Georeferenced phylogenetic analysis of a global collection of wild and cultivated Citrullus species

Enoch Gbenato Achigan-Dako[1], Herve Degbey[1], Iago Hale[2], and Frank Blattner[3]

[1]Affiliation not available
[2]University of New Hampshire, Durham, New Hampshire, USA
[3]Leibniz Institute of Plant Genetics and Crop Research (IPK)

August 17, 2020

Abstract

The geographical origin of watermelon (Citrullus lanatus) remains debated. While a first hypothesis suggests the center of origin to be west Africa, where a sister endemic species C. mucosospermus thrives, a second hypothesis suggests north-eastern Africa where the white-fleshed Sudanese Kordophan melon is cultivated. In this study, we infer biogeographical and haplotype genealogy for C. lanatus, C. mucosospermus, C. amarus, and C. colocynthis using non-coding cpDNA sequences (trnT-trnL and ndhF-rpl32 regions) from a global collection of 135 accessions. In total, we identified 38 haplotypes in C. lanatus, C. mucosospermus, C. amarus, and C. colocynthis; of these, 21 were found in Africa and 17 appear endemic to the continent. The least diverse species was C. mucosospermus (5 haplotypes) and the most diverse was C. colocynthis (16 haplotypes). Some haplotypes of C. mucosospermus were nearly exclusive to West-Africa, and C. lanatus and C. mucosospermus shared haplotypes that were distinct from those of both C. amarus and C. colocynthis. The results support previous findings C. mucosospermus to be the closest relative to C. lanatus (including subsp. cordophanus). West Africa, as a center of endemism of C. mucosospermus, is an area of interest in the search of the origin of C. lanatus. This calls for further historical and phylogeographical investigations and wider collection of samples in West and North-East Africa.

Introduction

Watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) is a horticultural species of high economic importance, accounting for nearly 103,9 million metric tons of global fruit production in 2018 from 3.2 million ha (Faostat, 2017). Over the last two decades, questions regarding the origin and taxonomy of*Citrullus* spp. have fuelled numerous studies to clarify phylogenetic relationships and nomenclature, identify wild relatives, and determine both centers of origin and divergence times (Jarret et al., 1997; Jarret & Newman, 2000; Levi et al., 2001; Dane et al., 2004; Levi et al., 2004; Levi & Thomas, 2005; Dane & Liu, 2007; Dane et al., 2007; Solmaz & Sari, 2009; Dje et al., 2010; Solmaz et al., 2010; Nesom, 2011; Levi et al., 2013; Mujaju et al., 2013; Hammer & Gladis, 2014; Chomicki & Renner, 2015; Renner et al., 2019; Chomicki et al., 2020). Despite these efforts, uncertainty vis-à-vis these questions remains as no wild relatives were found neither in west nor in northern east Africa; and comparatively few studies have focused on the distribution of the genetic variation within *Citrullus* or the likely colonization routes of various species within the genus.

The challenge of tracing the historical colonization routes of watermelon was for many years confounded by significant taxonomic confusion among species, subspecies, and varieties, all of which exhibit high morphological diversity. *Citrullus* Schrad. ex Eckl & Zeyh. is one of 95 genera of Cucurbitaceae (Jeffrey, 2005; Kocyan et al., 2007; Schaefer & Renner, 2011b; Schaefer & Renner, 2011a). To date there seems to be a consensus regarding its complex taxonomy. According to recent research, including phylogenetic analyses and nomenclatural reviews (Renner et al., 2014; Chomicki et al., 2020) as well as a phenetic comparison within the genus (Achigan-Dako et al., 2015),*Citrullus* encompasses the following seven species: 1) the widely

cultivated *C. lanatus* , a juicy fruit found in tropical and subtropical climates including var. *cordophanus* (Ter-Avan.) Fursa; 2) the tsamma melon *C. amarus* Schrad syn. *C. caffer*Schrad. or *C. lanatus* var. *citroides* (Bailey) Mansf., which grows in southern Africa (Whitaker & Bemis, 1976); 3) the egusi melon *C. mucoso-spermus* Fursa, previously referred to as a subtaxon of *C. lanatus* by many authors but which was raised to specific rank many decades ago (Fursa, 1972; Fursa, 1981; Fursa, 1983); 4) the bitter apple*C. colocynthis* (L.) Schrad., a perennial species growing in sandy areas throughout northern Africa and Near-East ; 5) *C. ecirrhosus*Cogn., another perennial wild species (De-Winter, 1990); 6) *C. rehmii* , a wild annual species, with small fruits used for feeding desert animals; and 7) *C. naudinianus* (Sond.) Hook.f. from the Namib-Kalahari region, previously placed in the genus *Acanthosicyos* Welw. ex Hook. f. and sister group to all other species. *Citrullus eccirhosus* , *C* . *rehmii,* and *C. naudinianus,*currently, are considered endemic and restricted to the desert region of Namibia with very little intraspecific variation (Dane & Lang, 2004); this understanding may change with more extensive sampling.

Given recent clarification of *Citrullus* taxonomy, it is appropriate to revisit the question of genealogy. In a recent phylogenetic study, Chomicki and Renner (2015) indicated west Africa as the possible center of origin of*C. lanatus* , a claim at odds with earlier assertions. Indeed, whereas some experts believe watermelon originated from southern Africa, based on the distribution of wild relatives in the Namibian desert (Bates & Robinson, 1995), others point to northern or north-east Africa, especially the Nile river area in Sudan, as the likely center of origin based on archaeological data (Wasylikowa & Van Der Veen, 2004; Paris, 2015; Renner et al., 2019). According to these latter studies, very few archaeological records of watermelon are known from southern Africa; and all date to a relatively recent period between the 8^{th} and 13^{th} centuries A.D. Furthermore, a cultigen is known to have been cultivated in the Nile Valley when farming was not yet practiced in southwest Africa (Zohary & Hopf, 2000). In contrast, archaeological records from West Africa are scanty, except for the presence of one endemic cultivated species (*C. mucosospermus*) previously deemed to be a subspecies or variety of *C. lanatus*(Nesom, 2011; Hammer & Gladis, 2014; Renner et al., 2014; Achigan-Dako et al., 2015).

The fundamental questions remain: how did watermelon spread throughout the world if it has originated from west or north-east Africa? How did the extant cultigens distribute throughout the world and how do they relate to wild types such as *C. colocynthis* or *C. amarus* ? To contribute to our understanding of these questions, this paper presents a chloroplast phylogeography of *Citrullus lanatus* and three related species, one cultivated (*C. mucosospermus*) and two wild (*C. amarus* and *C. colocynthis*), using a large sample size collected from four continents. The objective is to characterize the geographical distribution of *Citrullus* haplotypes and shed specific light of the chloroplast sequence evolution *of C. lanatus* , hypothesizing that such information will help clarify our understanding of the history of this globally significant agricultural species.

Materials and methods

Taxon sampling and total genomic DNA isolation

To investigate the geographical distribution of watermelon haplotypes, we included in the study the four most economically important*Citrullus* species: 1) *C. lanatus* , widely cultivated throughout the world (78 accessions from four continents out of which only 14 were from West Africa); 2) *C. mucosospermus* , restricted to West Africa and the closest sister species of cultivated watermelon (13 accessions)*;* 3) *C. amarus* , a wild species from Southern Africa that has spread to Europe and the closest relative to*C. ecirrhosus* (22 accessions); and *C. colocynthis* , a wild species found in northern Africa and East-Asia (22 accessions). In total, 135 accessions were assessed, including 53 from Africa, 41 from Asia, 25 from Europe, and 16 from North America (Table 1). Voucher specimens of all accessions were deposited in the herbarium of the Institute of Plant Genetics (Achigan-Dako et al., 2015) (IPK-Gatersleben).

As indicated in Table 1, a total of 53 accessions were received from the USDA National Plant Germplasm System, 66 were received from IPK-Gatersleben, and 16 were collected throughout West Africa as part of this study. Seeds of all accessions were germinated in a greenhouse at IPK-Gatersleben, and approximately

100 mg of leaf tissue was collected from one seedling per accession and dried with silica gel. Total genomic DNA was extracted from the dried leaf tissues using the QIAGEN DNAeasy Plant Kit, and one washing step was added according to the manufacturer's instructions to increase the quality of the DNA. Concentrations were estimated on 1% agarose gels stained with ethidium bromide. Samples exhibiting sub-optimal PCR amplification were purified via the QIAquick PCR Purification Kit (QIAGEN) and resuspended in 50 ml 1x TE buffer.

Choice of chloroplast regions

Based on the work of Shaw et al. (2007), the following nine non-coding chloroplast regions were chosen for initial screening of one accession each of *C. lanatus* , C. *mucosospermus, C. amarus,* and *C. colocynthis* :*rpl* 32-*trn* L, *trn* Q-5'*rps* 16, 3'*trn* V-*ndh* C, *ndh* F-*rpl* 32,*psb* D-*trn* T, *psb* J-*pet* A, 3'*rps* 16-5'*trn* K, *atp* I-*atp* H, and*trn* T-*trn* L. For most of these regions, total levels of variation were low and exclusively inter-specific. However, for*ndh* F-*rpl* 32 and *trn* T-*trn* L, polymorphisms were observed both within and among species; thus these two regions were selected for more in-depth investigation. These two regions of the chloroplast genome were amplified using the following primer pairs: 1)*ndh* F (5'-GAAAGGTATKATCAAYGMATATT-3') and *rpl* 32-R (5'-CCAATATCCCTTYYTTTTCCAA-3'); and 2)*trn* $L^{(UAG)}$ (5'-CTGCTTCCTAAGAGCAGCCT-3') and*trn* $T^{(GGU)}$ (5'-CCCTTTTAACTCAGTGGTAG-3').

