycarpa* [ID 1 and ID 2] and *C. indica* [ID 7]). The Structure analysis was performed with no population assignation. The optimal ΔK was 7 (Online Resource 6).

Structure analysis was compared with a NJ tree (Fig. 4) to validate the clustering. Genotypes included in each of the seven mandarin groups identified are presented in Table 3 in relation to the Tanaka classification. Nuclear group 1 (N1) included seven genotypes (all *C. deliciosa*); nuclear group 2 (N2) included three genotypes, two *C. nobilis* and one unknown; nuclear group 3 (N3) was formed by two genotypes of unknown origin [ID 83 and ID 188]; nuclear group 4 (N4) included seven genotypes (all *C. unshiu*); nuclear group 5 (N5) included 18 genotypes (94.44 % *C. reticulata*); nuclear group 6 (N6) included 13 genotypes (mainly *C. tangerina* and *C. reticulata*); and nuclear group 7 (N7) included ten genotypes (small mandarins, like *C. depressa*, *C. tachibana*,

Fig. 4 DARwin analysis comparing the Structure populations found to the groups observed with DARwin within the mandarin germplasm (166 genotypes). Bootstrap values over 50 are represented. One thousand resamplings were performed. Seven mandarin groups were identified at the nuclear level. (*N*) Nuclear group. *N1*, *light blue* (all *C. deliciosa*); *N2*, *purple* (*C. nobilis*); *N3*, *orange* (mandarin hybrids); *N4*, *red* (all *C. unshiu*); *N5*, *green* (*C. reticulata*); *N6*, *light red* (*C. tangerina* and others); *N7*, *grey* (*C. depressa*, *C. reshni*, *C. sunki*, *C. tachibana*)

Table 3 Parental mandarin groups identified at the nuclear level within the mandarin germplasm

Group	No. of genotypes	Tanaka species	Percent
N1	7	*C. deliciosa*	100.00
N2	2	*C. nobilis*	66.67
	1	Unknown	33.33
N3	1	*C. reticulata*	50.00
	1	*C.reticulata × C.sinensis*	50.00
N4	7	*C. unshiu*	100.00
N5	17	*C. reticulata*	94.44
	1	*C. tangerina*	5.56
N6	5	*C. tangerina*	38.46
	4	*C. reticulata*	36.36
	1	*C. deliciosa*	9.09
	1	*C. depressa*	9.09
	1	*C. paratangerina*	9.09
	1	*C. suhuiensis*	9.09
N7	2	*C. reticulata*	20.00
	4	*C. depressa*	40.00
	1	*C. reshni*	10.00
	1	*C. tachibana*	10.00
	1	*C. tangerina*	10.00
	1	*C. sunki*	10.00

Analysis based on the Structure and NJ tree analyses. Genotypes included in each group are compared with the Tanaka classification

N nuclear group

C. reshni or *C. sunki*). The genotypes included in N3 exhibited a high degree of mutual similarity and shared a high percentage of *C. sinensis* molecular marker data (ID 83 85.33 % and ID 188 96 %), and also exhibited high heterozygosity. Therefore, they are very probably interspecific hybrids, similar to sweet orange.

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

The contributions of the seven mandarin groups identified within the mandarin germplasm into the other mandarins under study is summarised in Fig. 5 and their population statistics in Table 4. H_o was higher than H_e for the seven groups for SSR, indel and SNP 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 (Online Resource 7).

The average contribution of each nuclear group to the genotypes not included in any defined population was 15 % for N1, 9 % for N2, 30 % for N3, 9 % for N4, 15 % for N5, 10 % for N6 and 13 % for N7.

Data from hybrids with known parents were checked in order to validate the analysis of the contributions from the mandarin groups identified (accessions ID 144 to ID 171). 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 'Palazzeli' ([ID 165]; *C. clementina* × 'King') had contributions from three different 'mandarin-like' genomes, defined as groups N1, N2 and N3 of the present Structure analysis, which come from its supposed parents, clementines ([ID 54, ID 55 and ID 56]; genomes from groups N1 and N3) and tangor 'King' ([ID 178]; included in group N2).

Another example is the hybrid mandarin 'Simeto' [ID 168], which was obtained from a cross between a *C. unshiu* and *C. deliciosa*. Our study confirms this cross (almost 50 % each from each parent).

