

Conservation Genetics – Heat Map Analysis of nuSSRs of aDNA of Archaeological Watermelons (*Cucurbitaceae*, *Citrullus l. lanatus*) Compared to Current Varieties

Gábor Gyulai^{1*} • Zoltán Szabó^{1,2} • Barna Wichmann¹ • András Bittsánszky^{1,3} • Luther Waters Jr.⁴ • Zoltán Tóth¹ • Fenny Dane⁴

¹ St. Stephanus University, School of Agricultural and Environmental Sciences, GBI, Gödöllő, H-2103 Hungary

² Agricultural Biotechnology Centre, Gödöllő, H-2100 Hungary

³ Plant Protection Institute, Hungarian Academy of Sciences, Budapest, H-1525 Hungary

⁴ Department of Horticulture, Auburn University, Auburn, Alabama AL 36849, USA

Corresponding author: * gyulai.gabor@mkk.szie.hu

ABSTRACT

Seed remains of watermelon (*Citrullus lanatus lanatus*) were excavated from two sites dating from the 13th (Debrecen) and 15th centuries (Budapest) Hungary. Morphological characterization, aDNA (ancient DNA) extraction, microsatellite analyses, and *in silico* sequence alignments were carried out. A total of 598 SSR fragments of 26 alleles at 12 microsatellite loci of DNAs were detected in the medieval and current watermelons. A heat map analysis using double dendrograms based on microsatellite fragment patterns revealed the closest similarity to current watermelons with red flesh (13th CENT) and yellow flesh (15th CENT) colors. *In silico* studies on cpDNA and mtDNA of watermelon revealed new data on *Citrullus* genome constitution. The results provide new tools to reconstruct and ‘resurrect’ extinct plants from aDNA used for conservation genetics.

Keywords: ancient DNA, *Citrullus*, SSR, watermelon

Abbreviations: aDNA, ancient DNA; SSR, microsatellite locus of DNA

INTRODUCTION

The aDNAs (ancient DNA) recovered from excavated remains of plants and animals supply unique genetic materials not only for the analysis of post-mortem DNA degradation (Brown 1999; Threadgold and Brown 2003), but also for tracing vegetation history and microevolution (Gugerli *et al.* 2005; Schlumbaum *et al.* 2008). Intact aDNA sequences (Gyulai *et al.* 2011a, 2011b) and complete genome (*mitome*) (Cooper *et al.* 2001; Pääbo *et al.* 2004) of the extinct organisms have been reconstructed. In this study we present the aDNA study of 800- and 600-year-old watermelons together with a comparison to current cultivars.

MATERIALS AND METHODS

Seed samples

Seed remains of watermelon (*Citrullus l. lanatus*) from the 13th century were excavated in Debrecen, Hungary (Tóth *et al.* 2007). In total, 95,133 seeds of 206 plant species were identified. Of these, 251 watermelon seeds were determined to have the same morphological characteristics. At the 15th century site (King’s Palace of Árpád Dynasty, Mansion Teleki, Buda Hill, Budapest; Hungary), 54,415 watermelon seeds were excavated (Gyulai *et al.* 2006, 2011b). Wet-sieved sediment samples were processed by floatation followed by seed sorting and identification (Schermann 1966; Hartványi and Nováky 1975). For comparative analysis, 38 current watermelon varieties were included.

Elimination of contamination

Seeds were incubated for seven days in an aseptic tissue culture medium (Gyulai *et al.* 2006) to eliminate contaminations before DNA extraction prior to washing with a detergent (3 min), rinsing

three times with distilled water (3 min), surface sterilization with ethanol (70% v/v) for 1 min, using a bleaching agent 8% Ca(OCl)₂ (w/v) for 1 min, followed by three rinses with sterile distilled water according to general aseptic culture technique (Gyulai *et al.* 2006). Seeds of the current varieties were also surface sterilized. Exogenously and endogenously contaminated seeds infected by fungi and bacteria were excluded from further analyses.

DNA extraction

Individual seeds were ground in an aseptic mortar with liquid nitrogen under aseptic conditions in the archaeobotanical laboratory of the St. Stephanus University, Gödöllő, Hungary. aDNA was extracted by the CTAB (cetyltrimethylammonium bromide) method according to Biss *et al.* (2003) and Yang (1997). Seed DNA of current cultivars (0.1 g) was also extracted in CTAB buffer, followed by an RNase-A treatment (Sigma, R-4875) for 30 min at 37°C. The quality and quantity of extracted DNA was measured (2 µl) by a NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Delaware, USA – BioScience, Budapest, Hungary). DNA samples were adjusted to a concentration of 30 ng/µl with ddH₂O and subjected to PCR amplification (Gyulai *et al.* 2011a, 2011b).

PCR amplification

Hot start PCR (Erlich *et al.* 1991) was combined with Touchdown PCR (Don *et al.* 1991) using AmpliTaq Gold™ Polymerase. Reactions were carried out in a total volume of 25 µl (containing genomic DNA of 30-50 ng, 1 x PCR buffer (2.5 mM MgCl₂), dNTPs (200 µM each), 20 pmol of each primer and 1.0 U of *Taq* polymerase. Touchdown PCR was performed by decreasing the annealing temperature by 1.0°C/cycle with each of the initial 12 cycles (PE 9700, Applied Biosystems), followed by a ‘touchdown’ annealing temperature for the remaining 25 cycles at 56°C for 30 s

Table 1 Nuclear SSR primer pairs applied in the analysis: 1-3 (Katzir *et al.* 1996; Danin-Poleg *et al.* 2001); 4-8 (Smith *et al.* 1997); 9-12 (Jarret *et al.* 1997), and the cpDNA primer pair # 13 (Al-Janabi *et al.* 1994). **Cl*-20 fragments were sequenced (Fig. 3a).

#	SSR loci	Primer pair sequences
1.	<i>CmTC</i> 51	attggggtttctttgaggtga ccatgctctctctctctctctctct
2.	<i>CmTC</i> 168	atcattggatgtgggattctc acagatggatgaaaccttagg
3.	<i>CmACC</i> 146	caaccaccgactactaagtc cgaccaaacccatccgataa
4.	<i>Bngl</i> 339	ccaaccgtatcagcatcagc gcagagctctctctctctctct
5.	<i>Bngl</i> 118-2	gccttcagccgcaacct cactgcatgcaaaggcaaccaac
6.	<i>Bngl</i> 161	gcttcctgcatcacacaactca atggagcatgagcttgcattt
7.	<i>Phi</i> 121	aggaaaatggagccgggtaacca ttggctggaccaagcacatacac
8.	<i>Phi</i> 118-2	atcgatcggtgctgcaaaa agacacgacggtgtccatc
9.	<i>Cl</i> 1-06	cacctctccagttgtcattcg aaggtcagcaaagcggcatagg
10.	* <i>Cl</i> 1-20	cgccgctgaggaccctata aacgcctcaatcaattgc
11.	<i>Cl</i> 2-23	gaggcggaggagttgagag acaaaacacgaaaccatagc
12.	<i>Cl</i> 2-140	cttttctctgattgactgg actgtttatcccgacttcaacta
13.	<i>Clp</i> 12	agttcagcctgattatccc gatgaacgctggcgcatgc

with a final cycle of 72°C for 10 min and held at 4°C. A minimum of three independent DNA preparations were used for each sample. Amplifications were assayed by agarose (1.8%, SeaKem LE, FMC) gel electrophoresis (Owl system), stained with ethidium bromide (0.5 ng/μl) after running at 80 V in 1 X TBE buffer. Each successful reaction with scorable bands was repeated at least twice. Transilluminated gels were analyzed by the Chemillimager v 5.5 computer program (Alpha Innotech Corporation - Bio-Science Kft, Budapest, Hungary). A negative control which contained all the necessary PCR components except template DNA was included in the PCR runs. Fragments were purified from agarose in a spin column (Sigma 5-6501) according to the manufacturer's protocol and subjected to sequencing. For SSR analysis 13 SSR (microsatellite) primer-pairs were used including 12 nuclear SSRs and one cpDNA SSR (Table 1).

