Theor Appl Genet (2006) 112: 1519–1531
DOI 10.1007/s00122-006-0255-9

ORIGINAL PAPER

Noelle A. Barkley · Mikeal L. Roose
Robert R. Krueger · Claire T. Federici

Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs)

Received: 13 September 2005 / Accepted: 24 February 2006 / Published online: 20 April 2006

Abstract Twenty-four simple sequence repeat (SSR) markers were used to detect molecular polymorphisms among 370 mostly sexually derived *Citrus* accessions from the collection of citrus germplasm maintained at the University of California, Riverside. A total of 275 alleles were detected with an average of 11.5 alleles per locus and an average polymorphism information content of 0.625. Genetic diversity statistics were calculated for each individual SSR marker, the entire population, and for specified *Citrus* groups. Phylogenetic relationships among all citrus accessions and putative non-hybrid *Citrus* accessions were determined by constructing neighbor-joining trees. There was strong support for monophyly at the species level when hybrid taxa were removed from the data set. Both of these trees indicate that *Fortunella* clusters within the genus *Citrus* but *Poncirus* is a sister genus to *Citrus*. Additionally, *Citrus* accessions were probabilistically assigned to populations or multiple populations if their genotype indicated an admixture by a model-based clustering approach. This approach identified five populations in this data set. These separate analyses (distance and model based) both support the hypothesis that there are only a few naturally occurring species of *Citrus* and most other types of *Citrus* arose through various hybridization events between these naturally occurring forms.

Communicated by J. S. Heslop-Harrison

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-006-0255-9 and is accessible for authorized users.

N. A. Barkley
Department of Botany and Plant Sciences and
Graduate Program in Genetics, Genomics, and Bioinformatics,
University of California, Riverside, CA 92521, USA

M. L. Roose · C. T. Federici
Department of Botany and Plant Sciences,
University of California, Riverside, CA 92521, USA

R. R. Krueger
USDA-ARS National Clonal Germplasm Repository for Citrus and Dates, Riverside, CA 92507, USA

N. A. Barkley (✉)
1109 Experiment Street, Griffin, GA 30223, USA
E-mail: nbarkley@ars-grin.gov
Tel.: +1-770-4124035
Fax: +1-770-2293323

Introduction

Citrus and its close relatives are represented by 28 genera in the tribe Citreae of the subfamily Aurantioideae in the family Rutaceae (Swingle and Reece 1967). Before the advent of molecular data, *Citrus* was classified based on morphology or biochemical techniques such as isozymes. There are currently two commonly used classifications of *Citrus*: Swingle (Swingle and Reece 1967) and Tanaka (Tanaka 1977). The Swingle system recognizes 16 species in the genus *Citrus*, whereas the Tanaka system recognizes 162 species in the genus *Citrus*. Scora (1975) and Barrett and Rhodes (1976) suggested that there are only three 'basic' true species of *Citrus* within the subgenus *Citrus* as defined by Swingle: citron (*C. medica* L.), mandarin (*C. reticulata* Blanco), and pummelo (*C. maxima* L. Osbeck). Other cultivated *Citrus* species within the subgenus *Citrus* are believed to be hybrids derived from these true species, species of the subgenus *Papeda*, or closely related genera. This idea has gained support in recent years from data derived from molecular markers (Federici et al. 1998; Nicolosi et al. 2000).

Phylogeny and taxonomy for certain *Citrus* cultivars have been somewhat debatable in the past; however, results from molecular marker technologies are helping to clarify some of these relationships. The difficulty in classifying *Citrus* taxa is mainly due to repeated cross-pollination and to adventitious nucellar embryony, which stabilizes and perpetuates hybrid taxa (Scora 1975). Another problem in *Citrus* taxonomy is disagreement on what degree of difference justifies species status,

and whether apparent hybrids among naturally occurring forms should be assigned species rank (Roose et al. 1995). Understanding taxonomy, phylogenetic relationships, and genetic variability in *Citrus* is critical for determining genetic relationships, characterizing germplasm, controlling genetic erosion, designing sampling strategies or core collections, establishing breeding programs, and the registration of new cultivars (Herrero et al. 1996).

Several previous studies have utilized various molecular markers (ISSR, RAPD, AFLP, and SSR) to fingerprint accessions, evaluate phylogenetic relationships among accessions, and examine the level of genetic diversity in *Citrus*. Many of these studies have targeted specific *Citrus* groups or sampled a few individuals of each taxon. For example, Bretó et al. (2001) examined the variability of 24 Clementine (*C. reticulata* Blanco) accessions by utilizing ISSR, RAPD, and AFLP markers and found that only two varieties of 24 could be distinguished. Gulsen and Roose (2001a) utilized ISSR, SSR, and isozymes to assess diversity, phylogenetic relationships, and parentage in lemon (*C. limon* (L.) Burm. f.) accessions and related taxa, finding little genetic variation among lemon accessions. In another study, Fang et al. (1997) employed isozymes, RFLP, and ISSR markers to classify 48 trifoliate orange (*Poncirus trifoliata* (L.) Raf.) accessions into four groups. Fang and Roose (1997) utilized ISSR markers to distinguish closely related cultivars, many of which had arisen by selection of spontaneous mutations. This study showed that ISSR markers could distinguish some (but not all) of these closely related accessions. Nicolosi et al. (2000) used RAPD, SCAR, and cpDNA markers to elucidate phylogenetic relationships and genetic origins of hybrids in 36 accessions of *Citrus* and one accession from each of four related genera. Federici et al. (1998) examined the phylogenetic relations of 88 accessions representing 45 *Citrus* species and six related genera by utilizing RFLP and RAPD markers. Overall, these previous studies demonstrated that molecular markers are powerful tools for elucidating genetic diversity, determining parentage, and revealing phylogenetic relationships among various *Citrus* species; however, accessions arising from spontaneous mutation are often difficult to distinguish.

The Citrus Variety Collection (CVC) located at the University of California, Riverside contains approximately 900 accessions of *Citrus* and related taxa, many of unknown parentage. At the time the studies described in this paper were carried out, the level of genetic diversity present in this collection was unknown because only small samples of it had been analyzed using quantitative methods. Simple sequence repeats (SSR), a type of microsatellite marker, are particularly useful for characterization of germplasm collections because they are highly polymorphic and usually codominant (Brown et al. 1996; Hokanson et al. 1998; Liu et al. 2003), but they have not been widely used in citrus. The objective of this study was to use 24 SSR markers to detect polymorphisms among *Citrus* and its near relatives as represented in the CVC in order to determine the level and organization of genetic diversity within this collection and elucidate phylogenetic relationships among accessions. Additionally, SSR markers were added to an existing citrus linkage map to produce a more detailed map and analyze their genome distribution in *Citrus*. Since SSR markers are highly informative, codominant, and have a high mutation rate, they are a good marker system for determining phylogenetic relationships among closely related taxa, fingerprinting varieties, and determining the level of genetic diversity within germplasm collections. This data will allow calculation of genetic diversity within various *Citrus* groups and for much of the collection, which has not been previously reported.

