ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization **8**(1); 16–25
doi:10.1017/S1479262109990098

Genetic reticulation and interrelationships among *citrullus* species as revealed by joint analysis of shared AFLPs and species-specific SSR alleles

Padmavathi Nimmakayala[1]*†, Yan R. Tomason[1,2‡], Jooha Jeong[1], Sathish K. Ponniah[1], Anoji Karunathilake[1], Amnon Levi[3], Ramasamy Perumal[4] and Umesh K. Reddy[1†]

[1]*Department of Biology and Gus R. Douglass Institute, West Virginia State University, Institute, WV 25112, USA,* [2]*Department of Selection and Seed Production, Dnepropetrovsk State Agrarian University, Voroshilov 25, Dnepropetrovsk 49600, Ukraine,* [3]*USDA, ARS, US Vegetable Laboratory, 2875 Savannah Highway, Charleston, SC 29414, USA and* [4]*Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX 77843-2132, USA*

Received 16 April 2009; Accepted 26 June 2009 – First published online 22 July 2009

Abstract

Thirty-one accessions of *Citrullus* spp. belonging to *Citrullus lanatus* var. *lanatus*, *C. lanatus* var. *citroides* and *Citrullus colocynthis* were subjected to phylogenetic analysis using combined datasets of amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs). Tree topologies inferred by neighbour-joining analysis have resolved the phylogenic relationships among the species with special reference to established taxonomic classification. In this study, we have clearly resolved species boundaries of various taxa of *citroides*, *lanatus* and *colocynthis* into three well-supported clusters. Clustering pattern of principal component analysis with the shared polymorphisms using the subsets of data between any two taxon combinations helped to elucidate the introgression and interrelationships among the species. We report two major groups of *C. lanatus* taxa, one of which has undergone wide introgressions with the taxa of *C. lanatus* var. *citroides* and *C. colocynthis*. In this paper, we identified 583 AFLP bands that are polymorphic within the var. *lanatus* of *C. lanatus*, which is the largest set ever reported. The species-specific diagnostic SSRs and polymorphic AFLPs that are informative within and between the taxa reported in this paper would be immensely useful for future studies of these economically important genera.

Keywords: AFLP; *Citrullus* species; microsatellites; molecular phylogeny; watermelon

Introduction

Watermelon is an important crop in the United States, whose farm value is estimated at $340 million (www.watermelon.org). Economic and nutraceutical importance of this crop is rapidly increasing throughout the world. Severe bottlenecks in the genetic background of cultivated watermelon have been reported based on the DNA marker analysis of genetic similarities (Navot and Zamir, 1987; Zhang *et al*., 1994; Lee *et al*., 1996; Levi *et al*., 2001, 2004).

According to Livingstone (1857), Meeuse (1962) and Pitrat *et al*. (1999), the species *Citrullus lanatus* (Thunb.) Matsum and Nakai originated in Kalahari region of Namibia and Botswana (Bates and Robinson, 1995;

* Corresponding author. E-mail: padma@wvstateu.edu
† Authors contributed equally.
‡ E-mail: yantomason75@mail.ru

Robinson and Decker-Walters, 1997; Ellul *et al*., 2007). The species *C. lanatus* includes two botanical varieties, namely var. *lanatus* (Bailey) and var. *citroides* (Mansf). Cultivated watermelons belong to var. *lanatus* and have endocarps in wide-ranging colours. The var. *citroides* is cultivated in southern Africa, and also called 'Tsamma' or 'citron' melon, whose rind is used as preservative in pickles (Burkill, 1985; Jarret *et al*., 1997; Jeffrey, 2001). The citron fruits have green- or white-coloured flesh and their taste may vary from bland to bitter. Seed production fields should be isolated from weedy citron types since these two botanical varieties cross readily (Wehner, 2007). The species *Citrullus colocynthis* (Schrad) is a perennial herb known as bitter apple and is a desert species with a rich history as a medicinal plant (Dane *et al*., 2007). T.W. Whitaker considered *C. colocynthis* to be a likely ancestor of watermelon as it is morphologically similar to *lanatus*, is freely intercrossable and produces fertile hybrids (Wehner, 2007). Dane *et al*. (2007) reported divergent lineages of *colocynthis* that are from tropical Asia and Africa, now widely distributed in the Saharo-Arabian phylogeographic region of Africa and in the Mediterranean region.

In earlier reports, isozyme and random amplified polymorphic DNA (RAPD) markers were used extensively in molecular diversity and phylogenetic analyses in *Citrullus* spp. (Zamir *et al*., 1984; Navot and Zamir, 1987; Biles *et al*., 1989; Levi *et al*., 2001a, b). Hillis (1994) and Harris (1995) argued against the use of RAPDs in phylogenetic analysis because of their questionable homology assessments. In contrast, the simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) markers are highly repeatable and in combination they are very efficient in resolving phylogenetic relationships, germplasm evaluation, quantitative trait locus (QTL) analysis and building integrated genetic maps (Ramakrishnan *et al*., 2004; Koopman *et al*., 2008). AFLP, in particular, has proven useful for estimating relationships among closely related species in a wide variety of lineage (Spooner *et al*., 2005; Pimentel *et al*., 2007; Thaler *et al*., 2008).

