Theor Appl Genet (2004) 108:958–966
DOI 10.1007/s00122-003-1512-9

ORIGINAL PAPER

F. Dane · P. Lang · R. Bakhtiyarova

Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species

Received: 10 April 2003 / Accepted: 25 September 2003 / Published online: 21 November 2003

Abstract PCR amplification and restriction site analysis of chloroplast (cp) DNA regions was used to detect inter- and intraspecific differences in the genus *Citrullus.* More than 55 *C. lanatus* and 15 *C. colocynthis* accessions collected from diverse geographical areas, *C. ecirrhosus* and *C. rehmii* were used. Most of the cpDNA variation within *Citrullus* was the result of large indels and transitions and transversions. Indels at the *ndh*A, *trn*S-*trnf*M and *trn*C-*trn*D regions and several substitutions at restriction enzyme sites can be used to separate *C. colocynthis* from the other *Citrullus* species. A nucleotide substitution at a restriction enzyme site at the 3′ flanking region of *ndh*F provided a diagnostic haplotype for *C. lanatus* var. *lanatus,* the cultivated watermelon. Similarly, a nucleotide substitution at an intergenic spacer region of the *trn*C-*trn*D region resulted in a diagnostic haplotype for citron, *C. lanatus* var. *citroides*. Several *C. lanatus* var. *citroides* accessions showed the var. *lanatus* haplotype. *C. rehmii* showed almost the same haplotype as *C. lanatus* var. *citroides* with the exception of a unique insertion at a cpSSR site. Since *C. ecirrhosus* lacks the derived diagnostic nucleotide substitutions of *C. lanatus,* it is probably the progenitor of the cultivated watermelon. Intraspecific haplotypes detected within *C. colocynthis* were associated with geographic origin.

Introduction

Watermelon, *Citrullus lanatus* (Thunb.) Matsum. & Nakai., is a member of the Cucurbitaceae, a large and diverse family containing several economically important domesticated species (Harlan 1992). Other prominent cucurbit crops are melon (*Cucumis melo*), cucumber (*C. sativus*), squashes, pumpkins and gourds (*Cucurbita spp*). While *Cucumis* and *Citrullus* were domesticated in the Old World, *Cucurbita* is a New World genus, providing important sources of food to native American economies. Watermelon has been produced and consumed throughout the world for centuries, with its cultivation going back to prehistoric times. It was grown by the ancient Egyptians (Robinson and Decker-Walters 1997). Old names for watermelon in Arabic, Berber, Sanskrit, Spanish and Sardinian are all unrelated linguistically, indicating the great antiquity of its culture in lands around the Mediterranean and west of Africa. Although the cultivation of watermelon spread quickly eastward to lands around the Mediterranean and India, it was apparently unknown to the ancient Greeks or Romans. Watermelons were ultimately introduced to Europe by the Moors during their invasion of Spain, but it never became a popular fruit. In contrast, watermelons reached China somewhere between the 10th and 12th century, where they were so well received that China is today a leading producer of watermelons. Watermelons reached the Americas in the 17th century on slave ships and have been cultivated ever since in the Western Hemisphere (Rubatzky 2001).

The domesticated watermelon is classified as *Citrullus lanatus* var. *lanatus*, whereas wild citron, which is common in central Africa, is classified as var. *citroides* (Bailey) Mansf. Citron is a preserving melon. The rind is used to make pickles, while its bland- to bitter-tasting fruit with white or green flesh is used as food for livestock (Wehner et al. 2001). In West Africa, especially Nigeria, Egusi-type watermelons with bitter fruit are cultivated for their large light-colored seeds with a high oil content. Three other species in the genus *Citrullus* are generally recognized: *C. ecirrhosus* Cogn., *C. colocynthis* (L.) Schrad. and *C. rehmii* De Winter. All are diploid ($2n$=22) and cross-compatible with each other to varying degrees (Robinson and Decker-Walters 1997). *C. colocynthis* is cultivated for "colocynth', a drug produced from the dried pulp of unripe but mature-sized fruits. The species can be found in the north and southwest areas of Africa and Asia. Different races are sometimes recognized: one is found on

Communicated by J. Dvorak

F. Dane (✉) · P. Lang · R. Bakhtiyarova
Department of Horticulture,
Auburn University,
Auburn, AL, USA
e-mail: fdane@acesag.auburn.edu
Fax: +1-334-8443131

the Mediterranean coast and in Israel; the other can be found in the deserts of the Negev and Sinai (Yanev et al. 1999). *C. ecirrhosus* is a desert perennial with white-fleshed and bitter fruit, which is endemic to the Namibian desert, while *C. rehmii* was recently discovered in the Namib desert in Southern Africa (De Winter 1990).

Molecular systematic and phylogeographic studies on plants have made extensive use of the chloroplast (cp) genome. CpDNA is evolutionary conserved in terms of genome size, structure, gene content and linear order of genes among plant lineages. Consequently, any change in the structure or arrangement of the chloroplast genome can have significant phylogenetic implications (Soltis et al. 1998). Both coding and non-coding cpDNA sequences are used either through direct sequencing or PCR-restriction fragment length polymorphism (RFLP) analysis (Demesure et al. 1995; Taberlet et al. 1991; Dumolin-Lapegue et al. 1997a). The high degree of sequence conservation has facilitated the use of PCR primers in unrelated species, thereby circumventing the need to clone cpDNA from every species under study (Olmstead and Palmer 1994). Universal primers have been designed to amplify the DNA located between the primer binding sites (Taberlet et al. 1991; Demesure et al. 1995; Dumolin-Lapegue et al. 1997a). Phylogenetically informative characters have been identified and used as species-diagnostic markers or for germplasm and population evaluations (Petit et al. 1997; Parducci and Szmidt 1999; Desplanque et al. 2000; Xu et al. 2001). Studies of cpDNA variability in several species have revealed regional patterns of genetic structure, which could be interpreted as reflecting routes of colonization in concert with human movement across different continents (Dumolin-Lapegue et al. 1997b; Mohanty et al. 2001).