ALF analysis

Microsatellite fragments were forwarded for ALF (*Automatic Laser Fluorometer*) analysis using ALF ExpressII (Pharmacia – Amersham, AP-Hungary). One strand of each of the SSR primer pairs was labeled with Cy5 dye. ALF analysis was carried out by the protocol of Röder *et al.* (1998) and Huang *et al.* (2002).

DNA sequencing

Amplified fragments of (CT)_n SSRs at *Cl*1-20 locus (Jarret *et al.* 1997) were isolated from the agarose gel with a spin column (Sigma, 56501) and subjected to automated fluorescent DNA sequencer (ABI PRISM 3100 Genetic Analyzer, Applied Biosystems, Hungary) (Gyulai *et al.* 2006). Authenticity of aDNA sequences were proved by BLAST analysis in each case using NCBI server (Altschul *et al.* 1997).

Sequence analysis

Sequence alignments were analyzed by programs of BioEdit Sequence Alignment Editor (North Carolina State University, USA) (Hall 1999), MULTALIN (Combet *et al.* 2000) and CLUSTAL W (Thompson *et al.* 1994). For SSRs, repeat search of FastPCR Program was used (Kalendar *et al.* 2009). For BLAST (*Basic*

Local Alignment Search Tool) analysis NCBI (*National Center for Biotechnology and Information*) server was used (Altschul *et al.* 1997). Molecular dendrograms were edited by computer programs of SPSS (*Statistical Package for the Social Sciences*; Stanford University, USA) and MEGA4 and 5 (Tamura *et al.* 2007). PIC value (*polymorphism index content*) of each SSR (Table 2) was calculated by the formula of Anderson *et al.* (1993):

$$PIC = 1 - \sum_{i=1}^n p_i^2$$

where p_i is the frequency of the i^{th} allele.

Heat map analysis

Double Cluster Analysis was carried out by R-program of Heat Map according to Ihaka and Gentleman (1996). (R stands for Ross Ihaka and Robert Gentleman).

RESULTS AND DISCUSSION

The genus *Citrullus* comprises four diploid species, including the annual watermelon (*Citrullus lanatus*), the perennial colocynth (*syn.*: bitter apple) (*Citrullus colocynthis*), and two wild species found growing in Kalahari desert, Africa, as the *Citrullus ecirrhosus* with bitter-tasting fruit, and the annual *Citrullus rehmii* with pink and olive green spotted, mandarin sized, non-edible fruits (Robinson and Decker-Walters 1997; Dane and Liu 2006).

Species watermelon (*Citrullus lanatus*) comprises diverse varieties, subspecies, mutants and feral forms such as the cultivated watermelon (*C. lanatus lanatus*) (*syn.*: *C. vulgaris*) with its ancient form of citron melon (the African *tsamma*) (*Citrullus lanatus citroides*), and the seed coat mutant egusi watermelon (*C. lanatus mucospermum*) (Kanda 1951; Gusmini *et al.* 2004). Watermelon, citron and colocynth have a long history of production in Europe (Wasylikowa and Veen 2004; Creamer 2005; Tóth *et al.* 2007; Gyulai *et al.* 2011b).

The oldest, 6,000-yr old watermelon (*Citrullus l. lanatus*) seeds were excavated in Helwan (Egypt, Africa), at a site 4,000 BCE (before current era) (Barakat 1990). Seeds were excavated in Uan Muhuggiag, Lybia, Africa from a site circa 3,000 BCE (Wasylikowa and Veen 2004), and were found in Pharaoh's tombs in Thebes (New Kingdom: 1,550-1,070 BCE; stored in the Agricultural Museum, Dokki, Giza, Egypt) (Warid 1995) and in the pyramid of Tutankhamun ca. 1,330 BCE (Hepper 1990; Vartavan and Amorós 1997; Kroll 2000).

The Greeks and Romans traveling to Egypt must have known of watermelons probably without discriminating it from colocynth and citron melon. Pliny II. wrote about a wild (probably the current colocynth) and two types (one with pale green, and the other with grass green rind) of cultivated colocynth (probably the current watermelon).

The Codex *De Materia Medica* (produced not too long after the time of Pliny) provides nearly 400 color paintings of different plants but no watermelon illustration, only a precise color painting of colocynth which looks very much like the current forms of colocynth (*Citrullus colocynthis*) (Dioscorides 1st CENT and the second 'edition' with color paintings from 512 CE (current era)).

Six hundred years later, when the Iberian Peninsula was conquered by the Berbers ('Moors') led by Tarik Ibn Ziyad in 711 CE, new watermelon types had entered Europe as recorded in the ancient *Book of Agriculture* (Al-Awwam 1158). In this book, two cultivated forms were compared, a black seed type (with dark-green rind which turns black when it ripens) and a red seed type (with green rind which turns to yellow) (Blake 1981). Watermelon might have also been introduced to Europe through Crusaders (Fischer 1929) led by either Richard I. the Lion-Hearted (*the 3rd Crusaders, 1190-1199*), or the Hungarian King, Endre II. of the Árpád Dynasty (*led the 6th Crusaders, 1228-1229*). Watermelon spread through Europe quickly and became a very popular and commonly cultivated fruit of the Renaissance Europe, with the first illustration on the frescos in the

Villa Farnesina, Rome, Italy, 1517 BCE (painted by Giovanni Martini da Udine) (Janick and Paris 2006). By 800 CE, watermelons had been introduced to India and by 1,100 CE to China. It reached the New World after Columbus' second voyage in 1493 and dispersed quickly among American natives (Blake 1981). One of the most ancient forms with small, round fruit and thin, green rind, red flesh and small black seeds has survived up to the recent times (Gilmore 1919).

One of the oldest watermelon herbarium samples is available from G Bauhin's (1560-1624) collections (about a hundred years earlier than Linnaeus's) (*pers. comm.*, Mark Spencer, *The Natural History Museum, London, UK*), who named it *Anguira citrullus*. No watermelon herbarium sample remained from C Linnaeus (1753) collections (*pers. comm.*, Arne Anderberg, *Swedish Museum of Natural History, Stockholm, S*), who named watermelon as *Cucurbita citrullus*, and colocynth as *Cucumis colocynthis* (The Linnean Collection, The Linnean Society of London, UK).

Microsatellite analysis

Watermelon seeds excavated at both medieval sites in Hungary and analyzed in this study appeared to be extremely well preserved due to the anaerobic conditions in the slime of a deep well covered by water, apparently used as dust holes in the Middle Ages (Gyulai *et al.* 2006, 2011a, 2011b).

Microsatellite (SSR) probes were used for morphological reconstruction of the ancient watermelons. Microsatellite analysis of the study presented revealed 598 SSR fragments in total of 26 alleles at 12 microsatellite loci (Table 2).

Allelic diversity of microsatellites (Orti *et al.* 1997; Tóth *et al.* 2000; Schulman 2007) were reliably detected in aDNAs of 300 – 1,100-yr old seagrass (*Posidonia oceanica*) (Raniello and Procaccini 2002). SSRs were also used to morphologically reconstruct 600-yr old melon (*Cucumis sativus*) (Szabó *et al.* 2005; Gyulai *et al.* 2011a) and millet (*Panicum miliaceum*) (Lágler *et al.* 2005; Gyulai *et al.* 2006, 2011c). SSR analysis was also applied to herbarium samples of common reed (*Phragmites australis*), about 100-yr old, to track plant invasion in North America (Saltonstall 2003).

Heat map

Double molecular dendrogram based on Heat Map of the 598 SSR fragments revealed that middle age samples show close lineages to ancient Hungarian landraces, the 13th CENT sample similar to cultivars with red flesh color, and the 15th CENT sample clustered with yellow flesh color cultivars (Fig. 1). Results of seed morphology correlated strongly with molecular results (Gyulai *et al.* 2011b).

Heat map is a graphical representation of molecular

Table 2 Variation of 598 SSR fragments of 26 alleles at 12 SSR loci of 40 watermelons (*Citrullus l. lanatus*) including medieval samples (13th and 15th centuries).

