1 **A 3500-year-old leaf from a Pharaonic tomb reveals that New Kingdom Egyptians were**
2 **cultivating domesticated watermelon**
3
4 **Authors**: Susanne S. Renner[1]*, Oscar A. Pérez-Escobar[2]*, Martina V. Silber[1], Mark Nesbitt[2],
5 Michaela Preick[3], Michael Hofreiter[3], Guillaume Chomicki[4,5]*
6
7 **Affiliations**:
8 [1]Department of Biology, Institute for Systematic Botany and Mycology, University of
9 Munich (LMU), Menzinger Str. 67, 80638 Munich, Germany. [2]Royal Botanic Gardens, Kew,
10 Richmond, TW9 3AE, UK. [3]Faculty of Mathematics and Natural Sciences, Institute for
11 Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany. [4]Department of
12 Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK; and [5]The
13 Queen's College, High Street, Oxford OX1 4AW, UK.
14
15 *These authors contributed equally to the study.
16
17 **Correspondence**: renner@lmu.de (SSR) or guillaume.chomicki@gmail.com (GC).
18
19 **Abstract**
20 Domestication of the watermelon (*Citrullus lanatus*) has alternatively been placed in South
21 Africa, the Nile valley, or more recently West Africa, with the oldest archeological evidence
22 coming from Libya and Egypt. The geographic origin and domestication of watermelons has
23 therefore remained unclear. Using extensive nuclear and plastid genomic data from a 3,560-
24 year-old *Citrullus* leaf from a mummy's sarcophagus and skimmed genomes for
25 representatives of the seven extant species of *Citrullus*, we show that modern cultivars and the
26 ancient plant uniquely share mutations in a lycopene metabolism gene (LYCB) affecting pulp
27 color and a stop codon in a transcription factor regulating bitter cucurbitacin compounds. This
28 implies that the plant we sequenced had red-fleshed and sweet fruits and that New Kingdom
29 Egyptians were cultivating domesticated watermelons. The genomic data also identify extant
30 Sudanese watermelons with white, sweet pulp as the closest relatives of domesticated
31 watermelons.
32
33 Archaeogenomics | plant domestication | ancient plant genome | Africa | New Kingdom Egypt
34 | Pharaonic tomb
35
36 **Significance statement [120 words]**
37 With some 197.8 million tons in 2017, watermelon, *Citrullus lanatus*, is among the World's
38 most important crops, yet its area of origin and domestication have remained unclear, with
39 competing hypotheses favoring South Africa, West Africa, Central Africa, or the Nile valley.
40 We generated extensive nuclear and plastid genomic data from a 3500-year-old leaf from a
41 Pharaonic sarcophagus and performed genome skimming for representatives of all other
42 *Citrullus* species to compare key genes involved in fruit bitterness and color. White-fleshed,
43 non-bitter melons from southern Sudan are the closest relatives of domesticated watermelon,
44 and the ancient genome shares unique alleles with a red-fleshed, non-bitter domesticated form
45 (but no wild forms), implying that 18th Dynasty Egyptians were cultivating domesticated
46 watermelon by 3500 years ago.

Introduction

48 Among the World's most important cucurbit fruit crops is the watermelon (*Citrullus lanatus*).
49 With a production of 197.8 million tons in 2017, watermelon accounts for over a third of the
50 global tropical fruit production (*1*). Its geographic origin and time of domestication, however,
51 remain contentious (*2-7*, *supplementary text*). Competing hypotheses have favored South
52 Africa (*7, 8*), Central Africa (*2*) or the Nile valley (*9-15*). More recently, molecular data have
53 pinpointed West Africa (*16, 17*) as the region of the crop's origin, yet archaeological evidence
54 is restricted to Egypt and Libya. Besides *C. lanatus*, the genus *Citrullus* contains six other
55 species, of which four (*C. amarus*, *C. ecirrhosus*, *C. naudinianus*, and *C. rehmii*) occur
56 mostly in the Namib-Kalahari region, one (*C. mucosospermus*) in West Africa (Benin, Ghana,
57 and Nigeria), and one (*C. colocynthis*) from northern Africa to west India (**Fig. 1A**) and
58 naturalized also in Australia (*6, 17*). All have white pulp that cannot be eaten raw due to the
59 presence of terpene bitter compounds called cucurbitacins. Only fruits of *C. mucosospermus*
60 are sometimes not bitter, but instead bland-tasting (*18*). *Citrullus lanatus* itself is cultivated
61 worldwide, but wild progenitors remained elusive despite the resequencing of 20 accessions
62 from germplasm collections (*16*). Putative wild forms with small fruits and sweet white pulp
63 have been reported since the 19th century from the Kordofan region, a former province of
64 Sudan, bordering North and South Darfur (*6, 9-15*) (**Fig. 1A, B**), but their genetic affinity to
65 domesticated and other watermelons has not been tested. Moreover, no wild species of
66 *Citrullus* with sweet pulp have been found in South Africa.
67   That the Egyptians ate domesticated watermelons is suggested by at least two wall
68 paintings (**Fig. 1C-D**; *Supplementary text*) as well as *Citrullus* leaves from a tomb at Deir el-
69 Bahari near Luxor (Theban Tomb number 320) first identified by the German archaeobotanist
70 Georg August Schweinfurth in the 19th century (*10, 11,19-21*) (**Fig. 1B**). The leaves were
71 laying on the mummy of "Unknown Man C" in a re-used coffin inscribed for a priest
72 Nebseni. On the evidence of embalming techniques and the coffin type the leaves may date to
73 the early 18th Dynasty (c. 3493-3470 BP; see *Supplementary text*). Finding leaves associated
74 with mummies, however, is insufficient evidence that Egyptians consumed raw, sweet
75 watermelons, since pharaonic tombs contain many kinds of plants that were not eaten as food,
76 instead used as medicine or for other purposes (*10*). The closest known relative of the modern
77 watermelon is a form from West Africa (*16, 17*) with white, usually bitter or bland pulp, and
78 seeds that are used in West African 'egusi' stews (*18*).
79   To determine whether ancient Egyptians already consumed sweet, red watermelons
80 and test whether sweet white-fleshed Sudanese *Citrullus* populations represent potential