Levi *et al*. (2001) used 662 RAPD markers to construct a dendogram with little resolution, which did not separate taxa from *lanatus* and *citroides* clearly. Later, Levi *et al*. (2004) used combined analysis of AFLP and inter-SSR among heirloom watermelons and concluded that this joint analysis resolved better genetic diversity than RAPDs. Valuable phylogenetic information pertaining to the genus *Citrullus* was generated using chloroplast-specific polymerase chain reaction–restriction fragment length polymorphism and comparative chloroplast gene sequence data (Dane, 2002; Dane *et al*., 2004; Dane and Liu, 2007). Although the resolution obtained in these studies was higher than that in RAPDs, several studies in other genera concluded that genome-wide character sets will have greater potential as phylogenetic markers (Eriksen and Töpel, 2006; Koopman *et al*., 2008). AFLPs were also widely used for resolving cytonuclear conflicts in several other organisms (Sullivan *et al*., 2004; Kyndt *et al*., 2005; Schönswetter *et al*., 2007).

The current study aims to use AFLPs and SSRs jointly to resolve genome-wide molecular phylogenies among the same *Citrullus* taxa that were analyzed previously using RAPDs by Levi *et al*. (2001) and further to subject the shared polymorphisms between two taxa combinations to principal component analysis (PCA) to resolve interrelationships. We also report 30 newly developed polymorphic SSRs and relevant primer sequence information for use of the watermelon-breeding community.

Materials and methods

Plant material and DNA isolation

Seeds of 31 accessions (Table 1) were kindly provided by Dr Robert Jarret, Plant Genetic Resources Conservation Unit, USDA–ARS, Griffin, GA, 30 223. DNA was extracted from leaf tissues using the method described in the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany).

AFLP analysis

AFLP analysis was carried out using the protocols and kits developed by LI-COR Biosciences (Lincoln, NE, USA, www.licor.com). The *EcoR*I and *Mse*I enzyme-digested products were ligated to respective restriction site (RS)-specific adapters and diluted tenfold. Diluted adapter-ligated templates were pre-amplified using adapter-specific primers with overhangs of A and C for *EcoR*I and *Mse*I, respectively. Pre-amplified products were further diluted 20-fold and subjected to selective amplification using IR-700 or IR-800 labelled *EcoR*I-AXX primers and unlabelled *Mse*I-CXX primers using standard touchdown PCR conditions (Vos *et al*., 1995). A total of 35 primer combinations were used (Table 2). Amplified products were denatured and resolved on a LICOR-4500 genotyper.

SSR development

Genomic DNA and fruit tissue-specific cDNA were simultaneously digested with a set of restriction enzymes. Purified, digested DNA was ligated to AP-11 (5′CTCTTGCTTAGATCTGGACTA3′) and AP-12 (5′*p*TAGTCCAGATCTAAGCA-AGAGCACA3′, where $p = 5'$ phosphate) adapters and hybridized with biotin-labelled oligos, which contain the repeat motifs. DNA fragments that are hybridized

https://doi.org/10.1017/S1479262109990098 Published online by Cambridge University Press

Table 1. Phenotypic classifications and origin of the *Citrullus* spp. used in the study

Entry	Accession name and seed colour	Origin
Citrullus lanatus var. *lanatus**		
PI 162667	Mendocina, very long fruits	Argentina
PI 165451	Reddish black and yellowish red seeds (mixed)	Oaxaca, Mexico
PI 169289	Yenidunya, yellow with black streaks	Bursa, Turkey
PI 169290	Kurbagi Alacasi, yellow with reddish black	Bursa, Turkey
PI 185635	Egusi, golden yellow	Ghana
PI 186975	Egusi, golden yellow with black spots	Ghana
PI 189316	Cream with black margin and yellow margin	Nigeria
PI 189317	Cream with two black spots	Nigeria
PI 192937	Greenskin red, greyish black	Shanghai, China
PI 203551	Greyish black	New Mexico, USA
PI 248178	Mangara, golden yellow	Zaire, central Africa
PI 249010	Golden yellow	Kaduna, Nigeria
PI 270306	Greyish black	Philippines
PI 270550	–	Ghana
PI 271778	Hybrid melon from Nelspruit	Transvaal, S. Africa
Charleston Grey	Cultivated (local)	USA
C. lanatus var. *citroides**		
PI 244018	Cream, cream with black and greyish black spots	Transvaal, S. Africa
PI 244019	Delagoa, skin dark green, hard, with spots	Transvaal, S. Africa
PI 248774	Tsamma, fruit globose; pulp bitter	Namibia
PI 255137	Large Tsamma, yellow with black spots	Transvaal, S. Africa
PI 270562	Greyish and black (mixed)	South Africa
PI 271779	Hybrid from Nelspruit, reddish black seeds	Transvaal, S. Africa
PI 299378	Green cattle melon, grey seeds	South Africa
PI 299379	Pale cattle melon, grey seeds	South Africa
PI 482252	Flesh yellow, seeds orange with yellow border	Zimbabwe
*Citrullus colocynthis**		
PI 346082	Ash and blackish ash coloured (mixed)	Helmand, Afghanistan
PI 386016	Blackish ash coloured	Iran
PI 386018	Ash, greyish black and yellow with black spots (mixed)	Iran
PI 386024	Ash, greyish black and yellow with black spots (mixed)	Iran
PI 386025	Ash, greyish black and yellow with black spots (mixed)	Iran
PI 386026	Ash, greyish black and yellow with black spots (mixed)	Iran

* As reported in the Germplasm Resource Information Network reference base.

with the repeat oligos were separated using streptavidin beads. These repeat motif-enriched fragments were separated from the beads in an alkaline buffer for purification using a Qiagen PCR purification kit. These enriched fragments were cloned (TOPO cloning kit; Invitrogen, Madison, WI, USA) and 96 randomly picked clones were sequenced. The sequences with repeat motifs were identified and used for designing SSR primer pairs. PCR conditions for SSRs were used as per Reddy *et al.* (2001), and gel electrophoresis was carried out using super fine resolution agarose (www.amresco-inc.com).