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 155] was reported to come from a cross between a

Fig. 5 Mandarins' genome structure. Contribution of the seven parental mandarin groups (N1–N7) into the mandarin genome portion of each 'mandarin-like' genotype under study. Contributions lower than 5 % were discarded. *N1, light blue* (all *C. deliciosa*); *N2, purple* (*C. nobilis*); *N3, orange* (mandarin hybrids); *N4, red* (all *C. unshiu*); *N5, green* (*C. reticulata*); *N6, light red* (*C. tangerina* and others); *N7, grey* (*C. depressa, C. reshni, C. sunki, C. tachibana*)
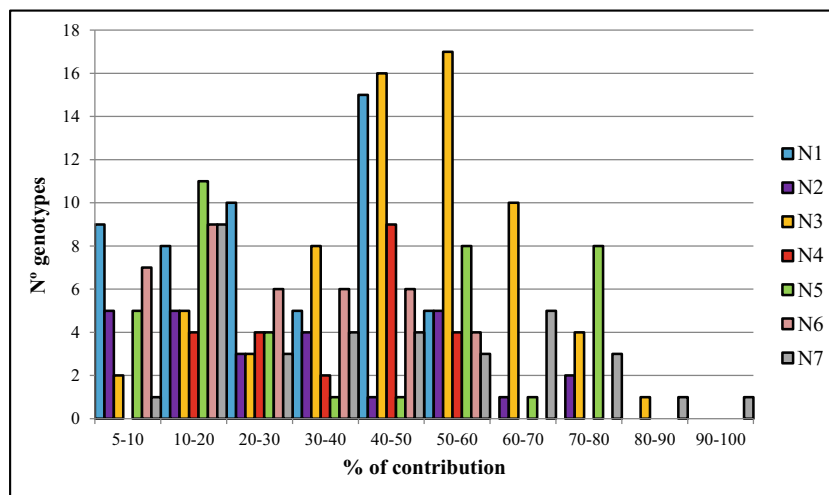
Table 4 Summary statistics of the whole mandarin population and the seven groups identified at the nuclear level (N1–N7)

Nuclear groups	N	Marker	A	H_o	H_e	F_w
All mandarins	166	Mean SSR	7.24	0.62	0.57	−0.10
		Mean indel	2.28	0.20	0.15	−0.31
		Mean SNP	1.90	0.29	0.22	−0.31
N1	7	Mean SSR	2.42	0.63	0.33	−0.89
		Mean indel	1.28	0.16	0.09	−0.76
		Mean SNP	1.34	0.22	0.11	−1.10
N2	3	Mean SSR	1.82	0.73	0.29	−1.54
		Mean indel	1.28	0.24	0.10	−1.49
		Mean SNP	1.40	0.40	0.15	−1.67
N3	13	Mean SSR	1.80	0.52	0.22	−1.40
		Mean indel	1.28	0.12	0.07	−0.84
		Mean SNP	1.09	0.09	0.03	−1.67
N4	18	Mean SSR	1.80	0.64	0.25	−1.57
		Mean indel	1.40	0.30	0.13	−1.37
		Mean SNP	1.25	0.24	0.09	−1.65
N5	7	Mean SSR	1.54	0.51	0.20	−1.58
		Mean indel	1.08	0.08	0.03	−1.67
		Mean SNP	1.28	0.28	0.10	−1.67
N6	2	Mean SSR	1.84	0.72	0.30	−1.42
		Mean indel	1.36	0.24	0.12	−0.94
		Mean SNP	1.62	0.49	0.21	−1.30
N7	8	Mean SSR	3.66	0.59	0.45	−0.28
		Mean indel	1.76	0.20	0.15	−0.36
		Mean SNP	1.45	0.14	0.10	−0.26

N number of genotypes, A number of alleles, H_o observed heterozygosity, H_e expected heterozygosity, F_w Wright fixation index

clementine and 'Dancy' [ID 26], made by Furr (1964). However, the Structure analysis showed that 'Fortune' has almost no contribution of N5 genome contribution while 'Dancy' has a 75 % contribution. The false parental origin was confirmed by individual locus checking: in 15 out of 50 SSR markers, one indel and four SNP markers, 'Fortune' possesses a specific allele absent in 'Dancy' and clementine. Barry et al. (2015) already showed that 'Dancy' is not a parent of 'Fortune' mandarin. Similar observations were made for 'Fremont' [ID 156], supposed hybrid between *C. clementina* and *C. reticulata* [Tan.] 'Ponkan' [ID 15], made by Furr (1964). Indeed, for 11 SSR markers, this hybrid possesses alleles that are not observed in its supposed parents. Moreover, 'Fremont' has almost no contribution of N5 genome, while 'Ponkan' belongs to this nuclear group.