81   watermelon progenitors, we generated skimmed genome data (~14 Gb) for all *Citrullus*

82   species, including the Unknown Man C mummy sample (~8 Gb) and accessions from North

83   and South Darfur (Supporting Information **Table S1**) hypothesized to be the wild progenitor

84   of watermelons (*6, 9-15*). The specific goals were to generate a strongly supported phylogeny,

85   probe the phylogenetic placements of the Unknown Man C mummy sample and Kordofan

86   melons, and searched bitterness and red-flesh markers in the ~3500-year-old sample.

87

88   **Results and discussion**

89   The 3500-year-old priest Unknown Man C mummy melon yielded good quality DNA with

90   relatively low rates of deamination (SI Appendix **Figs. S1, S2**). Because aDNA yields short

91   reads, we followed approaches used in other aDNA studies (*22, 23*) by mapping read data

92   against target sequences from the watermelon genome (*16*). The targets consisted of plastid

93   genes, a set of 143 nuclear genes for the phylogenomic analyses, and a set of 73 genes of

94   interest linked to fruit bitterness and color. We recovered a complete plastid genome for the

95   3500-year-old leaf (SI Appendix **Fig. S3**), and its nuclear sequences were sufficiently good so

96   that we could recover both our target genes for phylogenomics and our genes of interest

97   related to cucurbitacin and lycopene pathways. This enabled to probe the phylogenetic

98   position of the 3500-year-old *Citrullus* sample, and test whether its fruits were bitter or sweet,

99   and red or white-fleshed.

100   For the phylogenetic analyses of *Citrullus*, we assembled entire plastid genomes (121

101   genes and 33 spacers) and 143 single copy nuclear genes (**SI Appendix, dataset 1**) and

102   confirmed that nuclear genes are orthologs in reciprocal best alignment heuristic searches

103   (*Materials and Methods*, **SI Appendix, dataset 2).** Maximum Likelihood (ML) nuclear and

104   plastid phylogenies revealed the same topology (SI Appendix, **Fig. 2A; Fig. S4**), and a

105   backbone similar to an earlier phylogeny based on 11 genes that lacked any East African

106   samples (*17*). A coalescence-based tree also supported the same topology as ML analyses (SI

107   Appendix **Fig. S5**). These phylogenomic analyses revealed that the Unknown Man C mummy

108   sample is sister to modern domesticated watermelon (**Fig. 2A;** SI Appendix **Figs. S4, S5**) and

109   both together are sister to the Sudanese sweet and white-fleshed Kordofan melons (**Fig. 2A**;

110   *C. lanatus* subsp. *cordophanus*). All three form the sister clade to the West African *C.*

111   *mucosospermus* (**Fig. 2A**). These analyses revealed that the 3500-year-old Unknown Man C

112   mummy sample indeed represents a cultivated watermelon and that Kordofan melons (*C.*

113   *lanatus* subsp. *cordophanus*) are the closest known relative of the watermelon and likely

114   descendants of a progenitor population of the domesticated watermelon. All *Citrullus* species

115   are illustrated in **Fig. 2B-I**.

116   To find direct evidence of the timing of watermelon domestication, we next

117   investigated whether the 3500-year-old leaf belonged to a watermelon with bitter *vs*. sweet

118   fruits and white *vs*. red flesh. Bitterness in cucurbits depends on terpene compounds called

119   cucurbitacins. The *Bi* gene, which encodes an oxidosqualene cyclase (*OSC*) that catalyzes the

120   first committed step in cucurbitacin C biosynthesis is critical for determining bitterness (*24-*

121   *26*). Two bHLH transcription factors (*Bl* and *Bt*) regulate Cucurbitacin C biosynthesis by

122   upregulating *Bi* expression in the leaves (*Bl*) and fruits (*Bt*) directly via binding to the E-box

123   elements of the *Bi* promoter (*25, 26*). In addition to upregulating *Bi*, the watermelon *ClBl* and

124   *ClBt* upregulate most other cucurbitacin metabolism genes, including the two cytochrome

125   P450 enzymes that convert cucurbidienol in cucurbitacin precursor (*Cl890A* and *Cl890B*

126   genes in watermelon) and the acyltransferase that convert the cucurbitacin precursor in

127   cucurbitacin E (*26*). In cucumber (*Cucumis sativus*), honey melon (*Cucumis melo*), and

128   watermelon, examination of different lines with varying bitterness revealed that domestication

129   of non-bitter cucumber occurred via a nucleotide substitution leading to a premature stop

130   codon in the *Bt* gene resulting in a truncated, non-functional protein (*24-26*). A genomic

131   analysis comparing cucumber, honey melon, and watermelon (producing cucurbitacins C, B

132   and E, respectively), revealed a domestication sweep at the *Bt* locus, with the loss of

133   bitterness being due to convergent mutations at this locus (*26*).