Amplification and sequencing

PCR amplifications were performed using a Gene Amp 9700 PCR System (PE Biosystems) thermal cycler. For the *trn* T-*trn* L region, we used a reaction volume of 50 µl consisting of 26.6 µl H_2O, 5 µl of supply buffer (10x), an additional 2.5 µl of 15 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate, 10 µl Q-solution (Qiagen), 1.5 U Taq DNA polymerase (QIAGEN, Hilden, Germany), 50 pmol of each primer, and approximately 20 ng of genomic DNA. Cycling conditions for *trn* T-*trn* L region: 95°C for 3 mins; 10 cycles of 30 s at 95°C, 35 s at 56°C, and 90 s at 68°C; 35 cycles of 30 s at 95°C, 35 s at 53°C, and 90 s at 68°C; and a final extension of 10 min at 68°C. For the*ndh* F-*rpl* 32 region, PCR amplification was carried out using the Phusion Hot Start Kit (Thermo Scientific) in a reaction volume of 30 µl consisting of 17.7 µl H_2O, 6 µl of supply buffer (10x), an additional 1.5 µl of 15 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate, 50 pmol of each primer, and approximately 20 ng of genomic DNA. Cycling conditions for*ndh* F-*rpl* 32 region: 98°C for 3 mins; 35 cycles of 30 s at 98°C, 35 s at 58°C, and 80 s at 72°C; and a final extension of 15 min at 72°C. All PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN), following manufacturer's instructions, and re-suspended in 28 ml warmed 1x TE buffer. Sequencing was performed on either a MegaBACE 1000 (Amersham Biosciences) or an ABI 3730 XL (Applied Biosciences) capillary sequencer.

Sequence analysis and haplotype coding

For each chloroplast region, the forward and reverse sequences were manually edited and combined into a single sequence using Geneious 5.5.6 (Kearse et al., 2012); and these merged reads were submitted to NCBI GenBank to make them publicly available. Following merging, three alignments were generated: 1) Species-pairwise alignments of *C. lanatus* accessions with those of *C. mucosospermus, C. amarus* , and *C. colocynthis* for the chloroplast region *trn* T-L; 2) the same species-pairwise alignments for the region *ndh* F-*rpl* 32; and 3) a combined alignment of all species, containing both *trn* T-L and *ndh* F-*rpl* 32 regions, yielding a matrix of 1,611 aligned nucleotides. In the combined alignment, for the purpose of constructing coherent and parsimonious haplotypes, repeats and indels were re-coded as single bp polymorphisms. In the *trn* T-L region: 1) a microsatellite ACATA at position 366 was coded as A (repeat presence) or a single gap "-" (absence); 2) A TATT indel at position 405 was coded as a T (presence) or a single gap (absence); and 3) Another TTTATA microsatellite at position 423 was coded as T (presence) or a single gap (absence). In the*ndh* F-*rpl* 32 region: 1) a poly AT, usually six to eight units (position 1149), was just replaced by a single gap for 6*(AT), A for 7*(AT), and T for 8*(AT); and 6) A TGATT microsatellite at position 1198 was coded as a T (presence) or a single gap (absence).

Data analysis

Analysis of genetic diversity

Statistical parameters including sequence diversity, nucleotide diversity (Nei & Tajima, 1983; Nei, 1987), A+T content, and substitution, inversion, and transversion rates (Rozas & Rozas, 1997; Librado & Rozas, 2009; Baier, 2011; Chiu et al., 2013) were computed using DnaSP software version 5.10.01 (Librado & Rozas, 2009; Chiu et al., 2013). Pairwise intra- and inter-specific sequence divergences for each chloroplast region were computed as the mean number of nucleotide differences per site, following the formula:

100 x (Tv + Ts + ID)/L

where Tv is the number of transversions, Ts is the number of transitions, ID is the number of insertions/deletions, and L is the total length of the sequence (O'donnell, 1992; Dane et al., 2007). We used the PERMUT software package (Pons & Petit, 1996) to calculate the mean within-population gene diversity (Ching-Yi et al.) and the total gene diversity (h_T) (Martin et al., 2003; Guicking et al., 2011; Chiu et al., 2013; Sun et al., 2019; Zhao et al., 2019). Other intra-population metrics such as the number of haplotypes per population, the number of singleton haplotypes (haplotype that occurs only once in the study), the number of effective haplotypes, and the overall haplotype diversity were also estimated (Baier, 2011).

Population differentiation and genetic structure

To infer genetic differentiation parameters, haplotypes grouped by continent or sub-region were considered to comprise distinct geographic populations. We assessed the genetic differentiation among geographic populations by computing the gene differentiation statistic developed by Nei and Chesser (1983), an allele frequency-based approach that relies on estimates of genetic differentiation among geographic sub-populations. We further used Hudson et al. (1992)'s statistical test, a simple non-parametric method based on Monte Carlos permutations. That method, compared to the traditional Chi-square analysis of genetic differentiation estimates, helped understand whether the geographical populations are genetically different from one another. In addition, genetic differentiation among populations was estimated by computing a distance matrix based on the number of mutational steps between haplotypes (Nst) and by using haplotype frequencies (Gst). Phylogeographical structure was tested based on the difference between G_{ST} and N_{ST} using PERMUT 2.0 (Pons & Petit, 1996; Chiu et al., 2013) with 1000 permutations. In contrast to Gst, Nst considers sequence differences between the haplotypes. Thus, Nst > Gst indicates that closely related haplotypes are observed more often in a given geographical area than would be expected by chance (Pons & Petit, 1996; Burban et al., 1999; Grivet, 2002; Guicking et al., 2011; Chiu et al., 2013; Chavez-Pesqueira & Nunez-Farfan, 2016; Sun et al., 2019). Following Templeton (1996), we tested the null hypothesis of homogeneity of nucleotide mutations using Fisher's exact test to investigate haplotypic differentiation within the overall population. We also performed Fu's *Fs*(Fu, 1997) to analyze the expansion level of the population under the hypothesis of selective neutrality and population equilibrium. Tajima's D test also was implemented for comparison with the Fu's *Fs* .

Statistical parsimony network

Parsimony networks were constructed to infer phylogeographical relationships among haplotypes using TCS v1.21 (Clement et al., 2000). TCS estimates genealogical relationships of sequences and collapses identical sequences into haplotypes (HT). To account for the different mutation rates underlying base substitutions, indels, and microsatellites, we followed the two-step strategy described by Banfer et al. (2006) and performed by Guicking et al. (2011). The network was re-drawn from the TCS output using Adobe Illustrator.

Results

Nucleotide variations, intra- and interspecific diversity

The length of the amplified *trn* T-*trn* L region within*C. lanatus* ranged from 951-954 bp. No parsimony-informative site was found within *C. lanatus* , but 3 indels were found at positions 242, 295, and 296. The amplified *ndh* F-*rpl* 32 region ranged from 650-652 bp in the species, also with no parsimony-informative

site, though 5 indels were found at positions 970, 1028, 1143, 1178, and 1198 (Tables S1). The combined length of the two cpDNA regions was found equal to 1601-1605 bp and included 1 SNP (position 1399) and 1 microsatellite (position 366); but no polymorphisms were parsimony-informative. In total, the sampled accessions of this species comprise 12 distinct haplotypes, among which 10 were singletons, with an overall haplotype diversity of 0.5656 (Table 2).

Sequence lengths within *C. mucosospermus* were similar, with the combined length of the two regions spanning by 1601-1604 bp. One SNP (non-parsimony informative) was identified in the*ndh* F-*rpl* 32 region (position 1397), as well as two indels in *trn* T-*trn* L region (positions 242 and 296). Of the 5 haplotypes found among the sampled accessions of this species, three were singletons; and overall haplotype diversity is 0.5333.

The combined sequence length in *C. amarus* ranged between 1602-1604 bp (950-953 bp in *trn* T-*trn* L and 651-653 bp in*ndh* F-*rpl* 32) and contained ten polymorphic sites. Of those, 4 indels were observed in *trn* T-L (positions 295, 296, 297, 405) and 1 in *ndh* F-*rpl* 32 (positions 1198). Four SNPs were found at positions 918, 1149, 1397, and 1526; and there is a microsatellite at position 1149. *C. amarus* was characterized by eight haplotypes, among which six were private; and overall haplotype diversity is 0.81.

C. colocynthis was characterized by a combined sequence length of 1599-1605 bp (948-954 bp for *trn* T-*trn* L and 650-653 bp for*ndh* F-*rpl* 32) that features 10 SNPs (positions 406, 455, 487, 882, 918, 949, 1111, 1286, 1397, and 1526) and 3 microsatellites (positions 366, 423, 1149). In addition, there were 11 indels (positions 199, 242, 295, 296, 297, 972, 1179, 1180, 1200, 1262, and 1530), 7 of which were parsimony informative (6 within *trn* T-*trn* L and 1 within *ndh* F-*rpl* 32). The collection of this species contains 16 haplotypes, all private, and has an overall haplotype diversity of 0.96.

Based on the 29 polymorphic sites detected within the two cpDNA regions, 38 haplotypes were detected among the sampled accessions (Table 3). The most ancient haplotype (H1), according to TCS analysis, is exclusive to the cultivated species *C. lanatus* and *C. mucosospermus* . Of the 26 singleton haplotypes detected, 13 (50%) were found within*C. colocynthis* , indicating recent haplotype divergence in that species (Fig. 1).

Geographical distribution, genetic differentiation of haplotypes and population expansion

The pattern of polymorphism suggested non neutral selection as revealed by both Fu's *Fs* statistic and Tajima's *D* (Fs = -3.624, p = 0.016; D: -0.59858; not statistical significant, p > 0.10). Moreover, Ficher's exact test used to investigate haplotypic differentiation within the overall population suggested the rejection of the null hypothesis of homogeneity of nucleotide substitutions (LD = 0.1958, p < 0.001) following the neutral theory of molecular evolution.

Within-continent gene diversity (Hs) varied from 0.57 (in Europe) to 0.85 (in Africa), with the majority of haplotypes being specific to certain regions. For instance, of the 21 haplotypes found in Africa, 16 were specific to the continent; of the 14 haplotypes found in Asia, eight were specific; of the nine found in Europe, six were specific; and of the four recovered from America, two were specific to that region (see Figs. 2 - 5).