Although structural analysis of cpDNA has been carried out in a variety of plants, our knowledge of the chloroplast genome in the Cucurbitaceae is limited. Cucurbits are known to possess maternally inherited (Havey et al. 1998) genomes that are approximately equal with respect to size (150–155 kb) and structure based on the mapping of restriction enzyme sites in cucumber (Palmer 1982), squash (Lim et al. 1990) and melon (Perl-Treves and Galun 1985). Initial studies using PCR-RFLP analysis of cpDNA in watermelon and *C. colocynthis* were conducted using nine different chloroplast regions (Dane 2002) in which plant material collected from several geographical areas in the Americas (Chili, Mexico, USA), Africa, India and Afghanistan and Europe was analyzed. The results showed that the cucurbit chloroplast genome contains multiple variable sites, and while species diagnostic markers were detected, no intraspecific variation was observed.

The purpose of the investigation reported here was to further the study of the chloroplast genome to detect intraspecific variability in all *Citrullus* species in order to gain a better understanding of domestication events in watermelon, to analyze the geographic pattern of haplotypes and to study the evolution of the species in the genus.

Materials and methods

Plant materials

Seeds of watermelon cultivars were obtained from several commercial breeding companies, while seeds from plant introduction (PI) accessions of *Citrullus lanatus* var. *lanatus* and var. *citroides*, *C. colocynthis* and *Praecitrullus fistulosus* (Stocks) Pang., *Cucumeropsis mannii* Naudin and *Acanthosicyos naudinianus* (Sond.) C. Jeffrey were obtained from the Plant Genetic Resource Conservation Unit at Griffin, Georgia, USA (Table 1) or The Cucurbit Network. Seeds from *C. ecirrhosus* and *C. rehmii* × *C. lanatus* were obtained through B&T World Seeds. Cultivars and accessions were selected based on dendrograms using simple sequence repeat (SSR) polymorphism (Jarret et al. 1997) or random amplified polymorphic DNA (RAPD) marker polymorphisms (Levi et al. 2001). More than 55 *C. lanatus* accessions from widely different geographical areas were selected (Table 1). *P. fistulosus*, cultivated in India and Pakistan for its edible fruits, and *A. naudinianus*, native to southern Africa, are closely related to *Citrullus* and were once included in the genus.

PCR-RFLP analysis

Total DNA was extracted from 100 mg seed or seedling leaf tissue using the QIAGEN DNeasy kit (QIAGEN, Valencia, Calif.). More than 20 different cpDNA-specific primer pairs were used to amplify the corresponding cpDNA regions (Table 2; Dane 2002). Most of the cpDNA genes are located in the large single-copy region of the chloroplast genome, with the exception of *ndh*A and *ndh*F, which are located in the small single-copy region (Fig. 1). We used several primer pairs to detect variation at the *ndh*F, *rpl*16, *trn*S-*trn*fM, *trn*C-*trn*D and *trn*T-*trn*F regions (Table 2).

DNA was amplified by PCR in a 22-μl volume containing the following: 2.2 μl 10× PCR buffer (200 m*M* Tris-HCl, pH 8.4, 500 m*M* KCl), 1.6 μl (25 m*M*) $MgCl_2$, 0.4 μl (10 m*M*) dNTPs, 1.1 μl (4 μ*M*) forward primer, 1.1 μl (4 μ*M*) reverse primer, 0.2 μl *Taq* polymerase (5 U/μl; Sigma, St. Louis, Mo.) and 2 μl template DNA (25 ng/μl). DNA amplifications were performed in a Perkin Elmer 2400 thermal cycler (Perkin Elmer, Foster City, Calif.) programmed as follows: 4 min at 94°C; 35 cycles of 1 min at 94°C, 1 min at different annealing temperatures (dependent on the primer pair) (Table 2), 2–4 min at 65°C (dependent on the length of the fragments); a final 10-min extension at 65°C.

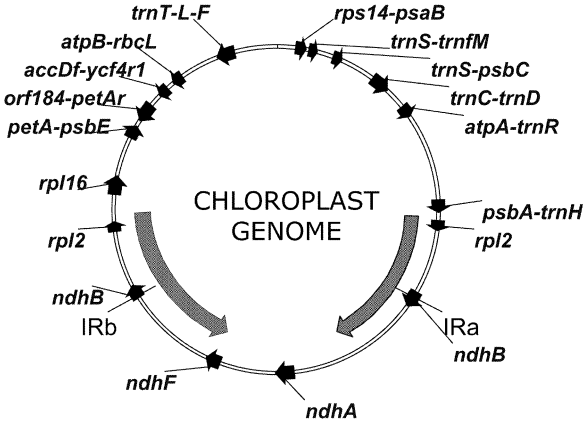

Fig. 1 Schematic representation of the chloroplast genome showing the location of the regions amplified with the primers used in the watermelon study. CpDNA variability within the genus *Citrullus* was detected at the *ndh*A, *ndh*F, *trn*S-*trn*fM, *trn*C-*trn*D and *atp*A-*trn*R regions

Table 1 List of investigated *Citrullus*, *Praecitrullus fistulosus*, *Acanthosicyos naudinianus* and *Cucumeropsis mannii* plant introductions (PI) and cultivars and their geographical origin. Morphological characteristics of most accessions are available in the GRIN database (*TCN* The Cucurbit Network)