134   We used a comparative-genomic approach to compare genes in the cucurbitacin

135   biosynthetic pathway and its regulation across *Citrullus* species and the ancient watermelon.

136   All cucurbitacin metabolic genes were conserved, with the same copy number across all

137   *Citrullus* species, including the 3500-year-old sample (**Supplementary Dataset S1**). To

138   probe whether the ~3500-year-old watermelon had bitter or sweet fruits, we compared its *Bt*

139   gene to that of the other *Citrullus* species. Most *Citrullus* species have two copies of *ClBt*,

140   with the exception of *C. rehmii* and the clade comprising *C. mucosospermus*, the Kordofan

141   melon, and the modern and ancient watermelon, which have a single copy (**Fig. 2J**). The

142   second copy appears non-functional (**Fig. 2J, Supplementary Dataset S1**). The analyses

143   further revealed that the 3500-year-old mummy sample has a single nucleotide substitution

144   leading to a premature stop codon in the *Bt* gene (**Fig. 2J,** SI Appendix **Fig. S6A and dataset**

145   **1**), implying that it had non-bitter fruits. The Kordofan melons (subsp. *cordophanus*) also

146   have the premature stop codon, matching their consistently sweet pulp (*13*). The fruits of the

147   West African *C. mucosospermus* are usually bitter, but over 20% have acid, plain, or sweet

148   pulp (*18: Table 5*), matching our finding of a stop coding in its *Bt* gene. The particular plant

149   that we sequenced, *E.G. Achigan-Dako 809AA603* (SI Appendix, Table S1), indeed had non-

150   bitter pulp (*18*; Achigan-Dako, pers. comm. SSR, 22 Nov. 2018). These results suggest a

151   scenario wherein loss of pulp bitterness was a pre-adaptation for the domestication of

152   watermelon, implying that early farmers probably took into cultivation non-bitter plants from

153   the wild. Thus, our comparative analyses of *ClBt* genes across all *Citrullus* species suggests

154   that the loss of bitterness resulted from initial selection of a non-bitter wild from, rather than

155   selection occurring during cultivation as may have been the case in cucumber (*25, 26*).

156   The red pulp color in modern watermelon is due to lycopene accumulation, likely by

157   blocking the conversion of lycopene into β-carotene, a step mediated by the enzyme LCYB

158   (lycopene β-cyclase) (*16, 27, 28*). LCYB is encoded by the *lcyb* gene, and watermelon

159   accessions with red flesh are characterized by having a unique A -> C substitution in the *lcyb*

160   gene on chromosome 4, resulting in a Valine instead of a Phenylalanine at position 226 (*27*).

161   To determine whether the 3500-old watermelon was red-fleshed, we sought all lycopene

162   metabolic genes across our *Citrullus* genomes. This revealed a high conservation of these

163   metabolic genes, with the same copy number for all genes in the lycopene pathway across all

164   species (**Supplementary Dataset S1**). We found that the 3500-year-old mummy sample and

165   the modern watermelon share the key substitution in LCYB linked to lycopene accumulation

166   (**Fig. 2K, SI Appendix Fig. S6B and SI dataset 1**), while all other *Citrullus* species

167   including the Kordofan watermelons lack this mutation and have white to greenish pulp. This

168   reveals that New Kingdom Egyptians were cultivating a red, sweet watermelon 3500 years

169   ago.

170

171   **Conclusion**

172   Our study reveals that New Kingdom Egyptians cultivated a red, sweet watermelon at least

173   3500 years ago. It also shows that likely progenitor populations still exist in the upper Nile

174   valley, in what is now the Darfur region of Sudan. The sweet, red watermelon may have been

175   domesticated there and the use of this plant then could have spread northward along the Nile

176   valley. Our finding that the 3500-year-old leaf belonged to a red-fleshed and sweet

177   watermelon is consistent with two Egyptian wall paintings dating back to 4350 BP that have

178   been interpreted as showing watermelons (**Fig. 1C-E**; *Supplementary text*). They show

179   elongate fruits with the characteristic white stripes of watermelons and were served on trays,

180   suggesting they were eaten raw. Watermelon may have been domesticated even earlier, given

181   a watermelon seed found at Uan Muhuggiag in the Tadrart Acacus area in Libya that is

182   between 5000 and 8650 years-old (*29*). Uan Muhuggiag lies 2240 km west of the 18th

183   Dynasty Theban necropolis of Deir el-Bahari where the genomic data here produced

184   originated. Recent work on barley (*22*) and maize (*23*), and the present study on the

185   watermelons, exemplifies the power of integrating collections-based phylogenomics with

186   archaeological data and comparative genomics to resolve the domestication history of modern

187   crops. The search for wild watermelon germplasm for engineering more disease-resistant

188   varieties now should concentrate on the Darfur region of Sudan.