SSR loci	cmtc168		bngl161		bngl118-2		phi118-2			bngl339		phi121		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Alleles	a	a	b	a	b	a	b	c	a	b	c	a	b	c
Length (SSR bp)	126	139	146	124	132	129	132	136	113	131	140	124	130	135
1	13 th CENT	•	•	•	•	•	•	•	•	•	•	•	•	•
2	15 th CENT	•	•	•	•	•	•	•	•	•	•	•	•	•
3	Bácsbokod	•	•	•	•	•	•	•	•	•	•	•	•	•
4	Belyj Dlinnij	•	•	•	•	•	•	•	•	•	•	•	•	•
5	Biri	•	•	•	•	•	•	•	•	•	•	•	•	•
6	Charleston gr.	•	•	•	•	•	•	•	•	•	•	•	•	•
7	Crimson sw.	•	•	•	•	•	•	•	•	•	•	•	•	•
8	Csárdaszállás	•	•	•	•	•	•	•	•	•	•	•	•	•
9	Debrecen	•	•	•	•	•	•	•	•	•	•	•	•	•
10	Dévaványa	•	•	•	•	•	•	•	•	•	•	•	•	•
11	Gyöngyös	•	•	•	•	•	•	•	•	•	•	•	•	•
12	Háromfa	•	•	•	•	•	•	•	•	•	•	•	•	•
13	Hevesi	•	•	•	•	•	•	•	•	•	•	•	•	•
14	Ilk	•	•	•	•	•	•	•	•	•	•	•	•	•
15	Kecskeméti v.	•	•	•	•	•	•	•	•	•	•	•	•	•
15	Kibéd	•	•	•	•	•	•	•	•	•	•	•	•	•
17	Klondike R7	•	•	•	•	•	•	•	•	•	•	•	•	•
18	Kömör	•	•	•	•	•	•	•	•	•	•	•	•	•
19	Korai kincs	•	•	•	•	•	•	•	•	•	•	•	•	•
20	Lipót	•	•	•	•	•	•	•	•	•	•	•	•	•
21	Marsowszky	•	•	•	•	•	•	•	•	•	•	•	•	•
22	Nagyecsed	•	•	•	•	•	•	•	•	•	•	•	•	•
23	Nagykálló	•	•	•	•	•	•	•	•	•	•	•	•	•
24	Nagyvárad	•	•	•	•	•	•	•	•	•	•	•	•	•
25	Napsugár	•	•	•	•	•	•	•	•	•	•	•	•	•
26	Nyírbátor	•	•	•	•	•	•	•	•	•	•	•	•	•
27	Nyíregyháza	•	•	•	•	•	•	•	•	•	•	•	•	•
28	Oros	•	•	•	•	•	•	•	•	•	•	•	•	•
29	Pusztadobos	•	•	•	•	•	•	•	•	•	•	•	•	•
30	Ráckeve	•	•	•	•	•	•	•	•	•	•	•	•	•
31	Rákóczifalva	•	•	•	•	•	•	•	•	•	•	•	•	•
32	Sándorfalva	•	•	•	•	•	•	•	•	•	•	•	•	•
33	Sibiriak	•	•	•	•	•	•	•	•	•	•	•	•	•
34	Sugar baby	•	•	•	•	•	•	•	•	•	•	•	•	•
35	Szentesi sh	•	•	•	•	•	•	•	•	•	•	•	•	•
36	Szirma	•	•	•	•	•	•	•	•	•	•	•	•	•
37	Taktaharkány	•	•	•	•	•	•	•	•	•	•	•	•	•
38	Tura	•	•	•	•	•	•	•	•	•	•	•	•	•
39	Túrkeve	•	•	•	•	•	•	•	•	•	•	•	•	•
40	Ukrainskij	•	•	•	•	•	•	•	•	•	•	•	•	•

Table 2 (Cont.)

SSR loci	<i>cmacc146</i>		<i>cmte51</i>		<i>cl1-06</i>			<i>cl 2-23</i>			<i>cl2-140</i>	<i>Clp-12</i>	Total
Alleles	15	16	17	18	19	20	21	22	23	24	25	26	# 598
	a	b	a	b	a	b	c	a	b	c	a	a	
Length (SSR bp)	124	136	306	309	130	145	150	178	183	187	190	301	
1	13. CENT.	•		•	•	•	•			•	•	•	17
2	15. CENT.	•	•	•	•			•	•	•	•	•	20
3	Bácsbokod	•		•		•	•	•	•	•	•	•	19
4	Belyj Dlinnij	•		•	•	•				•	•	•	15
5	Biri	•		•		•				•	•	•	12
6	Charleston gr.			•	•		•	•		•	•	•	15
7	Crimson sw.	•		•	•	•			•		•	•	18
8	Csárdaszállás	•	•	•	•	•	•	•	•		•	•	18
9	Debrecen		•			•				•	•	•	12
10	Dévaványa	•		•		•	•			•	•	•	16
11	Gyöngyös			•		•	•		•		•	•	11
12	Háromfa			•	•	•	•			•	•	•	14
13	Hevesi			•		•	•			•	•	•	13
14	Ilk	•			•	•				•	•	•	15
15	Kecskeméti v.	•		•	•	•	•	•		•	•	•	18
15	Kibéd			•		•					•	•	12
17	Klondike R7	•		•	•	•	•	•		•	•	•	21
18	Kömör	•		•	•					•	•	•	13
19	Korai kincs	•		•		•			•		•	•	14
20	Lipót	•	•	•		•			•	•	•	•	15
21	Marsowszky	•	•	•	•	•	•			•	•	•	17
22	Nagyecsed	•								•	•	•	12
23	Nagykálló	•		•		•	•		•	•	•	•	14
24	Nagyvárad	•		•	•	•				•	•	•	14
25	Napsugár	•		•		•	•			•	•	•	16
26	Nyírbátor	•			•					•	•	•	14
27	Nyíregyháza	•		•		•	•			•	•	•	13
28	Oros	•		•		•	•			•	•	•	14
29	Pusztadobos	•		•		•	•			•	•	•	16
30	Ráckeve			•		•	•			•	•	•	13
31	Rákóczi falva	•				•	•		•	•	•	•	13
32	Sándorfalva	•		•		•	•			•	•	•	17
33	Sibiriak	•		•		•	•	•	•	•	•	•	18
34	Sugar baby	•			•	•				•	•	•	13
35	Szentesi sh	•		•		•	•			•	•	•	13
36	Szirma	•		•		•	•			•	•	•	15
37	Taktaharkány				•	•	•		•	•	•	•	15
38	Tura			•	•		•			•	•	•	13
39	Túrkeve	•	•	•	•				•	•	•	•	16
40	Ukrainskij	•			•					•	•	•	14

data where cluster trees are joined to the rows and columns of the data matrix (Weinstein 2008). Four components were displayed in the heat map image (Fig. 1) with colour key (top left), a column dendrogram of genetic similarities of SSR loci (top centre), and row dendrogram of genetic similarities of watermelon cultivars, and the image plot in the middle. The colour key indicates the presence (# 1; red colour) or absence (# 0; green colour) status of the alleles. In both dendrograms the Euclidean distance calculation was applied, that helped to visualize the molecular differences among the watermelon cultivars based on SSR loci.

The analysis of SSR allele diversity according to Garvaka-G *et al.* (2008) indicated the most selective probes as *bngl339* (140 bp) which might be useful in the further studies (Fig. 2).

Sequences alignments

The SSR fragments amplified at the *Cl1-20* locus (Jarret *et al.* 1997) were isolated, sequenced and BLAST-ed (NCBI, Altschul *et al.* 1990). The results revealed orthologous SSR alleles in the monocot *Zoysia* and the dicot tree *Prunus* with different numbers of (CT)_n core sequence (Fig. 3A). The (CT)_n dinucleotide repeat was found to account for an unexpectedly high percent (46%) of the total SSRs of watermelon (Jarret *et al.* 1997). Data mining of watermelon SSRs (1 - 29) resulted in further SSR loci (Fig. 3B).