Taxon	PI number, TCN number or cultivar	Origin
C. lanatus var. *lanatus*	165451	Mexico
	176492	Turkey
	185636, 271751	Ghana
	211011	Afghanistan
	241689	Chili
	254742	Senegal
	273481	Ethiopia
	295845	South Africa
	385964	Kenya
	482251	Zimbabwe
	494527, 494529, 494531	Nigeria
	500314, 500324	Zambia
	507858	Hungary
	536453	Maldives
	549160	Chad
	AU Producer, Jubilee, MickyLee, Starbrite, Crimson Sweet	USA
C. lanatus var. *citroides*	179881, 288316	India
	189225, 532738	Zaire
	244018, 255136, 255137, 270563, 271769, 271779, 296334, 296335, 295842, 296341, 596665	South Africa
	248774, TCN 1126	Namibia
	254744	Senegal
	346082	Afghanistan
	379243	Yugoslavia
	482246, 482259, 482261, 482279, 482311, 482319	Zimbabwe
	512385, 512854	Spain
	525081	Egypt
	532664, 532667	Swaziland
	532819	China
	TCN 1360, 1337	USA
C. colocynthis	195927	Ethiopia
	220778, 296365, 374216	Afghanistan
	386024, 386026	Iran
	388770, TCN 955	Morocco
	432337	Cyprus
	525082	Egypt
	537277	Pakistan
	542616	Algeria
	549161	Chad
C. ecirrhosus	68444	Namibia
C. rehmii × *C. lanatus*	431727	Namibia
Praecitrullus fistulosus	174812, 180275, 381748, 427298, 536544	India
Acanthosicyos naudinianus	96690	South Africa
Cucumeropsis mannii	532721	Zaire

PCR products were digested with at least three of the following restriction endonucleases—*Hinf*I, *Rsa*I, *Taq*I, *Alu*I , *Hae*III, *Mbo*I, *Bgl*II—for at least 4 h at 37°C or 65°C (for *Taq*I). For each restriction, 5 μl of PCR product and 3 μl restriction-mix were used according to the manufacturer's instructions (Gibco BRL, Invitrogen, Carlsbad, Calif.). Digested cpDNA fragments were separated on 1.5% agarose gels, stained with ethidium bromide and photographed using a Bio-Rad (Hercules, Calif.) photodocumentation system. A 100-bp ladder (Invitrogen) was used as a size marker.

Results

The primers pairs successfully amplified the corresponding cpDNA regions in the *Citrullus*, *Praecitrullus*, *Cucumeropsis* and *Acanthosicyos* species investigated. The amplified products were similar in size, except for the products using *ndh*A primers, the primer pair (*orf*62-*trn*GM) covering a small section of the *trn*S-*trn*fM region and the primer pair *psb*M-*trn*DM covering a section of *trn*C-*trn*D. Larger amplification fragments were observed at the *ndh*A and *orf*62-*trn*GM region in *C. colocynthis* than in *C. lanatus*, *C. ecirrhosus* and *C. rehmii,* as a result of a large deletion (greater than 100 bp) in the latter three species. Larger fragments in both regions were also detected in *P. fistulosus*, *Cucumeropsis mannii* and *A. naudinianus*. *C. lanatus*, *C. ecirrhosus* and *C. rehmii* shared a large insertion (approx. 100 bp) at the *psb*M-*trn*DM region that was missing in *C. colocynthis* and *P. fistulosus.* These indels were easily discernable using different restriction enzymes (e.g. *trn*C-*trn*D using *Hinf*I, *Rsa*I or *Taq*I).

Table 2 DNA sequences of the primer pairs used in the present study, reference, approximate size of corresponding PCR products in base pairs and annealing temperature (T_a)