Data scoring and analysis

Minor AFLP polymorphisms that were not uniformly amplified (e.g. were faint or not distinct in some genotypes) were eliminated from the analysis. Similarly, stutter and background bands were not considered while scoring SSR markers. The presence or absence of each fragment was scored as a binary unit character (1 = present and 0 = absent) in the case of AFLPs, and in the case of SSRs, scoring was the presence or absence of putative alleles. Genetic similarities based on Jaccard's coefficients (Jaccard, 1908) were calculated using the SIMQUAL program of the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) Version 2.0 software package (Rohlf, 1987). The resulting genetic similarity indices were used to generate a tree using the neighbour-joining (NJ) method (Saitou and Nei, 1987). PAUP*4.0 was used to generate 1000 bootstrap replicates for testing the reliability of the dataset and to draw a consensus tree. Principal component analysis (PCA) based on the genetic similarity matrices was performed using of NTSYS-pc.

Results

A range of 146–324 bands were amplified per primer combination totalling 6879 AFLP markers that were

https://doi.org/10.1017/S1479262109990098 Published online by Cambridge University Press

Table 2. Total number of amplified and polymorphic fragments with 35 amplified fragment length polymorphism primer combinations

No.	Primer combination	Amplified fragments	Polymorphic fragments	Polymorphism (%)
1	E-ACA/M-CAC	199	85	42.7
2	E-ACC/M-CAC	162	63	38.9
3	E-ACG/M-CAC	199	52	26.1
4	E-AGG/M-CAC	240	78	32.5
5	E-AAC/M-CAC	212	95	44.8
6	E-AAG/M-CAC	205	35	17.1
7	E-ACT/M-CAC	146	115	78.8
8	E-AGC/M-CAC	247	85	34.4
9	E-ACA/M-CAA	134	87	64.9
10	E-ACC/M-CAA	159	63	39.6
11	E-ACG/M-CAA	156	95	60.9
12	E-AGG/M-CAA	324	117	36.1
13	E-AAC/M-CAA	212	105	49.5
14	E-AAG/M-CAA	309	112	36.2
15	E-ACT/M-CAA	235	132	56.2
16	E-AGC/M-CAA	160	76	47.5
17	E-ACA/M-CTT	151	76	50.3
18	E-ACC/M-CTT	149	89	59.7
19	E-ACG/M-CTT	178	115	64.6
20	E-AGG/M-CTT	160	123	76.9
21	E-AAC/M-CTT	152	85	55.9
22	E-AAC/M-CTA	232	104	44.8
23	E-AAG/M-CTA	167	82	49.1
24	E-ACT/M-CTA	246	66	26.8
25	E-AGC/M-CTA	176	85	48.3
26	E-ACA/M-CTC	252	58	23.0
27	E-ACC/M-CTC	249	105	42.2
28	E-ACG/M-CTC	162	63	38.9
29	E-AGG/M-CTC	187	52	27.8
30	E-ACA/M-CAG	206	85	41.3
31	E-ACC/M-CAG	154	91	59.1
32	E-ACG/M-CAG	183	79	43.2
33	E-AGG/M-CAG	176	146	83.0
34	E-AAC/M-CAG	183	112	61.2
35	E-AAG/M-CAG	199	78	39.2
Total (mean)		6879 (196)	3089 (88)	45.0

amplified from 35 primer combinations. Of these, 3089 bands (45%) were polymorphic among the 31 accessions sampled (Table 2). We counted specific bands that were polymorphic within the individual species and the shared bands between taxa. Five hundred and eighty-three amplicons were polymorphic within the taxa of *lanatus*, and 505 and 194 markers were polymorphic within *citroides* and *colocynthis* groups, respectively. The shared bands between any two taxa groups were also counted to determine how the shared polymorphisms will resolve interrelationships. Six hundred and fifty-two bands were shared between *lanatus* and *citroides*, 756 shared bands between *lanatus* and *colocynthis* and 620 bands between *colocynthis* and *citroides*. However, it is very difficult to identify species-specific bands using AFLP as they are dominant marker and individual bands are generally shared among species. The markers that are polymorphic within *lanatus* can be very important for the improvement of cultivated watermelon as they can be used for mapping and marker-assisted selection. In this study, AFLP markers were more polymorphic than SSRs within *lanatus* group.

Thirty SSRs amplified 169 alleles in 31 watermelon accessions. A range of 2–12 alleles were amplified per SSR. Primer sequence information and the range of amplified product sizes across the species are presented in Supplementary Table 1, available online only at http://journals.cambridge.org. The number of specific alleles was 50, 60 and 59 specific to var. *lanatus,* var. *citroides* and *C. colocynthis,* respectively.

To start with, diversity analysis was carried out for both AFLP and SSR separately, which interestingly resulted in similar grouping pattern of the selected accessions with the datasets when analyzed together (trees not

https://doi.org/10.1017/S1479262109990098 Published online by Cambridge University Press

presented). Combined AFLP and SSR analyses revealed that the genetic distances % (GDs) within the accessions of various species were 42, 38 and 34 for *lanatus, citroides* and *colocynthis,* respectively. Distance between the species groups was wider than within the species, as predicted. Our study revealed that the GD between the *lanatus* and *citroides* taxa was 40, and the GD between the species *colocynthis* and *citroides* was 43. The GD between *lanatus* and *colocynthis* was estimated to be 46, indicating that the var. *citroides* is closely related to the species *colocynthis* than var. *lanatus*.