## Cytoplasmic analysis (mitochondria and chloroplast)

The summary of the maternal origin information is shown in the Online Resource 1. In the whole population (199 genotypes, including the secondary species), the mitochondrial markers allowed discrimination of six mitotypes (Mito; Fig. 6a), previously described by Froelicher et al. (2011). One *Fortunella* genotype (*F. hindsii*; ID 48) was associated with the *Papeda* (*C. micrantha*) mitotype. In the mandarin group (166 genotypes), five mitotypes were distinguished: two of mandarins (Mito1 and Mito2), one identical to *C. maxima* (Mito3), one identical to *C. medica* (Mito4) and one identical to *C. micrantha* (*Papeda*, Mito5). The first mandarin mitotype (Mito1) included most of the genotypes studied. In the second mitotype (Mito2), 18 genotypes were present; 11 of them were acid mandarins (four *C. depressa* [ID 4, ID 67, ID 68, ID 69], three *C. sunki* [ID 22, ID 23, ID 128], *C. reshni* [ID 14], *C. daoxianensis* [ID 52] and two 'Sun chu sha' [ID1 6, ID 122]), and seven were sweet genotypes (*C. tankan* Hay. [ID 53], *C. kinokuni* [ID 71], *C. tangerina* [ID 134] and four *C. reticulata* [ID 92, ID 94, ID 108 and ID 110]). The *C. maxima* mitotype (Mito3) included mandarin cultivars [ID 83, ID 118, ID 140, ID 124, ID 127, ID 74, ID 18, ID 164 and ID 103], as well as a tangor ([ID 180]; *C. sinensis* × 'Dancy') or the tangelos. The *C. medica* mitotype (Mito4) included *C. indica* [ID 7]. The *Papeda* mitotype (Mito5) included 'Nicaragua' [ID 112] and the two *C. amblycarpa* [ID 1 and ID 2].

The chloroplastic markers discriminated two main mandarin chlorotypes (Chloro1 and Chloro 2; Fig. 6b). Chloro1 included most of the mandarins and chloro2 included the same 18 genotypes than in the Mito2. Besides these two main chlorotypes, there are other minor ones including only one genotype, e.g. mandarin 'Suntara' ([ID 118]; shared with two secondary species, *C. aurantium* and *C. limon*), *C. tachibana* [ID 24] and *C. indica* [ID 7].

In the whole population, the combination of chloroplastic and mitochondrial markers (Fig. 6c) differentiated 10 cytotypes (Cyto). Four of them corresponded to *C. maxima*, *C. medica*, *Papeda* and *Fortunella*. The *C. maxima* cytotype included eight mandarin cultivars [ID 83, ID 140, ID 124, ID 127, ID 74, ID 18, ID 164 and ID 103], as well as tangors and tangelos. The *Papeda* cytotype included 'Nicaragua' [ID 112] and *C. amblycarpa* [ID 1, ID 2]. Two main cytotypes were found within mandarins, one (Cyto1) that included most of the mandarins, mainly sweet genotypes. The second one (Cyto2) included 18 genotypes: *C. sunki* [ID 22, ID 23, ID128], 'Sun chu sha' [ID 16, ID 122], *C. daoxianensis* [ID 52], *C. depressa* [ID 4, ID 67, ID 68, ID 69], *C. reshni* [ID 14], *C. tankan* [ID 58] and other genotypes [ID 80, ID 92, ID 94, ID 108, ID 110 and ID 134]. Mandarin 'Suntara' [ID 118] appeared with an independent cytotype (included within the secondary species *C. aurantium* and *C. limon* cytotype), as well as *C. tachibana* [ID 24] or *C. indica* [ID 7].