Sequence analysis of the (CT)_n allele (*Cl1-20*) showed two sharp molecular events that might have occurred during the evolution of watermelon. First, due to a (CT)₃ deletion in the core sequences of (CT)_n two alleles developed (*Ia.* and *Ib.*) resulting in a heterozygote/heterozygote forms. Both medieval watermelons (Fig. 4) and current cultivars of # 6, 8, 11 and 30 were heterozygous at this locus (Gyulai *et al.* 2011b). All the rest of the current watermelons were homozygous. Second, a (CT)₅ partial sequence along the (CT)_n allele inverted resulting in a compound SSR (CT)₅-C-(CT)₄-T-(CT)₅ from a putative ancestral simple SSR (CT)₁₆ (Fig. 3A). These results indicate the first molecular steps in the endogenous way of elimination of long SSRs from the genomes through fragmentation, and it also indicates the birth (Messier *et al.* 1996) and the death (Taylor *et al.* 1999) of microsatellites (Gyulai *et al.* 2011b).

In silico analyses

The relatively small genome of watermelon ($2n = 2x = 22$; $n = 11$; $0.425-0.454 \times 10^9$ bp; $0.42-0.45$ pg DNS) has been characterized by total nuclear genome size of 1,357.74 cM - 1,514.26 cM - 1,144.057 cM (McGregor *et al.* 2011). Since phylogenies inferred from the highly conserved nuclear-encoded 18S rRNA genes which have been unable to unravel the interrelationships of major lineages of several algae and land plants the *in silico* studies might have crucial clues.

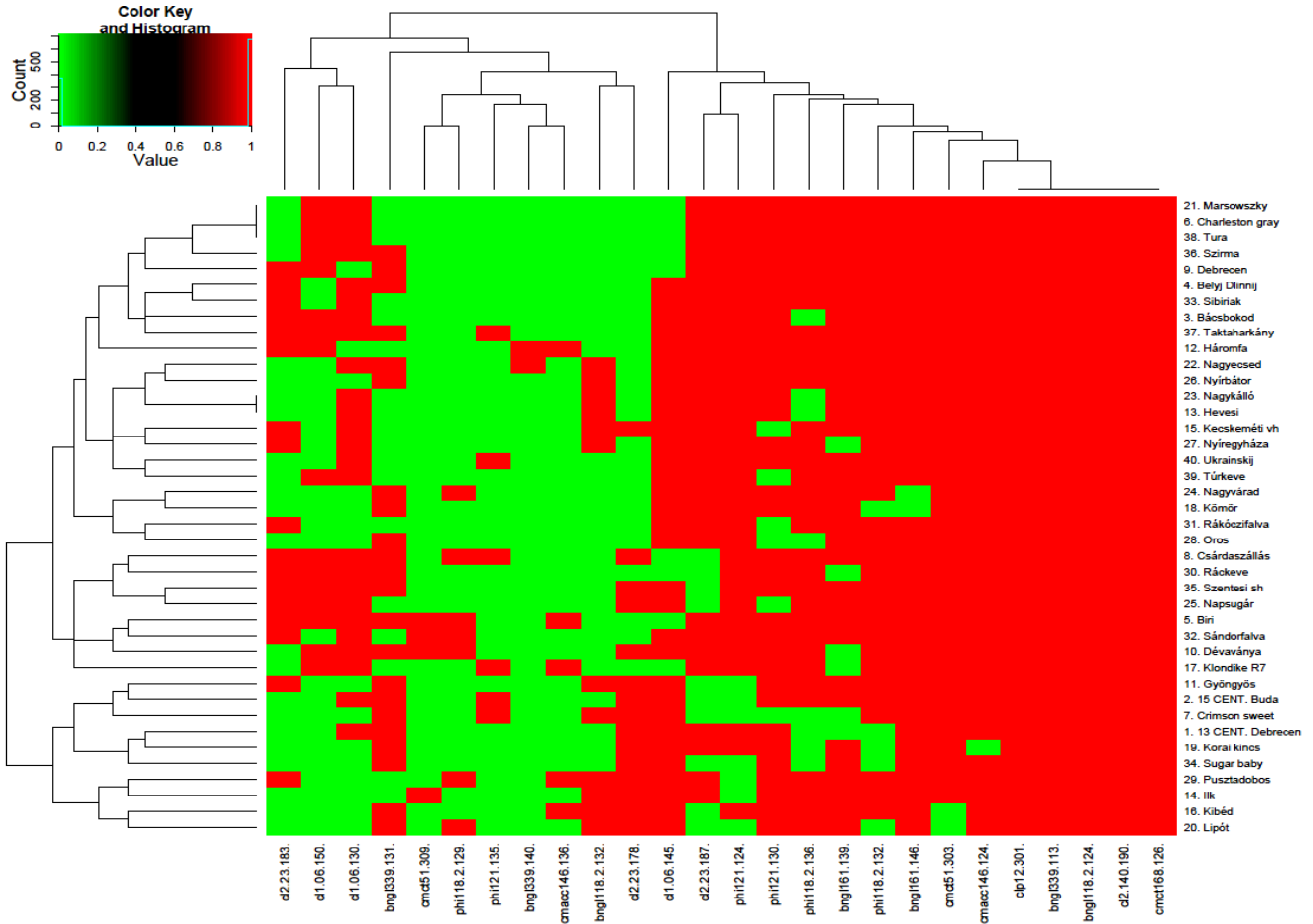


Fig. 1 Heat Map with double dendrograms. Genetic similarities (0.0 to 1; Color Key) of watermelon cultivars (1-38, and medieval samples; dendrogram left and cv. list right), and SSR allele frequencies (dendrogram top) are indicated.

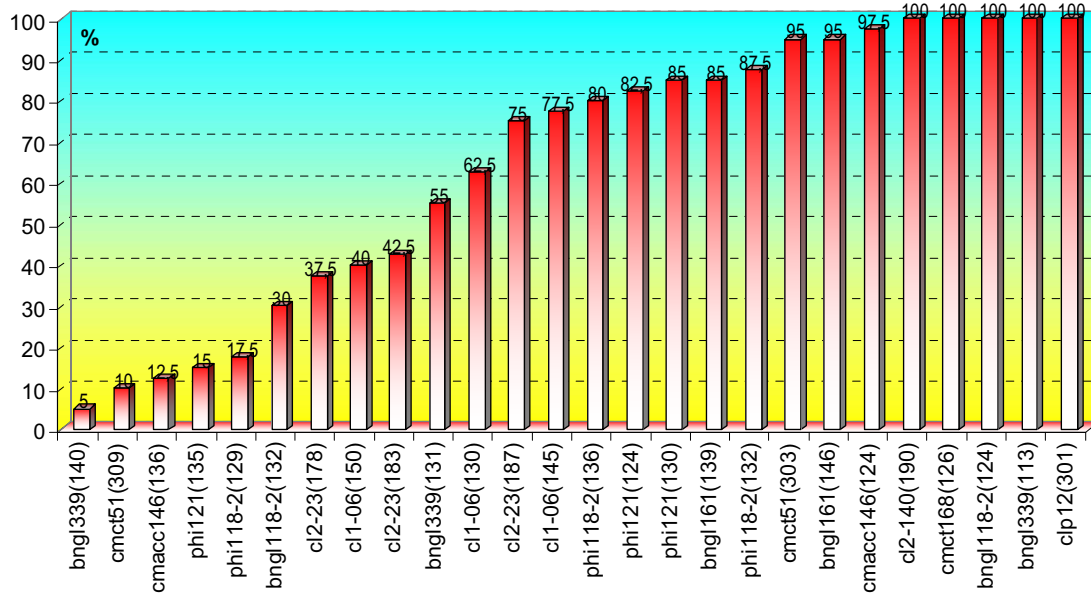


Fig. 2 Percentage of SSR alleles per loci of watermelons (*Citrullus lanatus*). Diversity (%) of he 598 SSR fragments of 26 alleles at 12 loci (listed in Table 2) identified in DNAs of 40 watermelon samples (*Citrullus l. lanatus*) including two archaeological remains (13th and 15th centuries) are indicated (notice the most selective locus at *bngl339*).

Complete cpDNA of watermelon (**Fig. 5A**) was supposed to be similar to cpDNA of *Cucumber* (*Cucumis sativus*) (155,293 bp; #NC 007144) and with close genetic similarity to *Morus indica* (Mansour *et al.* 2009). A cpDNA analysis of Cucurbitaceae genera has revealed species *Lagenaria* (*L. brevifolia* and *L. siceraria*) and *Peponium* (*P. calcedonicum* and *P. vogelii*) to have the closest genetic similarity to *Citrullus* (Renner *et al.* 2007).