A NJ phenogram was constructed using the combined datasets of AFLPs and SSRs (Fig. 1). Trees were rooted using the taxa *C. colocynthis* as the out-group. A monophyletic cluster of *colocynthis* was basally resolved with a bootstrap value support of 95. The second branch of the tree was of the species *C. lanatus*, which further split into two sister clades of var. *citroides* and var. *lanatus*. This split was supported with a breeding value (BV) of 97. The *lanatus* cluster had six *lanatus* subclusters, which were closely resolved in the middle of the tree with a BV range from 57 to 82. The sixth subcluster of *lanatus* was an exception in the whole analysis as two of the *citroides*' taxa from southern Africa were grouped into this subcluster. A well-supported monophyletic cluster composed of seven *citroides* accessions is resolved on top of the tree.

We separated the data into four subsets, which were polymorphic within different groups of (1) *lanatus* vs. *citroides* (Fig. 2), (2) *lanatus* vs. *colocynthis* (Fig. 3) and (3) *citroides* vs. *colocynthis* (Fig. 4). When PCA was done using the dataset of *lanatus* alone (figure not presented), the first three eigen vectors absorbed 36.90, 17.78 and 12.48 totalling 67.16% of total variation, indicating the robustness of the dataset and reliability of the analysis. The *lanatus* accessions PI 271778 (S. Africa), PI 169289

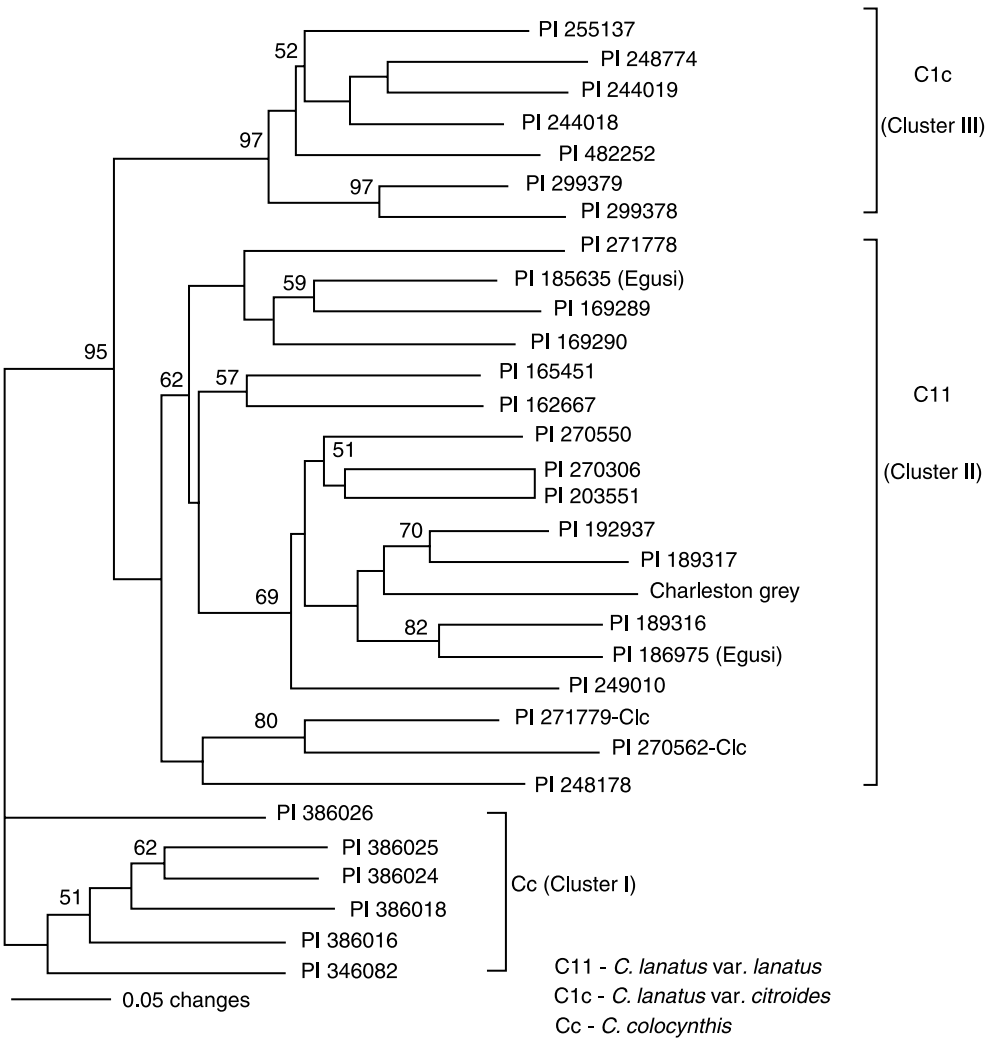

Fig. 1. NJ phenogram of 31 *Citrullus* spp. using AFLPs and newly captured SSRs. Numbers shown at different nodes represent percentage confidence limits obtained in the bootstrap analysis.

https://doi.org/10.1017/S1479262109990098 Published online by Cambridge University Press

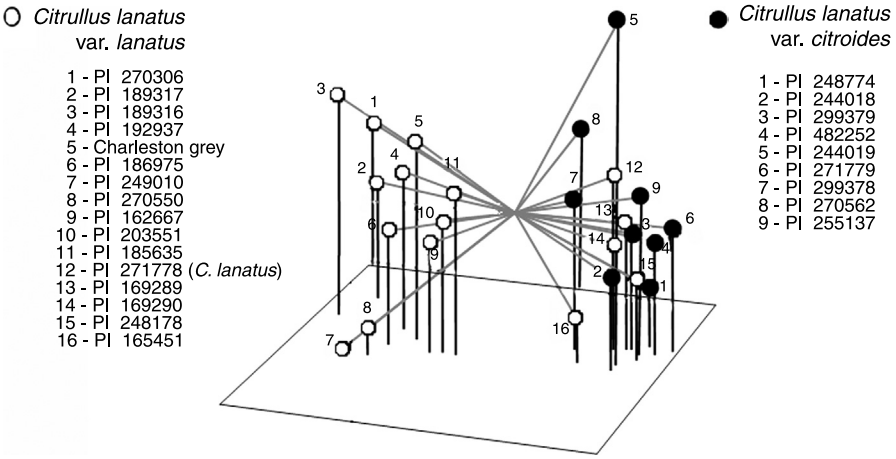

Fig. 2. Three-dimensional picture of principal component analysis estimated using AFLP and SSRs' genetic similarity matrix of 16 *C. lanatus* var. *lanatus* and 9 *C. lanatus* var. *citroides*' accessions.