Materials and methods

Plant samples

The accessions chosen from the genus *Citrus* consisted of 4 blood oranges, 3 sweet oranges, 2 navels, and 1 valencia orange (*Citrus sinensis* (L.) Osbeck), 105 mandarins (*C. reticulata* Blanco), 22 sour oranges (*C. aurantium* L.), 5 rough lemons (*C. jambhiri* Lush.), 21 limes (*C. aurantifolia* [Christm.] Swing.), 6 sweet limes (*C. aurantifolia* [Christm.] Swing.), 29 citrons (*C. medica* L.), 15 papedas (4 *Citrus species* of uncertain classification, 3 *C. ichangensis* [Swing.], 3 *C. hystrix* DC., 1 *C. latipes* [Swing.] Tan., 1 *C. micrantha* Wester, 1 *C. hanaju* Hort. ex Shirai, 1 *C. macrophylla* Wester, 1 *C. sphaerocarpa* Hort. ex Tan.), 76 pummelos (*C. maxima* [Burm.] Merrill), 3 calamondins (*C. reticulata hybrid* Lour.), 8 rangpurs (*C. limonia* Osbeck), 15 sour orange hybrids (*C. aurantium* L.), 4 grapefruit hybrids (*C. paradisi* Macf.), 20 lemon hybrids (*C. limon* (L.) Burm. f.), and 13 pummelo hybrids (*C. maxima* [Burm.] Merrill). Several *Citrus* relatives were also chosen: eight kumquats (three *Fortunella crassifolia* Swing., two *F. hindsii* (Champ.) Swing, one *F. japonica* (Thunb.) Swing., one *F. obovata* Tan., one *F. margarita* Lour.), and 10 trifoliate oranges (*Poncirus trifoliata* (L.) Raf.). The entire list of accessions used in this study along with cultivar names, genus and species classification can be viewed in Supplementary Table 1. (The binomials assigned to the accessions are those utilized at the USDA-ARS National Clonal Germplasm Repository for Citrus and Dates. This is a somewhat ad hoc system devised by the Citrus Crop Germplasm Committee but is more closely allied with the Tanaka than with the Swingle system).

DNA extraction and PCR separation

DNA was extracted from young leaves using a standard CTAB protocol (Webb and Knapp 1990). The DNA concentration was measured by a fluorometer (Hoefer Pharmacia Biotech, San Francisco, CA, USA), and the samples were diluted with TE buffer to 10 ng/μl for PCR reactions.

The PCR reaction consisted of 0.5 ng/μl of template DNA, 0.2 mM dNTPs, 0.025–0.05 μM forward and reverse primers, 1× Promega PCR buffer, 2–4 mM magnesium chloride and 0.1 U/μl Promega Taq DNA polymerase. The forward primer of each pair was labeled at the 5′ end with an infrared dye (LI-COR, Lincoln, NE, USA). Cycling conditions consisted of 95°C for 5 min; 38 cycles of: 95°C 1 min, 45–64°C for 30 s (annealing temperature was specific for each primer), and 72°C for 1 min; and one cycle of 72°C for 7 min. PCR reactions were performed on a Perkin Elmer 9600 thermocycler (Norwalk, CT, USA).

The PCR products were separated using a denaturing 7% Long Ranger (BMA, Rockland, ME, USA) polyacrylamide gel attached to a LI-COR IR^2 4200LR Global DNA sequencer dual dye system. PCR products were denatured (95°C) for 2 min before loading the gel. Approximately, 0.5 μl aliquot from each PCR amplification product was loaded in each lane of the gel, with 3 lanes (first, center, and last lane) containing a 50–350 bp size standard (LI-COR).

RFLPscan version 3.12 software (Scanalytics) was used to visualize gel images, analyze, and determine the length of the PCR products, and label lanes. Allele sizes were also manually scored to ensure proper scoring of band sizes.

SSR primers

The sequences of 14 citrus-specific primer pairs were obtained from Kijas et al. (1997) and 3 citrus-specific primer pairs were obtained from Ahmad et al. (2003). To obtain additional microsatellite primers, genomic libraries of *C. maxima* (pummelo) DNA enriched for five microsatellite repeats (CA, GA, ATG, AAG, and GGA) were purchased from Genetic Identification Services (GIS) (Chatsworth, CA, USA). Twenty-five primer pairs were designed from the sequence data provided by GIS using the Primer3 program (Rozen and Skaletsky 2000) at http://www.frodo.wi.mit.edu/primer3/primer3_code.html.

Optimization of annealing temperatures and $MgCl_2$ concentration for each primer designed from the enriched microsatellite library was accomplished by screening 1°C variation in annealing temperature using a Robocycler Gradient 96 (Strategene, La Jolla, CA, USA). Variations to the PCR reaction (i.e. annealing temperature, $MgCl_2$ concentration, primer concentration) for each marker can be viewed at http://www.plantbiology.ucr.edu/documents/files_of_Roose/rooselink2.html. Primer sequences for the markers designed from the enriched microsatellite library are listed in Table 1.

Genetic linkage mapping

An intergeneric cross between 'Troyer' citrange (*Citrus sinensis* L. Osb. × *Poncirus trifoliata* L. Raf.) × 'Sacaton' citrumelo (*C. paradisi* Macf. × *P. trifoliata*) was used to produce a mapping population consisting of 57 progeny (Jarrell et al. 1992). The population was treated as cross-pollinated. The cultivars used to represent grandparents were 'Pomeroy' trifoliate (*P. trifoliata*), 'Rubidoux' trifoliate (*P. trifoliata*), 'Duncan' grapefruit (*C. paradisi*), and 'Olinda' valencia (*C. sinensis*). Genotype data from the individuals of this population were determined by analyzing the segregation of alleles in progeny and subsequently used to determine the pairwise recombination frequencies. Joinmap version 3.0 (Stam 1993) with a Kosambi mapping function was utilized to add the new microsatellite markers to an existing linkage map (Roose et al. 2000).

Phylogenetic analysis

Microsat version 1.5 (Minch et al. 1997) was utilized to create a genetic distance matrix between all pairwise combinations and summarize information on unique alleles, frequency of alleles, and distribution of alleles. Microsat was also used to resample the data for bootstrap analysis with 500 replicates. The proportion of shared alleles [$D = -\ln(ps)$] was utilized as a genetic distance measure (Bowcock et al. 1994). PAUP version 4.0 beta (Swofford 1998) was used to construct a neighbor-joining tree from the genetic distance matrix generated from Microsat.

Diversity statistics

Cervus version 2.0 (Marshall et al. 1998) calculated polymorphic information content (PIC), observed and

Table 1 Forward and reverse primer sequences for 11 new SSR markers designed from sequences from microsatellite-enriched genomic libraries

Marker	Repeat motif	Forward sequence	Reverse sequence
AC01	CA/TA	TTTGACATCAACATAAAACAAGAAA	TTTTAAAATCCCTGACCAGA
AG14	GA	AAAGGGAAAGCCCTAATCTCA	CTTCCTCTTGCGGAGTGTTC
ATC09	TCA	TTCCTTATGTAATTGCTCTTTG	TGTGAGTGTTTGTGCGTGTG
CAG01	AGC	AACACTCGCACCAAATCCTC	TAAATGGCAACCCCAGCTTTG
CAT01	CAT/CTT	GCTTTCGATCCCTCCACATA	GATCCCTACAATCCTTGGTCC
CCT01	CCT	TCAACACCTCGAACAGAAGG	CCCACATGCTAGCACAAAGA
CT02	CT	ACGGTGCGTTTTGAGGTAAG	TGACTGTTGGATTTGGGATG
CT19	TC	CGCCAAGCTTACCACTCACTAC	GCCACGATTTGTAGGGGATAG
CT21	TC	CGAACTCATTAAAAGCCGAAAC	CAACAACCACCACTCTCACG
CTT01	CTT	TCAGACATTGAGTTGCTCG	TAACCACTTAGGCTTCGGCA
GT03	GT	GCCTTCTTGATTTACCGGAC	TGCTCCGAACTTCATCATTG

expected heterozygosity, allele frequencies, and tested Hardy–Weinberg equilibrium at each locus and for specific citrus groups. Botstein et al. (1980) originally defined PIC values as the probability of a given marker being informative in a random mating. The Cervus program also checked for identical accessions defined as those that had matching genotypes for all 24 SSR markers.