189

190   **Methods**

191   **Plant taxon sampling.** Material sequenced for this study is listed in **Supporting Information**

192   **Table S1**, which also reports GenBank accession number for the skimmed genomes. *Citrullus*

193   taxonomy follows Chomicki & Renner (*17*) and Renner et al. (*6*). We sampled all seven

194   *Citrullus* species using herbarium-verified samples all of which are linked to permanently

195   kept vouchers. This is essential, since very few –if any– germplasm accessions (for instance,

196   *Citrullus* accessions in the National Plant Germplasm System Germplasm Resources

197   Information Network [GRIN; http://www.ars-grin.gov/npgs]) are linked to herbarium voucher

198   specimens, leading to problems of unverifiability (see for instance Renner et al. (ref. *6*)).

199

200   **Ancient DNA extraction.** To document the leaf fragment received from Kew's economic

201   plant collection, we first took high resolution photographs (**Fig. 1B**). A small piece of a leaflet

202   (<1 cm^2, 5.8 mg) was ground with a Retsch mill (Retsch MM 400). DNA extraction was

203   performed following Wales et al. (*31*) (modified by Pedersen et al. ref. *32*) and Dabney at al.

204   (*33*). For the digestion treatment, a lysation buffer containing 0.5 % (w/v) N-Lauroylsarcosine

205   (Sigma Aldrich L9150-50G), 50 mM Tris-HCl (Thermo Fisher Scientific 15568025), 20 mM

206   EDTA (VWR E177-500MLDB) 150 mM NaCl (Thermo Fisher Scientific AM9760G), 3.3 %

207   2-Mercaptoethanol (Sigma Aldrich 63689-25ML-F), 50 mM DL-Dithiothreitol (Sigma

208   Aldrich D9779-250MG) and 0.25 mg/mL Proteinase K (Promega V3021) was applied to the

209   leaflet powder as described in Wales et al. (*31*). DNA purification was performed according

210   to Dabney at al. (*33*) with reduced centrifugation speed (450 x *g*) as described in Basler et al.

211   (*34*).

212

213   **Ancient DNA library preparation and sequencing.** DNA extract was converted to an

214   Illumina sequencing library using the single-stranded approach described in Korlević et al.

215   (*35*). The protocol included the treatment with Uracil-DNA-Glycolase (New England Biolabs

216   M0279) to remove Uracil residues and Endonuclease VIII (New England Biolabs M0299) to

217   cleave DNA strands at abasic sites. 2.5 U/µl of Circligase II (Biozym 131406) was used for

218   the fill-in reaction and carried out overnight. A quantitative PCR was performed on a

219   PikoReal 96 Real-Time PCR machine (Thermo Fisher Scientific TCR0096) using 0.2 % of

220   the unamplified library following this thermal profile: 10 min initial denaturation step at 95

221   °C, followed by 40 cycles of: 15 s at 95 °C, 30 s at 60 °C, and 1 min at 72 °C. The

222   quantitative PCR reaction mix contained a final volume of 10 µL: 1 µL of diluted library, 1 x

223   SYBR Green qPCR Master Mix (Applied Biosystems 4309155), 0.5 µM of each primer IS7

224   and IS8. Three replicates of each library were used. Indexing PCR was performed by the
225   appropriate number of cycles adding 8 bp indices to 5' and 3' adapters. PCR and final
226   concentrations are the same as described (*36*), but with a final volume of 80 µL using 20
227   µL of template. DNA sequencing was performed on an Illumina NextSeq 500 sequencing
228   platform, using the 500/550 High Output v2 kit (75 cycles, Illumina FC-404-2005), with a
229   custom read-1 (*36*) and a custom index-2 (*37*) sequencing primer. Extraction and library
230   building was performed in the ancient DNA facility of the University in Potsdam; negative
231   controls were included in all steps.
232
233   **Genome skimming of *Citrullus*.** We extracted genomic DNA from all *Citrullus* samples
234   except the 3500-year-old sample using the Quiagen DNeasy Plant kit, following the
235   manufacturer's protocol. Genome skimming was performed by the company Genewiz (New
236   Jersay, USA) in Germany. Paired (2x150 bp) DNA libraries were sequenced on an Illumina
237   HiSeq platform and 350 GB of sequence data were produced. An average of 100 million reads
238   were produced for each sample.
239
240   **Plastid data processing, assembly and annotation.** The Illumina raw reads were quality
241   filtered using Trim Galore v.0.4 (*38*), discarding sequences with averaged phred33 score
242   below 20. Pre- and post-trimming read quality was assessed using FASTQC v.0.1 (*39*).
243   Whole plastid genomes were assembled by blasting quality-filtered reads against a fully
244   annotated reference plastid genome of *C. lanatus* (GenBank accession NC032008) using the
245   tool *blastn* and a maximum number of target sequences of one. We obtained the best scoring
246   blast hits from the blastn output by filtering all hits by bitscore and query coverage values of
247   100 and 80%, respectively. Best scoring reads were mapped against the reference plastid
248   genome using the Geneious mapping tool (*40*) and the following parameters: five iterations,
249   minimum mapping quality of 30, maximum gap size of 100, and maximum mismatches per
250   read of 40%. Consensus genome sequences were assembled following a modified statistical
251   base-calling approach of Li et al. (*41*), i.e. minimum depth coverage of 10, and bases
252   matching at least 50% of the reference sequence. Plastid genomes were fully annotated by
253   transferring intron, exon, and spacer annotation from the reference plastid genome to the *de-
254   novo* assembled plastids.
255
256   **Plastid and nuclear phylogenomics of *Citrullus*.** De-novo plastid genomes were aligned
257   with Mauve using a progressive algorithm and assuming collinearity (*42*). The resulting
258   ~150,000 bp alignment was subjected to Maximum Likelihood (ML) tree inference in
259   RAxML v8.0 (*43*), using the GTR substitution model, 25 gamma categories, and 1,000
260   bootstrap replicates. Nuclear phylogenomic inferences relied on a selection of 143 out of 353
261   nuclear CDS known to be present in low copy numbers across angiosperms (*44*), all
262   assembled from the trimmed read data using the python package Hybpiper (*45*). The 143
263   genes (**SI dataset 1**) were selected based on the proportions of missing data (i.e. alignments
264   including more than five sequences with 50% of their length missing and less than five taxa
265   were excluded), and phylogenetic informativeness (PI) profiles (i.e. gene alignments with
266   disproportionate substitution rates were excluded) following Townsend (*46*). The PI profiles
267   are provided in SI Appendix, **Fig. S7**. Hybpiper-achieved gene assembly by mapping the
268   trimmed reads against the set of targeted CDSs using the Burrows-Wheeler alignment