Fig. 6 NJ trees of 199 varieties of *Citrus* and *Severinia buxifolia* used as an out-group with mitochondrial and chloroplastic markers. **a** Mitochondrial markers. **b** Chloroplastic markers. **c** Combination of mitochondrial and chloroplastic markers
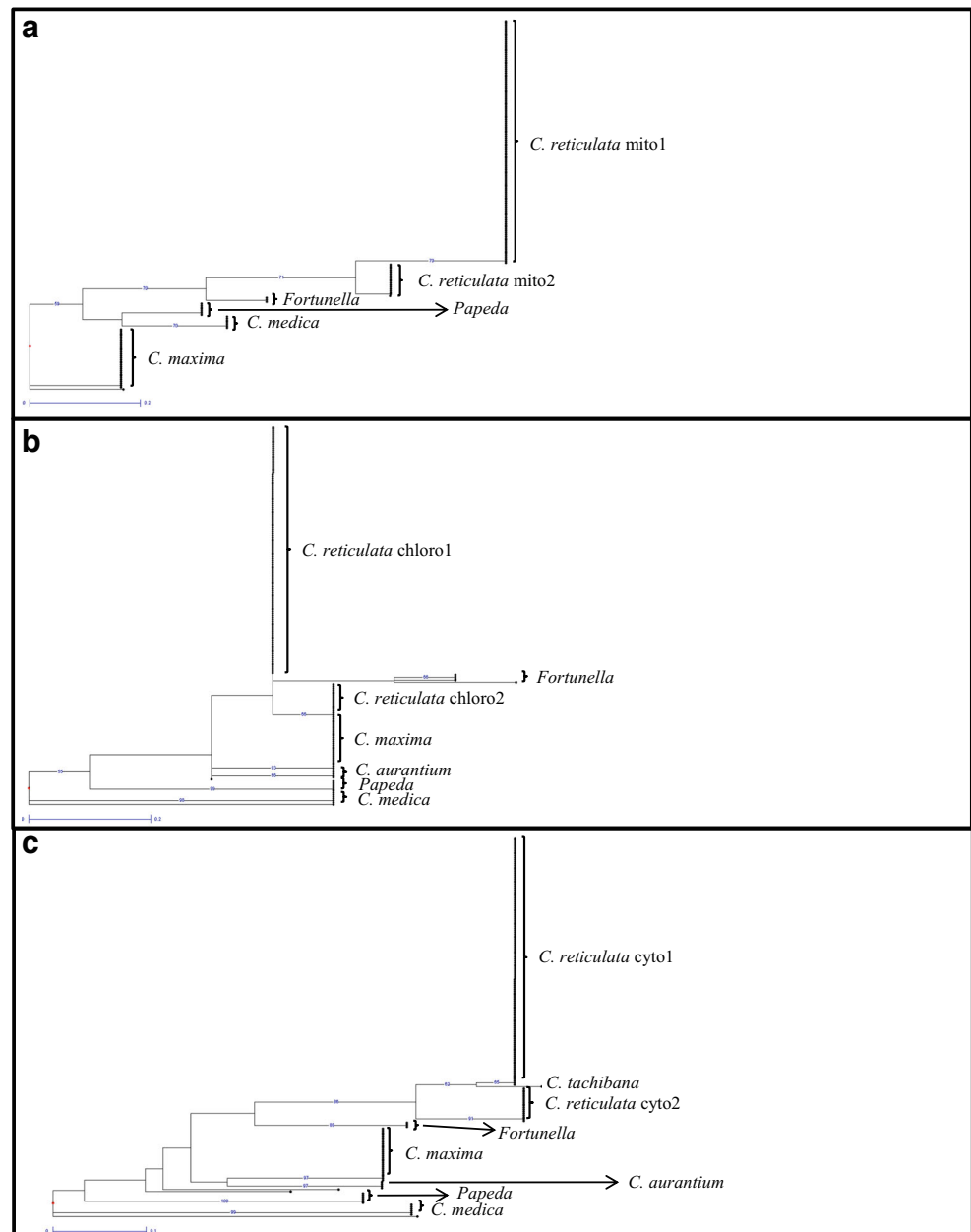
Discussion

Genetic structure of the studied population

SSR markers are more polymorphic than indel and SNP markers. The average numbers of alleles, gene diversity and heterozygosity were all higher in SSR markers. The combination of the three 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* (Fig. 1) has been described in several studies (Nicolosi et al. 2000; Barkley et al. 2006; Garcia-Lor et al. 2012, 2013a). 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; Fig. 1), perhaps due to the large number of hybrids.

Some accessions displayed the same genotype for the analysed markers. Among the groups of accessions with identical genotypes, the groups MLG1, MLG2 and MLG8 are known to have been diversified by the selection of natural mutations and are probably distinguished only by point mutations (satsumas and clementines). Therefore, the probability of distinguishing them by analysis of molecular markers such as

SSRs, indels or SNPs is very low. On the other hand, the clusters MLG3, MLG4, MLG5, MLG6 and MLG7 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 germ-plasm collections.

The overall F_w value among all loci and all genotypes was relatively low (0.02; 0.04 for SSR, 0.29 for indel and 0.12 for SNP) when compared to the high structuration (positive F_w) observed by Garcia-Lor et al. (2012) in the citrus genus. 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 0. 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).

Cytoplasmic 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 Mito1, mandarin mitotype Mito2 and *Fortunella*), all of them observed by Froelicher et al. (2011), who proposed a distinction between the acid mandarin and sweet mandarin mitotypes. The mandarin germplasm (166 genotypes) was represented in five of the six identified mitotypes; two of them included mandarin and 'mandarin-like' genotypes (Mito1, Mito2), and three corresponded to other ancestral species (Mito3, Mito4, Mito5). 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) may not be appropriate since we found sweet mandarin genotypes that share the supposed acid mitotype (7 out of 18 sweet mandarins in the mandarin mitotype 2).