Population structure

The software program Structure 2.0 (Pritchard et al. 2000) was utilized to infer population structure and assign individuals to populations based on the SSR genotypes. Structure uses model-based clustering in which a Bayesian approach is used to identify clusters based on fit to Hardy–Weinberg equilibrium and linkage equilibrium. Multiple runs of Structure were performed by setting K (the number of populations) from 1 to 10. The burn-in time and replication number was set to 500,000 for each run and each run was replicated five times.

There are several statistics to allow a user to decide what value of K best fits the data. First of all, the Pr (X/K) value should be a negative value or effectively zero, which was obtained at $K=5$. Additionally, if there are separate populations the inferred value of alpha, which is defined as the Dirichlet parameter for the degree of admixture, should remain constant (range ~0.2) while running the program. In our data set when K was set to 5, the alpha value varied by 0.015 during the entire run. A mean alpha value <1 implies that most individuals are essentially from one population or another. The mean alpha value for this data set was 0.0522. Another point in deciding the most appropriate value of K is that the proportion of individuals belonging to the various populations should not be equal. If the population membership is symmetric ($\sim 1/K$) most of the individuals will be fairly admixed and one should infer that there is no real population structure. The membership of individuals in the populations determined by Structure for this data set were not symmetric at $K=5$ (range 0.033–0.342), but became more symmetric as the value of K decreased. The authors state that while it may not be possible to know the true value of K, one should try to pick the smallest value of K that captures the major structure of the data (Pritchard et al. 2000).

Results and discussion

A total of 370 *Citrus* accessions were sampled (Supplementary Table 1) from the CVC located at The University of California, Riverside. Since citrus has been selected and bred for thousands of years, many varieties in this 97-year-old collection do not have thorough passport records; however, morphological data on many accessions is available in the CVC and the USDA-ARS National Clonal Germplasm Repository for Citrus and Dates [NCGRCD (Riverside, CA, USA)] archives. Moreover, many accessions in this collection were initially acquired from other collections around the world and their original geographic origin and parentage is unknown. The *Citrus* accessions and *Citrus* relatives chosen for this study have been assumed to be mainly sexually derived or of unknown origin. The samples included nearly all the *Citrus* accessions that are not known to be derived from other *Citrus* by apomixis, selection of bud mutations (limb sports), or controlled pollinations. Sexually derived accessions should be more genetically diverse than accessions arising apomictically or from controlled pollinations, and thus their ancestry is less well understood.

Certain accessions from the CVC were not included in this study, such as some of the known hybrid accessions developed in breeding programs. Additionally, only a few of the sweet oranges, navels, and valencias from the collection were sampled because nearly all sweet oranges originated through somatic mutations which altered horticultural characters, mostly fruit traits (Hodgson 1967), and they have been shown to be genetically similar (Luro et al. 1995; Fang and Roose 1997). Many of the grapefruit, lemon, Clementine mandarin, and Satsuma mandarin accessions were excluded for similar reasons. Only 10 of the 48 trifoliate orange accessions were included because of low polymorphism detected among this group of accessions (Fang et al. 1997). Lastly, this work did not include most of the accessions belonging to genera other than *Citrus* because SSR markers are often not conserved when transferred across genus borders (Jain et al. 2002). In total, 530 accessions from the CVC were excluded from this study because they had documented origins of arising via somatic mutation, had very low genetic diversity in previous studies, were known hybrids developed through breeding programs or were distant relatives of *Citrus*.

SSR markers

The number of alleles detected among the 24 markers studied ranged from 3 to 30 (Table 2). A total of 275 alleles were detected with a mean number of alleles per locus of 11.5. The dinucleotide repeat markers produced on average 13.6 alleles/locus, whereas the trinucleotide repeat markers had a mean of 10.2 alleles/locus. The observed heterozygosity was calculated for each individual marker as a measure of marker diversity. The percentage of heterozygotes per marker detected in our *Citrus* population ranged from 13% from marker AC01 to 71% from marker TAA41. The mean observed heterozygosity for all markers was 42.5%. The PIC values for the 24 markers ranged from 0.247 (CMS8) to 0.916 (TAA41) with a mean value of 0.625 (Table 2).

Eleven out of 25 primer pairs designed from sequence data from an enriched microsatellite library amplified products in *Citrus* (44% success rate). Out of the 11 successful markers, six (55%) were dinucleotide markers and five (45%) were trinucleotide markers. The number of

Table 2 Diversity statistics for 24 SSR markers studied in 370 citrus accessions. The first 10 markers listed are from Kijas et al. (1997), the next 3 from Ahmad et al. (2003), and the last 11 are previously unpublished

SSR loci	Allele sizes	Alleles observed	PIC value	H_{obs}
CAC23	237–270	9	0.600	0.560
TAA27	197–242	10	0.462	0.401
TAA15	141–204	19	0.713	0.634
TAA41	122–185	30	0.916	0.710
CAC39	147–180	8	0.346	0.236
TAA3	133–172	10	0.750	0.465
CAC33	132–168	6	0.499	0.498
CAC15	148–163	4	0.356	0.391
TAA33	107–125	7	0.516	0.385
CAGG9	103–121	6	0.478	0.291
CMS4	144–218	20	0.756	0.456
CMS7	140–162	11	0.724	0.548
CMS8	141–150	3	0.330	0.556
ATC09	169–202	8	0.600	0.564
GT03	149–197	19	0.834	0.478
CT19	117–171	14	0.844	0.619
AC01	135–167	15	0.799	0.130
CCT01	134–164	7	0.247	0.171
CAT01	122–164	13	0.736	0.387
AG14	119–163	20	0.858	0.484
CT21	139–163	12	0.795	0.341
CTT01	124–159	10	0.491	0.400
CT02	102–144	8	0.770	0.178
CAG01	114–132	6	0.580	0.327

alleles observed ranged from six to 20 (Table 2). The PIC values of the markers designed from the enriched library ranged from 0.247 to 0.858. The most informative markers were AG14, CT19, and GT03 with PIC values of 0.858, 0.844, and 0.834, respectively. There seemed to be no strong correlation between the PIC value and the repeat motif (perfect, imperfect, compound) or repeat length with these markers.

Overall, the 24 SSR markers were fairly successful at specifically fingerprinting the *Citrus* accessions. The primer sets scored amplified clear, well-resolved fragments with little stutter. The SSR markers could distinguish between the various *Citrus* species. However, these SSR markers could not distinguish between accessions within a few groups in which cultivars have arisen by apparent spontaneous mutation, such as sweet oranges (*C. sinensis*), Clementines (*C. reticulata*), Satsumas (*C.reticulata*), and small-fruited acid limes (*C. aurantifolia*). Previous studies also have found few molecular polymorphisms within groups like these, consisting of cultivars developed by spontaneous mutation (Fang and Roose 1997; Bretó et al. 2001). Therefore, certain clonally derived varieties in the CVC were not distinguishable by these 24 SSR markers.

