Australian Journal of Crop Science AJCS

AJCS 11(08):1033-1041 (2017) ISSN:1835-2707
doi: 10.21475/ajcs17.11.08.pne572

Microsatellite-based genetic diversity among accessions of *Citrullus* spp. collected from 36 countries

Chia M.V. Angui[1], Kouamé K. Koffi[1], Kouamé G. Koffi[1], Marie-Christine Flamand[2], Pierre Bertin[3], Jean-Pierre Baudoin[4], Yao Djè [1] and Bi I.A. Zoro[1]

[1]Université Nangui Abrogoua, Unité de Phytotechnie et Amélioration génétique, 02 BP 801 Abidjan 02, Côte d'Ivoire
[2]Institut des Sciences de la Vie (ISV), Université catholique de Louvain, Croix du Sud 4-5, (L7.07.14), 1348 Louvain-la-Neuve, Belgium
[3]Earth and Life Institute (ELI), Université catholique de Louvain, Croix du Sud 2, (L7.05.11), 1348 Louvain-la-Neuve, Belgium
[4]Laboratoire d'Agroécologie tropicale et Horticulture, Gembloux Agro-Bio Tech, Université de Liège, Gembloux, Belgium

Abstract

Watermelon (*Citrullus lanatus*) is one of the most economically important vegetable cucurbits. However, the genetic and taxonomic statuses of its closely related species remain scantly documented, limiting their full use for agronomic purposes. The genetic diversity and structure of 74 accessions covering 47 dessert type (*C. lanatus* subsp. *vulgaris*), 21 oilseed type (*C. mucosospermus*), and 6 citron melon (*C. lanatus* subsp. *lanatus* var *citroides*) collected from 36 countries throughout 4 continents, were analyzed using 18 polymorphic SSR markers. The mean values of proportion of polymorphic loci (*P* = 29.73), number of alleles per locus (*A* = 1.243), effective number of alleles per locus (*Ae* = 1.153) Shannon index (*I* = 0.191), observed and expected heterozygosities (*Ho* = 0.124; *He* = 0.149) confirmed the narrow genetic basis of *C. lanatus*. According to molecular variance analysis the most important component of the genetic variation was obtained among accessions (70%). On the contrary, lower genetic variation was noted among species (16%), countries (37%), and continents (14%). It is suggested that the cultivated forms of *Citrullus* spp. originated from or successive selection cycles aimed at few and/or common traits, in few ancestral populations. Clustering based on both Bayesian approach and an unweighted pair group method with arithmetic mean pointed out three groups of accessions corresponding to use types and collecting countries. Based on these results, future collecting missions could be focused mainly on representative ecological sites in *Citrullus* spp. distribution areas, and increasing the number of accessions and seeds per accession.

Keywords: *Citrullus lanatus*; *Citrullus mucosospermus*; Germplasm collection; Genetic structure; Null alleles; SSR; Watermelon.
Abbreviations: AMOVA_Analysis of Molecular Variance; DM_Dry Matter; GBS_ Genotyping by Sequencing; HWE_Hardy Weinberg Equilibrium; SSR_ Simple Sequences Repeat; SNP_Single Nucleotide Polymorphism.

Introduction

Watermelon [*Citrullus lanatus* (Thunb.) Matsum. and Nakai] belonging to Cucurbitaceae is one of the most economically important cultivated crop from this family. In 2012, the world acreage of watermelon was 3.47 million ha, representing 6.06% of the 57.27 million ha devoted to vegetable; its production is estimated to 105.4 million tons (FAOSTAT, 2016), representing 9.53% of the world vegetable production (1.1 billion tons).

In Africa, with 247,086 ha and 5.95 million tons, watermelon accounted for 3.46% and 8.02% of acreage and tonnage devoted to vegetable in 2012, respectively. In Sub-Saharan Africa, dried seed of the oleaginous form can be sold at urban markets up to 2-3 €/kg which is 1.5-2 times higher than the price of cocoa and coffee beans (Zoro Bi et al., 2005). The crop is characterized by a tremendous diversity in fruit and seed shape, color, and size (Achigan-Dako et al., 2015; Gbotto et al., 2016). Studies refining the taxonomic status are in progress (Achigan-Dako et al., 2015; Chomicki and Renner, 2015), but from a practical point of view, two types of *Citrullus* spp. are distinguished on the basis of its use. The most widely cultivated type, actually recognized as *C. lanatus* subsp. *vulgaris*, mainly consumed as dessert, produces fruit with sweet and red colored flesh containing hard unoiled seeds. The second type bears fruit with white, cream or yellow colored flesh containing tender and oleaginous seeds. Roasted and crushed kernels of this type are widely used to prepare cakes or as condiment to enhance taste and thickness of sauces. This type is recognized as *C. mucosospermus* (Achigan-Dako et al., 2015).

Nowadays, indigenous cultivated species of *Citrullus* are receiving growing interest from governments, plant genetic resources institutions and researchers (Nantoumé et al., 2013). This is due to its numerous agronomic attributes, particularly its adaptability to contrasted ecological zones (both temperate and tropical regions) and various cropping systems requiring minimal input (Jensen et al., 2011). In addition to these agronomic advantages, the crop is rich in proteins (36% DM) and lipids (45% DM) (Loukou et al., 2007). *Citrullus* species are therefore interesting plant models

for the implementation of breeding program and cropping systems in the context of climate change (Mc Gregor et al., 2014). However, contrary to the dessert type of *C. lanatus* involved in numerous extensive research programs aimed at breeding for pest-tolerant and high yielding varieties (Gusmini and Wehner, 2008) and crop husbandry optimization (Nerson, 2002), investigations devoted to breeding the oleaginous type for yield and local adaptation are scarce (Hashizume et al., 2003). Such investigations start by assembling a representative germplasm collection of the crop with a comprehensive documentation (Nass et al., 2012). Several studies aimed at collecting and genetically characterizing the oleaginous watermelon have been carried out and interesting results have been reported on the genetic richness, differentiation, and phylogenetic relationships (Djè et al., 2006, 2010; Gbotto et al., 2015; Nantoumé et al., 2013). However studies using reliable sample size to compare the genetic diversity in *Citrullus* spp. at macro geographical scale (e.g. country and continent levels) are scant (Nimmakayala et al., 2010). Results from such investigations are useful to implement reliable genetic resources collection and conservation strategy, including the plant material and collecting zones selection, as well as to define programs for genetic resources exchange and regeneration (Upadhyaya and Ortiz, 2001). Results should also help in defining appropriate sampling strategies for the sustainable conservation of *C. lanatus* genetic resources. The present study was undertaken to evaluate and compare the SSR-based genetic diversity in *C. lanatus* in several collecting zones (countries and continents).