(Turkey), PI 248178 (central Africa), PI 169290 (Turkey) and PI 165451 (Mexico) were clustered together with the *citroides*' taxa (cluster II), and the rest of the 11 *lanatus* taxa grouped as cluster I (Fig. 2). These *lanatus*' and *citroides*' PIs from cluster I may have been transitional between these two taxa. When we combined the dataset of *lanatus* with *citroides*, the PCA revealed that cluster I from the *lanatus*-specific PCA clustered with the six *citroides* types (Fig. 2). This PCA was supported by eigen vectors I (31.23), II (16.40) and III (9.75), cumulatively explaining 57.38% of the total possible variation. The PCA with *lanatus* and *colocynthis* explained about 65.63% of the total variation (eigen vectors I = 33.57, II = 19.58 and III = 12.48), and also reveals the clustering of *colocynthis* types with the cluster II from the *lanatus*-specific PCA (Fig. 3).

To understand the species relationship between the taxa *colocynthis* and *citroides*, we further separated the dataset of these two species and subjected them to PCA separately (Fig. 4). The analysis revealed that the first three eigen vectors that were used to construct the multidimensional spectrum contributed 32.37, 25.47 and 14.57% of variation with a total of 72.42% of the total possible variance. The pattern of clustering from this PCA was congruent with the results of GD analysis, confirming that these two species are closely interrelated.

Discussion

In the current study, accessions belonging to taxa *citroides*, *lanatus* and *colocynthis* were clearly resolved into three well-supported clusters that are consistent with the classical taxonomic nomenclature. The accessions in the current study are from various wide

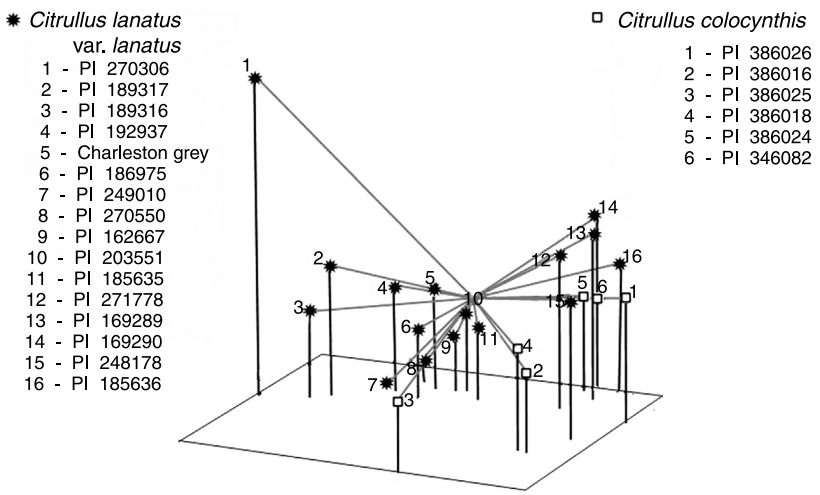

Fig. 3. Three-dimensional picture of principal component analysis estimated using AFLP and SSRs' genetic similarity matrix of 16 *C. lanatus* var. *lanatus* and 6 *C. colocynthis*' accessions.

https://doi.org/10.1017/S1479262109990098 Published online by Cambridge University Press

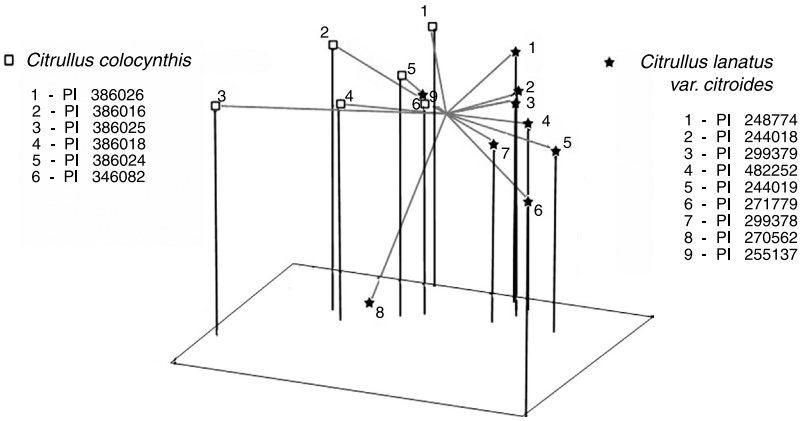

Fig. 4. Three-dimensional picture of principal component analysis estimated using AFLP and SSRs' genetic similarity matrix of nine *C. lanatus* var. *citroides* and six *C. colocynthis*' accessions.

geographic areas drawn from the countries Afghanistan, Argentina, China, Cyprus, Egypt, Ghana, Iran, Morocco, Namibia, Nigeria, Philippines, South Africa and Zaire. Irrespective of their geographical origin, in NJ analysis, all accessions were grouped with one of the three species-specific clusters. Geographic differentiation of genetic diversity of the genus *Citrullus* might have been considerably weakened by increased human interchange of cultivated watermelon and its associated weed species (*citroides* and *colocynthis*) among the countries in the past several years.