Rare alleles

Twenty accession-specific alleles (alleles detected in only one individual in the population) produced from 12 of the 24 SSR markers were identified (Supplementary Table 2). Markers TAA41, CAC23, and GT03 had the highest number of accession-specific alleles in the population with four, three, and three, respectively. Accessions displaying unique alleles may represent wild germplasm or wild derivatives. Of the four putative ancestral *Citrus* groups, mandarins and papedas had the highest proportion of unique alleles, whereas the pummelo and citron accessions had very few unique alleles (two and one, respectively). Additionally, several accession-specific alleles were produced from putative hybrid accessions, which suggest that the ancestral species or genotypes containing these alleles are not represented in the CVC. However, it is also possible that accession-specific alleles in these hybrid accessions were derived from a mutation event since SSR loci are known to have a high rate of mutation per locus per generation of 2.5×10^{-5} to 1×10^{-2} (Weber and Wong 1993).

Linkage mapping

Segregation information was collected for 24 SSR markers for the purpose of assigning these markers (including 10 previously scored by Kijas et al. 1997) to an existing *Citrus* linkage map (Roose et al. 2000). One hundred and eighty-three markers were used to produce this map, which consisted of SSR, isozymes, cDNA, RFLP, and genomic markers. At LOD 3.0, 15 of the 183 markers were either unlinked or could not be mapped. The deviation between map distance and two-point distance estimates had low mean chi-square values (range = 0.089–1.015, average = 0.424), which indicates a good map. Mean chi-square values are expected to be 1.0 when $P = 0.50$ (Roose et al. 2000). This map consists of 16 linkage groups with a total map distance of 679 cM (data not shown). The haploid number of chromosomes for *Citrus* is nine and perhaps with the addition of more markers to this map the linkage groups will be further reduced.

The SSR markers were assigned to 9 out of 16 (56.25%) of the linkage groups. Six SSR markers were mapped to linkage group number three making this linkage group the most saturated with SSR markers. Linkage group six had four SSR markers. Only one SSR marker was mapped to 7 of the 16 linkage groups. Seventeen of the SSR markers were assigned to one of the 16 linkage groups, while the remaining seven markers were either unlinked (one) or were monomorphic (six) in our mapping population. Assuming that a single marker covers 25 cM on either side (or to the end of a linkage group) the SSR markers covered approximately 50% of the total map (data not shown).

Population and group diversity

To examine the organization of genetic diversity within the total population and within specific *Citrus* groups, observed heterozygosities were calculated (Table 3). The total population was broken into groups (e.g. citrons) based on the taxonomic classifications used in the database at the NCGRCD. The population as a whole had an observed heterozygosity of 0.4318. Of the citrus groups that are thought to be true *Citrus* species, the citrons had the lowest observed heterozygosity. This low frequency may be explained by selfing, which would reduce the proportion of heterozygotes. R. R. Krueger and M. L. Roose (unpublished data) studied ISSR markers in open-pollinated *Citrus* seedlings obtained from seeds collected from single trees in India and China and found that a relatively high proportion of citron seedlings apparently originated from selfing. Barrett and Rhodes (1976) also reported that citrons produce vigorous selfed seedlings and tend to be highly homozygous, which is consistent with this data. The mandarins had the highest frequency of heterozygotes (0.4568) observed among the ancestral species. The pummelos had an observed heterozygosity of 0.4238. Pummelos are known to be self-incompatible and only reproduce sexually (Roose et al. 1995), which may explain why this group has a higher heterozygosity than the citrons.

Many of the groups thought to be hybrids of the naturally occurring forms of *Citrus* had a higher proportion of heterozygous loci than the groups classified as ancestral or citrus relatives. Limes, which are apparent hybrids of citrons and papedas (Scora 1975; Nicolosi et al. 2000) or a tri-hybrid cross of citron, pummelo and *Microcitrus* (Barrett and Rhodes 1976), had the highest observed heterozygosity of all the taxonomic groups at 0.6235. However, the oranges, long thought to be a natural hybrid of pummelo and mandarin (Scora 1975; Barrett and Rhodes 1976), had one of the lowest heterozygosity (0.4583) among the hybrid groups. Hybrid accessions are expected to lack unique alleles and be highly heterozygous, which is generally consistent with this data.

Population structure

Structure version 2.0 (Pritchard et al. 2000) was used to investigate population structure in the 370 accessions. A Bayesian clustering approach probabilistically assigns individuals to populations based on genotype. Individuals are assigned to populations or multiple populations if their genotype indicates admixture. This analysis makes no assumption about the particular mutation process and can utilize most genetic marker systems to infer population structure provided that the markers are not closely linked. Genotype data for 24 SSR markers were used to determine population structure among the various *Citrus* accessions (Fig. 1). When the microsatellite data set was trimmed to exclude four closely linked (<15 cM) markers, similar results were obtained (data not shown).

Structure analysis identified five populations in our *Citrus* data set: mandarins, pummelos, citrons, trifoliates, and a kumquat/papeda group. The trifoliate and kumquat accessions have the least amount of admixture within their groups. Mandarins, pummelos, citrons, and some papedas are thought to be true *Citrus* species, whereas kumquats (*Fortunella spp.*) and trifoliates (*Poncirus trifoliata*) are classified as separate but related genera. The rest of the *Citrus* "species", the apparent hybrids among naturally occurring forms, displayed admixture

Table 3 Observed heterozygosity listed by taxonomic citrus group (determined by NCGRCD database) and for the entire CVC population

Origin	Group	Number of individuals studied	Observed heterozygosity
Ancestral	Citrons	29	0.2805
Ancestral	Pummelos	89	0.4238
Ancestral	Papedas	15	0.4367
Ancestral	Mandarins	105	0.4568
Hybrid	Rough lemons	5	0.4336
Hybrid	Oranges	10	0.4583
Hybrid	Sweet limes	6	0.5330
Hybrid	Lemons	20	0.5366
Hybrid	Sour oranges	37	0.5140
Hybrid	Grapefruits	4	0.5604
Hybrid	Rangpurs	8	0.5661
Hybrid	Calamondins	3	0.5909
Hybrid	Limes	21	0.6235
Citrus Relative	Trifoliates	10	0.3047
Citrus Relative	Kumquats	8	0.4540
	Total population	370	0.4318