In angiosperms, cpDNA (circular 120–250 kb) consists of 2 copies of inverted repeats (IR; 20–28 kb) separating with 2 single copy (SC) regions of 80–90 kb (the large SC: LSC) and 16–27 kb (the small SC: SSC). The SSC of cp genome (*plastome*) usually encodes around 4 rRNAs, 30 tRNAs, and 80 proteins. There are also genes of rRNAs and 10–15 proteins encoded in duplicate copies in the IRs. The rate of silent (synonymous) mutations (nucleotide substitu-

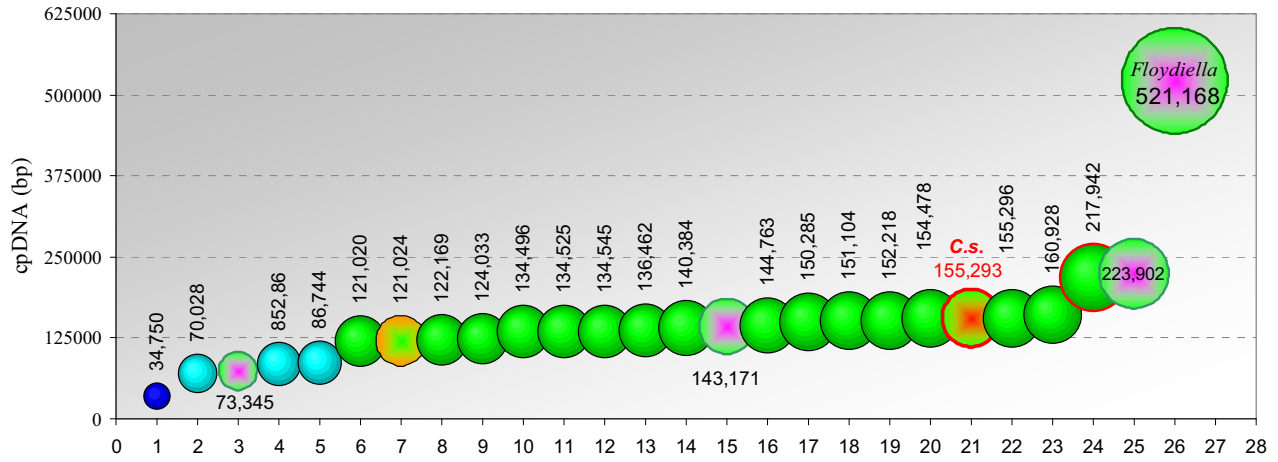


Fig. 5 (A) The cpDNA (plastome) sizes (bp) of plant genera including *Cucumis sativus* (21) (closest to *Citrullus* available) and the two green algae *Stigeoclonium helveticum* (223,902 bp) (25); and *Floydliella terrestris* (521,168 bp) (26) with the largest chloroplast genome ever sequenced (Brouard *et al.* 2010) (NCBI, Altschul *et al.* 1997). Parasite (1), parasite plants (2, 4, 5), algae (3, 15, 25, 26), moss liverworts (7), and *Cucumis* (21) are indicated. (1) *Eimeria tenella*, (2) *Epifagus virginiana* chloroplast, (3) *Euglena longa*, (4) *Cuscuta obtusiflora*, (5) *Cuscuta gronovii*, (6) *Lathyrus sativus*, (7) *Marchantia polymorpha*, (8) *Pisum sativum*, (9) *Medicago truncatula*, (10) *Oryza sativa Indica*, (11) *Oryza sativa Japonica*, (12) *Triticum aestivum*, (13) *Hordeum vulgare*, (14) *Zea mays*, (15) *Euglena gracilis*, (16) *Trifolium subterraneum*, (17) *Phaseolus vulgaris*, (18) *Helianthus annuus*, (19) *Glycine max*, (20) *Arabidopsis thaliana*, (21) *Cucumis sativus* (C.s.), (22) *Solanum tuberosum*, (23) *Vitis vinifera*, (24) *Pelargonium x hortorum* (with the largest genome of land plants; Chumley *et al.* 2006), (25) *Stigeoclonium helveticum*, and (26) *Floydliella terrestris*.

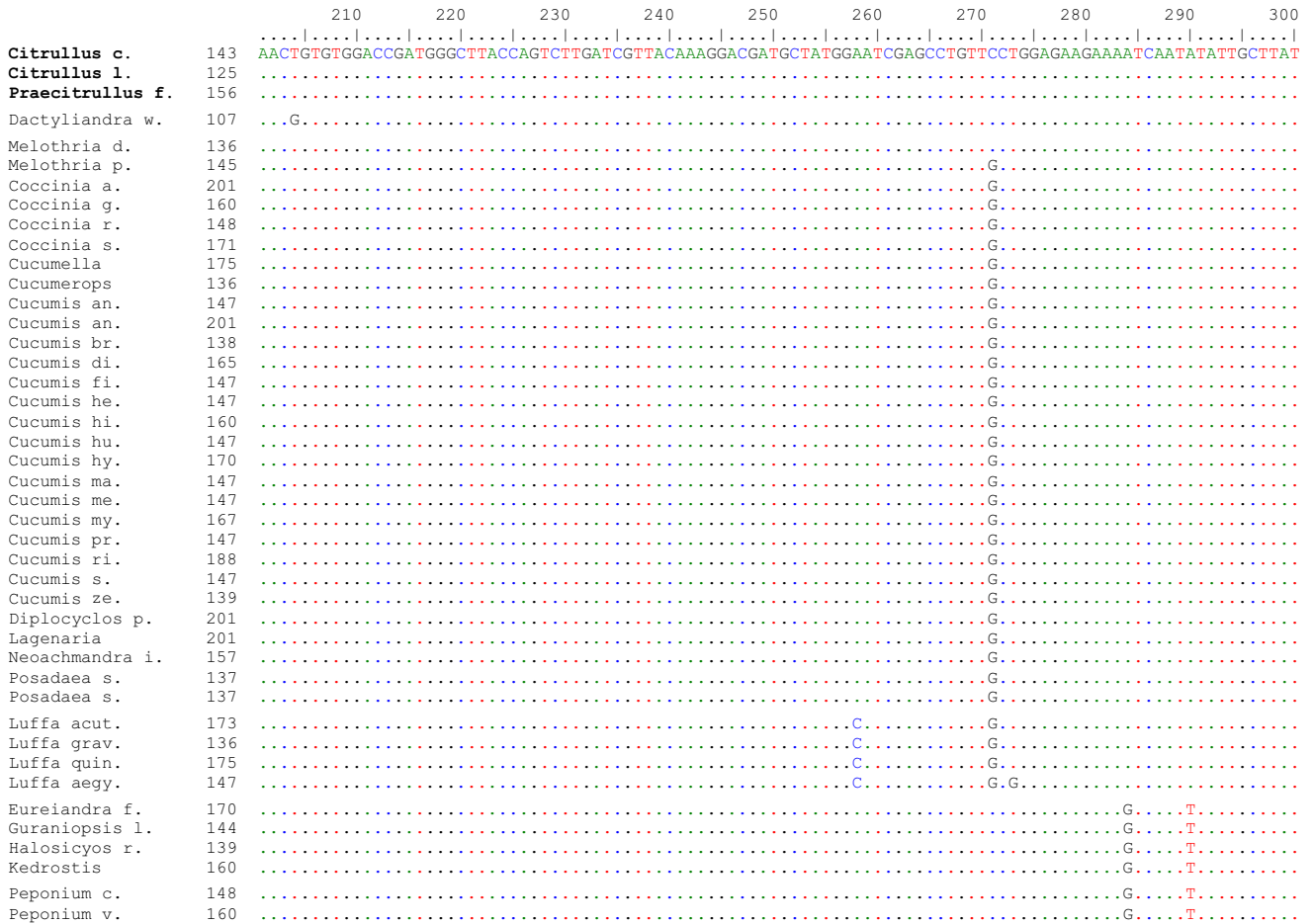


Fig. 5 (B) Consensus sequence alignments of *rbcL* genes of cpDNA. Partial sequences of the highly conserved *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) genes were blasted (NCBI, Altschul *et al.* 1997) and aligned (BioEdit, Hall *et al.* 1999) to DQ535791.1 (*Citrullus colocynthis*, 1384 nt). Notice *Citrullus lanatus* (*Citrullus l.*) and *Praecitrullus fistulosus*.

and the *in silico* analysis might provide further tools to reconstruct extinct plants from aDNA isolated from archaeological remains useful for conservation genetics.