Haplotypes of *C. mucosospermus* were almost uniquely restricted to West Africa, and *C. amarus* haplotypes appeared specific to southern Africa. Haplotypes of *C. colocynthis* shared by Namibia, Ethiopia, and northern Africa were also found widespread throughout Asia. Across that continent, some haplotypes of *C. colocynthis*were specific to different countries (Fig. 1). Six *C. colocynthis*haplotypes were specific to Asia, and six were specific to Africa. For this species, Iran contributed the highest number of haplotypes in Asia (Fig. 1), as Egypt did in Africa (Fig. 1).

Within *C. lanatus* , although all regions shared most haplotypes, Africa exhibited the highest number of singletons. The ancient haplotype H1 was found not only among West African countries but also in Europe (Georgia, Yugoslavia, Italy and Ukraine), Asia (Russia, Japan, China, India), and North America (USA and Canada). North Africa (Egypt) and southern Asia (India) shared *C. colocynthis* haplotype H12; and haplotype H4, specific to *C. amarus* , was shared by African countries (e.g. South-Africa and the Democratic

Republic of Congo) and Russia (Fig. 1). Haplotype H2 was found throughout West Africa (Benin, Burkina-Faso, and Ghana) as well as in Asia (China, Japan, Yemen, North-Korean Republic, Mongolia, and Armenia), France, and North America (USA and Canada). Haplotype H2 is shared by *C. lanatus* and*C. amarus* ; and haplotype H6 is shared by *C. mucosospermus*and *C. amarus* species (see Figs. 2 - 5).

Analysis of interspecific genetic differentiation revealed a high level of total genetic differentiation among continents (Tables 4 and 5). Coefficients of pairwise genetic differentiation values were highest between Africa and Europe, on the one hand, and Asia and Europe, on the other; Gst was lower between Africa and Asia (0.006). The coefficient of population differentiation Gst was 0.196, and the pairwise difference between haplotypes Nst = 0.374.

Discussion

Genetic diversity and sequence variation

Within the genus *Citrullus* genetic diversity analyses have been investigated since the second middle of the 20^{th}century (Hashizume et al. 1996) revealing various trends. Previous knowledge revealed lower genetic diversity in*Citrullus* for breeding purpose (Levi et al., 2001; Levi et al., 2004). Recent studies shed light on obvious genetic diversity within the genus. For instance, a study using High Frequency Oligonucleotide Target Active Genes (HFO-TAGs) revealed high genetic diversity among *Citrullus* spp. and highlighted the potential importance of PI accessions as sources of valuable traits like disease resistance (Levi et al., 2013).

Our findings revealed low cpDNA variability among *C. lanatus* and*C. mucosospermus* . This was also observed by Dane and Lang (2004) and Dane et al. (2004) who revealed low nucleotide variability based on a low number of parsimony-informative sites within each of the studied species. Most haplotypes were found within non-cultivated (*C. colocynthis*) rather than cultivated (*C. lanatus* and *C. mucosospermus*) species. Taxa were highly separated from one another with divergence based mainly on indels and transition events (Dane et al., 2004). However, there was sufficient resolution of the *trn* T-L and *ndh* F-*rpl* 32 non-coding regions to reveal intraspecific variability.

Chloroplast sequence analysis revealed that the *ndh* F-*rpl* 32 region exhibits comparatively higher variability within the two cultivated species than the *trn* T-L region. Dane and Lang (2004) analyzed four chloroplast regions (*nhd* F, *ycf* 6-*psb* M,*ycf* 9-*trn* G, and *atp* A-*trn* R) and found no variability within cultivated accessions, grouped either by morphological traits or geographical origin. In this study, we used a large number of *C. lanatus* accessions from a wide geographical range and observed low haplotype diversity within that species, as also revealed by Guo et al. (2013). While many factors can influence sequence diversity, selection is a major contributor via the imposition of bottlenecks that can substantially reduce diversity (Dane & Lang, 2004; Levi et al., 2013). The lack of haplotype divergence within *C. lanatus* and *C. mucosospermus* is likely the result of selection or other bottlenecks in the domestication histories of watermelon and egusi melon. Certainly, selection for sweet red-fleshed cultivars with high lycopene content or selection of seed type as source of protein/oil for consumption might contribute to current genetic structure in those cultivated species (Achigan-Dako et al., 2015; Renner et al., 2019).

C. colocynthis exhibited a relatively high number of parsimony-informative characters. Dane et al. (2004) revealed that haplotypes detected within *C. colocynthis* were associated with geographical origin and that was also confirmed by Levi et al. (2017). The haplotype diversity within *C. colocynthis* suggests cryptic evolution and calls for a comprehensive morphological comparison of Asian and African colocynths. Such an investigation is exemplified by the recent studies on *Cucumis melo* that revealed modern melon cultivars go back to two lineages and was domesticated at least twice: in Asia and in Africa (Endl et al., 2018).

Citrullus *haplotype evolution*

Thirty-eight haplotypes were detected among the cultivated and wild*Citrullus* accessions used in this study. Dane et al. (2004), found seven haplotypes within the genus, using 55 accessions of *C. lanatus*,15 accessions of *C. colocynthis* , and a total of seven cpDNA regions (HinfI, RsaI, TaqI, AluI, HaeIII, MboI, and BglII). With two cpDNA regions and 135 accessions carefully selected to represent a wide geographical region, we detected

an even higher haplotype diversity among *Citrullus* spp. This situation can be expected to continue to evolve as more watermelon accessions from Sudan or northeast Africa are sequenced, particularly, the Sudanese sweet white-fleshed melon. Unfortunately, sampling of *C. lanatus* from the Darfur region of Sudan has been scarce (Renner et al., 2019).

On average, we observed 9.5 haplotypes per species, varying from 5 to 16. In comparison with other species, Guicking et al. (2011) found 9.8 haplotypes per species in *Macaranga* and Jakob and Blattner (2006) found 2.83 haplotypes per species in *Hordeum* . In *Citrullus* spp., nucleotide substitutions appear to have evolved at different rates, an observation supported by the Fisher's test for homogeneity of nucleotide substitution. Fu's test *Fs* also rejected the null hypothesis of neutrality of evolution of nucleotide substitution, further supporting the hypothesis that the polymorphism pattern observed is non-random. Population expansions tend to produce significantly negative values of D, while population bottlenecks tend to produce significantly positive values of D. In our case the departure from neutrality might indicate that there is a high demographic expansion and a pattern of isolation by distance would be occurred between the continents (Jiang et al., 2016).

Genetic differentiation and geographical structure

The coefficient of population differentiation with no account the distances among haplotypes (Gst) and the coefficient of differentiation based on the pairwise difference between alleles that takes into account the distances among haplotypes (Nst) were found respectively, equal to 0.196 and 0.374; but the difference was not significant ($P > 0.05$). In *Citrullus* spp. Mujaju et al. (2011) found Gst = 0.56 and Nst = 0.49 for sweet watermelon and Gst = 0.71, Nst = 0.81 for cow watermelon. The fact that the differentiation parameter based on the pairwise difference between alleles is greater than the one calculated without permutation (i.e. Nst > Gst) indicates that the collection is characterized by clear geographic structure (Grivet, 2002; Dane et al., 2007; Guicking et al., 2011). Also the significant value of the total gene diversity across all four geographical regions (hT = 0.917, standard error = 0.0320) is indicating a strong structure in the population (Pons & Petit, 1996; Sun et al., 2019; Zhao et al., 2019).

Levi et al. (2017) observed that accessions of *C. colocynthis* were sub-divided into five groups in general agreement with their centres of diversification and origin. Our findings indicated that regional genetic differentiation statistics support Levi et al. (2017)'s conclusions, with sub-samples from different regions exhibiting genetic differentiation associated with their likely centers of diversification. Also, haplotypes of *C. amarus* were mostly grouped in Southern Africa, which is assumed to be the origin of that species (Dane & Liu, 2007; Chomicki & Renner, 2015).

Citrullus chloroplast sequences analysis with TCS 1.21 resulted in a network where haplotypes widely sampled throughout West Africa were placed at the root. While coalescence theory predicts that older alleles will prevail in a population due to a higher number of descending lineages and associated wider geographic distributions (Crandall & Templeton, 1993), such an observation may depend on sample sizes and evolutionary/domestication histories and also the lack of subs. *cordophanus* (from northeast Africa) in the germplasm studied. In this study, H1 is the most frequently sampled haplotype and has the most connections with other haplotypes; thus H1 may be considered the most ancient haplotype. This ancient haplotype was sampled most frequently in West Africa (i.e. Nigeria and Benin) and was highly shared by accessions of both *C. lanatus* and *C. mucosospermus* . These results support the findings of Chomicki and Renner (2015) and Renner et al. (2019) who used eleven gene regions to infer phylogeny among *Citrullus* species, and also a 3500-year-old leaf sample from the Egyptian tomb to infer close relationship between *C. lanatus* and *C. mucosospermus* . Our findings, based upon a large set of egusi melon and watermelon accessions from four continents, provide further evidence of that close relationship between these two species. However, they are indeed two different species, as previous crosses between them (e.g. Charleston Gray x PI 560006) resulted in high levels of sterility (Gusmini et al., 2004). The very limited haplotype diversity among the two species suggests an old split, with chlorotype fixation (Dane & Liu, 2007) and ancient types of *C . mucosospermus* originating from Western Africa (Renner et al., 2014). However, to the best of our knowledge, no wild populations have been confirmed in West Africa. Spontaneous plants may have been found earlier, but those

individuals certainly escaped from cultivation. A region-wide collecting mission by the first author yielded no wild population of *C. mucosospermus* in West Africa (Achigan-Dako et al., 2015) though, the presence in West Africa of the 'neri' type [Fig. 9f in Achigan-Dako et al. (2015) and Fig. 1 in Minsart et al. (2011)], another cultivated egusi melon that exhibits smaller seeds with yellow soft coat, should be highlighted as a contributor to the genepool of*Citrullus* is the region. While this neri type (*C. lanatus*) is morphologically distinct from *C. mucosospermus* , it has been rarely studied.