Three 'mandarins' ('Nicaragua' [ID 112] and the two *C. amblycarpa* [ID 1 and ID 2]) have a *Papeda* mitotype (Mito3), and nine have a *C. maxima* mitotype (Mito4). 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.

At chloroplastic level, most mandarins formed a main cluster, as it was already observed by Nicolosi et al. (2000), Cheng et al. (2005) and Yamamoto et al. (2013). We observed also a second chlorotype formed mainly by acid mandarins, as observed by Froelicher et al. (2011) and Yamamoto et al. (2013). Mandarins are well differentiated from the other citrus ancestral species (*C. maxima*, *C. medica* and Papeda), which are separated in different clusters in our study and others (Cheng et al. 2005; Bayer et al. 2009), but not in Yamamoto et al. (2013). Other genotypes, like *C. tachibana* [ID 24], appeared in a different sub-cluster (own chlorotype), as found by Nicolosi et al. (2000), Penjor et al. (2013) and Yamamoto et al. (2013), or mandarin 'Suntara' [ID 118] that appeared with an independent chlorotype may be related to *C. aurantium*, as it was observed by Garcia-Lor et al. (2012).

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). 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 approximately a 50 % of *Papeda* and *C. reticulata*, suggesting a potential origin from direct interspecific hybridisation between them at the nuclear level as proposed from a large SNP analysis (Ollitrault et al. 2012a).

Citrus indica clustered within the citron group at the nuclear level. It had contributions of 46 % from citron, 30 % from papeda and 24 % from mandarin genomes, as well as a very high observed heterozygosity (61.33 %), indicating that it was originated as an interspecific hybrid. In our study, *C. indica* had a citron mitotype, whereas Nicolosi et al. (2000) clustered *C. indica* with the citron on the basis of cpDNA markers. It has its own chlorotype, and therefore, cytotype separated from citron, although it is closed to it. This may be due that the female citron parent (pure or hybrid) is not present in our collection.

Citrus tachibana was considered native to Japan by Hirai et al. (1990) and to be a wild species of mandarin by Swingle and Reece (1967). Later on, it was clustered with the mandarins by Nicolosi et al. (2000). Our results point out that it is not a pure mandarin. Indeed, *C. tachibana* clustered with the mandarin mitotype Mito1, but it has its own chlorotype and cytotype, and displays contributions from different genomes, mainly *C. reticulata* (Sw) and *Fortunella* genomes at the nuclear level. It is included in the nuclear group 7 together with *C. sunki*, *C. reshni* and *C. depressa*. The high H_o (54.67 %) also suggests that *C. tachibana* is an interspecific hybrid. The *C. tachibana* genotype present in our collections may not be the original tachibana from Japan and may be a hybrid with a Chinese genotype (Hirai et al. 1990).

Citrus daoxianensis is mostly of *C. reticulata* origin (92 %, considering the first Structure analysis), with small introgressions of three different genomes (pummelo, papeda and kumquat) lower than 5 %, and therefore not considered significative. This result is in agreement with Li et al. (1992), who considered *C. daoxianensis* to be a wild mandarin and with Curk et al. (2015) who did not found introgression from other genomes. On the other hand, in the second Structure analysis, we found an 80 % contribution of the acidic group and the rest from other mandarin groups, which indicates that it is a hybrid between mandarins.

Introgressions from other genomes were also suggested by the Structure analysis for other genotypes considered to be pure mandarin species by Tanaka and that come from the area of origin. The *Fortunella* genome (~15 %) is present in *C. depressa* and the *C. maxima* (~8 %) genome 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* (~81 %) and contribution from *Fortunella* (~15 %). It is also remarkable that some other genotypes, like *C. unshiu* and *C. tankan*, have contribution from *C. maxima*. *Citrus nobilis* was considered also as a species by Tanaka (group1), but other authors (Nicolosi et al. 2000; Garcia-Lor et al. 2012, 2013a; Coletta-Filho et al. 1998) considered it as a tangor, with introgression from the *C. maxima* genome. Our results confirm this pummelo introgression in the various *C. nobilis* analysed ('Geleking' ~22 % and 'Campeona' ~13 %), also displayed by 454 amplicon sequencing (Curk et al. 2014). Cytoplasmic data include it within the mandarins. At the nuclear level, it appears as an independent group (N2).