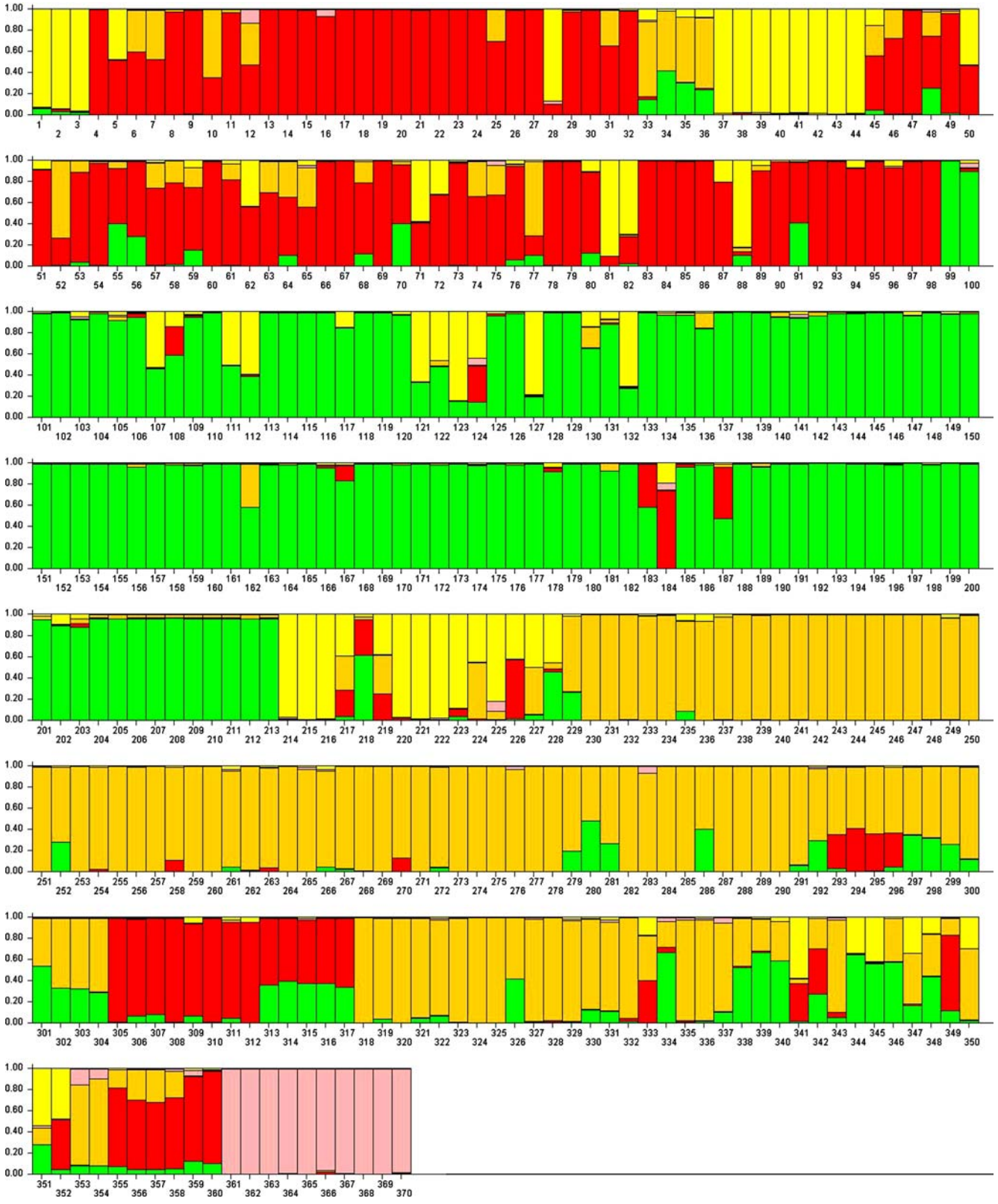

Fig. 1 Assignment of 370 *Citrus* accessions to five populations by Structure version 2.0. Each *individual bar* represents an accession from the CVC (see Supplementary Table 1 for cultivar information). Numbers 1–3 = Calamondins, 4–32 = Citrons, 33–36 = Grapefruits, 37–44 = Kumquats, 45–77 = Lemons, 78–98 = Limes, 99–203 = Mandarins, 204–213 = Oranges, 214–228 = Papedas, 229–304 = Pummelos, 305–312 = Rangpurs, 313–317 = Rough Lemons, 318–354 = Sour Oranges, 355–360 = Sweet Limes, and 361–370 = Trifoliates. The *Y-axis* displays the estimated membership of each individual in a particular cluster or population. *Red* = citron population, *Green* = mandarin, *Yellow* = kumquat/papeda, *Gold* = pummelo, *Pink* = trifoliate

between two or more of these five populations (Fig. 1). This analysis is consistent with previous claims that there are only a few true species of *Citrus*, which gave rise to many other *Citrus* species through hybridization (Scora 1975; Barrett and Rhodes 1976).

This analysis suggests that some accessions currently assigned to non-hybrid *Citrus* groups are actually hybrids or hybrid derivatives. For example, 'Djerok' citron (CRC 2456) (#7) (all numbers in this discussion refer to Fig. 1) was classified as a citron but it was suspected that this variety could be a hybrid (CVC archives). Structure analysis indicated that this variety shows admixture between the citron and the pummelo groups. Another example is #10 'Bengal' citron (CRC 3055), which displayed an admixture between the citron and pummelo groups with the majority of its alleles being donated from the pummelo group. This particular variety has always been classified as a citron but has also been suspected as a possible hybrid. Upon examination of the morphology, 'Bengal' has small-winged petioles, a pointed leaf apex, and smooth leaf blades, all of which are not characteristic of citron foliage but are features of pummelo morphology. Additionally, 'India' lime (#91) (CRC 2450) has always been classified as a lime; however, Structure analysis shows admixture between the citron and mandarin groups. The foliage of this variety has distinctive citron and mandarin characteristics. The fruit has a hollow core and the puffy rind characteristic of mandarins; however, the flowers are consistent with citron morphology, having purple coloration. Therefore, the morphology of these varieties is consistent with the admixture between groups that were identified by Structure using molecular data.

This analysis also correctly identified some known hybrid accessions and provides further support for their hybrid origin. For example, *C. halimii* (#28) (CRC 3780) has been classified as a citron but has been thought to be a hybrid of a citron and a kumquat based on morphological and phytochemical data (Scora et al. 1976). This variety displayed an admixture between the kumquat and the citron group with the majority of its genetic makeup being derived from kumquat. Since *C. halimii* did not have any unique alleles, this suggests that it may not be a valid species, but a wild hybrid. Another example is seen in the variety 'Hassaku' (CRC 3907) (#302), which is classified as a hybrid of pummelo by a mandarin (CVC archives). The SSR marker data reflects this ancestry and displays admixture between these groups. This clustering approach also identified several putative hybrid *Citrus* groups and affirmed the previously suspected parentage. For example, rough lemons (#313–317) were confirmed by this analysis as being natural hybrids between mandarin and citron.

An interesting result from this analysis is seen in the sweet orange, which is thought to be a natural hybrid, predominately mandarin with some pummelo traits (Scora 1975; Barrett and Rhodes 1976). Although the chloroplast genome of the sweet orange is derived from pummelo (Green et al. 1986; Nicolosi et al. 2000; Gulsen and Roose 2001b), this analysis suggests that sweet orange (#204–213) has a majority of its genetic makeup derived from mandarin and only a small proportion from pummelo. If the sweet orange was derived from one or more backcrosses to the mandarin, this would explain why sweet oranges had the lowest observed heterozygosity of all the hybrid *Citrus* groups. However, Nicolosi et al. (2000) found that sweet oranges had 50% of RAPD and SCAR markers derived from pummelo and the other 50% from mandarin.

Since mandarins, pummelos, citrons, and papedas are thought to be ancestral *Citrus* species and have given rise to many *Citrus* species through hybridization events, we expected that six populations would be identified: mandarin, pummelo, citron, papeda, kumquat, and trifoliate. The analysis was run multiple times without obtaining an output where the papedas clustered as a distinct population. This may be due to the low number of papeda accessions included in this data set and the high amount of admixture that most of these accessions display. It is possible that since most papeda accessions apparently have alleles derived from more than one of the larger populations identified, the few papeda accessions with little admixture are forced into another population instead of clustering as a separate small population. Lastly, it is also possible that additional molecular markers are needed to recognize the identification of a separate papeda population. It is notable that 6 of the 20 unique alleles occur in papeda accessions (Supplementary Table 2), suggesting papedas' distinction from other *Citrus* species.