Some accessions of *lanatus* in the USDA-ARS germplasm show particular phenotype usually known as 'egusi' seed types. The egusi watermelon is commonly cultivated in Ghana, Nigeria and Congo, where the protein- and carbohydrate-rich seeds are used as a regular part of the human diet and fruits as cattle feed. Sometimes these fruit types are confused with *C. colocynthis* type, but they are cultivated watermelons (Wehner, 2007). In the current study, both of the egusi types were clustered with the *lanatus* group.

The current study indicated that the joint analysis of AFLP and SSR is very effective for phylogenetic analysis of *Citrullus* spp. Studies such as Špunarová *et al.* (2005) in barley, Gillaspie *et al.* (2005) in *Vigna*, Maluf *et al.* (2005) in *Coffea*, Saini *et al.* 2004 in rice, Perumal *et al.* (2007) in sorghum are few examples of copious published works, where joint analysis of AFLP and SSR datasets proved that these markers analyzed together have unprecedented utility for phylogenetic studies. Instead of generating a particular sequence-specific tree that does not necessarily reflect the true species tree, especially among the closely related and potentially interbreeding species such as *Citrullus* spp., where reticulate evolution might have occurred, the simultaneous analysis of many loci representing the whole genome has the potential to generate true phylogenies. This is an advantage of the genotyping techniques that scan diversity at loci across the genome, as the accuracy of measurements on GDs increases with the number of loci used (Travis *et al.*, 1996; Schmidt and Jensen, 2000). Our study indicated that the GD between *lanatus* and *colocynthis* is wider than the GD between the species *citroides* and *colocynthis*. These results are consistent with the GDs estimated by Levi *et al.* (2001).

SSRs are simple to use multiallelic and co-dominant marker systems that are sequence based and produce highly repeatable amplifications. The SSRs in this study generated important diagnostic markers that are species specific and can be of immense use for resolving species conflicts that are reported to exist between *lanatus* and *citroides*. Dane *et al.* (2004) identified diagnostic markers using cpDNA haplotypes for *lanatus* and *citroides* types and used them to track lineages with *Citrullus rehmii* and *Citrullus ecirrhosus*. Jarret *et al.* (1997) developed seven SSRs and used them to amplify 32 watermelon genotypes; they found that SSR-derived polymorphisms are very efficient in discriminating among various species. Guerra-Sanz (2002) identified 19 SSRs from cDNA sequence data. The AFLP technique is highly efficient in detecting polymorphisms at RSs and flanking sequences around RS by digestion and PCR amplification. The advantage of this technique is that AFLP can simultaneously amplify several polymorphic loci without requiring prior sequence information. In the current study, 18.8% of polymorphic markers that were generated using AFLP were informative within the cultivated watermelons. These markers could be of immense use in genetic mapping, QTL location and marker-assisted programs. Despite their advantages, SSRs and AFLPs can generate only binary character states (0–1), and hence they can introduce homoplasy into datasets. AFLP and SSR datasets are analyzed phenetically rather than cladistically. The lack of resolution is less likely

https://doi.org/10.1017/S1479262109990098 Published online by Cambridge University Press

to result from homoplasies than from the fact that the most of these polymorphisms are not species specific, i.e. there is retention of ancestral polymorphisms in derived lineages. However, in the current study, the tree topologies are well supported and in congruence with the tree topologies generated using cladistic methods by Dane and Liu (2007) and Dane and Lang (2004).

Evolutionary forces such as population bottlenecks, genetic drift through founder effects, adaptive radiation and recurrent gene flow due to cross compatibility between the species might have contributed to the genetic variation (Ellstrand *et al*., 1999; Dane and Lang, 2004). Various species in the genera of *Citrullus* spp. are freely crossable (Wehner, 2007). The shared polymorphisms that were subjected to PCA further resolved interrelationships at the molecular level and the distribution of genetic lineages among these freely interbreeding taxa. These data can provide insights into different factors that shape genetic diversity (Avise, 2000). In this paper, we resolved population structure in the taxa of *lanatus*. For NJ analysis, we used entire data that were generated along with a broad category of taxa including *colocynthis*. Clearly, when applied such a wide-ranging dataset that is not entirely informative within the *lanatus* cluster, the essential phylogenetic signals that are needed to resolve population substructures within *lanatus* cluster are shrouded. This is because only 18.8% of the total data is informative within the *lanatus* group and when used along with the total data for analysis, population structure within the *lanatus* group is masked due to sampling error in the selection of markers. To overcome this, we did separate PCAs, first with the dataset that showed polymorphism within the *lanatus* group, and then two additional PCAs using the shared polymorphisms with *citroides* and *colocynthis* separately to understand genetic reticulation and introgression histories. In all three PCAs (PCA-1, PCA-2 and PCA-3; *lanatus* itself, *lanatus* vs. *citroides* and *lanatus* vs. *colocynthis*), the taxa of *lanatus* grouped into two clusters. Clustering patterns in PCA-2 and PCA-3 analyses indicated that a subcluster of *lanatus* accessions had undergone introgression and gene flow with the both *citroides* and *colocynthis* taxa. The first cluster in both the PCAs had 11 *lanatus* types from northern Africa (Ghana, Nigeria), USA, Argentina and China, whereas the other five accessions of *lanatus* were from southern Africa, Turkey. The accessions from the Philippines and Mexico were grouped as the second cluster. Since Africa is the centre of origin for the cultivated watermelon, our results indicate that there are two predominant groups of *lanatus*. One group is from northern Africa (Nigeria and Ghana) and it spread across the world; the second group from the central (Zaire) and southern Africa, which introgressed considerably with the taxa *citroides* and *colocynthis*. However, a separate study should be undertaken with the large number of *lanatus* collections. PCA of *citroides* and *colocynthis* suggests that *citroides*' types have undergone extensive introgression with *colocynthis*, and *citroides* species are genetically more closely related to *colocynthis* than the species *lanatus*. The presence of divergent lineages in *C. colocynthis*, which form different clusters in the PCA involving taxa of *colocynthis*, is congruent with the results of cpDNA studies by Dane *et al*. (2007).