This introgression of other taxa in varieties considered as pure mandarin was previously found in molecular marker studies (Nicolosi et al. 2000; Barkley et al. 2006) and more recently from NGS data (Curk et al. 2014; Wu et al. 2014) and in a 454 sequencing work (Curk et al. 2015). Within our discrete markers' dispersion throughout the genome (124 markers), it is probable that for some varieties, we have miss small introgressions, as the 4 % of *C. maxima* introgression identified by Whole Genome Sequencing (WGS) (Wu et al. 2014) in 'Ponkan' and 'Willowleaf' not detected in our study was like. On the other hand, in this study, many more mandarin genotypes have been used than in previous studies with molecular markers and the recent ones of WGS.

The other mandarin hybrids, tangors, tangelos and clementines that 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.

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* (~9 %), instead of the *Papeda/Fortunella* group reported by Barkley et al. (2006). Moreover, we extend the study to a much higher number of genotypes than in previous published works.

Organisation of the mandarin germplasm

The two main *Citrus* classification systems of Swingle and Reece (1967) and 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 (Nicolosi et al. 2000; Barkley et al. 2006; Federici et al. 1998). 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 (1954) and Webber's (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). Kaçar 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 two analyses with the Structure software (Figs. 2 and 4), it is clear that there is introgression of ancestral genomes within many mandarins, mainly *C. maxima*. Moreover, seven groups were identified within the mandarin germplasm at the nuclear level (N1–N7; Fig. 4, Table 3), although there is not a strong differentiation between nuclear groups, confirmed by the F_{st} value (0.445). These mandarin groups exhibited higher allelic diversity for SSRs than for indel and SNP markers. The negative F_w values observed in these groups should be due to fixated heterozygosity resulting from apomixis and vegetative propagation of citrus varieties. The global mandarin population has an F_w value close to 0, reflecting strong intergroup gene flow.

Five nuclear groups, N1, N2, N4, N5 and N6, share the same mandarin mitotype (Mito1). The N3 group is clustered with the pummelo mitotype. Genotypes from N7 were present in three mitotypes (Mito1, Mito2 and *Papeda*). Most of the genotypes sharing the mandarin mitotype (Mito2) are also

differentiated at the nuclear level, having a high contribution of mandarin nuclear group N7, like *C. daoxianensis* (80 %). Regarding the N7, it was dispersed in different chlorotype and cytotype groups, being the most heterogeneous chloroplastic group. Tanaka (1954) divided the acid mandarin genotypes into two groups, with *C. reshni*, *C. sunki* and *C. tachibana* in group 4 and *C. depressa* in group 5, which are joined in our analysis at the nuclear level, but separated into two different groups at the cytoplasm level, including *C. reshni*, *C. sunki* and *C. depressa* in one cytotype and *C. tachibana* in another one.

Tanaka (1954) grouped the 36 mandarin species that he considered into five clusters. These clusters are compared with the nuclear groups found in our study (Table 5). One cluster included the species *C. nobilis* and *C. unshiu*, which seem to be of interspecific hybrid origin mandarin×pummelo in our analysis and are separated into two nuclear groups (N2 and N4). The second cluster included species not analysed in our study. The third cluster had 14 species, including *C. clementina* (considered in our study as a hybrid and not a pure mandarin species), *C. deliciosa*, *C. reticulata* (Tan.) and *C. tangerina*, which appear in our work as different mandarin groups (N1, N5 and N6). The fourth Tanaka group was formed by *C. reshni*, *C. sunki*, and *C. tachibana*, and the fifth group included *C. depressa* and *C. lycopersicaeformis*. These two groups are included in a unique nuclear group in our study (N7). From these species, *C. tachibana* is separated at the cytoplasmic level. Other Tanaka species, such as *C. erythrosa* and *C. suhuiensis*, seem to have originated from hybridisation between mandarin groups.

Hodgson (1967) divided the mandarins into four groups: *C. unshiu*, *C. reticulata* (Tan.) ('Ponkan', 'Dancy', clementines), *C. deliciosa* and *C. nobilis* ('King'). Three groups are fully in agreement with our results, *C. deliciosa* (N1), *C. nobilis* (N2) and *C. unshiu* (N4). The fourth group defined by Hodgson, *C. reticulata* (Tan.), included a known hybrid (*C. clementina*, mix of group N1 and N3) and two genotypes, 'Ponkan', which is within the *C. reticulata* (Tan.)

group (N5), and 'Dancy', with a big contribution of this group (75 %).

The contributions of the seven mandarin groups identified in the mandarin germplasm (Fig. 4), besides the contributions of the other ancestral taxa and *Fortunella* (Fig. 2), were estimated for the entire 'mandarin-like' collection. These analyses revealed that the genomes of most 'mandarin-like' genotypes are complex admixtures of the 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).

For some accessions, the 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.