Results

Genetic diversity and mating pattern

The number of loci detected among the 74 accessions using the 19 SSR primers ranged from 3 for MCPI-07 to 15 for MCPI-13 (Table 1). The polymorphic information content values of the 19 primers ranged from 0.249 for TJ-10 to 0.788 for MCPI-28 with an average of 0.586. Among the 19 SSR primers used for DNA amplification, 132 different alleles were identified, only one primer being monomorphic (MCPI14-01). The SSR primer MCPI-28 showed the highest number of accessions (25) in which null alleles were observed (Table 1). Only two markers (MCPI-07 and MCPI-14) did not show any null allele among the 74 accessions studied. Null alleles were detected in 10 accessions per SSR marker on average.

The proportion of polymorphic loci per accession (*P*) varied from 5.26 to 73.68% with a mean of 29.73 ± 2.17% (Table 2). Only 12 accessions (16.21%) showed *P* values higher than 50%. The mean number of alleles per locus (*A*) varied from 0.789 (PI 482283 and PI 596671) to 1.789 (PI 176487 and PI 246559) with an average mean of 1.243 ± 0.018, indicating a low allelic richness. The effective number of alleles varied from 0.768 (PI 596671) to 1.556 (PI 270563) with an average of 1.153 ± 0.015. The mean accession diversity using the Shannon information index (*I*) was 0.191 ± 0.008, with PI 506439 the most diverse accession (*I* = 0.470) and PI 596671 the least diverse (*I* = 0.030).

A low genetic diversity was also observed. Indeed, the average observed heterozygosity (*Ho*) was 0.124, ranging from 0.026 (PI 596671) to 0.298 (PI 482343); and the average unbiased expected heterozygosity (*He*) was 0.149 ± 0.006, ranging from 0.026 (PI 596671) to 0.367 (PI 176487).

The average fixation index of population (*F*) was 0.058 ± 0.020. According to the HWE test this mean value was significantly different from zero (*p*-value for all loci and all accessions < 0.01). The Student *t* test revealed no significant difference between observed and expected heterozygosities ($t = 1.880$; $P = 0.062$): the observed genotype frequencies for each locus in each accession were consistent with expectation values. However, HWE tests revealed that only 3 (16.67%) of the 18 polymorphic loci were in equilibrium ($P > 0.05$); 15 polymorphic loci departed significantly from HWE, indicating a reproductive pattern of non-random mating.

Genetic structure among and within accessions, countries and continents

AMOVA (Table 3) revealed that the accessions were structured and highly differentiated ($F_{ST} = 0.705$; $P = 0.010$). The largest proportion of genetic variation (70%) was found among accessions, and only 30% within accessions. Genetic differentiation analyses among species, countries, and continents showed opposed trends. Thus, lower genetic diversity was observed among species (16%), compared to higher diversity within species (84%). Similarly, variation among countries (27%) was lower than variation within countries (73%). Indeed, the genetic differentiation at country level was significant ($F_{ST} = 0.269$; $P = 0.010$). Similar trend was observed at continent level: 14% and 86% of genetic variation among and within continent levels, respectively, with significant genetic differentiation ($F_{ST} = 0.135$; $P = 0.010$).

Genetic relationships among accessions

The Bayesian structure clustering analysis was carried out for K = 1-10. The ΔK values based on the method of Evanno et al. (2005) were plotted against the K numbers of the subgroups. The maximum ΔK was obtained at K = 3. This value, considered as the best from which the genetic structure could be depicted among the 74 accessions, suggested the occurrence of three genetic groups (Fig 1). The first group (blue color) contained of accessions of citron type and one of dessert type (PI 225557). The second group (green color) mainly contained accessions of dessert type. The third group (red color) contained the oilseed type. The UPGMA-based dendrogram topology (Fig 2) supported mainly the clustering into three major genetic groups, in relation with the geographic origins and varietal distinction.

According to these results the markers SSR reveal mainly higher genetic diversity among the accessions, with low value for allelic richness but high gene flow between accessions. We recommend collecting accessions from as many distinctive ecological sites as possible. However the largest proportion of diversity among accessions at continental and country level suggests to use an increasing the number of seeds per accessions to capture a maximum of diversity.

Discussion

Polymorphism information content (PIC) is considered as the most reliable statistics indicating molecular marker discrimination power (Kesari et al., 2010). The three SSR markers that showed the highest PIC values in the present

Table 1. SSR marker names and polymorphism statistics resulting from the analysis of 74 accessions of *Citrullus* spp.

Name of markers	No alleles	No private alleles[e]	No alleles common in 25% accessions	No alleles common in 50% accessions	No accessions with null alleles[f]	PIC
C.1.1-20[a]	10	2	40	104	9	0.786
TJ10[b]	4	0	16	16	7	0.249
MCPI-04[c]	9	3	53	53	9	0.715
MCPI-05[c]	9	3	27	54	9	0.680
MCPI-07[c]	3	0	18	18	0	0.571
MCPI-12[c]	6	2	23	23	1	0.727
MCPI-13[c]	15	1	73	73	13	0.739
MCPI-14[c]	5	2	7	33	0	0.470
MCPI-16[c]	6	0	34	58	14	0.707
MCPI-21[c]	7	4	22	22	6	0.408
MCPI-24[c]	7	1	30	67	7	0.595
MCPI-26[c]	8	2	7	57	24	0.686
MCPI-28[c]	7	1	29	103	25	0.788
MCPI-30[c]	4	1	3	35	6	0.497
MCPI-32[c]	5	2	10	43	14	0.548
MCPI-33[c]	13	3	41	106	16	0.744
MCPI-37[c]	5	2	21	21	9	0.411
MCPI-39[c]	4	1	14	14	20	0.507
MCPI-42[c]	5	1	25	25	6	0.299
Total	132	31	-	-	-	-
Average	6.94	1.63	25.95	48.68	10.26	0.586
SE[d]	3.15	1.12	17.18	30.55	7.26	0.164

[a] Jarret et al. ; [b] Gonzalo et al. (2005); [c] Joobeur et al. (2006). [d] SE: standard error.[e] Number of allele unique in a single accession (Slatkin, 1985).
[f] Estimated using the maximum likelihood estimator based on the EM algorithm (Dempster et al., 1977).