CpDNA region	Primers	Primer pair sequence (5′–3′)	Reference	Region length (bp)	T_a (°C)
*acc*D-*ycf4*	aacDF	GCA GGT AAA AGA GTA ATT GAA C	Heinze 2002 [b]	1,600	52
	ycf4R1	CTA ATA AGA AGC CTA ATG AAC C			
*atp*B-*rbcL*	atpB	GTG TCA ATC ACT TCC ATT CC	Fofana et al. 1997	1,700	55
	rbcL	GTA AAA TCA AGT CCA CCG CG			
*atp*A-*trn*R	ccSSR4F	AGG TTC AAA TCC TAT TGG ACG CA	Chung et al. 2003	600	55
	ccSSR4R	TTT TGA AAG AAG CTA TTC ARG AAC			
*ndh*A	ndhAF	GGW CTT CTY ATG KCR GGA TAT RGM TC	Small et al. 1998	1,300	42
	ndhAR	CTG YGC TTC MAC TAT ATC AAC TGT AC			
*ndh*B	9F	ATG GTT TCT CTT GGC TAT ATG G	Graham and Olmstead 2000	1,100	52
	13R	GCA TAC GTT TCA TGC TTG TTT GAG			
*ndh*F	rpl32	GAA GTR CGY TTT TTT GGA ACT GCC	Olmstead and Sweere 1994	1,500	52
	803R	GAA AAA TWC CCG CCG CTA CCA TAG			
	803F	CTA TGG TAG CGG CGG GAW TTT TTC	Olmstead and Sweere 1994	500	55
	1318R	CGA AAC ATA TAA AAT GCR GTT AAT CC			
	1318F	GGA TTA ACY GCA TTT TATTAT ATG TTT CG	Olmstead and Sweere 1994	640	55
	1955R	CGA TTA TAT GAC CAA TCA TAT A			
	1955F	TAT ATG ATT GGT CAT ATA ATC G	Olmstead and Sweere 1994	750	55
	607	ACC AAG TTC AAT GTT AGC SAG ATT AGT C			
*orf*184-*pet*A	orf184F	TGG CGA TCA GAA CAY ATA TGG ATA G	Grivet et al. 2001	1,800	55
	petAR	CAT AHY CYT GYT GBG CRA AAA TDG G	Heinze 2002		
*psb*C-*trn*S	psbC	GGT CGT GAC CAA GAA ACC AC	Demesure et al. 1995	1,600	52
	trnS	GGT TCG AAT CCC TCT CTC TC			
*rpl*2	20F	AAA GGT CGT AAT GCC AGA GGA AT	Graham and Olmstead 2000	900	55
	25R	TTC CAA GYG CAG GAT AAC CCC A			
*rpl*16	rpl16R1516	CCC TTC ATT CTT CCT CTA TGT TG	Kelchner and Wendel 1996	2,700	52
	ccmp10r	TTC GTC GDC GTA GTA AAT AG	Weising and Gardner 1999		
	exon1	AAT AAT CGC TAT GCT TAG TG	Downie et al. 2000	1,100	54
	exon2	TCT TCC TCT ATG GTT GTT TAC G			
*psb*A-*trn*H	psbAF	GTT ATG CAT GAA CGT AAT GCT C	Sang et al. 1997	500	52
	trnHR	GCG CAT GGT GGA TTC ACA AAT C			
*trn*C-*trn*D	trnCF	CCA GTT CAA ATC TGG GTG TC	Demesure et al. 1995	3,200	52
	trnD-M (R)	GGG ATT GTA GTT CAA TTG GT			
	trnC	CCA GTT CAA ATC TGG GTG TC		700	50
	ycf6R	CAT TAA AGC AGC CCA AGC	Heinze 2002		
	ycf6F	CTT GGG CTG CTT TAA TGG	Heinze 2002	1,300	50
	psbMR	GTA AAT ATT CTT GCA TTT ATT GC			
	psbMF	AAT AGT GCA GTA GCA ATA AAT GC	Heinze 2002	1,200	52
	trnDM (R)	GGG ATT GTA GTT CAA TTG GT			
*trn*T-*trn*L	A	CAT TAC AAA TGC GAT GCT CT	Taberlet et al. 1991	700	56
	B	TCT ACC GAT TTC GCC ATA TC			
*trn*L intron	C	CGA AAT CGG TAG ACG CTA CG	Taberlet et al. 1991	600	55
	D	GGG GAT AGA GGG ACT TGA AC			
*trn*L-*trn*F	E	GGT TCA AGT CCC TCT ATC CC	Taberlet et al. 1991	450	52
	F	ATT TGA ACT GGT GAC ACG AG			
*trn*S-*trn*fM(SfM)	trnSP (F)	GAG AGA GAG GGA TTC GAA CC	Demesure et al. 1995	1,230	50
	trnfM (R)	CAT AAC CTT GAG GTC ACG GG	Dumolin-Lapeque et al. 1997a		
	orf62 (F)	CTT GCT TTC CAA TTG GCT GT	Heinze 2002	750	55
	trnGMR	AAC CCG CAT CTT CTC CTT GG			
	trnGP (F)	GCC AAG GAG AAG ATG CGG G	Heinze 2002	300	52
	trnfM (R)				
*rps*14-*psa*B	rps14	CAT TTC ACG AAG TAT GTG TCC G	Fofana et al. 1997	600	55
	psaB	TGG CGT GGA TAT TGG CAG GA			
*pet*A-*psb*E	petA	GCA TCT GTT ATT TTG GCA CA	Fofana et al. 1997	1,800	52
	psbE	TAC CTT CCC TAT TCA TTG CG			

[a] W, AT; Y, CT; K, GT; R, AG; M, AC; S, GC
[b] Website: http://fbva.forvie.ac.at/200/1892.html

Table 3 *Citrullus* PI accessions with informative haplotype patterns at specific cpDNA regions

PI station classification	PI number	cpDNA region/restriction enzyme	Haplotype
Citrullus colocynthis	549161	*ndh*F803F–1318R/*Taq*I	a
C. colocynthis	195767, 220778, 269365, 374216, 386024, 537277	*trn*C-*ycf*6R/*Taq*I	b
C. lanatus var. *citroides*	346082	*trn*C-*ycf*6R/*Taq*I	b
C. colocynthis	388770, 432337, 525082, 542616, 955	*trn*C-*ycf*6R/*Taq*I	c
C. lanatus var. *citroides*	525081	*trn*C-*ycf*6R/*Taq*I	c
C. ecirhosus	68444	*ycf*6-*psb*MR/*Taq*I *ndh*F1955F-607R/*Alu*I	d
C. lanatus var. *lanatus*	165451, 176492, 185636, 211011, 241689, 254742, 271751, 271778, 273481, 295845, 385964, 482251, 494527, 494529, 494531, 500314, 500324, 507858, 536453, 549160, AU-Producer, Starbrite, Micky Lee, Crimson Sweet, Jubilee	*ndh*F1955F-607R/*Alu*I	e
C. lanatus var. *citroides*	189225, 244018, 248774, 255137, 270563, 271769, 288316, 296334, 296335, 296341, 379243, 482246, 482259, 482261, 482279, 482311, 512385, 512854, 532664, 532667, 532738, 532819, 596665, 1126, 1337, 1360	*ycf*6-*psb*MR/*Taq*I	f
C. lanatus var. *citroides*	179881, 254744, 255136, 271779, 295842, 482319	*ndh*F1955F-607R/*Alu*I	e
C. rehmii × *C. lanatus*	431727	ccSSR4/*Taq*I	g

The total length of the amplified regions was 23.5 kb, which accounts for 15% of the total *Citrullus* genome, assuming an average size of 155 kb (Havey et al. 1998). Using PCR-RFLP we could only identify five variable sites (covering 32% of regions studied) within *Citrullus*, while most of the regions showed polymorphisms across different genera. Using different fragment/restriction enzyme combinations we were able to detect a total of seven haplotypes (Table 3). The *C. colocynthis* haplotype can be observed using four different cpDNA regions (Table 4) and a variety of restriction enzymes (for example *ndh*A and *Taq*I, *Hae*III or *Bgl*II; *trn*S-*trn*fM using *Hinf*I or *Alu*I; *trn*C-*trn*D using *Hinf*I or *Alu*I; *ndh*F803F-607R using *Taq*I).