ACKNOWLEDGEMENTS

Medieval seed samples were excavated, determined and kindly provided by Dr. Ferenc Gyulai.

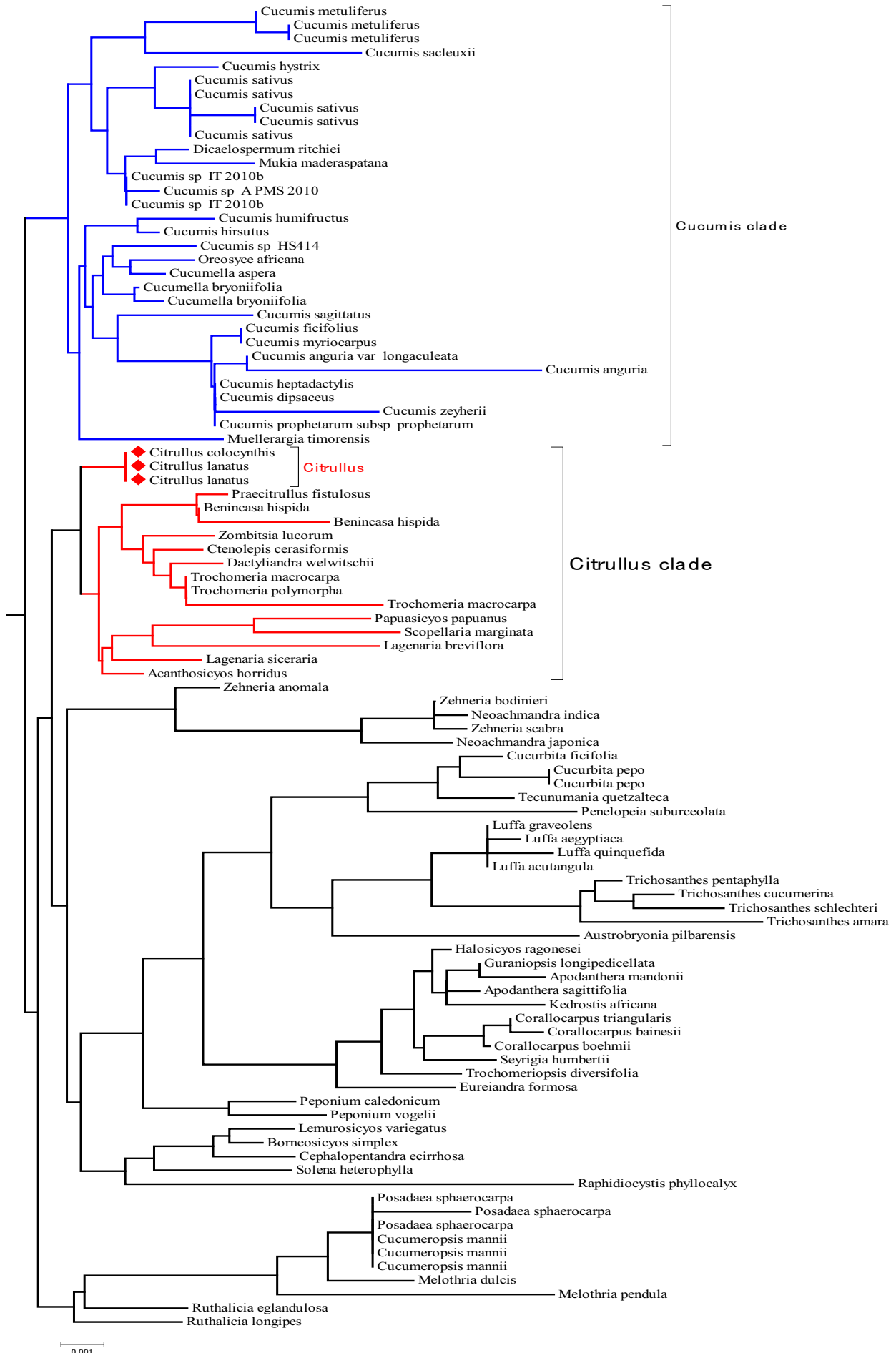


Fig. 5 (C) Phylogenetic tree of *Cucurbitaceae* species based on DNA sequences of the highly conserved gene *rbcL* of cpDNAs. Genera of *Cucurbitaceae* including species *Citrullus* were blasted (NCBI, Altschul *et al.* 1997) to DQ535791.1 (*Citrullus colocynthis*, 1384 nt). Relative genetic distance (0.001) is indicated (MEGA5, Tamura *et al.* 2007).

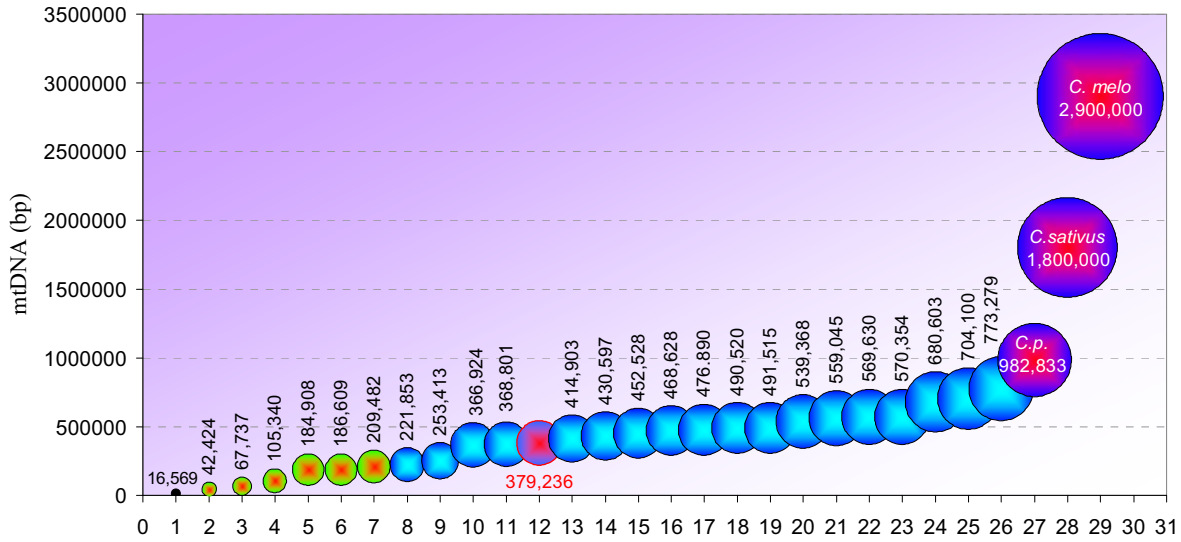


Fig. 6 (A) The mtDNA (*mitome*) sizes (bp) of plant genera (*Streptophyta*) including species of *Cucurbitacea* *Citrullus l.* (12), *Cucurbita p.* (27), *Cucumis s.* (28), and *Cucumis melo* (29). Human mtDNS (1), and lower plants (2 -7) are indicated (NCBI accession #, Altschul *et al.* 1997). (1) *Homo s.* NC_012920, (2) *Mesostigma v.* NC_008240, (3) *Chara v.* NC_005255, (4) *Physcomitrella p.* NC_007945, (5) *Megaceros ae.* NC_012651, (6) *Marchantia p.* NC_001660, (7) *Phaeoceros l.* NC_013765, (8) *Brassica n.* NC_008285, (9) *Silene l.* NC_014487, (10) *Arabidopsis th.* NC_001284, (11) *Beta v.* NC_002511, (12) *Citrullus l.* NC_014043, (13) *Cycas t.* NC_010303, (14) *Nicotiana t.* NC_006581, (15) *Triticum ae.* NC_007579, (16) *Sorghum b.* NC_008360, (17) *Carica p.* NC_012116, (18) *Oryza s.l.* NC_011033, (19) *Oryza s.l.* NC_007886, (20) *Zea lux.* NC_008333, (21) *Oryza r.* NC_013816, (22) *Zea m.* NC_007982, (23) *Zea pren.* NC_008331, (24) *Zea parv.* NC_008332, (25) *Tripsacum d.* NC_008362, (26) *Vitis v.* NC_012119, (27) *Cucurbita p.* NC_014050, (28) *Cucumis sativus*, and (29) *Cucumis melo* with the largest mtDNA (Alverson *et al.* 2010).