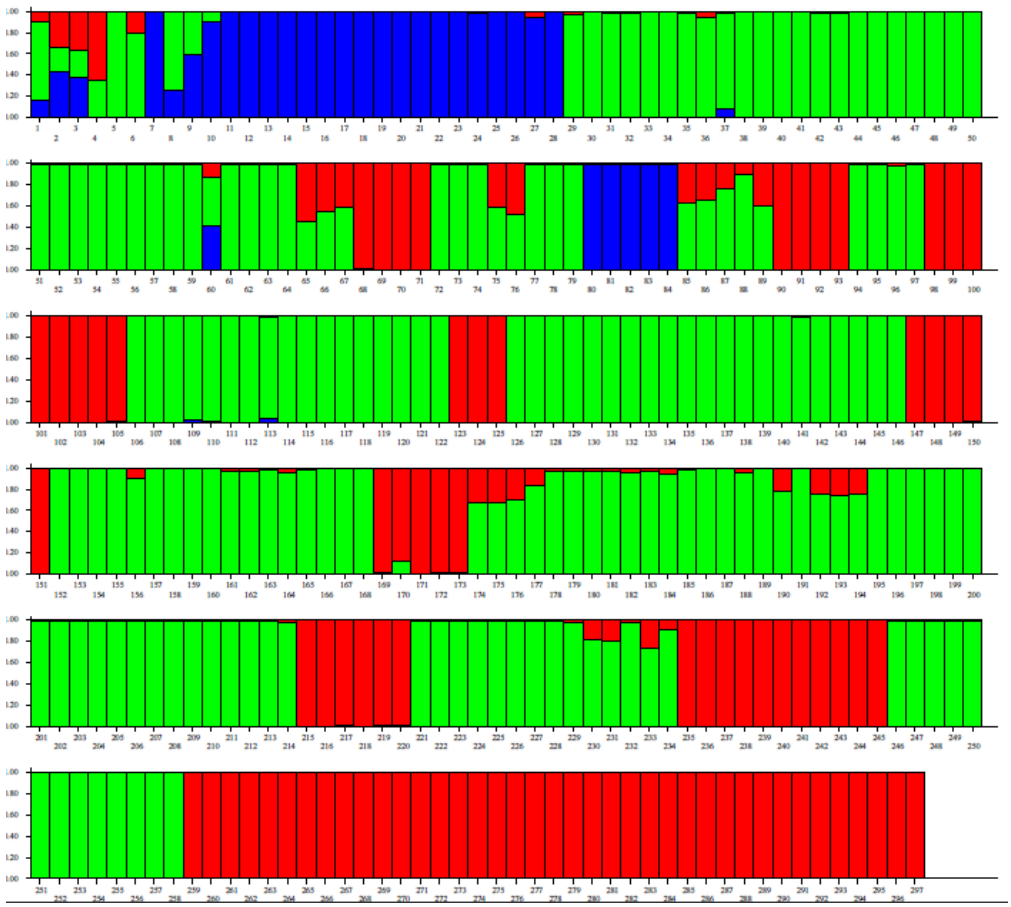

Fig 1. Population structure as inferred from Bayesian approach implemented in Structure software (Pritchard et al., 2000), based on 74 accessions of *Citrullus* spp. and analyze of 18 SSR markers. The values on the Y-axis correspond to membership coefficients of each genotype and the number on the X-axis indicate the individual of each accession as indicated in table S1. The blue population is composed of *C. lanatus* var. *citroides*; the green population concerns *C. lanatus* subsp. *vulgaris* (dessert type) and the red corresponds to *C. mucosospermus* (oilseed type).

Table 2. Estimates of SSR marker statistics and genetic diversity indices of 74 accessions of *Citrullus* spp.