Archaeological evidence indicates north-east Africa as a center of origin and domestication (Chomicki et al., 2020). Authors reported wild dessert watermelon in that region (Paris, 2015) or the genetic affinity with the *C. lanatus* var. *cordophanus* (a sweet white-fleshed cultivar) (Renner et al., 2019). However, within the genus *Citrullus mucosospermus* remains the closest relative species to *C. lanatus* . The presence of an ancient haplotype in West Africa on the one hand and the close relationship between *C. lanatus* and subsp. *cordophanus* of Darfur in north-eastern Africa as revealed by Renner et al. (2019) on the second hand, calls for further molecular and archaeological investigations to generate sufficient knowledge on newly published results, including those reported here. New molecular investigations should include more materials from Sudan and neighbouring countries where wild populations of watermelon have been found (Paris, 2015). Moreover, our data showed that one of the Egyptian accessions (PI 525083), indicated to be*C. amarus* and observed by Levi et al. (2013) to cluster with dessert watermelon, exhibits a unique haplotype (H32). That accession is several mutations away from *C. colocynthis* and closer to watermelon and egusi melon haplotype. Previous findings of Levi et al. (2017) showed that PI 525083 rather clustered with *C. lanatus* var. *lanatus* . In addition, the hypothesis that watermelon is from north-eastern Africa does not explain how an endemic species such as*C. mucosospermus* shares the same haplotype with dessert watermelon, while other accessions from the region (e.g. PI 525083) shows unique haplotype. If *C. lanatus* did indeed spread to the world from west or north-eastern Africa, how and when was it domesticated in those region as New Kingdom Egyptians were cultivating sweet red-fleshed watermelon more than 3500 years ago? From which species was *C. mucosospermus* domesticated? Through what mechanisms was *C. lanatus* spread to Asia, and when? More germplasm collections from all continents are necessary to fully understand the phylogeographical relationships among *Citrullus*species. In Africa the focus should be on both west and north-eastern regions to resolve the domestication history of modern cultivars.

Conclusion

The genus *Citrullus* includes seven species that may originate from different parts of the world, according to previous and current data. Our results reveal 38 distinct chloroplast haplotypes among*Citrullus* spp. and the distribution of those haplotypes across the world. The close relationship of egusi melon and Kordofan melon to watermelon raised new questions regarding the colonization routes of major crops and the current status of extant genetic diversity of wild relatives in places of origin.

Acknowledgement

This study was financially supported by the Vavilov-Frankel Fellowship to the first author under grant CONT/08/136/RF and the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). We acknowledge contribution and technical support from Christina Koch, Petra Oswald, Birgit Wohlbier, R.S. Vodouhe, and Adam Ahanchede. We thank Augustin Ganse for assistant in map making and Hanno Schaefer for useful comments on the manuscript.

References

Achigan-Dako, E. G., Avohou, E. S., Linsoussi, C., Ahanchede, A., Vodouhe, R. S., & Blattner, F. R. (2015). Phenetic characterization of Citrullus spp.(Cucurbitaceae) and differentiation of egusi-type (C. mucosospermus). *Genetic Resources and Crop Evolution, 62* (8), 1159-1179.

Baier, C. (2011). *Comparative phylogeographic and population genetic analyses of three tropical pioneer trees, Macaranga winkleri, M. winkleriella and M. tanarius (Euphorbiaceae).* Martin-Luther-Universitat Halle-Wittenberg,

Banfer, G., Moog, U., Fiala, B., Mohamed, M., Weising, K., & Blattner, F. R. (2006). A chloroplast genealogy of myrmecophytic *Macaranga* species (Euphorbiaceae) in Southeast Asia reveals hybridization, vicariance and long-distance dispersals *Molecular Ecology, 15* , 4409-4424.

Bates, D. M., & Robinson, R. (1995). Cucumber, melon and watermelons: *Cucumis* and *Citrullus*(Cucurbitaceae), P. 89-96. In J. Smart & N. M. Simmonds (Eds.). Evolution of crop plants 2nd ed. Longman, London

Burban, C., Petit R. J, Carcreff E., & H., J. (1999). Range wide variation of the maritime pine bast scale Matsucoccus feytaudi Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular ecology, 8* , 1593-1602.

Chavez-Pesqueira, M., & Nunez-Farfan, J. (2016). Genetic diversity and structure of wild populations of Carica papaya in Northern Mesoamerica inferred by nuclear microsatellites and chloroplast markers. *Annals of botany, 118* (7), 1293-1306.

Ching-Yi, L., Hsin-Mei Ku, Yi-Hua Chiang, Hsiu-Yin Ho, Tsong-Ann Yu, & Jan, F.-J. (2012). Development of transgenic watermelon resistant to Cucumber mosaic virus and Watermelon mosaic virus by using a single chimeric transgene construct. *Transgenic Research 21* (5), 983-993.

Chiu, Y.-W., Bor, H., Tan, M.-S., Lin, H.-D., & Jean, C.-T. (2013). Phylogeography and genetic differentiation among populations of the moon turban snail Lunella granulata Gmelin, 1791 (Gastropoda: Turbinidae). *International journal of molecular sciences, 14* (5), 9062-9079.

Chomicki, G., & Renner, S. S. (2015). Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics. *New Phytologist, 205*526-532. doi:10.1111/nph.1363

Chomicki, G., Schaefer, H., & Renner, S. S. (2020). Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New Phytologist, 226* (5), 1240-1255.

Clement, M., Posadu, D., & Crandall, K. (2000). a computer program to estimate gene genealogies. *Molecular Ecology, 9* , 1657-1660.

Crandall, K. A., & Templeton, A. R. (1993). Empirical tests of some predictions from coalescent theory with aplications to intraspecific phylogeny reconstruction. . *Genetics, 143* , 959-969.

Dane, F., & Lang, P. (2004). Sequence variation at cpDNA regions of watermelon and related wild species: implications for the evolution of *Citrullus* haplotypes. *American Journal of Botany, 91* (11), 1922-1929.

Dane, F., Lang, P., & Bakhtiyarova, R. (2004). Comparative analysis of chloroplast DNA variability in wild and cultivated Citrullus species. *Theoretical and Applied Genetics, 108* (5), 958-966.

Dane, F., & Liu, J. (2007). Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genetic Resources and Crop Evolution, 54* (6), 1255-1265.

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

De-Winter, B. (1990). A new species of *Citrullus* (Benincaseae) from the Namib desert, Namibia. *Bothalia, 3* (20), 209-211.

Dje, Y., Tahi, C. G., Bi, A. I., Baudoin, J.-P., & Bertin, P. (2010). Use of ISSR markers to assess genetic diversity of African edible seeded< i> Citrullus lanatus</i> landraces. *Scientia Horticulturae, 124* (2), 159-164.

Endl, J., Achigan-Dako, E. G., Pandey, A. K., Monforte, A. J., Pico, B., & Schaefer, H. (2018). Repeated domestication of melon (Cucumis melo) in Africa and Asia and a new close relative from India. *American Journal of Botany, 105* (10), 1662-1671.

FAOSTAT. (2017). Watermelon production in 2014. from Crops/Regions (World list)/Production quantity (from pick lists)

Fu, Y. X. (1997). Statistical neutrality of mutations against population growth, hitchhiking and background selection *Genetics, 147* , 915–925.

Fursa, T. B. (1972). K sistematike roda *Citrullus* Schrad. [On the taxonomy of genus*Citrullus* Schrad.]. *Botanicheskii Zhurnal, 57* , 31–41.

Fursa, T. B. (1981). Intraspecific classification of water-melon under cultivation. *Die Kulturpflanze, 29* (1), 297-300.

Fursa, T. B. (1983). Novyi vid arbuza*Citrullus mucosospermus* (Fursa) Fursa. (A new species of watermelon *Citrullus mucosospermus* (Fursa) Fursa.) *Trudy po prikladnoi botanike genetike i selektsii1* (81), 108-112.

Grivet, D. (2002).*Phylogeographie et evolution moleculaire comparee d'arbres forestiers a l'aide des marqueurs chloroplastiques.* (PhD), Universite Henri Poincare, Nancy-I,

Guicking, D., Brigitte Fiala, Franck R. Blattner, Ferry Slik, Maryati Mohamed, & Weising, K. (2011). Comparative chloroplast DNA phylogeography of two tropical pioneer tree,*Macaranga gigantea* and *Macaranga pearsonii*(Euphorbiaceae). *Tree Genetics & Genomes 7* , 573-585.

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z., Min, J., Guo, X., Murat, F., Ham, B.-K., Zhang, Z., Gao, S., Huang, M., Xu, Y., Zhong, S., Bombarely, A., Mueller, L. A., Zhao, H., He, H., Zhang, Y., Zhang, Z., Huang, S., Tan, T., Pang, E., Lin, K., Hu, Q., Kuang, H., Ni, P., Wang, B., Liu, J., Kou, Q., Hou, W., Zou, X., Jiang, J., Gong, G., Klee, K., Schoof, H., Huang, Y., Hu, X., Dong, S., Liang, D., Wang, J., Wu, K., Xia, Y., Zhao, X., Zheng, Z., Xing, M., Liang, X., Huang, B., Lv, T., Wang, J., Yin, Y., Yi, H., Li, R., Wu, M., Levi, A., Zhang, X., Giovannoni, J. J., Wang, J., Li, Y., Fei, Z., & Xu, Y. (2013). The draft genome of Watermelon (*Citrullus lanatus)* and resequencing of 20 diverse accessions *NATURE GENETICS, 45* , 51-58.

Gusmini, G., Whener, C. T., & Jarret, L. R. (2004). Inheritance of egusi seed type in watermelon*Journal of Heredity, 95* , 268-270.

Hammer, K., & Gladis, T. (2014). Notes on infraspecific nomenclature and classifications of cultivated plants in Compositae, Cruciferae, Graminae (with a remark on*Triticum dicoccon* Schrank) and Leguminosae *Genetic Resources and Crop Evolution, 61* (8), 1455-1467.