In summary, this work shows that the mandarin horticultural varietal group is highly polymorphic and that 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. These data point out that many mandarins (including ancient genotypes coming from the mandarin area of origin) are really hybrids and not pure mandarins. This idea is in agreement with the recent results obtained from a few mandarins' sequenced genomes, leading to the uncertainty of which is/are the ancestral mandarin/s, which has/have not been identified yet.

This work has provided new insights into mandarin structuration, and it will help for a better management of the two germplasm collections studied, it will allow the

Table 5 Comparison of the mandarin organisation between the Tanaka groups and the nuclear groups found in this study

Tanaka\nuclear groups (*N*)	N1 (*C.deliciosa*)	N2 (*C.nobilis*)	N3 (mandarin hybrids)	N4 (*C.unshiu*)	N5 (*C.reticulata*)	N6 (*C.tangerina*)	N7 (acid mandarins)
Tanaka group 1		X		X			
Tanaka group 2	–	–	–	–	–	–	–
Tanaka group 3	X				X	X	
Tanaka group 4							X
Tanaka group 5							X

Tanaka group 1 included *C. nobilis*, *C. unshiu* and *C. yatsushiro*; group 2 included *C. keraji*, *C. oto* and *C. toragayo*; group 3 included 14 species, such as *C. reticulata*, *C. deliciosa*, *C. clementina* and *C. tangerina*; group 4 included *C. reshni*, *C. sunki* and *C. tachibana*; and group 5 included *C. depressa* and *C. lycopersicaeformis*. (−) Species within this group not present in our study. Genotypes present in nuclear group (N3) not present in the Tanaka classification

implementation of a database with all the molecular characterisation and is already helping to select the proper genotypes to perform new crosses in order to generate new diversity and to define new strategies for the citrus triploid breeding programme that is being carried out.

References

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. Ann Bot 108:37–50

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 Theor Appl Genet 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 Syst Evol 282:71–86

Barrett H, Rhodes A (1976) A numerical taxonomic study of affinity relationships in cultivated citrus and its close relatives. Syst Bot 1:105–136

Barry GH, Gmitter FG Jr, Chen C, Roose ML, Federici CT, McCollum GT (2015) Investigating the parentage of 'Orri' and 'Fortune' mandarin hybrids. Acta Hortic 1065:449–456. doi:10.17660/ActaHortic.2015.1065.55

Bayer RJ, Mabberley DJ, Morton C et al (2009) A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. Am J Bot 96:668–685

Belkhir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (2002) *GENETIX v. 4.03, logiciel sous windows pour la génétique des populations*. In Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France). Available at [http://www.univ-montp2.fr)].

Cheng Y, De Vicente MC, Meng H, Guo W, Tao N et al (2005) A set of primers for analyzing chloroplast DNA diversity in citrus and related genera. Tree Physiol 25:661–672

Coletta-Filho HD, Machado MA, Targon MLPN, Moreira MCPQDG, Pompeu Júnior J (1998) Analysis of the genetic diversity among mandarins (citrus spp.) using RAPD markers. Euphytica 102:133–139

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

Curk et al (2014) Next generation haplotyping to decipher nuclear genomic interspecific admixture in Citrus species: analysis of chromosome 2. BMC Genet 15:1272

Curk F, Ancillo G, Ollitrault F, Perrier X, Jacquemoud-Collet JP, Garcia-Lor A, Navarro L, Ollitrault P (2015) Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern Citrus varieties. PLoS ONE 10: e0125628. doi:10.1371/journal.pone.0125628

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

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, Tomi F, Luro F, Desjobert JM, Casanova J (2006) Chemical variability of peel and leaf oils of mandarins. Flavour Fragrance J 21:359–367

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. Theor Appl Genet 96:812–822

Food and Agriculture Organization of the United Nations, FAOSTAT database (2013) The state of the world's plant genetic resources for food and agriculture. FAO, Rome, Italy. Available at http://faostat.fao.org/.

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). Mol Ecol Resour 8:119–122

Froelicher Y, Mouhaya W, Bassene JB, Costantino G, Kamiri M et al (2011) New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. Tree Genet Genomes 7:49–61

Furr JR (1964) New tangerines for the desert. Calif Citrog 49:266–276

Garcia-Lor A, Luro F, Navarro L, Ollitrault P (2012) Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. Mol Gen Genomics 287:77–94

Garcia-Lor A, Curk F, Snoussi-Trifa H, Morillon R, Ancillo G et al (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. Ann Bot 111:1–19

Garcia-Lor A, Ancillo G, Navarro L, Ollitrault P (2013b) Citrus (Rutaceae) SNP markers based on Competitive AlleleS pecific PCR; transferability across the Aurantioideae subfamily. Applications in Plant Sciences 1: doi:10.3732/apps.1200406.