269 software (*47*) with a coverage threshold of 8x. Contig assembly relied on the software
270 SPAdes (*48*), relying on default settings. Whenever multiple copies of the target genes are
271 present in the genome, HybPiper raises paralog warnings during the contig assembly step.
272 Here, no paralog warnings were issued for any of the 143 low copy nuclear genes. Because
273 the Illumina sequencing of the ancient *C. lanatus* sample produced very short reads (max
274 length of 75 bp; SI Appendix, **Fig. S1**), we carried out low-copy-gene assembly using the
275 Geneious mapping algorithm, using customized settings. Gene mappings and assemblies were
276 all checked by eye and carefully curated. To determine whether the selected set of low copy
277 nuclear genes are in fact orthologs, we carry out orthology analyses in the software
278 ProteinOrtho (*49*), which relies on a reciprocal best alignment heuristic approach. No
279 paralogues were detected by ProteinOrtho (**SI Appendix, dataset 2**).
280   Phylogenomic analyses of the unlinked gene alignments matrix relied on ML and
281 coalescence-based methods. We generated ML trees for every nuclear gene alignment as well
282 as the concatenated nuclear matrix using the same settings as employed for plastid tree
283 inference. Species tree analyses relied on the coalescence-based method ASTRAL III (*50*),
284 using as input the annotated consensus gene trees derived from RAxML, with branches
285 having Likelihood Bootstrap Support (LBS) < 10 collapsed. Gene tree conflict was visualized
286 by plotting the local posterior probabilities derived from ASTRAL as pie diagrams at nodes.
287 The final normalized quartet score is: 0.73.
288

289 **In-silico capture of genes relating to cucurbitacin biosynthesis and its regulation,**
290 **lycopene synthesis and sugar metabolism and its regulation in *Citrullus* species**
291   We followed the same read-data filtering and trimming procedure as used for the
292 plastid data. Nucleotide deamination analysis and filtering of the ancient watermelon reads
293 were carried out by the pipeline PALEOMIX (*51*) and using as a reference a genome
294 assembly of a modern accession of *C. lanatus* (No. 97103, available at
295 http://cucurbitgenomics.org/). The pipeline relied on the software Bowtie2 (*52*) for read
296 mapping and on mapDamage2 (*53*) for nucleotide deamination analysis. Genes involved in
297 the cucurbitacin and lycopene synthesis were also captured in-silico following the same
298 approach employed to data-mine low copy nuclear genes in the pipeline HybPiper. To attain
299 this, we build a customized set of target genes obtained from Guo et al. (*16*) and Zhou et al.
300 (*26*) (gene IDs are provided in **SI Appendix, dataset 3**). Gene copies in the *ClBt* gene were
301 detected by HybPiper during the contig assembly step. No paralog warnings were produced
302 for the remaining genes involved in the cucurbitacin and lycopene pathways (except for the
303 *ClBt* gene for which some species had two copies– cf. *Results and discussion*).
304

314

315 **Author Contribution**

316 SSR and GC conceived the study with subsequent methodological contribution from OAPE;

317 SSR, MVS, MP gathered data; OAPE, GC, SSR analyzed data, SSR, GC, MH provided

318 funding; SSR and GC wrote the manuscript with edits from all authors.

319

320 **Additional information**

321 **Supplementary Information** accompanies this paper at https: xxx

322

323 **Competing interests:** The authors declare no competing financial interests.