Phylogenetic analysis of all accessions

The proportion of shared alleles (ps) was used to calculate genetic distances as −ln (ps) between all pairwise combinations in this population. This genetic distance measure assumes that all alleles are equally related (Bowcock et al. 1994), which is appropriate since there is no sequence information on the evolution (stepwise mutation or infinite allele) of these alleles. The main advantage of this distance measure is that it makes no assumptions about the structure of the population or the allele frequency within the population (Dangl et al. 2001). Therefore, this distance is appropriate for unnatural populations such as the CVC. The genetic distance matrix generated by Microsat was utilized to construct a neighbor-joining tree using PAUP (Supplementary Figure 1). Overall, this dendrogram indicates that accessions in the genus *Citrus* are not distant from *Fortunella*; however, *Citrus* is distant from *Poncirus*.

The neighbor-joining tree clustered the majority of the accessions into five main groups. The first group was the trifoliates (*P. trifoliata*) which were analyzed as the outgroup. There was strong support of monophyly for the trifoliate accessions, with a bootstrap value of 98%. The remaining groups in this tree did not have strong bootstrap support, perhaps, due to the large proportion of hybrids in this data set. The ten trifoliate varieties have little genetic variation and thus clustered closely

together with short branch lengths between accessions. In a previous study of 48 trifoliate orange accessions using isozyme, RFLP, and ISSR markers, Fang et al. (1997) also found that these same accessions were very similar.

The second group consisted of kumquats, calamondins, and *C. halimii* (CRC 3780). The kumquats were supported with a bootstrap value of 55%, and the calamondin accessions were highly supported at 98%. *C. halimii*, which is thought to be a hybrid between a kumquat and a citron (Scora et al. 1976), clustered with the kumquats as opposed to the citrons suggesting that *C. halimii* is more closely related to the kumquats than to the citrons. The three accessions of calamondins used in this study were very similar to one another. The calamondins are believed to be a natural hybrid of a kumquat and a mandarin (Barrett and Rhodes 1976). Since the calamondins are clustered with kumquats as opposed to the mandarins, they may be more closely related to the kumquats. Analysis with Structure 2.0 suggested that *C. halimii* and the calamondin accessions derived large portions of their alleles from the kumquat group. Both of these analyses (phylogenetic and Structure) support a hybrid origin and a close relationship of these accessions to the kumquats.

The next main cluster consists of citrons, rough lemons, lemon hybrids, limes, a papeda hybrid, a few pummelos, sour orange hybrids, and rangpurs. Citrons (*C. medica*) are thought to be an ancestral *Citrus* species that gave rise to lemons, limes, and rough lemons through various hybridization events. For example, Scora (1975) suggested that the rough lemon was a natural hybrid of a mandarin and a citron. The rough lemons clustered in the citron group as opposed to the mandarin group suggesting a close relationship to the citrons, as observed previously (Federici et al. 1998; Nicolosi et al. 2000). The Structure analysis suggests that the majority of their alleles were donated from citrons with the rest donated from mandarins.

All of the lemon and lime accessions clustered with the citrons. This same clustering pattern was also obtained in previous studies (Federici et al. 1998; Nicolosi et al. 2000). Lemons are thought to be natural hybrids of a citron and a lime (Scora 1975; Barrett and Rhodes 1976) or a hybrid of citron and sour orange (Gulsen and Roose 2001b). Limes are apparent hybrids of citrons and papedas (Scora 1975) or a tri-hybrid cross of citron, pummelo, and *Microcitrus* (Barrett and Rhodes 1976). *C. macrophylla*, which is sometimes classified as a papeda hybrid, also clustered with the lemons in the citron grouping. 'Cuban Shaddock' (CRC 1462), which is classified as a pummelo hybrid, clustered with lemon varieties in the citron group, but it has been documented in the CVC archives as having lemon foliage and some citron characteristics. 'Bergamot' (CRC 2881) and *C. vulgaris* (CRC 760) clustered within the citron group but are currently classified as sour orange hybrids despite being documented in the CVC archives as having lemon and citron characteristics. Although 'Cuban Shaddock' and 'Begamot' also clustered within the citron group in another marker study (Federici et al. 1998), *C. macrophylla* and limes did not. However, Nicolosi et al. (2000) found that lemons, limes, rangpurs, rough lemons, and 'Bergamot' all clustered within the citron group, which is consistent with these data.

The rangpurs were originally classified as a mandarin–lime. It was believed that the lime part of the name was a misnomer since they are only similar to limes in that they both have small flowers and are highly acidic (Swingle and Reece 1967). Webber (1943) believed that rangpurs were more similar to mandarins but thought they possibly were hybrids of limes and mandarins or possibly hybrids of limes and sour mandarins. Barrett and Rhodes (1976) reported that rangpurs are a *C. reticulata* (mandarin) genotype introgressed with a few genes of *C. medica* (citron). Singh and Schroeder (1962) suggested that the possible parentage of rangpurs was a cross between a mandarin and a rough lemon. More recently, Nicolosi et al. (2000) suggested that the hybrid origin of the rangpur lime was a cross of citron and mandarin hybrid. Therefore, the origin and parentage of the rangpurs has been unclear and still has not been resolved, but they have generally been classified as being similar to mandarins in most previous studies. However, the rangpurs in this study clustered with sweet limes with fairly strong bootstrap support (73%) within the citron group as opposed to the mandarin group, suggesting that the rangpurs are not as closely related to the mandarins as assumed previously. Structure analysis (Fig. 1) shows that the rangpurs received most of their genetic contribution from the citron group and only a small proportion from the mandarin group. This explains why the rangpurs clustered within the citron group with the limes as opposed to clustering with the mandarins. Overall, the SSR data suggest that the rangpurs may have resulted from a cross of lime and a mandarin with subsequent backcrossing to the lime parent. Federici et al. (1998) also obtained a phylogenetic tree using RFLP and RAPD markers that clustered rangpur accessions with a lime within the citron group.

The next major group was the mandarins, which did not form a well-defined clade, as did some of the other groups. The mandarin group split into several subclusters. Federici et al. (1998) also found that the mandarin group did not form a unified clade when hybrid and non-hybrid accessions were analyzed. When hybrid accessions were removed from the data set, the mandarins formed a discrete monophyletic clade (Fig. 2). Overall, the mandarin grouping consisted of mandarins and a few sour orange hybrids. Mandarins are considered to be a true *Citrus* species. The sour oranges are thought to be natural hybrids of a mandarin and a pummelo (Scora 1975; Barrett and Rhodes 1976). *C. miaray* (CRC 3574), *C. maderaspatana* (CRC 3225), and an unnamed variety (CRC 3175) which are all thought to be sour orange hybrids, all clustered within the mandarin group as opposed to the pummelo group suggesting a closer relationship to the mandarins.