Hudson, R. R., Boos, D. D., & Kaplan, N. L. (1992). A Statistical Test for Detecting Geographic Subdivision. *Molecular Biology and Evolution, 9* (1), 138-151.

Jakob, S. S., & Blattner, F. R. (2006). A chloroplast Genealogy of Hordeum (Poaceae): Long-Term Persisting Haplotypes, Incomplete Lineage Sorting, Regional Extinction, and the Consequence for Phylogenetic Inference. *Mol. Biol. Evol., 23* (8), 1602-1612.

Jarret, R. L., Merrick, L. C., Holms, T., Evans, J., & Aradhya, M. K. (1997). Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. &Nakai).*Genome, 40* , 433-441.

Jarret, R. L., & 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. *Genetic Resources and Crop Evolution, 47* (2), 215-222.

Jeffrey, C. (2005). A new system of Cucurbitaceae. *Bot Zhur, 90* , 332-335.

Jiang, N., Man, L., Zhang, W., Dong, H.-X., Wang, H.-Y., Li, M.-R., Shi, F.-X., & Sun, M.-Z. (2016). Chloroplast view of the population genetics and phylogeography of a widely distributed shrub species, Rhododendron dauricum (Ericaceae).*Systematic Botany, 41* (3), 626-633.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28* (12), 1647-1649.

Kocyan, A., Zhang, L.-B., Schaefer, H., & Renner, S. S. (2007). A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. *Molecular Phylogenetics and Evolution, 44* (553-577).

Levi, A., Simmons, A. M., Massey, L., Coffey, J., Wechter, W. P., Jarret, R. L., Tadmor, Y., Nimmakayala, P., & Reddy, U. K. (2017). Genetic diversity in the desert watermelon Citrullus colocynthis and its relationship with Citrullus species as determined by high-frequency oligonucleotides-targeting active gene markers. *Journal of the American Society for Horticultural Science, 142* (1), 47-56.

Levi, A., Thies, J. A., Wechter, P. W., Harrison, H. F., Simmons, A. M., Reddy, U. K., Nimmakayala, P., & Fei, Z. (2013). High frequency oligonucleotides: targeting active gene (HFO-TAG) markers revealed wide genetic diversity among *Citrullus*Spp. accession useful for enhancing diseases or pest resistance in watermelon cultivars *Genetic Resources and Crop Evolution, 14* (60), 427-440.

Levi, A., & Thomas, C. E. (2005). Polymorphisms among chloroplast and mitochondrial genomes of*Citrullus* species and subspecies. *Genetic Resources and Crop Evolution, 52* (5), 609-617.

Levi, A., Thomas, C. E., Keinath, A. P., & Wehner, T. C. (2001). Genetic diversity among watermelon (Citrullus lanatus and Citrullus colocynthis) accessions. *Genetic Resources and Crop Evolution, 48* (6), 559-566.

Levi, A., Thomas, C. E., Newman, M., Reddy, O., Zhang, X., & Xu, Y. (2004). ISSR and AFLP markers differ among American watermelon cultivars with limited genetic diversity.*Journal of the American Society for Horticultural Science, 129* (4), 553-558.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive of analysis of DNA polymorphism data *Bioinformatics, 25* , 1451-1452.

Martin, J.-F., Claude Novoa, Stephanie B. Lanc-manel, & Taberlet, P. (2003). Les populations de perdrix grise des Pyrenees (*Perdix perdix hispaniensis*) ont-elles subi une introgression genetique a partir d'individus d'elevage ? Analyse du polymorphisme de l'ADN mitochondrial. *Les Actes du BRG, 4* , 115-126.

Minsart, L. A., Dje, Y., Baudoin, J.-P., Jacquemart, A.-L., & Bertin, P. (2011). Set up of simple sequence repeat markers and first investigation of the genetic diversity of West-African watermelon (*Citrullus lanatus* ssp. vulgaris oleaginous type). *Genetic Resources and Crop Evolution, 58* (6), 805-814.

Mujaju, C., Sehic, J., & Nybom, H. (2013). Assessement of EST-SSR Markers for Evaluating Genetic Diversity in Watermelon Accessions from Zimbabwe. *American Journal of Plant Sciences, 4* , 1448-1456.

Mujaju, C., Zborowska, A., Werlemark, G., Garkava-Gustavssson, L., Andersen, S. B., & Nybom, H. (2011). Genetic diversity among and within watermelon (*Citrullus lanatus*) landraces in Southern Africa. *Journal of Horticculture Sciences and Biotechnology, 86* , 353–358.

Nei, M. (1987). Molecular evolutionary genetics. *Columbia University Press*

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Ann. Hum. Genet., 47* , 253-259.

Nei, M., & Tajima, F. (1983). Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. *Genetics, 105* , 207-217.

Nesom, G. L. (2011). Toward consistency of taxonomic rank in wild/domesticated Cucurbitaceae.*Phytoneuron, 13* , 1-33.

O'Donnell, K. (1992). Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*).*Curr. Genet., 22* , 213-220.

Paris, H. S. (2015). Origin and emergence of the sweet dessert watermelon, *Citrullus lanatus* .*Annals of botany, 116* , 133-148.

Pons, O., & Petit, R. J. (1996). Measuring and testing genetic differentiation with ordered versus unordered allells. *Genetics, 144* , 1237-1245.

Renner, S. S., Chomicki, G., & Greuter, W. (2014). Proposal to conserve the name *Momordica lanata* (*Citrullus lanatus*) (watermelon, Cucurbitaceae), with a conserved type, against *Citrullus battich* . *Taxon, 63* (4), 941-942.

Renner, S. S., Perez-Escobar, O. A., Silber, M. V., Nesbitt, M., Preick, M., Hofreiter, M., & Chomicki, G. (2019). A 3500-year-old leaf from a Pharaonic tomb reveals that New Kingdom Egyptians were cultivating domesticated watermelon.*bioRxiv* , 642785.

Rozas, J., & Rozas, R. (1997). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics, 2* (15), 174-175.

Schaefer, H., & Renner, S. S. (2011a). Cucurbitaceae In K. Kubitzki (Ed.), *The families and genera of vascular plants, vol. 10, Sapindales, Cucurbitales, Myrtaceae* . 112-174 Berlin: Springer.

Schaefer, H., & Renner, S. S. (2011b). Phylogenetic relationships in the order Cucurbitales and a new classification of the gourd family (Cucurbitaceae). *TAXON, 60* (1), 122-138.

Shaw, J., Edgar B. Lickey, Edward E. Schilling, & Randall L. Small. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperm: the tortoise and hare III. *American Journal of Botany, 94* (3), 275-288.

Solmaz, I., & Sari, N. (2009). Characterization of watermelon (*Citrullus lanatus*) accessions collected from Turkey for morphological traits. *Genetic Resources and Crop Evolution, 56* (2), 173-188.

Solmaz, I., Sari, N., Aka-Kacar, Y., & Yalcin-Mendi, N. Y. (2010). The genetic characterization of Turkish watermelon (*Citrullus lanatus*) accessions using RAPD markers.*Genetic Resources and Crop Evolution, 57* (5), 763-771.

Sun, R., Lin, F., Huang, P., Ye, X., Lai, J., & Zheng, Y. (2019). Phylogeographical structure of Liquidambar formosana Hance revealed by chloroplast phylogeography and species distribution models. *Forests, 10* (10), 858.

Templeton, A. R. (1996). Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in hominoid primates. . *Genetics, 144* , 1263–1270.

Wasylikowa, K., & Van der Veen, M. (2004). An archaeobotanical contribution to the history of watermelon,*Citrullus lanatus* (Thunb.) Matsum. & Nakai (syn. *C. vulgaris* Schrad.). *Vegetation history and archaeobotany, 13* (4), 213-217.

Whitaker, T. W., & Bemis, B. W. (1976). Cucurbits *Simmonds N. W (e.d), Evolution of crop plants 6* , 64-69.

Zhao, Y., Pan, B., & Zhang, M. (2019). Phylogeography and conservation genetics of the endangered Tugarinovia mongolica (Asteraceae) from Inner Mongolia, Northwest China.*PloS one, 14* (2), e0211696.

Zohary, D., & Hopf, M. (2000). Domestication of plants in the old World *Oxford University Press, Oxford, UK.* .

Author contributions

E.G.AD and F.R.B. planned and designed the research. E.G.AD assembled plant materials. E.G.AD. performed experiments, conducted field and laboratory work. E.G.AD and H.D. analysed data. E.G.AD., H.D., and I.H. wrote the manuscript. F.R.B supervised the data collection and analysis. All authors read and approved the final manuscript.

Data accessibility

DNA sequences: NCBI Genbank accession numbers are provided in Table S1.

Figures captions

Fig. 1 . TCS network of 38 *Citrullus* spp. haplotypes. Circle size is proportional to haplotype frequency. Taxon names are abbreviated with two or three letters. Clv: *C. lanatus* subsp.*vulgaris* ; Cll: *C. lanatus* subsp. *lanatus* ; Cm:*C. mucosospermus* ; Cam: *C. amarus* ; and Cco: *C. colocynthis* . The numbers are arbitrary haplotype ID numbers (see Table S2), and the colors indicate geographical distribution: Africa (green), Asia (yellow); Europe (red), and North America (blue).

Fig. 2 . Distribution and frequencies of *Citrullus* spp. haplotypes in Africa.

Fig. 3 . Distribution and frequencies of *Citrullus* spp. haplotypes in Asia.

Fig. 4 : Distribution and frequencies of *Citrullus* spp. haplotypes in Europe.

Fig. 5 : Distribution and frequencies of *Citrullus* spp. haplotypes in North America.

Table 1: List of *Citrullus* accessions, their geographical origin, and accession numbers.