324

325 **References:**

326   1. FAO, http://www.fao.org/faostat/en/#data/QC, accessed 29 Jan. 2019

327   2. Hancock JF (2012) *Plant evolution and the origin of crop species*. 2nd edn.
328   Wallingford, UK & Cambridge, MA, USA: CABI.

329   3. Dane F, Liu J. (2007) Diversity and origin of cultivated and citron type watermelon
330   (*Citrullus lanatus*). *Genet Resources Crop Evol* 54:1255–1265.

331   4. Paris HS (2015) Origin and emergence of the sweet dessert watermelon, *Citrullus*
332   *lanatus*. *Ann Bot* 116:133–148.

333   5. Paris HS (2016) Overview of the origins and history of the five major cucurbit crops:
334   issues for ancient DNA analysis or archaeological specimens. *Veg Hist Archaeobot*
335   25:405-414.

336   6. Renner SS, Sousa A, Chomicki G (2017) Chromosome numbers, Sudanese wild
337   forms, and classification of the watermelon genus *Citrullus*, with 50 names allocated
338   to seven biological species. *Taxon* 66:1393-1405.

339   7. Bailey LH (1930) Three discussions in Cucurbitaceae. *Gentes Herb* 2:175–186.

340   8. Dane F, Lang P (2004) Sequence variation at cpDNA regions of watermelon and
341   related wild species: Implications for the evolution of *Citrullus* haplotypes. *Amer J*
342   *Bot* 91:1922–1929.

343   9. Schweinfurth G (1873) Sur l'origine de quelques plantes cultivées en l'Égypte. *Bull.*
344   *de l'Inst. Egypt.* 12:200-206.

345   10. Schweinfurth G (1883) La flore de l'ancienne Egypte. *Rev Sci* 32:72-77.

346   11. Schweinfurth G (1883) The flora of ancient Egypt. *Nature* 28:109–114.

347   12. Schweinfurth G (1884) Ueber Pflanzenreste aus altaegyptischen Gra□bern. *Ber.*
348   *Deutsch. Bot. Ges*. 2:351–371.

349   13. Ter-Avanesyan DV (1966) Arbuz Kordofanskyi *Citrullus lanatus* Mansf. ssp.
350   *cordophanus* Ter-Avan. *Bot Zhurnal* 51:423–426.

351   14. Fursa TB (1972) K sistematike roda *Citrullus* Schrad. [On the taxonomy of genus
352   *Citrullus*] Schrad. Bot. Zhurn. (Moscow & Leningrad) 57:31–41.

353   15. Fursa TB, Gavrilyuk IP (1990) Phylogenetic relations of the genus *Citrullus* Schrad.
354   based on the immunochemical analysis of seed proteins. *Sborn Nauchn Trudov Prikl*
355   *Bot Genet Selekts* 133:19–26. [In Russian with English summary]

356   16. Guo S. *et al*. (2013) The draft genome of watermelon (*Citrullus lanatus*) and
357   resequencing of 20 diverse accessions. *Nat Genet* 45:51–58.

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

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

19. Maspero G (1889) *Les momies royales de Deir El-Bahari.* Collection XIX.

20. Keimer L. (1924) *Die Gartenpflanzen im alten Ägypten.* Vol. I. Hamburg-Berlin: Hoffmann und Campe Verlag.

21. Germer R (1988) Katalog der Altägyptischen Pflanzenreste der Berliner Museen. Ägyptologische Abhandlungen 47. Wiesbaden: Otto Harrassowitz.

22. Mascher M *et al.* (2016) Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat Genet* 48:1089-1093.

23. Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, Wales N. (2016) Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* **26**, 3195-3201.

24. Qi J *et al.* (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat Genet* 45:1510–1515.

25. Shang Y *et al.* (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346:1084-1088.

26. Zhou Y *et al.* (2016) Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat Plants* 2:16183.

27. Bang H, Kim S, Leskovar D, King S (2007) Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β-cyclase (LCYB) gene. *Mol Breed* 20:63-72.

28. Grassi S *et al.* (2013) Comparative genomics reveals candidate carotenoid pathway regulators of ripening watermelon fruit. *BMC Genomics* 14:781.

29. Wasylikowa K, van der Veen M (2004) An archaeobotanical contribution to the history of watermelon, *Citrullus lanatus* (Thunb.) Mats. & Nakai (syn. *C. vulgaris* Schrad.). *Veget Hist Archaeobot* 13:213–217.

30. Naville E (1912) Papyrus funéraires de la XXI dynastie. Vol. 1. Le papyrus hiéroglyphique de Kamara, le papyrus hiératique de Neskhonsou au Musée de Caire. Paris, Ernest Leroux.

31. Wales N, Andersen K, Cappellini E, Ávila-Arcos MC, Gilbert PT (2014) Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PloS ONE* 9:e86827.

32. Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, Andersson R. (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **24**, 454-466.

33. Dabney J *et al.* (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort fragments. *Proc Natl Acad Sci USA* 110:15758-15763.