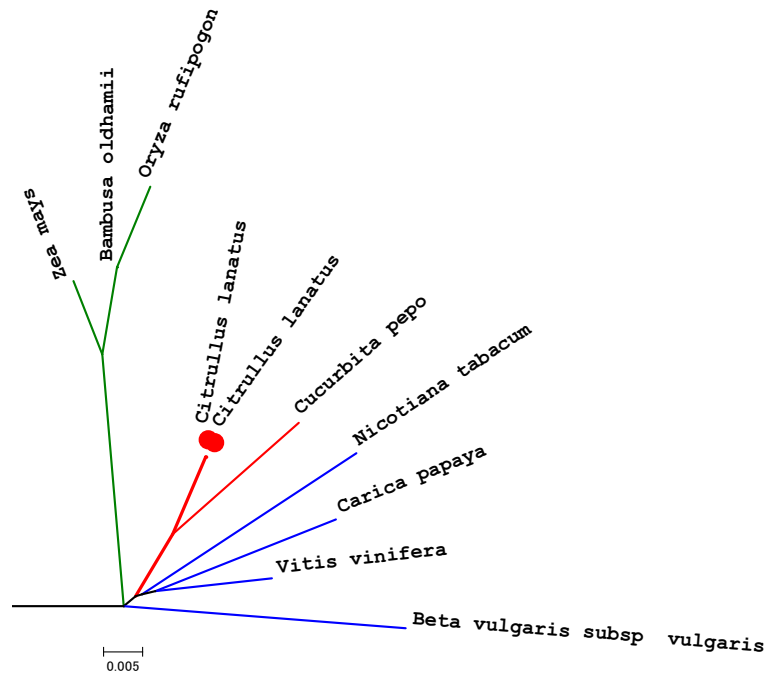


Fig. 6 (B) Phylogenetic tree of the total mtDNA genomes (*mitomes*) of *Citrullus* and related species. *Citrullus lanatus* (379,236 bp, #GQ856147.1 and NC_014043.1). *Vitis vinifera* cv Pinot noir (773,279 bp, #FM179380.1). *Cucurbita pepo* (982,833 bp, #GQ856148.1). *Carica papaya* (476,890 bp, #EU431224.1). *Nicotiana tabacum* (430,597 bp, #BA000042.1). *Beta vulgaris* (368,801 bp, #BA000009.3). *Bambusa oldhamii* (509,941 bp, #EU365401.1). *Oryza sativa* Japonica (490,520 bp, #BA000029.3). *Zea mays* (569,630 bp, #AY506529.1). (NCBI, Altschul *et al.* 1997). Relative genetic distance (0.005) is indicated (MEGA5, Tamura *et al.* 2007).

		1310	1320	1330	1340	1350	1360
<i>Citrullus lanatus</i> (#EU069546.1)	1247	CCA TGG GTT TCG CTA TGA AGA ATC GAC TAC C--	--- --	TC AAC TAA GTA TTC GAG TAA CTA ATA			
		Thr Ile Gly Phe Ala Met Lys Asn Arg Leu Pro		Gln Leu Ser Ile Arg Val Thr Asn			
<i>Cucumis metuliferus</i> (#EU069548.1)	1253 C.. TC.
		Thr Ile Ala Phe Ala Met Lys Asn Leu Pro		Gln Leu Ser Ile Arg Val Thr Asn			
<i>Cucumis melo</i> (#EU069547.1)	1244
		Thr Ile Gly Phe Ala Met Lys Asn Arg Leu Pro		Val Arg Gly Ser Leu Met Ile Arg Val Thr Asn			

Fig. 6 (C) Consensus sequence alignment and protein translate (BioEdit, Hall *et al.* 1999) of partial sequences of *cox1* genes of mtDNAs of watermelon (*Citrullus lanatus*) and two related species. Silent (synonymous) mutations (e.g. C), non-synonymous mutations (e.g. CTT/Ala), consensus nucleotides (•) deletions (-), three-letter codes of amino acids (e.g. The – threonine) and NCBI accession # (Altschul *et al.* 1997) are indicated.