Accessions	Marker statistics[a]			Genetic diversity indices[b]			
	P	*A*	A_e	*I*	H_o	H_e	*F*
PI 271778	47.37	1.474	1.356	0.314	0.088	0.256	0.611
PI 482343	47.37	1.316	1.241	0.304	0.298	0.288	-0.356
PI 270563	68.42	1.737	1.556	0.462	0.162	0.361	0.520
PI 299378	5.26	0.895	0.895	0.036	0.053	0.030	-1.000
PI 482283	5.26	0.789	0.789	0.036	0.053	0.032	-1.000
PI 500354	10.53	0.947	0.915	0.060	0.070	0.049	-0.600
PI 532624	21.05	1.053	0.929	0.088	0.084	0.061	-0.333
PI 596671	5.26	0.789	0.768	0.030	0.026	0.026	-0.333
Grif 1734	52.63	1.684	1.354	0.348	0.157	0.242	0.231
Grif 5597	42.11	1.368	1.203	0.232	0.083	0.179	0.457
PI 113326	36.84	1.316	1.211	0.244	0.162	0.190	-0.013
PI 163203	47.37	1.474	1.365	0.313	0.079	0.244	0.604
PI 163574	57.89	1.632	1.416	0.335	0.208	0.246	-0.010
PI 164634	5.26	0.947	0.947	0.036	0.053	0.029	-1.000
PI 165448	21.05	1.105	1.030	0.117	0.088	0.095	-0.100
PI 176487	73.68	1.789	1.531	0.463	0.254	0.367	0.145
PI 179883	47.37	1.474	1.324	0.275	0.105	0.225	0.378
PI 184800	21.05	1.105	1.074	0.137	0.123	0.116	-0.250
NI 186489	21.05	1.211	1.198	0.143	0.184	0.117	-0.800
PI 193963	52.63	1.579	1.399	0.322	0.168	0.240	0.187
PI 219691	52.63	1.368	1.262	0.331	0.193	0.277	0.180
PI 225557	10.53	0.895	0.867	0.063	0.053	0.048	0.000
PI 246559	63.16	1.789	1.535	0.445	0.200	0.313	0.336
PI 249008	52.63	1.632	1.441	0.371	0.184	0.283	0.243
PI 254623	21.05	1.158	1.088	0.119	0.066	0.094	0.214
PI 254735	5.26	1.000	1.000	0.036	0.053	0.030	-1.000
PI 254740	5.26	1.053	1.053	0.036	0.053	0.030	-1.000
PI 269677	10.53	0.947	0.937	0.070	0.088	0.060	-0.750
PI 270545	26.32	1.316	1.256	0.187	0.116	0.139	0.177
PI 270551	36.84	1.421	1.311	0.263	0.162	0.202	0.067
PI 271981	10.53	0.947	0.943	0.072	0.053	0.057	0.000
PI 306782	26.32	1.316	1.250	0.183	0.211	0.168	-0.667
PI 307750	52.63	1.579	1.453	0.353	0.123	0.291	0.517
PI 344395	5.26	1.000	1.000	0.036	0.053	0.032	-1.000
PI 345545	21.05	1.105	1.014	0.110	0.083	0.083	-0.086
PI 346787	21.05	1.053	1.032	0.140	0.105	0.119	-0.042
PI 357656	36.84	1.316	1.148	0.213	0.133	0.154	-0.036
PI 435991	26.32	1.158	1.104	0.164	0.088	0.137	0.227
PI 438675	63.16	1.684	1.471	0.429	0.271	0.319	0.006
PI 458739	42.11	1.421	1.245	0.295	0.154	0.212	0.180
PI 470249	21.05	1.105	1.078	0.137	0.066	0.111	0.367
PI 482272	21.05	1.211	1.179	0.135	0.105	0.106	0.000
PI 487476	5.26	1.000	1.000	0.036	0.053	0.032	-1.000
PI 490386	31.58	1.316	1.194	0.175	0.053	0.132	0.667
PI 491265	42.11	1.368	1.150	0.237	0.145	0.177	0.054
PI 500301	10.53	1.000	0.994	0.071	0.092	0.081	-0.800
PI 500307	15.79	1.053	1.046	0.108	0.118	0.088	-0.511
PI 505584	52.63	1.632	1.391	0.346	0.263	0.250	-0.185
PI 506439	68.42	1.737	1.518	0.470	0.284	0.347	0.129
PI 507862	10.53	1.000	0.989	0.070	0.088	0.060	-0.750
PI 512359	36.84	1.421	1.327	0.273	0.118	0.213	0.410
PI 512368	36.84	1.316	1.239	0.224	0.137	0.175	0.146
PI 512381	31.58	1.368	1.216	0.185	0.158	0.140	-0.315
PI 525095	36.84	1.368	1.238	0.234	0.137	0.173	0.256
PI 532723	47.37	1.526	1.431	0.337	0.281	0.274	-0.232
PI 532730	36.84	1.421	1.324	0.247	0.193	0.200	-0.119
PI 536451	42.11	1.316	1.122	0.211	0.132	0.158	0.012
PI 536459	10.53	1.000	0.972	0.063	0.053	0.048	0.000
PI 542617	47.37	1.474	1.315	0.310	0.174	0.224	0.122
PI 559994	42.11	1.368	1.250	0.273	0.145	0.209	0.232
PI 560024	36.84	1.263	1.167	0.225	0.132	0.177	0.219
PI 595203	21.05	1.105	1.084	0.140	0.105	0.140	0.000
PI 612464	5.26	0.947	0.947	0.036	0.053	0.030	-1.000
PI 658680	31.58	1.158	1.039	0.201	0.088	0.158	0.327

NI 208	21.05	1.053	0.999	0.127	0.070	0.105	0.200
NI 209	5.26	0.842	0.842	0.036	0.053	0.032	-1.000
NI 213	15.79	1.053	0.974	0.107	0.126	0.078	-0.725
NI 214	10.53	1.000	1.000	0.073	0.105	0.060	-1.000
NI 216	36.84	1.368	1.220	0.230	0.158	0.170	0.128
NI 217	26.32	1.211	1.094	0.168	0.116	0.125	0.113
NI 218	10.53	0.947	0.947	0.073	0.105	0.059	-1.000
NI 219	26.32	1.158	1.126	0.171	0.158	0.135	-0.207
NI 223	21.05	1.158	1.066	0.140	0.147	0.103	-0.555
NI 225	5.26	0.895	0.895	0.036	0.053	0.029	-1.000
Mean	29.73%	1.243	1.153	0.191	0.124	0.149	0.058
SE[c]	2.17%	0.018	0.015	0.008	0.008	0.006	0.020

[a] *P*: percentage of polymorphic loci; *A*: number of alleles per polymorphic locus; A_e: effective number of alleles per polymorphic locus
[b] *I*: Shannon's information index; *Ho*: observed heterozygosity; *He*: unbiased expected heterozygosity; *F*: Wright's (1965) fixation index
[c] SE: standard error