Knowledge regarding the path of domestication, however, is fragmentary and various scenarios have been proposed for the origin of the domesticated watermelon from its progenitor, wild *C. lanatus* (Bates and Robinson, 1995; Robinson and Decker-Walters, 1997; Maggs-Kölling *et al*., 2000). Landraces of the Kalahari region of southern Africa (Taylor, 1985) are early forms of domestication, and several others (Mallick and Masui, 1986) have proposed that the domestication process might also have occurred in northern Africa. Our results support both views, i.e. the *lanatus* accessions in the current study are two distinct groups, one being from the northern and other from central and southern Africa.

Our phylogenetic study could have been completed if we had included some accessions of *C. rehmii* and *C. ecirrhosus*, as these have been shown to be direct descendents of the species *C. colocynthis* (Dane and Lang, 2004; Dane and Liu, 2007). We made crosses between *lanatus* accessions from different clusters and developing recombinant inbred line populations, so that we can map the *lanatus*-specific polymorphisms that we identified in the current study. This might broaden the narrow genetic diversity, which is currently a bottleneck for watermelon improvement (Levi *et al*., 2001, 2004). Our future endeavor is to study a large number of *C. lanatus* var. *lanatus* for resolving population structure using model-based (Bayesian) clustering algorithms (Pritchard *et al*., 2000; Falush *et al*., 2003; Falush *et al*., 2007), as Bayesian inference has proven to be useful in resolving population structure and species boundaries in several plant species that are closely related and intermating (Koopman *et al*., 2008).

Acknowledgements

The authors are grateful to Dr Jarret, Plant Genetic Resources Conservation Unit, USDA-ARS, Griffin, GA, 30 223, for providing the seeds of germplasm accessions. The authors thank Drs Wolfe and Collins for their critical comments. Funding support is provided by USDA-CSREES Research (2007-38 814-18 472 and 2004-38 814-15 118).

References

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard University Press.

Bates DM and Robinson RW (1995) Cucumbers melon and watermelons. In: Smart J and Simmonds NW (eds) *Evolution of Crop Plants*. 2nd edn., London, UK: Longman, pp. 89–96.

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

Burkill HM (1985) *The Useful Plants of West Tropical Africa*. vol. 1, 2nd edn. Richmond, Surrey, UK: Royal Botanic Gardens, Kew.

Dane F (2002) Chloroplast DNA investigations in *Citrullus* using PCR-RFLP analysis. *Cucurbitaceae* 2002: 100–108.

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

Dane F and Liu J (2007) Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genetic Resources and Crop Evolution* 54: 1255–1265.

Dane F, Lang P and Bakhtiyarova R (2004) Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species. *Theoretical and Applied Genetics* 108: 958–966.

Dane F, Liu J and Zhang C (2007) Phylogeography of the bitter apple, *Citrullus colocynthis*. *Genetic Resources and Crop Evolution* 54: 327–336.

Ellstrand NC, Prentice HC and Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecological Systems* 30: 539–563.

Ellul P, Lelivelt C, Naval MM, Noguera FJ, Sanchez S, Atarés A, Moreno V, Corella P and Dirks R (2007) Watermelon biotechnology. *Agriculture and Forestry, Transgenic Crops*. vol. 60. Berlin/Heidelberg, Germany: Springer Verlag Publication, pp. 129–165.

Eriksen B and Töpel MH (2006) Molecular phylogeography and hybridization in members of the circumpolar Potentilla sect. Niveae (Rosaceae). *American Journal of Botany* 93: 460–469.

Falush D, Stephens M and Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.

Falush D, Stephens M and Pritchard J (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7(4): 574–578.

Gillaspie AG Jr, Hopkins MS Jr and Dean RE (2005) Determining genetic diversity between lines of *Vigna unguiculata* subspecies by AFLP and SSR markers. *Genetic Resources and Crop Evolution* 52: 245–247.

Guerra-Sanz JM (2002) *Citrullus* simple sequence repeats markers from sequence databases. *Molecular Ecology Notes* 2: 223–225.

Harris SA (1995) Systematic and randomly amplified polymorphic DNA in the genus *Leucaena* (Leguminosae, Mimosoideae). *Plant Systematics and Evolution* 197: 197–208.

Hillis DM (1994) Homology in molecular biology. In: Hall BK (ed.) *Homology: The Hierarchical Basis of Comparative Biology*. San Diego, CA: Academic Press, pp. 339–368.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bull Soc Vaudoise Sci Nat* 44: 223–270.

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

Jeffrey C (2001) Cucurbitaceae (*citrullus*). In: Hanelt P (ed.) *Mansfeld's Encyclopedia of Agricultural and Horticultural Crops*. New York: Springer, pp. 1533–1537.

Koopman WJM, Wissemann V and Cock KD (2008) AFLP markers as a tool to reconstruct complex relationships: a case study in *Rosa* (Rosaceae). *American Journal of Botany* 95: 353–366.