REFERENCES

- Al-Awwam Ibn (1158) Le Livre de l'agriculture, translated by Clement-Mullet; 2 Tomes in 3 Vols., Paris (1864-1867). Reprints (1802, 1988). Libro de agricultura / su autor el doctor excelente Abu Zacaria Iahia; [traducido al castellano y anotado por Josef Antonio Banqueri; estudio preliminar y notas, J. E. Hernández Bermejo y E. García Sánchez]. Madrid
- Al-Janabi SM, Honeycutt RJ, Peterson C, Sobral BWS (1994) Phylogenetic analysis of organellar DNA sequences in the Andropogoneae: *Saccharum*. *Theoretical and Applied Genetics* **88**, 933-944
- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402
- Alverson AJ, XX Wei, DW Rice, DB Stern, K Barry, JD Palmer (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* **27**, 1436-1448
- Anderson, JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* **36**, 181-186
- Barakat H (1990) Appendix IV – plant remains from El-Omari. In: Debono F, Mortensen B (Eds) El-Omari. *A Neolithic Settlement and Other Sites in the Vicinity of Wadi Hof*, Helwan. Verlag Philipp von Zabern ISBN 3-8053-1119-2, Mainz am Rhein, 1990, pp 109-114
- Biss P, Freeland J, Silvertown J, McConway K, Lutman P (2003) Successful amplification of rice chloroplast microsatellites from century-old grass samples from the park grass experiment. *Plant Molecular Biology Reporter* **21**, 249-257
- Blake LW (1981) Early Acceptance of Watermelon by Indians of the United States. *Ethnobiology* **1**, 193-199
- Brouard JS, Otis C, Lemieux C, Turmel M (2010) The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. *Genome Biology and Evolution* **2**, 240-256
- Brown TA (1999) How ancient DNA may help in understanding the origin and spread of agriculture. *Proceedings of the Royal Society of London, Series B* **354**, 89-98
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium x hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* **23**, 2175-2190
- Combet C, Blanchet C, Geourjon C, Deléage G (2000) NPS@: Network Protein Sequence Analysis. *Trends in Biochemical Sciences* **25**, 147-150
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707
- Creamer J (2005) Searching for a better melon (by Gyulai G and Dane F). AAES Impact, Augustus, Auburn, USA, p 2
- Dane F, CJ Liu (2006) Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genetic Resources and Crop Evolution* **54**, 1255-1265
- Danin-Poleg Y, Reis N, Tzuri G, Katzir N (2001) Development and characterization of microsatellite markers in *Cucumis*. *Theoretical and Applied Genetics* **102**, 61-72
- Dioscorides (512) (*true name Pedianos Dioskurides*) De Materia Medica (with nearly 400 color paintings) (1st century, Painting from the year 512). Der Wiener Dioskurides: Codex medicus Graecus I der Österreichischen Nationalbibliothek (1998-1999) commentary by Otto Mazal (published in a reduced two-volume facsimile by Akademische Druck-und Verlagsanstalt)
- Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**, 4008
- Erllich HA, Gelfand D, Sninsky JJ (1991) Recent advances in the polymerase chain reaction. *Science* **252**, 1643-1651
- Fischer H (1929) *Mittelalterliche Pflanzenkunde*, Verlag der Münchner Drucke, München, pp 257-326
- Gilmore MR (1919) *Uses of Plants by the Indians of the Missouri River Region*. University of Nebraska Press, Lincoln, pp 1-27 (reprinted from the 33rd Annual Report of the Bureau of American Ethnology, Washington, 1977)
- Garkava-G L, Brantestam AK, Sehic J, Nybom H (2008) Molecular characterization of indigenous Swedish apple cultivars based on SSR and S-allele analysis. *Hereditas* **145**, 99-112
- Gugerli F, Parducci L, Petit RJ (2005) Ancient plant DNA: Review and prospects. *New Phytologist* **166**, 409-418
- Gusmini G, Wehner TC, Jarret RL (2004) Inheritance of Egusi seed type in watermelon. *Journal of Heredity* **95**, 268-270
- Gyulai G, Humphreys M, Lágler R, Szabó Z, Tóth Z, Bittsánszky A, Gyulai F, Heszky L (2006) Seed remains of common millet from the 4th (Mongolia) and 15th (Hungary) centuries; AFLP, SSR, and mtDNA sequence recoveries. *Seed Science Research* **16**, 179-191
- Gyulai G, Szabó Z, Törjék O (2011a) Genetical and morphological reconstruction of medieval melon (*Cucumis melo*) (15th CENT). In: Gyulai G (Ed) *Plant Archaeogenetics*, Nova Science Publisher Inc., New York, USA, pp 89-104
- Gyulai G, Tóth Z, Bittsánszky A (2011b) Medieval *Citrullus* DNAs – unlocking domestication events (13th and 15th CENT). In: Gyulai G (Ed) *Plant Archaeogenetics*, Nova Science Publisher Inc., New York, USA, pp 69-88
- Gyulai G, MO Humphreys, R Lágler (2011c) aDNA extraction and molecular analysis from seed remains of ancient common millets (*Panicum miliaceum*) (4th and 15th CENT). In: Gyulai G (Ed) *Plant Archaeogenetics*, Nova Science Publisher Inc., New York, USA, pp 51-68
- Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series No. 41*, 95-98
- Hartyányi BP, Nováki Gy (1975) Samen- und fruchtfunde in Ungarn von der neusteeinzeit bis zum 18. Jahrhundert. *Agrártört Szemle, Budapest* **17**, 1-88
- Hepper FN (1990) *Pharaoh's Flowers. The Botanical Treasures of Tutankhamun*, Royal Botanic Gardens, Kew, London, 80 pp
- Huang XQ, Börner A, Röder MS, Ganai MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theoretical and Applied Genetics* **105**, 699-707
- Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314
- Janick J, Paris HS (2006) The Cucurbit Images (1515–1518) of the Villa Farnesina, Rome. *Annals of Botany* **97**, 165-176
- Janick J, Paris HS, Parrish DC (2007) The Cucurbits of Mediterranean Antiquity: Identification of Taxa from Ancient Images and Descriptions. *Annals of Botany* **100**, 1441-1457
- Jarret RL, Merrick LC, Holms T, Evans J, Aradhya MK (1997) Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai). *Genome* **43**, 433-441
- Kalendar R, D Lee, AH Schulman (2009) FastPCR Software for PCR primer and probe design and repeat search. *Genes, Genomes and Genomics* **3**, 1-14. [http://primerdigital.com/fastpcr.html]
- Kanda T (1951) The inheritance of seed-coat colouring in the watermelon. *Japanese Journal of Genetics* **7**, 30-48
- Katzir N, Danin-Poleg Y, Tzuri G, Karchi Z, Lavi U, Cregan PB (1996) Length polymorphism and homologies of microsatellites in several *Cucurbitaceae* species. *Theoretical and Applied Genetics* **93**, 1282-1290
- Kroll H (2000) Literature on archaeological remains of cultivated plants (1998/1999) *Vegetation History and Archaeobotany* **9**, 31-68
- Lágler R, Gyulai G, Humphreys M, Szabó Z, Horváth L, Bittsánszky A, Kiss J, Holly J, Heszky L (2005) Morphological and molecular analysis of common millet (*P. miliaceum*) cultivars compared to an aDNA sample from the 15th century (Hungary). *Euphytica* **146**, 77-85
- Linnaeus C (1753) *Species Plantarum*. 2 vols. Salvius, Stockholm. Facsimile edition (1957–1959), Ray Society, London, 176 pp
- Mannich L (1989) *An Ancient Egyptian Herbal*, University of Texas Press, Austin, USA, 176 pp
- Mansour A, Teixeira da Silva JA, Gyulai G (2009) Assessment of molecular (dis)similarity: The role of multiple sequence alignment (MSA) programs in biological research. *Genes, Genomes and Genomics* **3**, 23-30
- Messier W, Li S-H, Stewart C-B (1996) The birth of microsatellites. *Nature* **381**, 483
- McGregor CE, Sandlin K, Prothro J, Heesacker AF, Khalilian N, Okashah R, Xiang W, Bachlava E, Caldwell D, Seymour D, White V, Chan E, Tolla G, White C, Safran D, Graham E, Knapp SJ (2011) Comparative mapping within *Citrullus lanatus*. In: *International Plant and Animal Genomes XIX Conference*, January 15-19, Town & Country Convention Center, San Diego, CA, USA, W187
- Orti G, Pearce DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Sciences USA* **94**, 10745-10749
- Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**, 645-679
- Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on presence of inverted repeat. *Journal of Molecular Evolution* **55**, 501-508
- Raniello R, Procazzini G (2002) Ancient DNA in the seagrass *Posidonia oceanica*. *Marine Ecology – Progress Series* **227**, 269-273
- Renner SS, Schaefer H, Kocyan A (2007) Phylogenetics of *Cucumis* (Cucurbitaceae): Cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). *BMC Evolutionary Biology* **7**, 58
- Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganai MW (1998) A microsatellite map of wheat. *Genetics* **149**, 2007-2023
- Robinson RW, Decker-Walters DS (1997) *Cucurbits*. CAB International, Cambridge, 240 pp
- Saltonstall K (2003) Microsatellite variation within and among North American lineages of *Phragmites australis*. *Molecular Ecology* **12**, 1689-1702
- Schermann Sz (1966) Magismeret (Seed morphology), I, II. Akadémiai Kiadó, Budapest, pp 209, 861
- Schlumbaum A, Tensen M, Jaenicke-Despres V (2008) Ancient plant DNA in archaeobotany. *Vegetation History and Archaeobotany* **17**, 233-244
- Schulman AH (2007) Molecular markers to assess genetic diversity. *Euphytica*

158, 313-321

- Szabó Z, Gyulai G, Humphreys M, Horváth L, Bittsánszky A, Lágler R, Heszky L** (2005) Genetic variation of melon (*C. melo*) compared to an extinct landrace from the Middle Ages (Hungary) I. rDNA, SSR and SNP analysis of 47 cultivars. *Euphytica* **146**, 87-94
- Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresowitch S, Ziegler J** (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): Comparison with data from RFLPs and pedigree. *Theoretical and Applied Genetics* **95**, 163-173
- Tamura K, Dudley J, Nei M, Kumar S** (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599
- Taylor JS, Durkin JMH, Breden F** (1999) The death of a microsatellite: A phylogenetic perspective on microsatellite interruptions. *Molecular Biology and Evolution* **16**, 567-572
- Threadgold J, Brown TE** (2003) Degradation of DNA in artificially charred wheat seeds. *Journal of Archaeological Sciences* **30**, 1067-1076
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680
- Tóth G, Gáspári Z, Jurka J** (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**, 967-981
- Tóth Z, Gyulai G, Szabó Z, Horváth L, Heszky L** (2007) Watermelon (*Citrullus l. lanatus*) production in Hungary from the Middle Ages (13th CENT). *Hungarian Agricultural Research* **2007/4**, 14-19
- Vartavan C de, Amorós AV** (1997) *Codex of ancient Egyptian plant remains*. Codex des restes végétaux de l'Égypte ancienne. London, 401 pp
- 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**, 213-217
- Warid WA** (1995) Vegetable species known to the ancient Egyptians. *Acta Horticulturae* **391**, 273-290
- Weinstein J** (2008) A postgenomic visual icon. *Science* **319**, 1772-1773
- Yang H** (1997) Ancient DNA from Pleistocene fossils: Preservation, recovery, and utility of ancient genetic information for quaternary research. *Quaternary Science Reviews* **16**, 1145-1161