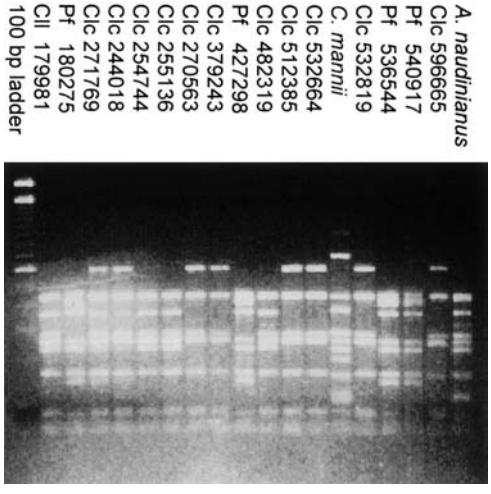

Fig. 2 Restriction fragment patterns of cpDNA from *Citrullus, Praecitrullus, Cucumeropsis* and *Acanthosicyos* accessions, detected using the fragment/enzyme combination of *trn*C-*trn*D/*Taq*I, and showing the loss of a *Taq*I restriction enzyme site in *Citrullus lanatus* var. *citroides* in comparison to *C. lanatus* var. *lanatus*, *P. fistulosus* and *A. naudinianus*. *Clc C. lanatus* var. *citroides*, *Cll C. lanatus* var. *lanatus*, *Pf P. fistulosus*

Molecular variation at *ndh*F

Using the *ndh*F803F-607R/*Alu*I fragment/enzyme combination, we detected a unique haplotype for *C. lanatus* var. *lanatus* that distinguishes the cultivated watermelon from citron, *C. lanatus* var. *citroides*, *C. ecirrhosus, C. rehmii* and *C. colocynthis*. PI's 179981, 254744, 255136, 296842 and 482319, which are classified as *C. lanatus* var. *citroides*, showed the *C. lanatus* var. *lanatus* haplotype. The variability was due to a mutation at the 3′ flanking region of *ndh*F since the unique *C. lanatus* var. *lanatus* haplotype was also detected using the primer pair 1955F-607R/*Alu*I combination.

When the 803F-1318R primer pair was used to amplify part of the coding region of the *ndh*F gene, a restriction enzyme site mutation was discovered within *C. colocynthis*. Accessions 269365, 525082, 432337, 542616, 388770, 386024, 374216, 269395, 195727and 220778, and *C. lanatus* var. *citroides* accessions 346082 and 525081, with the exception of PI 549161, show a *Taq*I site, which is missing in all other *Citrullus* species, *P. fistulosus* and *A. naudinianus* (PI's 618817 and 596670).

Molecular variation at *trn*C-*trn*D

Inter- and intraspecific differences were also detected at *trn*C-*trn*D using PCR-RFLP (Fig. 2). *C. colocynthis* can be distinguished from the other *Citrullus* species using *Hinf*I or *Rsa*I or *Alu*I. A unique haplotype was discovered using *Taq*I that separates *C. lanatus* var. *citroides* and *C. rehmii* from the other *Citrullus* accessions. The haplotype of

Table 4 Description of haplotypes identified in the genus *Citrullus* using PCR-RFLP[a]

Haplo-type	*ndh*A	*orf*62P-*trn*GM	*psb*M-*trn*DM	*psb*M-*trn*DM/ *Mbo*I	*ndh*F 1955F-607R/ *Taq*I	*ndh*F 803F-1318R/ *Taq*I	*ndh*F 1955F-607R/ *Alu*I	*trn*C-ycf6R/ *Taq*I	*ycf*6-*psb*MR/ *Taq*I	ccSSR4/ *Taq*I	Taxon
a	1	1	2	-	-	+	+	+	+	2	*Citrullus colocynthis*
b	1	1	2	-	-	-	+	-	+	2	*C. colocynthis*
c	1	1	2	-	-	-	+	+	+	2	*C. colocynthis*
d	2	2	1	+	+	+	+	-	+	2	*C. ecirrhosus*
e	2	2	1	+	+	+	-	-	+	2	*C. lanatus* var. l*anatus*
f	2	2	1	+	+	+	+	-	-	2	*C. lanatus* var. *citroides*
g	2	2	1	+	+	+	+	-	-	1	*C. rehmii*
h	3	3	2	-	-	-	+	-	+	3	*P. fistulosus*

[a] 1, Longer fragment due to insertion versus 2, shorter fragment due to deletion. +, gain, and -, loss of restriction enzyme site

PI 482319, 255136, 254744, 296842 and 179981 (classified as *C. lanatus* var. *citroides*) was similar to that of *C. lanatus* var. *lanatus* (Fig. 2). The unique haplotype was the result of a loss of a *Taq*I site at the *ycf*6-*psb*M region.