Green RM, Vardi A, Galun E (1986) The plastome of Citrus. Physical map, variation among citrus cultivars and species and comparison with related genera. Theor Appl Genet 72:170–177

Hirai M, Mitsue S, Kita K, Kajiura I (1990) A survey and isozyme analysis of wild mandarin, tachibana (Citrus tachibana (Mak.) Tanaka) growing in Japan. J Japan Soc Hort Sci 59:1–7

Hodgson RW (1967) Horticultural varieties of citrus. In: Reuther W, Webber HJ, Batchelor LD (eds) The Citrus industry. University of California, Berkeley, pp 431–591

Kaçar YA, Uzun A, Polat I, Yesiloglu T, Yilmaz B et al (2013) Molecular characterization and genetic diversity analysis of mandarin genotypes by SSR and SRAP markers. J Food, Agric Environ 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 Rep 30:1415–1425

Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of Citrus. Theor Appl Genet 94:701–706

Koehler-Santos P, Dornelles ALC, Freitas LB (2003) Characterization of mandarin citrus germplasm from Southern Brazil by morphological and molecular analyses. Pesq Agrop Brasileira 38:797–806

Krueger RR, Roose ML (2003) Use of molecular markers in the management of citrus germplasm resources. J Am Soc Hortic Sci 128:827–837

Krueger RR, Navarro L (2007) Citrus germplasm resources. In: Khan I (ed) Citrus genetics, breeding and biotechnology. CABI Publishing, Wallington, pp 45–140

Li WB, Liu GF, He SW (1992) Leaf isozymes of mandarin. Proc Int Soc Citricult 1:217–220

Liu K, Muse S (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Lota ML, Serra DR, Tomi F, Casanova J (2000) Chemical variability of peel and leaf essential oils of mandarins from citrus reticulata blanco. Biochem Syst Ecol 28:61–78

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

Mabberley D (1997) A classification for edible citrus (rutaceae). Telopea 7:167–172

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. Org Divers Evol 9:52–68

Motohashi T, Matsuyama T, Akihama T (1992) DNA fingerprinting in citrus cultivars. Proc Int Soc Citriculture: 221–224.

Nicolosi E, Deng ZN, Gentile A, Malfa S, Continella G et al (2000) Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theor Appl Genet 100:1155–1166

Ollitrault P, Terol J, Garcia-Lor A, Berard A, Chauveau A et al (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 C, Lotfy S et al (2012b) A reference genetic map of C. clementina hort. ex Tan.; citrus evolution inferences from comparative mapping. BMC Genomics 13:593

Penjor T, Yamamoto M, Uehara M, Ide M, Matsumoto N, Matsumoto R, Nagano Y (2013) Phylogenetic relationships of citrus and its relatives based on matK gene sequences. PLoS ONE 8:e62574

Perrier X, Jacquemoud-Collet J (2006) DARwin software. Available at http://darwin.cirad.fr/darwin 5.0.158.

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

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Swingle W, Reece P (1967) The botany of citrus and its wild relatives. In: Reuther W, Webber HJ, Batchelor LD (eds) The Citrus industry. University of California, Berkeley, pp 190–430

Tanaka T (1954) Species problem in Citrus (Revisio aurantia cearum IX). Japanese Society for Promotion of Science, Tokyo, Japan, pp 1–157

Tanaka T (1977) Fundamental discussion of citrus classification. Studia Citrologia 14:1–6

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, Available at: http://www.redalyc.org/articulo.oa?id=33911005 ISSN 0378–1844

Webber HJ (1943) Cultivated varieties of citrus. In: Anonymous the Citrus industry. History, World Distribution, Botany and Varieties, California, pp 475–668

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F et al (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol 32:656–663

Wright S (1969) The theory of gene frequencies. In: Anonymous evolution and the genetics of populations. University of Chicago Press, Chicago

Yamamoto M, Tominaga S (2003) High chromosomal variability of mandarins (citrus spp.) revealed by CMA banding. Euphytica 129:267–274

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. Res Sta, Okitsu, Japan, pp 39–46

Yamamoto M, Tsuchimochi Y, Ninomiya T, Koga T, Kitajima A, Yamasaki A, Inafuku-Teramoto S, Yang X, Yang X, Zhong G, Nasir N, Kubo T and Tominaga S (2013) Diversity of chloroplast DNA in various mandarins (Citrus spp.) and other citrus demonstrated by CAPS analysis. J. Japan. Soc. Hort. Sci. 82: 106-113