No	Taxon	Haplotype number	Accession number	Origin	So
1	*Citrullus lanatus* var. *lanatus*	9	PI 494527	Nigeria	US
2	*Citrullus mucosospermus*	1	PI 559993	Nigeria	US
3	*Citrullus mucosospermus*	26	PI 559994	Nigeria	US
4	*Citrullus mucosospermus*	9	PI 560000	Nigeria	US
5	*Citrullus lanatus* var. *lanatus*	17	PI 560002	Nigeria	US
6	*Citrullus mucosospermus*	1	PI 560008	Nigeria	US
7	*Citrullus mucosospermus*	1	PI 560010	Nigeria	US
8	*Citrullus mucosospermus*	1	PI 560013	Nigeria	US
9	*Citrullus mucosospermus*	1	PI 560018	Nigeria	US
10	*Citrullus lanatus* var. *lanatus*	1	PI 560024	Nigeria	US
11	*Citrullus mucosospermus*	1	849 BSN 001	Benin	Pr
12	*Citrullus mucosospermus*	1	975 MAT 007	Benin	Pr
13	*Citrullus mucosospermus*	1	977 MAT 008	Benin	Pr
14	*Citrullus mucosospermus*	1	1068 SN 045	Benin	Pr
15	*Citrullus lanatus* var. *lanatus*	19	GRIF 12336	China	US
16	*Citrullus lanatus* var. *lanatus*	1	GRIF 14199	India	US
17	*Citrullus lanatus* var. *lanatus*	1	GRIF 17300	China	US
18	*Citrullus lanatus* var. *lanatus*	2	GRIF 17310	China	US
19	*Citrullus lanatus* var. *lanatus*	1	GRIF 17330	China	US
20	*Citrullus mucosospermus*	6	PI 186975	Ghana	US
21	*Citrullus lanatus* var. *lanatus*	1	PI 192937	China	US
22	*Citrullus mucosospermus*	1	PI 249010	Nigeria	US
23	*Citrullus lanatus*	1	PI 271778	South Africa	US
24	*Citrullus lanatus* var. *lanatus*	10	GRIF 55960	India	US
25	*Citrullus lanatus* var. *lanatus*	1	GRIF 55990	India	US
26	*Citrullus amarus*	3	PI 596662	South Africa	US
27	*Citrullus amarus*	4	GRIF 15896	Russia	US

No	Taxon	Haplotype number	Accession number	Origin	So
28	*Citrullus amarus*	4	GRIF 15897	Russia	US
29	*Citrullus amarus*	6	PI 179881	India	US
30	*Citrullus amarus*	4	PI 189225	Democratic Republic of Congo	US
31	*Citrullus amarus*	3	PI 299378	South Africa	US
32	*Citrullus amarus*	4	PI 299379	South Africa	US
33	*Citrullus amarus*	3	PI 244018	South Africa	US
34	*Citrullus amarus*	3	PI 244019	South Africa	US
35	*Citrullus amarus*	4	PI 255137	South Africa	US
36	*Citrullus amarus*	4	PI 270563	South Africa	US
37	*Citrullus amarus*	6	PI 271779	South Africa	US
38	*Citrullus amarus*	32	PI 525083	Egypt	US
39	*Citrullus amarus*	8	PI 596659	South Africa	US
40	*Citrullus amarus*	8	PI 596669	South Africa	US
41	*Citrullus amarus*	14	PI 596671	South Africa	US
42	*Citrullus amarus*	3	PI 596676	South Africa	US
43	*Citrullus amarus*	15	CIT 101	Ukraine	IP
44	*Citrullus amarus*	4	CIT 139	Russia	IP
45	*Citrullus amarus*	3	CIT 152	Zimbabwe	IP
46	*Citrullus amarus*	3	CIT 310	South Africa	IP
47	*Citrullus amarus*	2	CIT 313	Yemen	IP
48	*Citrullus lanatus* subsp. *vulgaris*	2	CIT 207	France	IP
49	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 31	Ukraine	IP
50	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 44	Yugoslavia	IP
51	*Citrullus lanatus* subsp. *vulgaris*	18	CIT 60	Croatia	IP
52	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 67	Italy	IP
53	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 69	Italy	IP
54	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 86	Greece	IP
55	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 97	Hungary	IP
56	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 99	China	IP
57	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 102	USA	IP
68	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 103	Russia	IP
59	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 105	Ukraine	IP
60	*Citrullus lanatus* subsp. *Vulgaris*	1	CIT 107	Russia	IP
61	*Citrullus lanatus* subsp. *Vulgaris*	1	CIT 109	Russia	IP
62	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 112	Ukraine	IP
63	*Citrullus lanatus* subsp. *vulgaris*	2	CIT 126	Armenia	IP
64	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 128	Mongolia	IP
65	*Citrullus lanatus* subsp. *vulgaris*	18	CIT 130	Yugoslavia	IP
66	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 135	Bulgaria	IP
67	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 142	Bulgaria	IP
68	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 143	Bulgaria	IP
69	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 156	Georgia	IP
70	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 158	Georgia	IP
71	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 160	Georgia	IP
72	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 164	Russia	IP
73	*Citrullus lanatus* subsp. *vulgaris*	2	CIT 167	North Korea	IP
74	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 235	USA	IP
75	*Citrullus lanatus* subsp. *vulgaris*	2	CIT 237	Japan	IP
76	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 239	USA	IP
77	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 242	USA	IP

No	Taxon	Haplotype number	Accession number	Origin	So
78	*Citrullus lanatus* subsp. *vulgaris*	11	CIT 244	USA	IP
79	*Citrullus lanatus*	11	CIT 259	USA	IP
80	*Citrullus lanatus* subsp. *vulgaris*	22	CIT 253	Japan	IP
81	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 303	Turkey	IP
82	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 306	Portugal	IP
83	*Citrullus lanatus* subsp. *vulgaris*	1	06 NIA 224	Mali	Pr
84	*Citrullus lanatus* subsp. *vulgaris*	2	06 NIA 567	Benin	Pr
85	*Citrullus lanatus* subsp. *vulgaris*	2	07 NIA 995	Ghana	Pr
86	*Citrullus lanatus* subsp. *vulgaris*	1	846 BAX1	Mali	Pr
87	*Citrullus lanatus* subsp. *vulgaris*	1	1005 SE 032	Mali	Pr
88	*Citrullus lanatus* subsp. *vulgaris*	1	CIT 168	North Korea	IP
89	*Citrullus lanatus*	24	CIT 175	Italy	IP
90	*Citrullus lanatus*	2	CIT 182	Mongolia	IP
91	*Citrullus lanatus*	1	CIT 193	Ukraine	IP
92	*Citrullus lanatus*	1	CIT 195	Georgia	IP
93	*Citrullus lanatus*	1	CIT 200	Tajikistan	IP
94	*Citrullus lanatus*	1	CIT 203	Tunisia	IP
95	*Citrullus lanatus*	2	CIT 206	China	IP
96	*Citrullus lanatus*	1	CIT 226	USA	IP
97	*Citrullus lanatus*	1	CIT 230	Israel	IP
98	*Citrullus lanatus*	1	CIT 234	USA	IP
99	*Citrullus lanatus*	1	CIT 260	USA	IP
100	*Citrullus lanatus*	2	CIT 264	USA	IP
101	*Citrullus lanatus*	21	CIT 270	USA	IP
102	*Citrullus lanatus*	1	CIT 271	Canada	IP
103	*Citrullus lanatus*	1	CIT 273	USA	IP
104	*Citrullus lanatus*	1	CIT 278	USA	IP
105	*Citrulus lanatus* subsp. *lanatus*	16	CIT 309	South Africa	IP
106	*Citrullus colocynthis*	36	CIT 150	Canary Island	IP
107	*Citrullus colocynthis*	28	CIT 154	Turkmenistan	IP
108	*Citrullus colocynthis*	33	CIT 166	Cape Verde	IP
109	*Citrullus colocynthis*	35	CIT 190	Morocco	IP
110	*Citrullus colocynthis*	12	CIT 192	India	IP
111	*Citrullus colocynthis*	12	CIT 199	Egypt	IP
112	*Citrullus colocynthis*	38	CIT 281	Cyprus	IP
113	*Citrullus colocynthis*	13	CIT 307	Namibia	IP
114	*Citrullus colocynthis*	30	PI 195927	Ethiopia	US
115	*Citrullus colocynthis*	7	PI 220778	Afghanistan	US
116	*Citrullus colocynthis*	7	PI 346082	Afghanistan	US
117	*Citrullus colocynthis*	5	PI 386014	Iran	US
118	*Citrullus colocynthis*	5	PI 386015	Iran	US
119	*Citrullus colocynthis*	5	PI 386016	Iran	US
120	*Citrullus colocynthis*	5	PI 386018	Iran	US
121	*Citrullus colocynthis*	7	PI 386021	Iran	US
122	*Citrullus colocynthis*	27	PI 386024	Iran	US
123	*Citrullus colocynthis*	29	PI 386026	Iran	US
124	*Citrullus colocynthis*	37	PI 432337	Cyprus	US
125	*Citrullus colocynthis*	34	PI 525082	Egypt	US
126	*Citrullus colocynthis*	31	PI 537277	Pakistan	US
127	*Citrullus lanatus* subsp. *vulgaris*	2	824 AE 60	Burkina Faso	Pr

No	Taxon	Haplotype number	Accession number	Origin	So
128	*Citrullus lanatus* subsp. *vulgaris*	23	825 AE 60	Burkina Faso	Pr
129	*Citrullus lanatus* subsp. *vulgaris*	2	831 AE 031	Burkina Faso	Pr
130	*Citrullus colocynthis*	25	962 KU 026	Burkina Faso	Pr
131	*Citrullus lantus* cv. neri	1	06 NIA 095	Ghana	Pr
132	*Citrullus lantus* cv. neri	20	06 NIA 103	Ghana	Pr
133	*Citrullus lantus* cv. neri	1	06 NIA 111	Ghana	Pr
134	*Citrullus lanatus vulgaris* sugar baby	2	GRIF 15895	Canada	US
135	*Citrullus lanatus vulgaris* sugar baby	2	GRIF 15898	USA	US

Table 2: Genetic statistics based on the *trn* T-L,*ndh* F-*rpl* 32 and their combination in *Citrullus* spp.