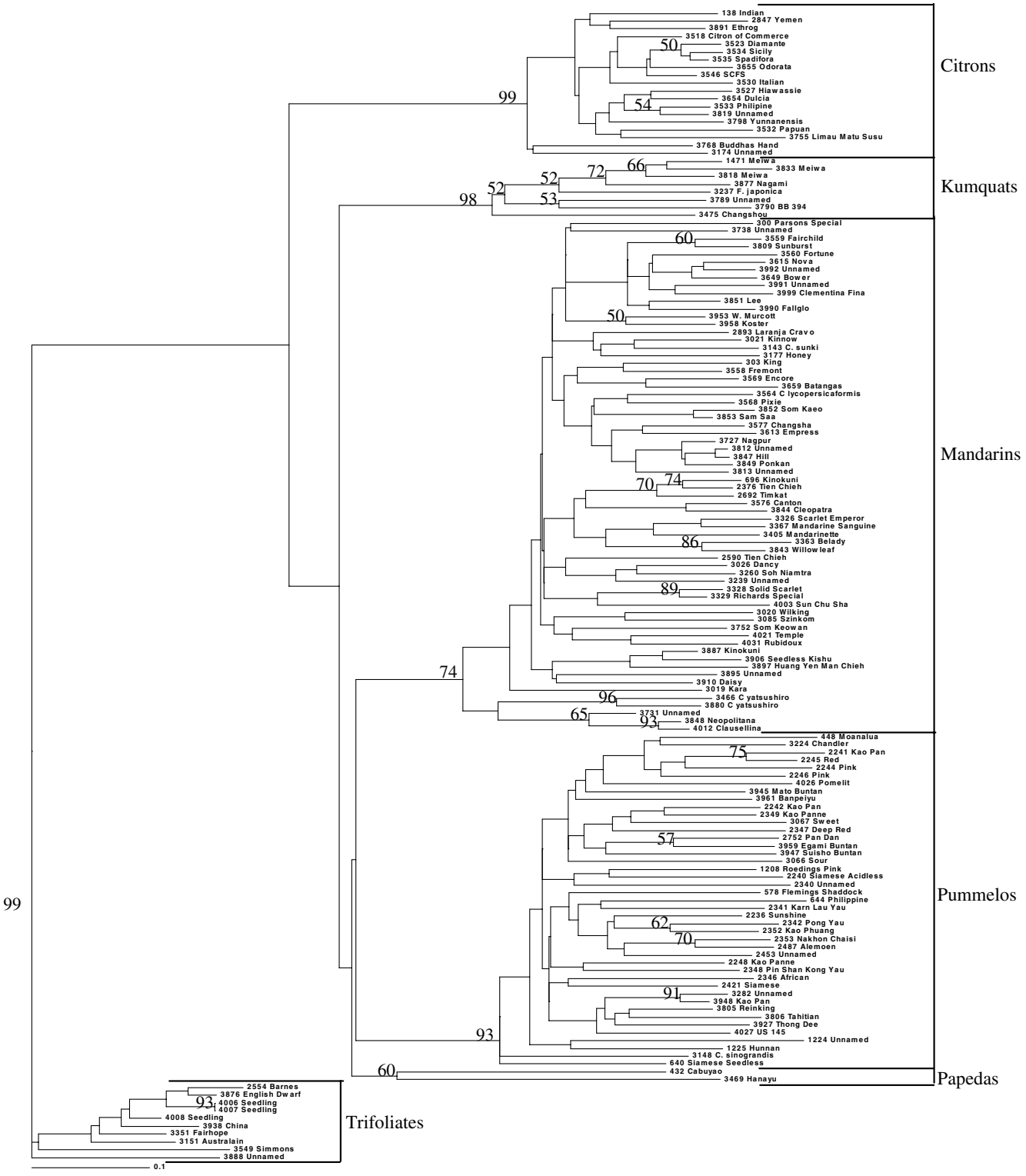

Fig. 2 Neighbor-joining tree showing the relationships of nonhybrid *Citrus* and accessions of related genera (*Fortunella* and *Poncirus*) as determined by PAUP using a distance matrix calculated from the proportion of shared alleles. Bootstrap analysis was performed with 500 replications. Bootstrap values greater than 50% are placed on branches

Several varieties that were classified as mandarins (*C.reticulata*) by Swingle (Swingle and Reece 1967), but were classified as separate species (*C. depressa*, *C. keraji*, *C. tachibana*, *C. nippokoreana*, and *C. leiocarpa*) by Tanaka (1977), clustered with papeda accessions (Hanayu CRC 3469 and Kabosu CRC 3943) in the mandarin group. The Structure analysis suggests that these accessions (107, 111, 121–123, and 127) have some

(approximately 50% or greater) of their genes derived from the papeda/kumquat group and a smaller proportion from the mandarin group. In previous studies, Federici et al. (1998) studied *C. keraji*, *C. tachibana*, and *C. nippokoreana*, and Nicolosi et al. (2000) examined *C. tachibana*. Both of these reports also found that these accessions clustered within the mandarin group.

The last main group consists of pummelos, pummelo hybrids, most of the sour oranges, a few sour orange hybrids, and grapefruit hybrids. Pummelo is also thought to be a true *Citrus* species (Scora 1975), which gave rise to sour oranges and grapefruits through hybridization (Scora 1975; Barrett and Rhodes 1976). The sour oranges clustered together with very short branch lengths between accessions and thus were very similar to one another. The sour oranges, some sour orange hybrids, grapefruit hybrids, and some pummelo hybrids have evidence of ancestry with the pummelos. The neighbor-joining tree provides further support of this suspected ancestry. Because numerous accessions derived by hybridization between mandarins and pummelos were included in this data set, neither the mandarin nor pummelo group in this tree had strong bootstrap support or was a well-resolved clade; however, some clusters within these groups were highly supported.

Phylogenetic analysis of nonhybrid accessions

To determine if the low bootstrap for many clusters was due to the inclusion of many hybrid accessions, the data set was trimmed to a smaller subset containing only 151 accessions of probable non-hybrid species of *Citrus*, and related genera *Fortunella* and *Poncirus*. (Hybrid accessions were identified based on an admixture obtained in the Structure analysis and from CVC archival data). This neighbor-joining tree divided the accessions into six clades: citrons, kumquats, mandarins, pummelos, papedas, and trifoliates (Fig. 2). Many clusters in this tree had higher bootstrap support than in the dendrogram that included hybrid accessions. Previous reports of the relationships of *Citrus* and its relatives have generally shown that bootstrap values tend to be higher within species than between species (Federici et al. 1998; Nicolosi et al. 2000). Overall, there was strong support for monophyly at the species level. However, the relationships between the citron, kumquat, mandarin, pummelo, and papeda clades are still not clear because low bootstrap support and short branch lengths were obtained between these clades.

The trifoliate, kumquat, pummelo, and citron monophyletic clades were highly supported with bootstrap values of 99, 98, 93, and 99%, respectively. The trifoliates were supported as distant from the rest of the clade (which included citrons, kumquats, mandarins, pummelos, and papedas) by a bootstrap value of 99% as also seen in the tree that included all taxa. Once again this demonstrates that *Poncirus* is distant from *Citrus*. However, *Fortunella* (kumquats) clustered within the *Citrus* clade and thus is not a distant relative. The clustering pattern of the kumquat accessions was very similar to the pattern seen in the previous tree. However, when hybrid accessions were removed, the bootstrap support for this clade increased substantially from 55 to 98%. The pummelo clade appears to be a sister taxon of the mandarin clade in both trees. However, the branch lengths between the pummelo and mandarin clades were short and bootstrap values were low between these clades.

The mandarin clade was also fairly well supported in this nonhybrid neighbor-joining tree with a bootstrap value of 74%. When hybrid accessions were removed the mandarin accessions clustered in a monophyletic clade instead of splitting into subgroups as it was seen when hybrid accessions were included. The relationships between accessions within the mandarin group were similar to those in the previous tree. Many of the mandarin accessions that had significant bootstrap support ($>70\%$) in the previous tree also had high bootstrap support in this tree. For example, 'Belady' (CRC 3363) and 'Willowleaf' (CRC 3843) clustered together in both trees and were supported with bootstrap values of 83 and 86% in the trees with and without hybrid taxa. Furthermore, *C.yatsushiro* (CRC 3466) *and C. yatsushiro* (CRC 3880) clustered together as previously and were supported with a bootstrap value of 96%.