Kyndt T, Romeijnpeeters E, Droogenbroeck BV, Romeromotochi JP, Gheysen G and Goetghebeur P (2005) Species relationships in the genus *Vasconcellea* (Caricaceae) based on molecular and morphological evidence. *American Journal of Botany* 92: 1033–1044.

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

Levi A, Thomas CE, Keinathand AP and Wehner TC (2001*a*) Genetic diversity among watermelon (*Citrullus lanatus* and *Citrullus colocynthis*) accessions. *Genetic Resources and Crop Evolution* 48: 559–566.

Levi A, Thomas CE and Wehner TC (2001*b*) Low genetic diversity indicates the need to broaden the genetic base of cultivated watermelon. *Horticultural Science* 36: 1096–1101.

Levi A, Thomas CE, Newman M, Reddy OUK, Zhang X and Xu Y (2004) ISSR and AFLP markers differ among American watermelon cultivars with limited genetic diversity. *Journal of American Society of Horticultural Science* 129(4): 553–558.

Mallick MFR and Masui M (1986) Origin, distribution and taxonomy of melons. *Scientia Horticulturae* 28: 215–261.

Maluf MP, Silvestrini M, Ruggiero LMC, Filho OG and Colombo CA (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola* 62: 366–373.

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

Meeuse ADJ (1962) The Cucurbitaceae of southern Africa. *Bothalia* 8: 1–111.

Navot N and Zamir D (1987) Isozyme and seed protein phylogeny of the genus *Citrullus* (Cucurbitaceae). *Plant Systematics and Evolution* 156: 61–68.

Perumal R, Krishnaramanujam R, Menz MA, Katilé S, Dahlberg J, Magill CW and Rooney WL (2007) Genetic diversity among sorghum races and working groups based on AFLPs and SSRs. *Crop Science* 47: 1375–1383.

Pimentel M, Sahuquillo E and Catalán P (2007) Genetic diversity and spatial correlation patterns unravel the biogeographical history of the European sweet vernal grasses (*Anthoxanthum* L. Poaceae). *Molecular Phylogenetics and Evolution* 44(2): 667–684.

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

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

Ramakrishnan AP, Meyer SE, Waters J, Stevens MR, Coleman CE and Fairbanks DJ (2004) Correlation between molecular markers and adaptively significant genetic variation in

Bromus tectorum (Poaceae), an inbreeding annual grass. *American Journal of Botany* 91: 797–803.

Reddy OUK, Pepper AE, Abdurakhmonov I, Saha S, Jenkins JN, Brooks T, Bolek Y and El-Zik KM (2001) New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *The Journal of Cotton Science* 5: 103–113.

Robinson RW and Decker-Walters DS (1997) *Cucurbits*. Wallingford, UK: CAB International.

Rohlf FJ (1987) NTSYSpc. Numerical Taxonomy and Multivariate Analysis System, Version 1.30, Exeter Software, Setauket, NY, USA.

Saini N, Jain N, Jain S and Jain RK (2004) Assessment of genetic diversity within and among Basmati and non-Basmati rice varieties using AFLP, ISSR and SSR markers. *Euphytica* 140: 133–146.

Saitou N and Nei M (1987) The neighborjoining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.

Schmidt K and Jensen K (2000) Genetic structure and AFLP variation of remnant populations in the rare plant *Pedicularis palustris* (Scrophulariaceae) and its relation to population size and reproductive components. *American Journal of Botany* 87: 678–689.

Schönswetter P, Suda J, Popp M, Weissschneeweiss H and Brochmann C (2007) Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. *Molecular Phylogenetics and Evolution* 42: 92–103.

Spooner DM, Mclean K, Ramsay G, Waugh R and Bryan GJ (2005) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences of the United States of America* 102(41): 14694–14699.

Špunarová M, Ovesná J, TvarŮžek L, Kučera L, Špunar J and Hollerová I (2005) The use of molecular markers for characterisation of spring barley for breeding to *Fusarium* head blight resistance. *Plant Soil and Environment* 51(11): 483–490.

Sullivan JP, Lavoue S, Arnegard ME and Hopkins CD (2004) AFLPs resolve phylogeny and reveal mitochondrial introgression within a species flock of African electric fish (Mormyroidea: Teleostei). *Evolution* 58: 825–841.

Taylor FW (1985) The potential for the commercial utilization of indigenous plants in Bostwana. In: Wickens GE, Goodin JR and Field DV (eds) *Plants for Arid Lands*. London, UK: George Allen & Unwin.

Thaler R, Brandstätter A, Meranera A, Chabicovskia M, Parsonc W, Zelgerd R, Dallaviad J and Dallinge R (2008) Molecular phylogeny and population structure of the codling moth (*Cydia pomonella*) in Central Europe: II. AFLP analysis reflects humanaided local adaptation of a global pest species. *Molecular Phylogenetics and Evolution* 48(3): 838–849.

Travis SE, Maschinski J and Keim P (1996) An analysis of genetic variation in *Astragalus cremnophylax* var. cremnophylax, a critically endangered plant, using AFLP markers. *Molecular Ecology* 5: 735–745.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Paleman J, Kuiper M and Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21): 4407–4414.

Wehner TC (2007) Watermelon. In: Prohens J and Nuez F (eds) *Handbook of plant breeding; Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae*. New York, NY: Springer Science Business LLC, pp. 381–418.

Zamir D, Navot N and Rudich J (1984) Enzyme polymorphism in *Citrullus lanatus* and *C. colocynthis* in Israel and Sinai. *Plant Systematics and Evolution* 146: 163–170.

Zhang XP, Rhodes BB and Skorupska HS (1994) RAPD molecular markers in watermelon. *Cucurbit Genetic Cooperative Report* 17: 116119.