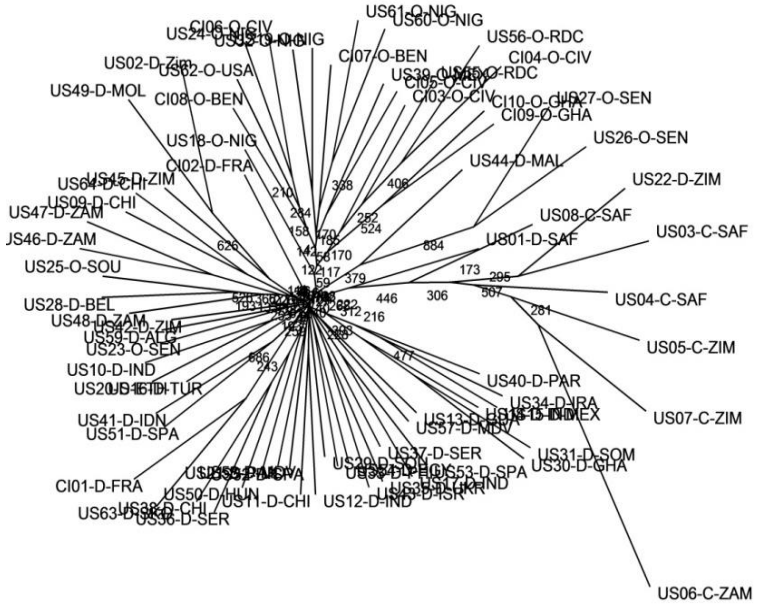

Fig 2. Dendogram of 74 accessions of *Citrullus* spp. produced by UPGMA cluster analysis of Nei's (1978) genetic distance matrix based on 18 SSR markers. Numbers shown at different nodes represent percentage obtained in the bootstrap analysis. US: accession from USDA-ARS; CI: accession from Côte d'Ivoire; C: citron type, D: dessert type, and O: oilseed type.ZIM: accession from Zimbabwe; SAF: accession from South of Africa; ZAM: accession from Zambia; CHI: accession from China; IND: accession from India; GUA: accession from Guatemala; MEX: accession from Mexico; TUR: accession from Turkey; NIG: accession from Nigeria; ETH: accession from Ethiopia; PAK: accession from Pakistan; SEN: accession from Senegal; SOU: accession from Soudan; BEL: accession from Belize; GHA: accession from Ghana; SOM: accession from Somalia; PHI: accession from Philippines; IRA: accession from Iran; UKR: accession from Ukraine; SER: accession from Serbia-Montenegro; PAR: accession from Paraguay; IDN: accession from Indonesia; ISR: accession from Israel; MAL: accession from Mali; MOL: accession from Moldova; HUN: accession from Hungary; SPA: accession from Spain; EGY: accession from Egypt; RDC: accession from RD Congo; MLD: accession from Maldives; ALG: accession from Algeria; USA: accession from USA; SKO: accession from South of Korea; FRA: accession from France; CIV: accession from Côte d'Ivoire; BEN: accession from Benin.

Table 3. Analysis of molecular variance of 74 accessions of *Citrullus* spp. collected in 36 countries and four continents based on data from 19 microsatellite loci

Hierarchical level	Source of variation	d.f.	M.S.	%Var.	F_{ST}	*P*
Partioning all accessions	Among accessions	73	65.596	70%	0.705	0.010
	Within accessions	223	6.209	30%		
	Total	296		100%		
Partioning species	Among species	2	305.336	16%	0.164	0.010
	Within species	294	18.920	84%		
	Total	296		100%		
Partioning per countries	Among countries	9	118.272	27%	0.269	0.010
	Within countries	155	16.878	73%		
	Total	164		100%		
Partioning per continent	Among continents	3	193.981	14%	0.135	0.010
	Within continent	293	19.083	86%		
	Total	296		100%		

d.f.: degree of freedom, M.S.: mean squares, %Var.: distribution of total variance, F_{ST}: genetic differentiation, *P*: probability.

study (MCPI-13, MCPI-33, and C.1.1-20) have also been identified as the most variable in previous studies (Joobeur et al., 2006; Nantoumé et al., 2013; Gbotto et al., 2015). However, the PIC values obtained in the present study were higher than those reported by these authors. Such a difference could be explained by the difference in the origin and size of the samples. However, Mujaju et al. (2011) reported a higher PIC mean value (0.92) in a study using 10 SSR markers to analyze 25 accessions of *C. lanatus* collected from 5 Southern African countries (Botswana, Namibia, South Africa, Zambia, and Zimbabwe). This difference could be attributed to the origin of plant material sampled, as accessions analyzed by Mujaju et al. (2011) were obtained from farmers located in the center of origin of *C. lanatus*, whereas those used in the present study were collected from countries mainly located outside of this area. In addition, they were obtained from genebanks where they have certainly undergone several cycles of regeneration. Depletion of genetic diversity throughout successive regenerations of accessions has been demonstrated elsewhere (Cieslarová et al., 2011).

Genetic diversity indices (*Ho* and *He*) and fixation index (*F*) estimated in this study were close to those reported from several previous studies based on SSR markers (Nantoumé et al., 2013; Gbotto et al., 2015) and were generally low. These results suggested that although SSR markers appeared to be suitable for investigation on genetic diversity in *Citrullus* genus, they did not reflect its tremendous morphological variability, namely the plant architecture as well as fruit and seed shape, size, and color (Achigan-Dako et al., 2015; Gbotto et al., 2016). This is probably due to the complex and multigenic inheritance of these traits in *C. lanatus* (Guner and Wehner, 2004). Discrepancy between the trend of variation revealed by morphological and molecular markers has been observed in several cultivated crops such as melon (Zhang et al., 2012), and zucchini (Ferriol et al., 2003).

No significant difference between observed and expected heterozygosities was noted, indicating that the observed genotype frequencies for each locus in each accession were consistent with expectations. However, HWE tests revealed that the majority of polymorphic loci (15 out of 18) were in disequilibrium ($P < 0.05$), indicating a reproductive pattern of non-random mating. The departure of genotypic frequencies from HWE could be attributed to founder effects, a higher selfing rate, assortative mating (homogamy), or selection favoring homozygote individuals. However, the predominantly outcrossing mating system of *C. lanatus* does not fit with homogamy and homozygotes selection hypotheses. Indeed, *C. lanatus* that is a monoecious or andromonoecious species is bound to experiment insect-mediated crosspollination, which promotes random mating, buffering homogamy and homozygotes selection. Two main causes could explain the HWE departure of genotypic frequencies observed in this study. The first cause is the founder effect (bottleneck), due to farmers' seed selection approaches coupled with sampling strategy for germplasm regeneration, resulting in unbalanced gene frequencies. Indeed, usually rural farmers take their seeds (generally small in number) from the previous harvest, or obtained them from neighboring farmers or local markets, resulting in genetic variability depletion. The second cause is the relative high prevalence of null alleles observed in the accessions analyzed. Null alleles affects generally the estimation of population genetic indices (Chapuis and Estoup, 2007), leading to an underestimation of observed heterozygosity and overestimation of population differentiation due to gene diversity reduction (Chapuis and Estoup, 2007). Based on a simulation study, Dakin and Avise (2004) showed that null alleles can introduce substantial errors into empirical assessments of specific mating events by leading to high frequencies of false parentage exclusions.

Results from genetic structure based on AMOVA showed that *C. lanatus* maintained 70% of the genetic variability among accessions and only 30% within accessions. Nantoumé et al. (2013) have reported similar trend in accessions of *C. lanatus* subsp. *vulgaris* and *C. lanatus* subsp. *mucosospermus*. AMOVA performed by these authors revealed a significant differentiation among accessions (51% of the total molecular variation). As demonstrated in the previous paragraph, the trend observed in the present study should be attributed to farmer's seed management strategy or the germplasm regeneration approaches generally adopted by genebank curators. In addition, during germplasm regeneration, accessions are spatially or temporally isolated to avoid undesirable pollination, consequently increasing genetic differentiation among them (Suso et al., 2008).