Phylogenetic and structure analysis

The phylogenetic tree and the Structure analysis of the varieties in the CVC allowed the determination of parentage or classification of varieties that had some questionable passport information. An example of this is 'Reinking' pummelo (CRC 3805), which is documented in the CVC archives as a controlled pollination of 'Kao Phuang' (pummelo) × 'Shamouti' (sweet orange). It has been suspected that this controlled pollination was possibly contaminated with pummelo pollen. The fruit from 'Reinking' is described as a medium-sized pummelo. 'Reinking' clustered with 'Thong Dee' pummelo (CRC 3927) and 'African' pummelo (CRC 2346) and was close to the 'Kao Panne' accessions and 'Kao Phuang'. The Structure analysis shows that 'Reinking' (# 275) has no admixture indicating that this supposed controlled pollination was contaminated with pummelo pollen as opposed to the intended 'Shamouti' sweet orange.

Another example of questionable passport information is CRC 3163, which is supposed to represent *C.indica. C. indica* was classified by Swingle (Swingle and Reece 1967) as a mandarin (*C. reticulata*), but classified as a separate species by Tanaka (1977). This variety clustered with 'Unnamed' (CRC 3056 and CRC 3797) papeda accessions suggesting a closer relationship with papedas as opposed to mandarins. In the CVC archives, this accession was characterized and documented as probably not representing *C. indica* as described by Swingle and Reece (1967) and was thought to be a hybrid of *C.indica*. It was originally imported as a seed from India, and is monoembryonic. The fruit was also characterized in the CVC archives as mandarin-like but having a mammiform apex like a rough lemon and being extremely

sour. Additionally, the seeds were not typical of citrus seeds in that they were large and had bright green cotyledons (CVC archives). The Structure analysis of *C. indica* (#124) displays an admixture between four of the five populations with most of its genome (~80%) equally contributed from the citron and kumquat/papeda groups and a small proportion from the mandarin group. This accession contained one unique allele of TAA3 suggesting that while it may not be *C. indica*, it could represent unique germplasm. Currently, the collection does not have another accession of *C. indica*; therefore, CRC 3163 could not be compared to a true accession of *C. indica* to determine if it was a hybrid of *C. indica*. However, based on molecular and morphological analysis this accession should not be classified as a mandarin.

The neighbor-joining tree and Structure analysis of the SSR data, which has led to a deeper understanding of the relationships of citrus taxa and their hybrid origins, are approaches with different assumptions. However, the results complement one another to provide a robust analysis. The neighbor-joining tree is a distance-based method that utilizes the proportion of shared alleles to calculate distances between taxa and subsequently plot these distance relationships in the form of a tree. Structure utilizes a Bayesian clustering approach to probabilistically assign individuals to populations based on their genotypes and attempts to find population structure in which each population is in linkage equilibrium and Hardy–Weinberg equilibrium. These two analyses seem to support the hypothesis that there are only a few naturally occurring forms of *Citrus* (*C. medica*, *C. maxima*, *C. reticulata*) as previously suggested by Scora (1975) and Barrett and Rhodes (1976). Additionally, these analyses also support the idea that most other *Citrus* "species" are hybrids derived from these taxa and provide further support of their previously suspected ancestry.

Acknowledgements We would like to thank the U.S. Department of Agriculture for the funding to support this project. In addition, we thank the Genetic Graduate Group at the University of California, Riverside for providing partial funding of this project. Also, thanks to Dr. Mark Springer and Dr. Angela Burk-Herrick (University of California, Riverside) for providing assistance on PAUP and phylogenetic interpretation. Thanks to Dr. Tracy Kahn and Ms. Toots Bier for their support in providing historical and morphological information on *Citrus* varieties from the Citrus Variety Collection (CVC archival data). Lastly, thanks to colleagues at USDA-ARS Plant Genetic Resource Conservation Unit in Griffin, GA, USA for reviewing this manuscript.

References

Ahmad R, Struss D, Southwick SM (2003) Development and characterization of microsatellite markers in *Citrus*. J Amer Soc Hort Sci 128:584–590

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

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Bretó MP, Ruiz C, Pina JA, Asins MJ (2001) The diversification of *Citrus clementina* Hort. Ex Tan., a vegetatively propagated crop species. Mol Phyl and Evol 21:285–293

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum. Theor Appl Genet 93:190–198

Dangl GS, Mendum ML, Prins BH, Walker MA, Meredith CP, Simon CJ (2001) Simple sequence repeat analysis of a clonally propagated species: a tool for managing a grape germplasm collection. Genome 44:432–438

Fang DQ, Roose ML (1997) Identification of closely related *Citrus* cultivars with inter-simple sequence repeat markers. Theor Appl Genet 95:408–417

Fang DQ, Roose ML, Krueger RR, Federici CT (1997) Fingerprinting trifoliate orange germplasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. Theor Appl Genet 95:211–219

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 94:812–822

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

Gulsen O, Roose ML (2001a) Lemons: diversity and relationships with selected *Citrus* genotypes as measured with nuclear genome markers. J Amer Soc Hort Sci 126:309–327

Gulsen O, Roose ML (2001b) Chloroplast and nuclear genome analysis of the parentage of lemons. J Amer Soc Hort Sci 126:210–215

Herrero R, Asins MJ, Carbonell EA, Navarro L (1996) Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. Theor Appl Genet 92:599–609

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

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* × *domestica* borkh. Core subset collection. Theor Appl Genet 97:671–683

Jain MS, Brar DS, Ahloowalia BS (2002) Microsatellites and molecular breeding: exploitation of microsatellite variability for the analysis of a monotonous genome. In: Molecular techniques in crop improvement. Kluwer, London, pp 85–135

Jarrell DC, Roose ML, Traugh SN, Kupper RS (1992) A genetic map of *Citrus* based on the segregation of isozymes and RFLPs in an intergeneric cross. Theor Appl Genet 84:49–56

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

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among Maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

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*. Hort Sci 30:1063–1067

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Kidd JR, Cavalli-Sforza LL (1997) Microsat 1.5: a computer program for calculating various statistics on microsatellite allele data. Available via http://www.hpgl.stanford.edu/projects/microsat/programs/

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

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Roose ML, Soost RK, Cameron JW (1995) Citrus. In: Smart J, Simmonds NW (eds) The evolution of crop plants. 2nd edn, Longman, Essex, pp 443–449
Roose ML, Fang D, Cheng FS, Tayyar RI, Federici CT, Kupper RS (2000) Mapping of the *Citrus* genome. Acta Hortic 535:25–32
Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386
Scora RW (1975) On the history and origin of *Citrus*. Bull Torr Bot Club 102:369–375
Scora RW, Kumamoto J, Esen A, Stone BC (1976) A phytochemical investigation of *Citrus halimii*. Biochem Sys and Ecol 4:255–258
Singh R, Schroeder CA (1962) Taxonomic and physiological relationships of the so-called mandarin-lime group of citrus. Proc Amer Soc Hort Sci 80:291–295
Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744
Swingle WT, Reece PC (1967) The botany of Citrus and its wild relatives. In: Reuther W, Webber HJ, Batchelor LD (eds) The Citrus Industry. Vol 1. University of California Press, Berkeley, pp 190–430
Swofford DL (1998) PAUP*. Phylogenetic analysis using parismony (*and other methods). Version 4. Sinauer Associates, Sunderland
Tanaka T (1977) Fundamental discussion of Citrus classification. Stud Citrologia 14:1–6
Webb DM, Knapp SJ (1990) DNA extraction from a previously recalcitrant plant genus. Plant Mol Bio 8:180–185
Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128
Webber JH (1943) Cultivated varieties of citrus. In: Webber HJ, Batchelor LD (eds) The Citrus Industry. Vol 1. University of California Press, Berkeley. pp 475–668