Intraspecific differences in *C. colocynthis* were detected using the *trn*C-*ycf*6R/*Taq*I combination. PI's 388770, 549161, 432337, 525082, 542616 and 525081 show a haplotype different from all other *Citrullus* accessions, which was the result of the gain of a *Taq*I site. These accessions originated around the Mediterranean region (Egypt, Cyprus, Algeria, Morocco, and Chad), while accessions lacking the *Taq*I site originated in Ethiopia, Afghanistan, Iran, or Pakistan (Table 3).

At the *psb*MF-*trn*DM region, *C. colocynthis* and *P. fistulosus* accessions can be distinguished from the other *Citrullus* species by a large (greater than 100 bp) insertion and the gain of a *Mbo*I site.

Discussion

PCR-RFLP can be used to identify *Citrullus* species and provide diagnostic haplotypes for the different watermelon forms (citron versus cultivated), making the technique extremely useful for large germplasm collections such as the one harbored at the PI Station in Griffin, Ga. with more than 1,500 *Citrullus* accessions. Sequence analysis is not practical for population genetics and germplasm evaluations, in which a large number of samples often need to be surveyed. A single base variation at particular regions can now be detected by the amplification of specific regions using PCR and digestion with the appropriate restriction enzyme (Table 3). The unique haplotype of cultivated watermelon (var. *lanatus*) was the result of a loss of *Alu*I site at the 3′ flanking region of *ndh*F. Similarly, the unique haplotype of citron (var. *citroides*) was the result of a loss of a *Taq*I site at *ycf*6. *C. rehmii* showed the loss of *Taq*I site at *ycf*6 and a unique insertion using ccSSR 4/*Taq*I, suggesting that *C. rehmii* might be considered the progenitor of citron melon. The loss of restriction enzyme binding sites at *ndh*F and *ycf*6 must have been relatively recent occurrences since all other *Citrullus* species as well as *P. fistulosus* do not show nucleotide substitutions at these sites. *C. ecirrhosus*, which shares almost all of its restriction enzyme binding sites (approx. 200 detected in this study) with *C. lanatus* might be considered the ancestral species of cultivated watermelon (Table 4).

Only seven haplotypes were detected within the genus *Citrullus*. The range of cpDNA diversity encountered in plant species varies from nil, as in pearl millet (Gepts and Clegg 1989), to low in soybean (Xu et al. 2002), European chestnut (Fineschi et al. 2000) and pear (Katayama and Uematsu 2003), with high polymorphisms in wild beet (Forcioli et al. 1998) and many tree species such as *Prunus* (Mohanty et al. 2001) and olive (*Olea europaea*; Besnard et al. 2002). Studies using cpDNA polymorphism on European trees such as oaks (*Quercus* spp), beech (*Fagus*) and black alder (*Alnus glutinosa*) showed strong east-west

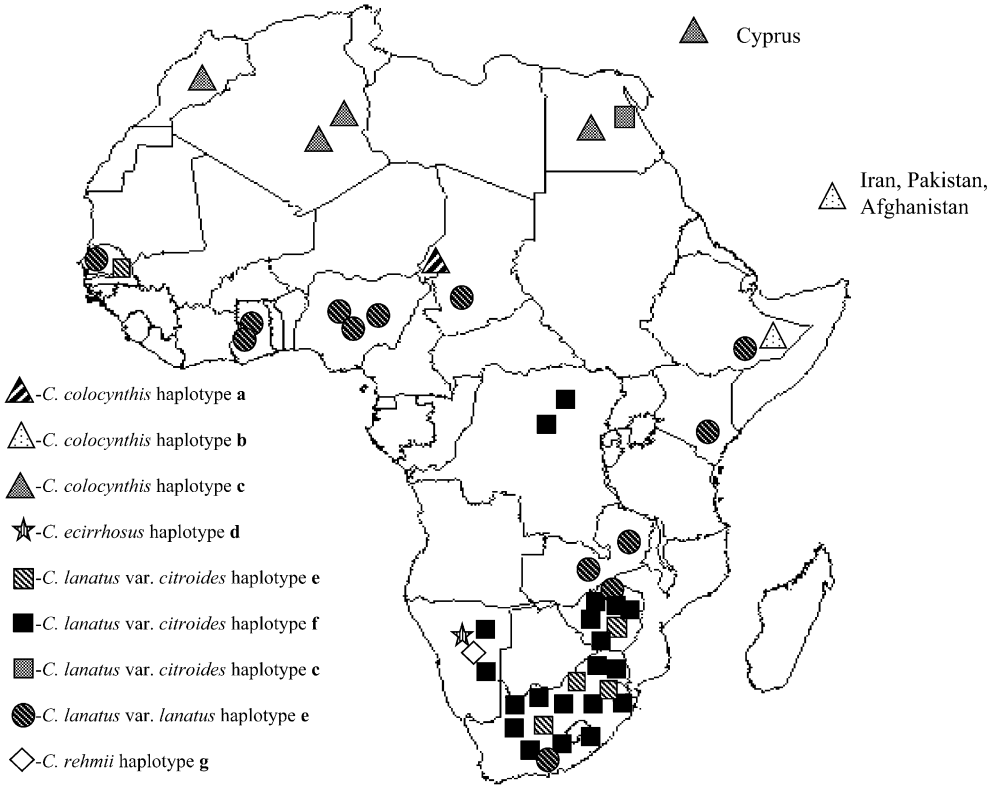

Fig. 3 Geographical distribution of *Citrullus* haplotypes on the African continent

clines in variation, which could be interpreted to be a result of post-glacial migration from the same glacial refugia, leading to concordance of variation patterns among species (Demesure and Comps 1996; Dumolin-Lapegue et al. 1997b; King and Ferris 1998). A similar concordance in phylogeographic patterns associated with post-glacial spread is observed in plant species in the Pacific Northwest of North America (Soltis et al. 1992).