403   34. Basler N, Xenikoudakis G, Westbury MV, Song L, Sheng G, Barlow A (2017)
404   Reduction of the contaminant fraction of DNA obtained from an ancient giant panda
405   bone. *BMC Res Notes* 10:754.
406   35. Korlević P *et al.* (2015) Reducing microbial and human contamination in DNA
407   extractions from ancient bones and teeth. *Biotechniques* 59:87-93.
408   36. Gansauge M-T, Meyer M (2013) Single-stranded DNA library preparation for the
409   sequencing of ancient or damage DNA. *Nat Protoc* 8:737-748.
410   37. Paijmans JLA, Baleka S, Henneberger K, Barlow A (2017) Sequencing single-
411   stranded libraries on the Illumina NextSeq 500 platform. arXiv:1711.11004v1.
412   38. Krueger F Trim Galore v.0.4 (2015) A wrapper tool around Cutadapt and FastQC to
413   consistently apply quality and adapter trimming to FastQ files.
414   39. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.
415   40. Kearse M *et al.* (2012) Geneious Basic: an integrated and extendable desktop
416   software platform for the organization and analysis of sequence data. *Bioinformatics*
417   28:1647-1649.
418   41. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling
419   variants using mapping quality scores. *Genome Res* 18:078212.
420   42. Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of
421   conserved genomic sequence with rearrangements. *Genome Res* 14:1394-1403.
422   43. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-
423   analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
424   44. Johnson M, L. Pokorny L, Dodsworth S, Botigue LR, Cowan RS, Devault A,
425   Eiserhardt WL, Epitawalage N, Forest F, Kim JT, Leebens-Mack JH (2019) A
426   universal probe set for targeted sequencing of 353 nuclear genes from any flowering
427   plant designed using k-medoids clustering. *Syst Biol* in press; bioRxiv, p.361618.
428   45. Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJ,
429   Wickett NJ (2016) HybPiper: Extracting coding sequence and introns for
430   phylogenetics from high-throughput sequencing reads using target enrichment. *Appl*
431   *Plant Sci* 4:1600016.
432   46. Townsend JP (2007) Profiling phylogenetic informativeness. *Syst Biol* 56:222-231.
433   47. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler
434   transform. *Bioinformatics* 25:1754-1760.
435   48. Bankevich A *et al.* (2012) SPAdes: a new genome assembly algorithm and its
436   applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
437   49. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ (2011)
438   Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC Bioinformatics*
439   12:124.
440   50. Zhang C, Rabiee M, Sayyari E, Mirarab S (2018) ASTRAL-III: polynomial time
441   species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*
442   19:153.
443   51. Schubert M *et al.* (2014) Characterization of ancient and modern genomes by SNP
444   detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat*
445   *protoc* 9:1056.
446   52. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat*
447   *Methods* 9:357.