CpDNA regions	Taxonomic groups	Number of accessions	Total Length (bp)	Parsimony informative sites	Number of haplotypes	Haplotypes diversity	Nucleotide diversity (Pi)	Average number of nucleotide difference (k)	Indel events
*trn*T-L	*Citrullus lanatus*	78	951-954	0	4	0.44	0	0	3
	C. mucosospermus	16	950-953	0	3	0.34	0	0	2
	C. amarus	22	950-953	0	5	0.52	1 x 10^{-4}	0.09	4
	C. colocynthis	22	948-954	6	12	0.92	28 x 10^{-4}	2.65	5
*ndh*F-*rpl*32	*C. lanatus*	78	650-652	0	8	0.24	0.4 x 10^{-4}	0.027	5
	C. mucosospermus	16	651-652	0	3	0.25	1.9 x 10^{-4}	0.125	0
	C. amarus	22	651-653	2	6	0.71	10.5 x 10^{-4}	0.68	1
	C. colocynthis	22	650-653	1	11	0.80	7 x 10^{-4}	0.45	6
*trn*T-L & *ndh*F-*rpl*32	*C. lanatus*	78	1601-1605	0	12	0.56	0.2 x 10-4	0.025	8
	C. mucosospermus	16	1601-1604	0	5	0.53	0.8 x 10-4	0.125	2
	C. amarus	22	1602-1604	2	8	0.81	4 .8 x 10-4	0.78	6

CpDNA regions	Taxonomic groups	Number of accessions	Total Length (bp)	Parsimony informative sites	Number of haplotypes	Haplotypes diversity	Nucleotide diversity (Pi)	Average number of nucleotide difference (k)	Indel events
	C. colocynthis	22	1599-1605	7	16	0.96	19.5 x 10-4	3.10	12

Parsimony-informative sites : Polymorphic sites with a minimum of two alleles that are each present at least twice in the population.

Non-informative sites: Polymorphic sites that are unique in the population (singleton sites).

Haplotype diversity: The probability that two given sequences from two different haplotypes belong to two different regions or populations.

Nucleotide diversity : The average number of nucleotide substitutions per site between two sequences (Lynch and Crease 1990).

Average number of nucleotide differences : The average number of nucleotide differences (either Indels or SNPs) within a given population.

Indel events : The number of insertions/deletions in the genomic region.

A + T (%) : A+T content in the genomic region.

Table 3: Haplotype codes for the combined *trn* T-L and*ndh* F-*rpl* 32 chloroplast regions for the global collections of the four *Citrullus* species in this study.

ID	Haplotype	Species	Origin
1	T–TT-TGTGTAAACACAAA—ATTAGA-	*C. lanatus ; C. mucosospermus*	Africa ; Asia, Europe, America
2	T-**T**TT-TGTGTAAACACAAA—ATTAGA-	*C. lanatus ; C. amarus*	Africa ; Asia, Europe, America
3	T–**-**T-TGTGTAAACACAAA—ATTA**TC**-	*C. amarus*	Southern Africa
4	T–**-**T-TGTGTAAACACAAA—ATTA**TCC**	*C. amarus*	Africa ; Asia
5	**-–-**T**A**TGTGT**T**AA**A**ACAAA-**T**-A**-**TA**T**A-	*C. colocynthis*	Near Eastern
6	T–TT-TGTGTAAACACAAA—ATTAGA**C**	*C. mucosospermus ; C. amarus*	Africa ; Asia
7	T–**-**T**A**TGTG**GT**AA**A**ACAAA-**T**-A**-**TA**T**A-	*C. colocynthis*	Near Eastern
8	T–TT-TGTGTAAACACAAA—ATTA**TCC**	*C. amarus*	South-Africa
9	T**G**-TT-TGTGTAAACACAAA—ATTAGA-	*C. mucosospermus*	West-Africa
10	T–TT-TGTGTAAACACAAA—**-**TTAGA-	*C. lanatus*	Europe ; Asia
11	T–TT-TGTGTAAAC**-**CAAA—ATTAGA-	*C. lanatus*	America
12	T**G--**T**A**TGTG**GT**AA**A**ACAAA-**T**-A**-**TA**T**A-	*C. colocynthis*	Northern Africa
13	T**G**–T**A**TGTGTAAACACAAA—ATTA**TC**-	*C. colocynthis*	Southern Africa
14	T–TT-–TGTAAACACAAA—ATTA**TCC**	*C. amarus*	Southern Africa
15	T–**-**T-TGTGTAAACACAAA—ATTA**T**A-	*C. amarus*	Europe
16	T–**-**T-TGTGTAAAC**-**CAAA—ATTA**T**A-	*C. lanatus*	Southern Africa
17	T–TT-TGTGTAAACACAAA—ATTA**T**A-	*C. mucosospermus*	Africa
18	T**G**-TT-TGTGTAAACACAAA—**-**TTAGA-	*C. lanatus*	Europe
19	T-**T**TT-TGTGTAAACACAAA—ATTAGA**C**	*C. lanatus*	Asia

ID	Haplotype	Species	Origin
20	T–TT-TGTGTAAACACAA**-**—ATTAGA-	*C. lanatus*	Africa
21	T–TT-TGTGTAAACA**-**AAA—ATTAGA-	*C. lanatus*	America
22	T-**T**TT-TGTGTAAACACAAA—A**-**TAGA-	*C. lanatus*	Asia
23	T-**T**TT**A**TGTGTAAACACAAA—ATTAGA-	*C. lanatus*	Africa
24	T**GT**TT-TGTGTAAACACAAA—ATTAGA-	*C. lanatus*	Europe
25	T-**T**TT-TGTGTAAACAC**-**AA—ATTAGA-	*C. colocynthis*	Africa
26	T—-TGTGTAAACACAAA—ATTAGA-	*C. mucosospermus*	Africa
27	T–TT**A**TGTG**GT**AAA**A**ACAAA-**T**-A**-**TA**T**A-	*C. colocynthis*	Asia
28	T–**-**T**A**TGTG**GT**AAA**A**ACAAA–**A**A**-**TA**T**A-	*C. colocynthis*	Asia
29	T–**-**T**A**TGTG**GT**AAA**A**ACAAA-**T**-A**-**TAGA-	*C. colocynthis*	Asia
30	T–**-**T**A**TGTGT**T**AACACA**C**A-**T**-A**-**TA**T**A-	*C. colocynthis*	Africa
31	T-**T**TT**A**TGTGTA**G**ACACAAA-**T**-–TA**T**A-	*C. colocynthis*	Asia
32	T—-TGTGTAA**G**CACAAA**AT**-A**-**TAGA**C**	*C. amarus*	Africa
33	T**G**-–**A**T**A-A**TAA**GA**ACAAA**ATA**A**-**TA**T**A-	*C. colocynthis*	Africa
34	T—**A**T**A-A**TAA**GA**ACAAA**ATA**A–**CT**A-	*C. colocynthis*	Africa
35	T**G**-–**A**T**A-A**TAA**GA-**CAAA–**A**A-TA**T**A-	*C. colocynthis*	Africa
36	T—**A**T**A-A**TAA**GA**ACAAA–**A**A-TA**T**A-	*C. colocynthis*	Europe
37	T—**A**T**A-A**TAA**G**C**-**CAAA**ATA**A**-**TA**T**A-	*C. colocynthis*	Europe
38	T**G**-–**A**TGT**A**TAA**GA**ACAAA**ATA**A**-**TA**T**A-	*C. colocynthis*	Europe

Table 4: Diversity and differentiation statistics for the four *Citrullus* spp. in this study, based on combined cpDNA haplotypes, according to Pons and Petit (1996) and adapted from Guicking et al. (2011).

Genetic parameters	Value	Standard error
Expected mean within-population gene diversity (h_S)	0.737	0.0671
Expected total gene diversity (h_T)	0.917	0.0320
Expected coefficient of genetic differentiation (Gst)	**0.196**	0.0812
Observed mean within-population gene diversity (Vs)	0.668	0.1878
Observed total gene diversity, accounting for similarities among haplotypes (V_T)	1.067	0.1609
Observed coefficient of genetic differentiation (Nst)	**0.374**	0.1274

h_S : The average permuted value of gene diversity within the four geographical regions (Africa, America, Asia, and Europe).

h_T : The permuted value of gene diversity across all four geographical regions.

G_{St} : The permuted value of genetic differentiation among the four geographical regions.

V_S: The average observed value of gene diversity within the four geographical regions.

V_T: The observed value of gene diversity across all four geographical regions.

N_{St}: The observed value of genetic differentiation among the four geographical regions.

Table 5: Pairwise genetic differentiation between continents (a), between African regions (b) and between Asian regions (c)

5-a: Pairwise genetic differentiation between continents (Hudson, 1993)	5-a: Pairwise genetic differentiation between continents (Hudson, 1993)	5-a: Pairwise genetic differentiation between continents (Hudson, 1993)	5-a: Pairwise genetic differentiation between continents (Hudson, 1993)	5-a: Pairwise genetic differentiation between continents (Hudson, 1993)	5-a: Pairwise genetic differentiation between continents (Hudson, 1993)	5-a: Pairwise genetic differentiation between continents (Hudson, 1993)
Region 1	**Region 2**	**Hs**	**Ks**	**Kxy**	**Gst**	**Chi-square**
Africa	Asia	0.85	0.85	4.78	0.006	Chi2 = 135.067 P-value = 0.05
Africa	Europe	0.76	0.76	3.84	0.035	
Africa	America	0.81	0.81	2.92	0.023	
Asia	Europe	0.73	0.73	4.41	0.038	
Asia	America	0.77	0.77	3.43	0.014	
Europe	America	0.57	0.57	2.12	0.0079	
5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)	5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)	5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)	5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)	5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)	5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)	5-b: Pairwise genetic differentiation between African regions (Hudson, 1993)
Region 1	**Region 2**	**Hs**	**Ks**	**Kxy**	**Gst**	**Chi-square**
West-Africa	South-Africa	0.73	1.92	3.79	0.12	Chi2 = 84.02 P-value = 0.0001
West-Africa	South-Africa	0.72	3.14	9.02	0.043	
South-Africa	North-Africa	0.85	3.88	9.34	0.05	
5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)	5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)	5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)	5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)	5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)	5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)	5-c: Pairwise genetic differentiation between Asian regions (Hudson, 1993)
Region 1	**Region 2**	**Hs**	**Ks**	**Kxy**	**Gst**	**Chi-square**
East-Asia	West-Asia	0.77	3.50	6.30	0.04	Chi2 = 65.75 P-value = 0.0047
East-Asia	South-Asia	0.76	2.65	4.73	0.06	
East-Asia	North-Asia	0.64	1.30	2.37	0.09	
West-Asia	South-Asia	0.89	6.20	6.20	0.014	
West-Asia	North-Asia	0.78	4.97	6.64	0.08	
South-Asia	North-Asia	0.77	4.19	5.11	0.07	

Hs : The mean within-continent gene diversity.