Our results in *C. lanatus* show high intra-varietal variability (84%) compared to inter-varietal (16% of total variation), indicating high genetic similarity between the three forms of watermelon studied. Using four species of *Citrullus* genus, Jarret and Newman (2000) reported similar result from the analysis of internal transcribed spacer (ITS) sequences data. Indeed, the cladistics (PAUP) analysis performed by these authors showed no significant differentiation between the dessert and the oilseed forms. Furthermore, Nantoumé et al. (2013) reported low inter-varietal genetic differentiation in watermelon landrace types from Mali. Natural pollination between varieties or cultivars (Ferreira et al., 2000) and hybrid vigor (Bansal et al., 2002; Singh et al., 2009) largely observed in *Citrullus* genus supported such results.

AMOVA also revealed that the geographical origin of the accessions explained a relatively low percent of molecular variation: only 27% and 14% of variation at country and continent levels, respectively. Such a trend could be explained by three genetic events: high level of macro-geographical gene flow, few independent migrations from the center of origin, and selection based on few and/or common agronomic traits. Country cross-border transportation of crop seeds for trade and/or agronomy purposes is a common practice, particularly in Africa (Asfaw et al., 2009; Mokuwa et al., 2014). This phenomenon has been reported from recent studies investigating the genetic structure of *C. lanatus* collected in several African countries (Nantoumé et al., 2013; Gbotto et al., 2016). The presence of few private alleles (1.63 ± 1.12 per SSR marker), coupled with a relatively high prevalence of common alleles (7-73 alleles were common in 25% of accessions analyzed) supported the first hypothesis. Concerning the second hypothesis, it is worth noting that the domestication of watermelon outside its center of origin is well documented (Jarret and Newman, 2000). After migration, selection pressure in *C. lanatus* was based on few and/or common desirable traits, leading to low geographical genetic differentiation. Indeed, breeding *C.lanatus* for high yielding was generally devoted to non-exhaustive farmer- and consumers-preferred traits, such as pest tolerance and desirable fruit, flesh, and seed characteristics, according to use purposes (Rhodes and Zhang, 2000).

The clustering based on both Bayesian approach and UPGMA resulted in grouping accessions into three clusters, corresponding to the use forms (dessert, oilseed, and citron). The branch separating citron form was well supported by the bootstrap values (44-100%). One citron accession (PI 270563) was classified in the oilseed group. Such result indicated that citron and oilseed forms that are generally

indigenous landraces, share a large part of the genome, as highlighted by studies conducted by Jarret and Newman (2000). In fact, it is sometimes difficult to establish phenotypic difference between citron and oilseed types (Nantoumé et al., 2013), demonstrating that the probability of their misclassification occurs. The clear classification of *Citrullus* spp. is still under debate (Dane and Liu, 2007; Chomicki and Renner 2015; Paris, 2015). To our knowledge, simple, clear and reliable morphological traits discriminating indisputably dessert from oilseed forms are scarce in literature (Achigan-Dako et al., 2015).

Materials and methods

Plant materials

Seventy-four open pollinated accessions of *Citrullus* spp. originating from 36 countries were used in this study (Table S1). This plant material was composed of 47 dessert type (*C. lanatus* subsp. *vulgaris*), 21 oilseed type (*C. mucosospermus*), and 6 citron melon (*C. lanatus* suspp. *lanatus* var *citroides*), called citron in this study. Sixty-four of these accessions were obtained from United States Department of Agriculture - Agricultural Research Service (USDA-ARS) while ten were provided by Nangui Abrogoua University (UNA, Abidjan, Côte d'Ivoire). The accessions received from the USDA-ARS genebank originated from different countries of Africa, Asia, America, and Europe. Those from UNA included four accessions collected in different regions of Côte d'Ivoire, four from two other West African countries (Benin and Ghana), and two from France (Table S1). The seeds were sown to obtain young leaves for molecular analysis. From three to five plantlets per accession were analyzed, with a total of 297.

DNA extraction and PCR amplification

DNA was extracted from young leaf tissue using CTAB method according to the protocol of Murray and Thompson (1980) with few modifications. Dried leaf (0.075 g) tissue was finely ground in 1.5 mL microtubes (Eppendorf) in liquid nitrogen and resuspended in 700 µL CTAB extraction buffer [0.1 M Tris-Base, 1.4 M NaCl, 2.5% cetyltrimethylammonium bromide (CTAB), 20 mM EDTA-dissodium, 0.2% sodium dodecyl sulfate (SDS), 0.5% Sarkosyl, 250 mg polyvinylpyrrolidone (MW 40 (PVP-40) and 250 mg polyvinylpyrrolidone (PVP)]. Each tube was mixed by gentle agitation and then incubated for 30 min at 65˚C. The supernatant was taken and 350 µL of isopropanol were added to precipitate the DNA. The DNA pellet was washed in absolute ethanol and dried. Then the pellet was resuspended in TE to a final concentration of 100 ng μL^{-1} containing 10 g mL^{-1} RNAse. The DNA solution was stored at −20˚C until use. DNA concentration was measured by a NanoDrop ND-1000 spectrophotometer (NanoDrop Tech., Wilmington, Delaware, USA).

Nineteen microsatellite primer pairs were selected and combined in three multiplex. Forward primers were fluorescently labeled with 6-FAM, HEX or TAMRA. The PCR was performed in a 5µL solution containing 0.5µL of primers (2µM), 1 µL of demineralized water, 1µL diluted DNA to 20 ng μL^{-1} and 2.5 µL of QIAGEN Multiplex PCR Master Mix.

The reactions were realized with a thermal cycler Mastercycler pro S (Eppendorf AG, Hamburg, Germany). The first step (15 min at 95°C) allowed the activation of the Taq polymerase and alter DNA long fragment. The second step was constituted of 30 cycles, each including a denaturation step (30 s at 94 °C), followed by primer hybridization (90 s at 58 °C) and elongation (60 s at 72 °C). In the last step (30 min at 72 °C) adenine (A) was added to obtain accurate data from high-resolution analysis using DNA capillary sequencers (Qiagen Multiplex PCR Handbook, 10/2010). At the end, thermal cycler temperature was reduced to 4°C in order to maintain the reaction products intact.

PCR products were separated and analyzed using capillary gel electrophoresis. Chromatograms were analyzed using Peak Scanner software V1.0 (Applied Biosystems, Foster City, CA, USA).

Data analysis

Genetic diversity analysis

To evaluate the informativeness of each marker, polymorphic information content (PIC) of each SSR locus was calculated, based on the alleles frequencies, using the formula PIC = 1- $\sum p_i^2$, in which p_i is the frequency of the i^{th} allele of the locus considered.