The three haplotypes detected within *C. colocynthis* are clearly associated with geographic origin (Fig. 3). Haplotype a was detected in one accession from Chad, haplotype b was found in accessions from southeastern Ethiopia and South East Asia (Iran, Afghanistan, Pakistan and India) while haplotype c was detected in accessions collected around the Mediterranean from Morocco to Cyprus. These results support the studies of Yanev et al. (1999) who described different races of *C. colocynthis*, one on the Mediterranean coast and another in the deserts of Negev and Sinai.

Only two haplotypes—e and f—were detected within *C. lanatus*, and these were associated with the classification as cultivated var. *lanatus* versus citron-type var. *citroides*. The morphological characteristics of many of the citron PI's (*C. lanatus* var. *citroides*), described in the GRIN database, indicate high phenotypic variability. Fruit size varies from small (10×10 cm for PI 244018) to medium (30×30 cm for PI 270563), fruit shape varies from round to oblong, flesh color is mostly white or yellow and seed size can vary from 5×8 mm to 8×15 mm. Citron seeds generally lack the flatness of cultivated watermelon seeds. Several of the citron accessions showed the *C. lanatus* var. *lanatus* haplotype. The morphological characteristics of these accessions are variable, with fruit flesh color ranging from white to yellow and red, seed color varying from red to brown and seed size ranging from 7×10 mm to 8×14 mm. Cultivated watermelons (haplotype e) similarly have variable fruit types, ranging from small fruit (10×10 cm for PI 494527) with white flesh and Egusi-type edible seeds (5×27 mm) to the large, round or oblong watermelon fruit with red or yellow flesh presently available on the US market.

Phylogeography and plant domestication has been studied in several plant species using DNA variability. In wild and cultivated olive (*Olea europaea*), 15 chlorotypes were detected which could be clustered into five different clades located in distinct geographic regions (Besnard et al. 2002). Little geographic structure was detected in European chestnut species as a result of the human impact on its distribution and the long period of cultivation experienced in the last 1,000 years (Fineschi et al. 2000). CpDNA studies in wild and cultivated soybean using cpSSRs showed considerably higher genetic diversity in the wild soybean (*Glycine soja*) than in the cultivated species (*G. max*). The predominant haplotype of the cultivated species could be traced to a rare haplotype of the wild soybean presently distributed in southern Japan and China, while other haplotypes probably originated independently in different regions (Xu et al. 2002). Extremely low levels of polymorphism were recently detected among cultivated tomato (var. *esculentum)*

cultivars using sequence information at the fruit weight (*fw2.2*) locus. Only a single nucleotide substitution in one var. *esculentum* accession was observed in a sample of more than 7 kb (Nesbitt and Tanksley 2002). This lack of diversity was attributed to three population bottlenecks in the history of modern tomato cultivars: initial domestication, transfer of varieties to Europe by Spanish explorers and subsequent breeding efforts, primarily by US breeders.

The lack of haplotype divergence within *C. lanatus* is similarly indicative of the existence of a bottleneck in the history of watermelon, probably as a result of human selection. Domestication was one of the most profound and rapid events in plant evolution, irreversibly altering the distribution of plant species on earth and enabling human colonization to come into existence. Domestication of individual plant species was the result of one or more dramatic changes in the anatomy of the species, allowing the fruit to become greatly exaggerated. Over recent years, evidence has been accumulated to support the hypothesis that the majority of these anatomical changes can be attributed to a few loci and that selection for these loci by our ancestors rendered alterations in overall genetic diversity of the species (Nesbitt and Tanksley 2002). Other than in maize and tomato (Gepts 2002; Nesbitt and Tanksley 2002), molecular events accompanying domestication are relatively unknown.

C. lanatus appears to have diverged relatively recently since it contains unique restriction enzyme sites not found in other *Citrullus* species or *P. fistulosus*. It was probably derived from *C. ecirhosus* and is not as clearly delimited from this species using cpDNA markers as it is from *C. colocynthis* and *P. fistulosus*. *C. ecirrhosus* is a desert perennial native to Namibia that is consumed by gemsbok, but not humans. Nucleotide divergence between *C. ecirrhosus* and *C. lanatus* was considerably less than divergence between *C. colocynthis* and *C. lanatus*. Three large indels and four restriction enzyme site differences at four cpDNA regions (*ndh*A, *trn*S-*trn*fM, *ndh*F and *trn*C-*trn*D, Table 4) separated *C. lanatus*, *C. ecirrhosus* and *C. rehmii* from *C. colocynthis* and *P. fistulosus*. However, fragment/restriction enzyme site differences at almost all of the cpDNA regions studied were detected between *Citrullus* and *P. fistulosus*, indicative of wide divergence between the different cucurbit genera. Low divergence within *C. lanatus* points to population bottlenecks within the cucurbit species. It might also be related to their low levels of outcrossing and low effective population size as compared to maize, for example, where outcrossing rates are high and long-term effective population size is estimated at roughly a million plants (Buckler et al. 2001). Only a few plants are needed for cultivation, thus bottlenecks are very severe in the cucurbits. Also, human selection has moved the species towards similar phenotypes and adaptations. Understanding why domestication succeeded in the past should provide important knowledge on how to exploit the diversity and genome structure for future agricultural improvement in this and other cultivated species.