Ks: A weighted average of the number of differences between sequences from continents i and j .

Kxy: The average number of differences between two samples, regardless of their provenance.

GST : The coefficient of genetic differentiation between continents

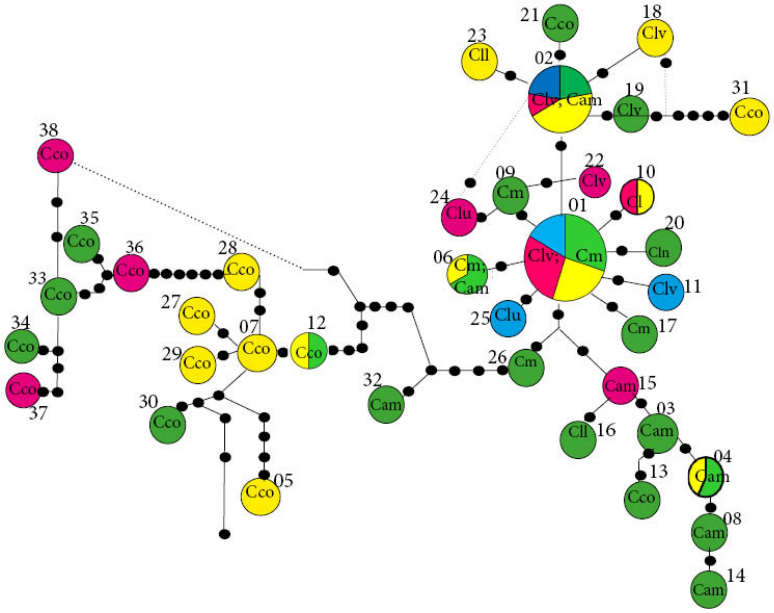

Fig. 1 . TCS network of 38 *Citrullus* spp. haplotypes. Circle size is proportional to haplotype frequency. Taxon names are abbreviated with two or three letters. Clv: *C. lanatus* subsp.*vulgaris* ; Cll: *C. lanatus* subsp. *lanatus* ; Cm:*C. mucosospermus* ; Cam: *C. amarus* ; and Cco: *C. colocynthis* . The numbers are arbitrary haplotype ID numbers (see Table S2), and the colors indicate geographical distribution: Africa (green), Asia (yellow); Europe (red), and North America (blue).

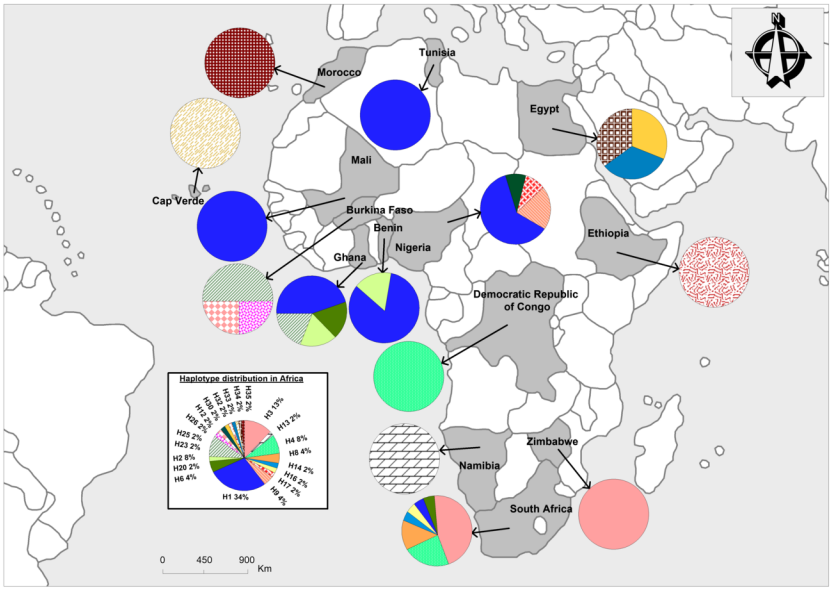

Fig. 2 . Distribution and frequencies of *Citrullus* spp. haplotypes in Africa.

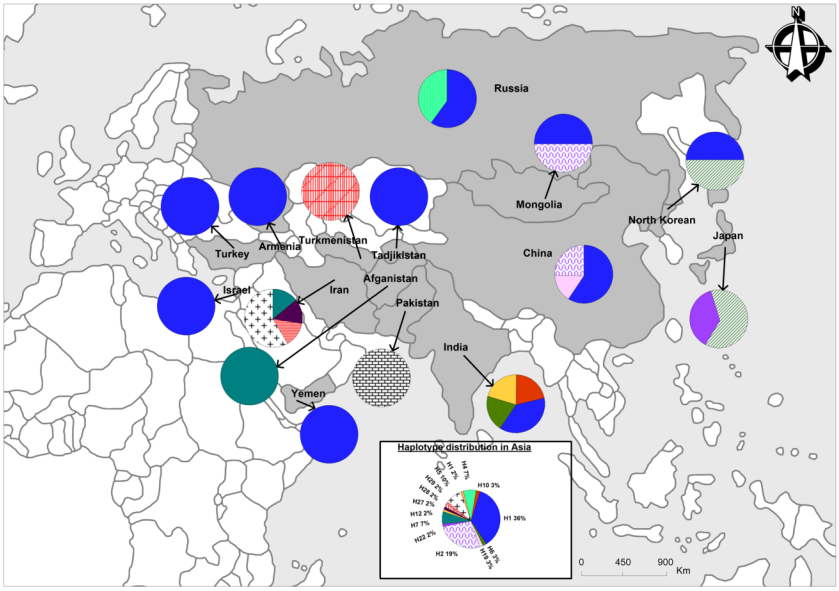

Fig. 3 . Distribution and frequencies of *Citrullus* spp. haplotypes in Asia.

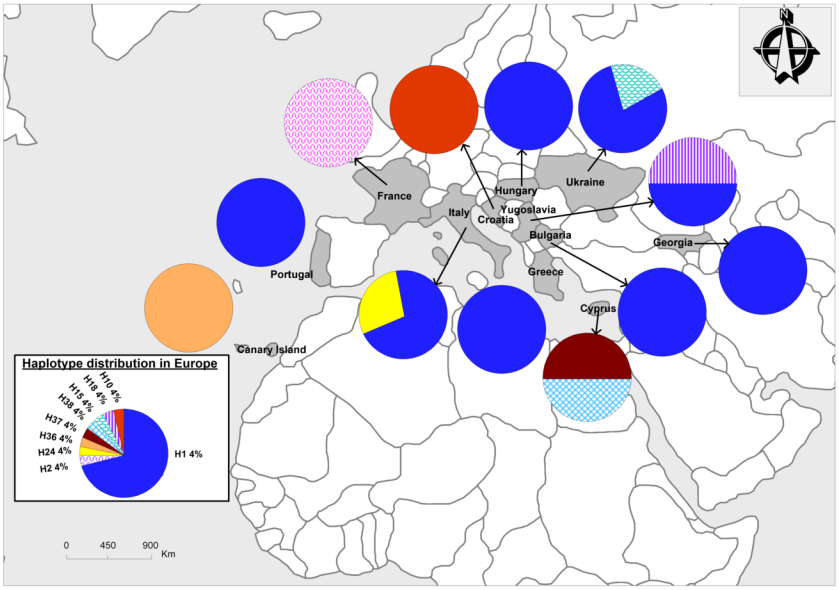

Fig. 4 : Distribution and frequencies of *Citrullus* spp. haplotypes in Europe.

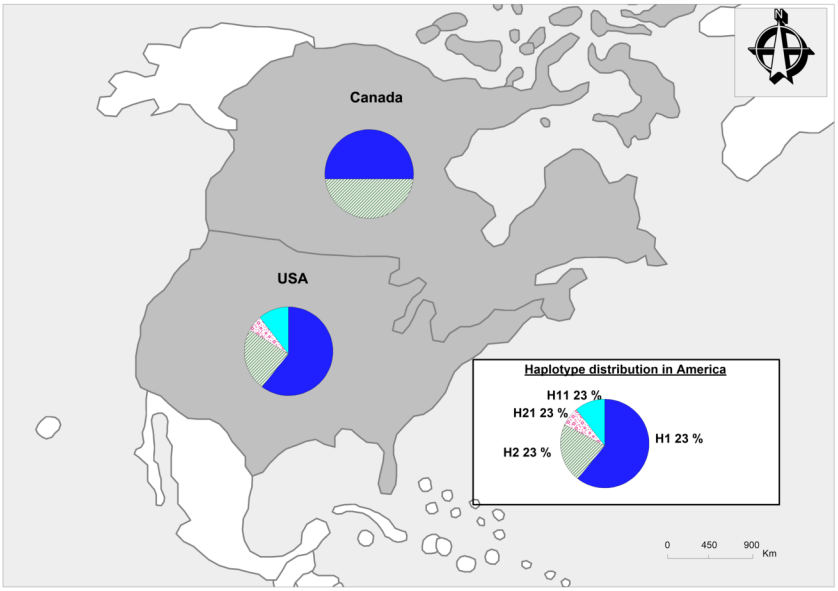

Fig. 5 : Distribution and frequencies of *Citrullus* spp. haplotypes in North America.