The program GenAlEx v. 6.4 (Peakall and Smouse, 2012) was used to calculate the percentage of polymorphic loci (*P*), the number of alleles per polymorphic locus (*A*), the effective number of alleles per polymorphic loci (*Ae*), the Shannon's information index (*I*), the observed heterozygosity (*Ho*), the unbiased expected heterozygosity (*He*) calculated following Nei (1978), and the Wright's (1965) fixation index (*F*) following Weir and Cockerham (1984). Null allele frequency estimation based on the maximum likelihood algorithm proposed by Dempster et al. (1977) and implemented in Genepop software (Rousset, 2008) was used to calculate the frequencies of null alleles within each accession.

The occurrence of evolutionary forces acting on individuals and accessions was examined for each locus through Hardy–Weinberg equilibrium (HWE) tests performed by the Genepop software. In addition, observed and expected heterozygosities were compared according to Student *t* test. This statistic test was performed after checking normal distribution and variances equality assumptions (Dagnelie, 2011) using the R statistical package (R Development Core Team, 2011).

Genetic structure analysis

The degree of genetic differentiation within and among the accessions, forms, countries, and continents was determined by analysis of molecular variance (AMOVA), based on the estimates of F_{ST} (Excoffier and Smouse, 1994). Then, the hierarchical level upon which genetic variation could be attributed was investigated. Only the ten countries in which at least three accessions were sampled were included in country-based AMOVA. Significance of AMOVA was tested using permutation test (Excoffier et al., 1992) with 999 permutations.

Genetic relationships among accession

The most likely number of populations and the assignment of each accession to populations were determined following a model based on clustering algorithm implemented in the Structure software (Pritchard et al., 2000). This algorithm also examines heterogeneity within accession. From 1 to 10 putative populations (K1-K10) were tested with 50,000 burn-in iterations and 1,000,000 iterations after burn-in or Markov

Chain Monte Carlo simulation, using the admixture model. The best value of K was evaluated following Evanno et al. (2005).

An unweighted pair group method with arithmetic mean (UPGMA) tree based on the Nei's genetic distance (Nei, 1978) matrix was constructed in Phylip package version 3.69 (Felsenstein, 1995). The relationships among and within accessions considered at both and countries levels were described using cluster analysis. First, 1000 bootstraps were performed on Seqboot program to generate confidence in the dataset. Then, Gendist program was used to compute genetic distance from gene frequencies (Nei, 1978). The cluster analysis tree was produced with the Neighbor program which uses a matrix of pairwise distances between all pairs of accessions and Consensus program. Confidence in tree topology was assessed by 1000 bootstrapping over SSR markers and the phylogenetic tree was visualized in Treeview (Page, 1996).

Conclusion

Results of the present study and those from previous ones confirmed the reliability of SSRs as genetic markers for effective genetic diversity and structure analysis in *C. lanatus*. A higher proportion of genetic diversity was found among than within accessions. This trend was inverted when partitioning were based on species and collecting zones, where higher proportions of genetic diversity were found within than among species, countries and continents. Consequently, we recommend focusing collecting missions on representative ecological sites in the distribution area of *C. lanatus*. However, as the largest proportion of diversity was attributable to the within component at both species and geographical (country and continent) levels, increasing the number of accessions in each collecting site, with high number of seeds per accession can capture a maximum of diversity. The study also has potential implications for oilseed type breeding that actually is at the starting point. Indeed, given the genetic proximity of dessert type to oilseed type, crosses between them to explore hybrid vigor appear as an interesting perspective. In each of the main clusters, although not supported by significant bootstrap values, accessions were grouped according to their origins, indicating that diversity in *C. lanatus* is potentially structured in space. The use of next generation molecular markers (e.g. SNP, GBS, etc.) should also clarify the understanding of the genetic structure of watermelon. In particular, it can help in understanding the remarkable morphological diversity characterizing *C. lanatus* genetic resources especially in terms of fruit and seed agronomic traits.

Acknowledgments

The authors would like to thank Mister Bob Jarret from the United the States Department of Agriculture - Agricultural Research Service (USDA-ARS) for providing accessions for this study.

Funding: This research (PIC 2010) and the fellowship (SOFT 2015) of the first author were financially supported by the Académie de Recherche et d'Enseignement Supérieur – Commission de la Coopération pour le Développement (ARES-CCD, Brussels, Belgium).

References

Achigan-Dako EG, Avohou ES, Linsoussi C, Ahanchede A, Vodouhe RS, Blattner FR (2015) Phenetic characterization of *Citrullus* spp (Cucurbitaceae) and differentiation of egusi-type (*C mucosospermus*). Genet Resour Crop Ev. 62:1159–1179.

Asfaw A, Blair MW, Almekinders C (2009) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L) landraces from the east african highlands. Theor Appl Genet. 120:1–12.

Bansal R, Sooch BS, Dhall RK (2002) Heterosis in watermelon (*Citrullus lanatus* (Thunb (Mats). Environ Ecol. 20:976–979.

Chapuis MP, Estoup A (2007) Microsatellites null alleles and estimation of population differentiation. Mol Biol Evol. 24:621–631.

Chomicki G, Renner SS (2015) Watermelon origin solved with molecular phylogenetics including linnaean material: another example of museomics. New Phytol. 205:526–532.

Cieslarová J, Smýkal P, Dočkalová Z, Hanáček P, Procházka S, Hýbl M, Griga M (2011) Molecular evidence of genetic diversity changes in pea (*Pisum sativum* L) germplasm after long-term maintenance. Genet Resour Crop Ev. 58:439–451.

Dagnelie P (2011) Statistique théorique et appliquée Tome 2 Inférence statistique à une et à deux dimensions 3rd ed De Boeck Bruxelles (Belgique).

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity. 93:504–509.

Dane F, Liu J (2007) Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). Genet Resour Crop Ev. 54:1255–1265.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B. 39:1–38.

Djè Y, Tahi CG, Zoro Bi AI, Baudoin JP, Bertin P (2010) Use of ISSR markers to assess genetic diversity of African edible seeded *Citrullus lanatus* landraces. Sci Hortic-Amsterdam. 124:159–164.

Djè Y, Tahi CG, Zoro Bi AI, Malice M, Baudoin JP, Bertin P (2006) Optimization of ISSR markers for African edible-seeded Cucurbitaceae species genetic diversity analysis. Afr J Biotechnol. 5:83-87.

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

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. Genetics. 136:343–359.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131:479–491.

FAOSTAT (2016) Crop production in 2012: watermelon and primary vegetable data.