References

Besnard G, Khadari B, Baradat P, Berville A(2002) *Olea europaea* (Oleraceae) phylogeography based on chloroplast DNA polymorphism. Theor Appl Genet 104:1353–1361

Buckler ES IV, Thornberry JM, Kresovich S (2001) Molecular diversity, structure and domestication of grasses. Genet Res 77:213–218

Chung S-M, Decker-Walters DS, Staub JE (2003) Genetic relationships within the Cucurbitaceae as assessed by ccSSR marker and sequence analysis. Can J Bot 81:814–832

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

Demesure B, Comps B (1996) Chloroplast DNA phylogeography of the common beech *(Fagus sylvatica* L) in Europe. Evolution 50:2515–2520

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

Desplanque B, Viard F, Bernard J, Forcioli D, Saumitou-Laprade P, Cuguen J, van Dijk H (2000) The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* spp. *maritima* (L.): The usefulness of both genomes for population genetic studies. Mol Ecol 9:141–154

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

Downie SR, Katz-Downie DS, Watson MF (2000) A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl*16 and *rpo*C1 intron sequences: towards a suprageneric classification of subfamily Apioideae. Am J Bot 87:273–292

Dumolin-Lapegue S, Pemonge M-H, Petit RJ (1997a) An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 6:393–397

Dumolin-Lapegue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997b) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487

Fineschi S, Taurchini D, Villani F, Vendramin GG (2000) Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* mill. (Fagaceae) throughout southern European countries. Mol Ecol 9:1495–1503

Fofana B, Harvengt L, Baudoin JP, du Jardin P (1997) New primers for the polymerase chain amplification of cpDNA intergenic spacers in *Phaseolus* phylogeny. Belg J Bot 129:118–122

Forcioli D, Saumitou-Laprade P, Velero M, Vernet P, Cuguen J (1998) Distribution of chloroplast DNA diversity within and among populations in gynodioecious *Beta vulgaris* ssp *maritima* (Chenopodiaceae). Mol Ecol 7:1193–1204

Gepts P (2002) Ten thousand years of crop evolution. In: Chrispeels MJ, Sadava DE (ed) Plants, genes and crop biotechnology. Jones and Bartlett, Sudbury, Mass., pp 328–359

Gepts P, Clegg MT (1989) Genetic diversity in pearl millet (*Pennisetum glaucum* L. R.Br.) at the DNA sequence level: consequences of genetic conservation. J Hered 81:183–188

Graham SW, Olmstead RG (2000) Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am J Bot 87:1712–1730

Grivet D, Heinze B, Vendramin GG, Petit RJ (2001) Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Mol Ecol Notes 1:345–349

Harlan JR (1992) Crops and man. American Society of Agronomy, Madison, Wis.

Havey MJ, McCreight JD, Rhodes B, Taurick G (1998) Differential expression of the *Cucumis* organellar genomes. Theor Appl Genet 97:122–128

Heinze B (2002) http://fbva.forvie.ac.at/200/1892.html

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

Katayama H, Uematsu C (2003) Comparative analysis of chloroplast DNA in *Pyrus* species: physical map and gene localization. Theor Appl Genet 106:303–310

Kelchner SA, Wendel JF (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr Genet 30:259–262

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* L. Gaertn. Mol Ecol 7:1151–1161

Levi A, Thomas CE, Wehner TC, Zhang X (2001) Low genetic diversity indicated the need to broaden the genetic base of cultivated watermelon. Hortscience 36:1096–1101

Lim H, Gounaris I, Harrison RC, Boyer CD (1990) Restriction site and genetic map of *Cucurbita pepo* chloroplast DNA. Curr Genet 18:273–275

Mohanty A, Martin JP, Aguinagalde I (2001) Chloroplast DNA study in wild populations and some cultivars of *Prunus avium* L. Theor Appl Genet 103:112–117

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus *Lycopersicon*: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162:365–379

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. Am J Bot 81:1205–1224

Olmstead RG, Sweere JA (1994) Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. Syst Biol 43:467–481

Palmer JD (1982) Physical and gene mapping of chloroplast DNA from *Atriplex triangularis* and *Cucumis sativus*. Nucleic Acids Res 10:1593–1605

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

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

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. Proc Natl Acad Sci USA 94:9996–10001

Robinson RW, Decker-Walters DS (1997) Cucurbits. CAB Int, Wallingford, UK

Rubatzky VE (2001) Origin, distribution and uses. In: Maynard DN (ed) Watermelons. Characteristics, production and marketing. ASHS Press, Alexandria, Va., pp 21–26

Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am J Bot 84:1120–1136

Schnabel A, Wendel JF (1998) Cladistic biogeography of *Gleditsia* (Leguminosae) based on *ndhF* and *rpl*16 chloroplast gene sequences. Am J Bot 85:1753–1765

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301–1315

Soltis DE, Soltis PS, Milligan BG (1992) Intraspecific chloroplast DNA variation: sytematic and phylogenetic implications. In: Soltis DE, Soltis PS, Doyle JJ (ed) Molecular systematics of plants. Kluwer, Dordrecht, pp 117–150

Soltis DE, Soltis,PS, Doyle JJ (1998) Molecular systematics of plants II. DNA sequencing. Kluwer, Dordrecht

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Wehner TC, Shetty NV, Elmstrom GW (2001) Breeding and seed production. In: Maynard DN (ed) Watermelons. Characteristics, production and marketing. ASHS Press, Alexandria, Va., pp 27–73

Weising K, Gardner R (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Xu DH, Abe J, Kanazawa A, Gai JY, Shimamato Y (2001) Identification of sequence variation by PCR-RFLP and its application to the evaluation of cpDNA diversity in wild and cultivated soybeans. Theor Appl Genet 102:683–688

Xu DH, Abe J, Gai JY, Shimamato Y (2002) Diversity of chloroplast SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet 105:645–653

Yanev Z, Shabelsky E, Schafferman D (1999) Colocynth: potential arid land oilseed from an ancient cucurbit. In: Janick J (ed) Perspectives on new crops and new uses. ASHS Press, Alexandria, Va., pp 257–261