Felsenstein J (1995) Phylogeny inference package (phylip) University of Washington Seattle (USA).

Ferreira MAJD, Vencovsky R, Vieira MLC, Queiróz MA (2000) Outcrossing rate and implications for the improvement of a segregating population of watermelon. Acta Hortic. 510:47–54.

Ferriol M, Pico B, Nuez F (2003) Genetic diversity of a germplasm collection of Cucurbita pepo using SRAP and AFLP markers. Theor Appl Genet. 107:271–282.

Gbotto AA, Koffi KK, Bi NDF, Bi STD, Tro HH, Baudoin J-P, Bi IAZ (2016) Morphological diversity in oleaginous watermelon (*Citrullus mucosospermus*) from the nangui abrogoua university germplasm collection. Afr J Biotechnol. 15:917–929.

Gbotto AA, Koffi KK, Koffi KG, Baudoin J-P, Zoro Bi IA (2015) Genetic structure in oleaginous *Citrullus lanatus* from the nangui abrogoua university germplasm collection. Philipp Agric Sci. 98:105–110.

Guner N, Wehner TC (2004) The genes of watermelon. HortScience. 39:1175–1182.

Gusmini G, Wehner TC (2008) Fifty-five years of yield improvement for cucumber melon and watermelon in the united states. HortTechnology. 18:9–12.

Hashizume T, Shimamoto I, Hirai M (2003) Construction of a linkage map and QTL analysis of horticultural traits for watermelon [*Citrullus lanatus* (Thunb) Matsum & & Nakai] using RAPD, RFLP and ISSR markers. Theor Appl Genet. 106:779–785.

Jarret RL, Newman M (2000) Phylogenetic relationships among species of *Citrullus* and the placement of *C rehmii* De Winter as determined by Internal Transcribed Spacer (ITS) sequence heterogeneity. Genet Resour Crop Ev. 47:215–222.

Jensen BD, Hamattal MA, Touré FA, Nantoumé AD (2011) Watermelons in the sand of Sahara: Cultivation and use of indigenous landraces in the tombouctou region of Mali. Ethnobot Res Appl. 9:151–162.

Joobeur T, Gusmini G, Zhang X, Levi A, Xu Y, Wehner T, Oliver M, Dean R (2006) Construction of a watermelon BAC library and identification of SSRs anchored to melon or arabidopsis genomes. Theor Appl Genet. 112:1553–1562.

Kesari V, Madurai Sathyanarayana V, Parida A, Rangan L (2010) Molecular marker-based characterization in candidate plus trees of *Pongamia pinnata* a potential biodiesel legume. AoB Plants 2010 doi:101093/aobpla/plq017.

Loukou AL, Gnakri D, Djè Y, Kippré AV, Malice M, Baudoin JP, Zoro Bi IA (2007) Macronutrient composition of three cucurbit species cultivated for seed consumption in Côte d'Ivoire. Afr J Biotechnol. 6:529–533.

McGregor CE, Waters V, Vashisth T, Abdel-Haleem H (2014) Flowering time in watermelon is associated with a major quantitative trait locus on chromosome 3. J Am Soc Hortic Sci. 139:48–53.

Mokuwa A, Nuijten E, Okry F, Teeken B, Maat H, Richards P, Struik PC (2014) Processes underpinning development and maintenance of diversity in rice in west africa: Evidence from combining morphological and molecular markers. PLoS ONE 9 e85953.

Mujaju C, Zborowska A, Werlemark G, Garkava-Gustavsson L, Andersen SB, Nybom H (2011) Genetic diversity among and within watermelon (*Citrullus lanatus*) landraces in Southern Africa. J Hortic Sci Biotech. 86:353–358.

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8:4321–4325.

Nantoumé AD, Andersen SB, Jensen BD (2013) Genetic differentiation of watermelon landrace types in mali revealed by microsatellite (SSR) markers. Genet Resour Crop Ev. 60:2129–2141.

Nass LL, Sigrist MS, Ribeiro CS da C, Reifschneider FJB (2012) Genetic resources: the basis for sustainable and competitive plant breeding. Crop Breed Appl Biot. 12:75–86.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.

Nerson H (2002) Relationship between plant density and fruit and seed production in muskmelon. J Am Soc Hortic Sci. 127:855–859.

Nimmakayala P, Tomason YR, Jeong J, Ponniah SK, Karunathilake A, Levi A, Perumal R, Reddy UK (2010) Genetic reticulation and interrelationships among Citrullus species as revealed by joint analysis of shared AFLPs and species-specific SSR alleles. Plant Genet Resour-C. 8:16–25.

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 12:357–358.

Paris HS (2015) Origin and emergence of the sweet dessert watermelon Citrullus lanatus. Ann Bot-London. 116:133–148.

Peakall R, Smouse PE (2012) GenAlEx 65: genetic analysis in Excel Population genetic software for teaching and research—an update bioinformatics. 28: 2537–2539.

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

R Development Core Team (2011) R: A language and environment for statistical computing R foundation for statistical computing R foundation for Statistical Computing Vienna (Austria).

Rhodes B, Zhang X (2000) Hybrid seed production in watermelon. J New Seeds. 1:69–88.

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour. 8:103–106.

Singh SP, Dadwadia G, Annapurna (2009) Analysis of heterosis and combining ability status among diallel set of hybrids for yield and quality traits in watermelon (*Citrullus lanatus* Thunb). Veg Sci. 36:323–326.

Suso M, Nadal S, Roman B, Gilsanz S (2008) *Vicia faba* germplasm multiplication–floral traits associated with pollen mediated gene flow under diverse between plot isolation strategies. Ann Appl Biol. 152:201–208.

Upadhyaya H, Ortiz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor Appl Genet. 102:1292–1298.

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

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 19:395–420.

Zhang C, Pratap AS, Natarajan S, Pugalendhi L, Kikuchi S, Sassa H, Senthil N, Koba T (2012) Evaluation of morphological and molecular diversity among South Asian germplasms of *Cucumis sativus* and *Cucumis melo*. ISRN Agron. 2012:11.

Zoro Bi IA, Koffi KK, Djè Y, Malice M, Baudoin J-P (2005) Biodiversity of cucurbits consumed as sauce thickener in Côte d'Ivoire: a capital resource for the economic prosperity of rural women in: Segers H (Ed) Tropical Biodiversity: Science Data Conservation Global Biodiversity Information Facility (GBIF) Brussels (Belgium). pp 158–167.