448   53. Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L (2013) mapDamage2. 0:
449   fast approximate Bayesian estimates of ancient DNA damage parameters.
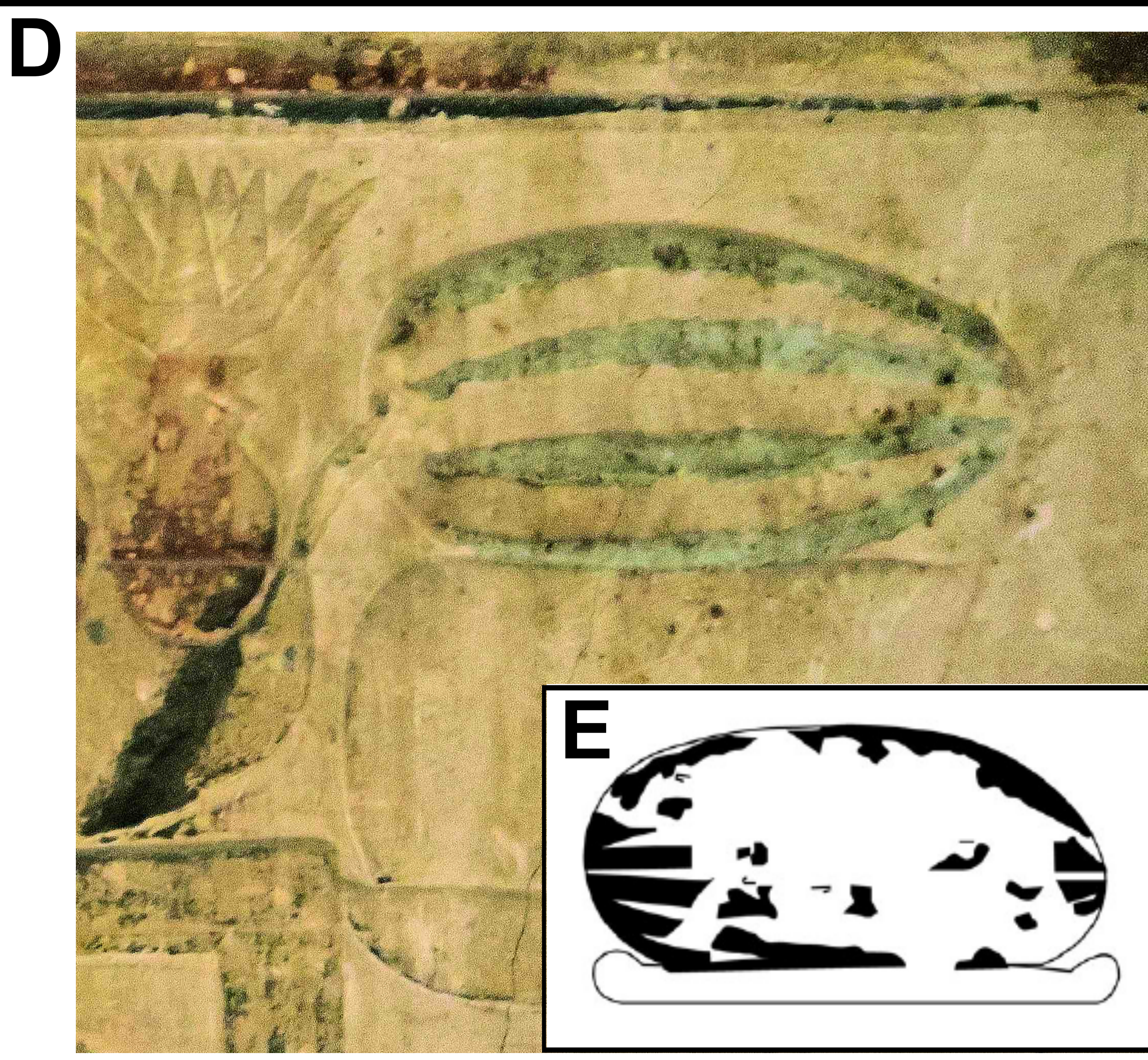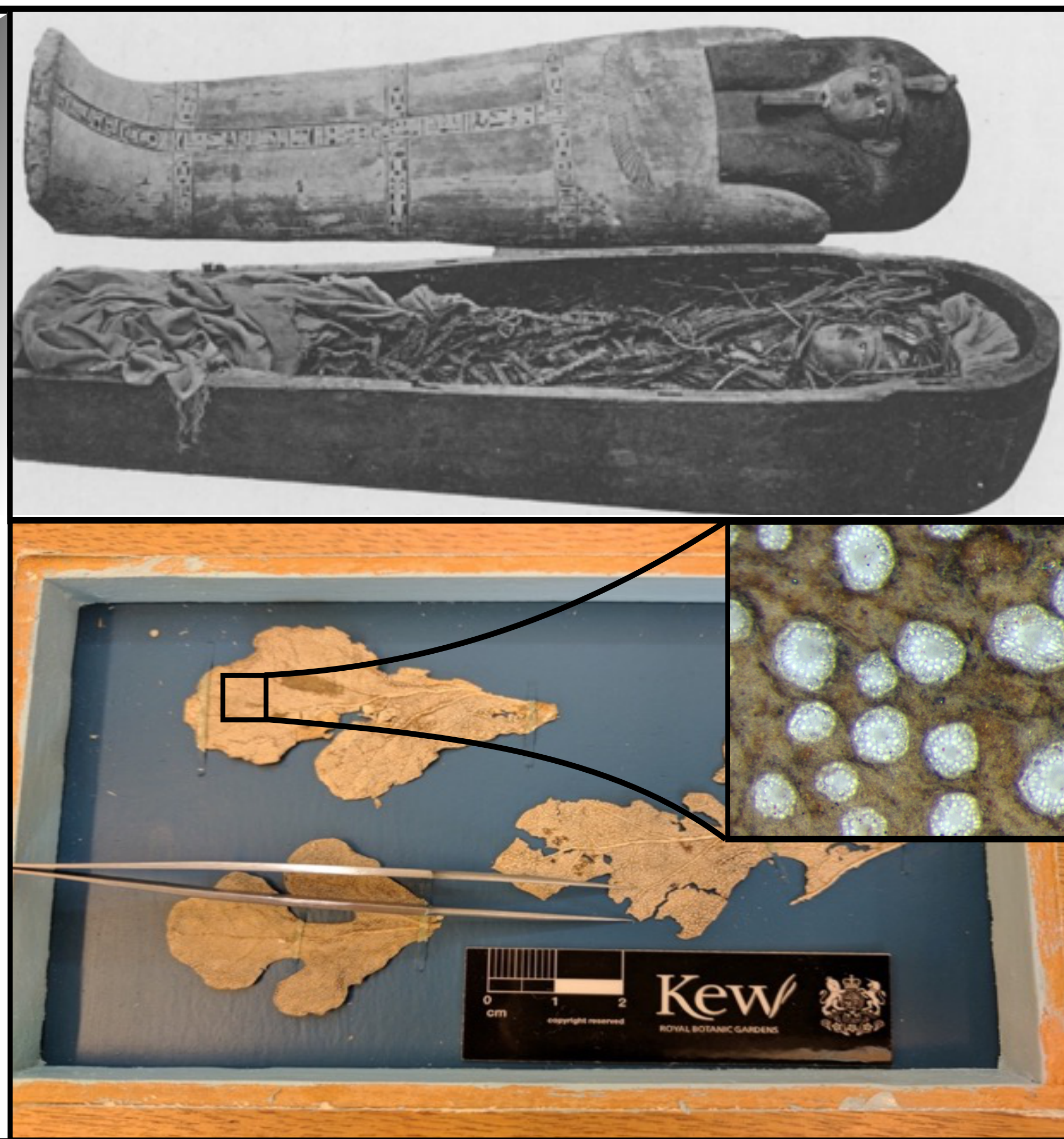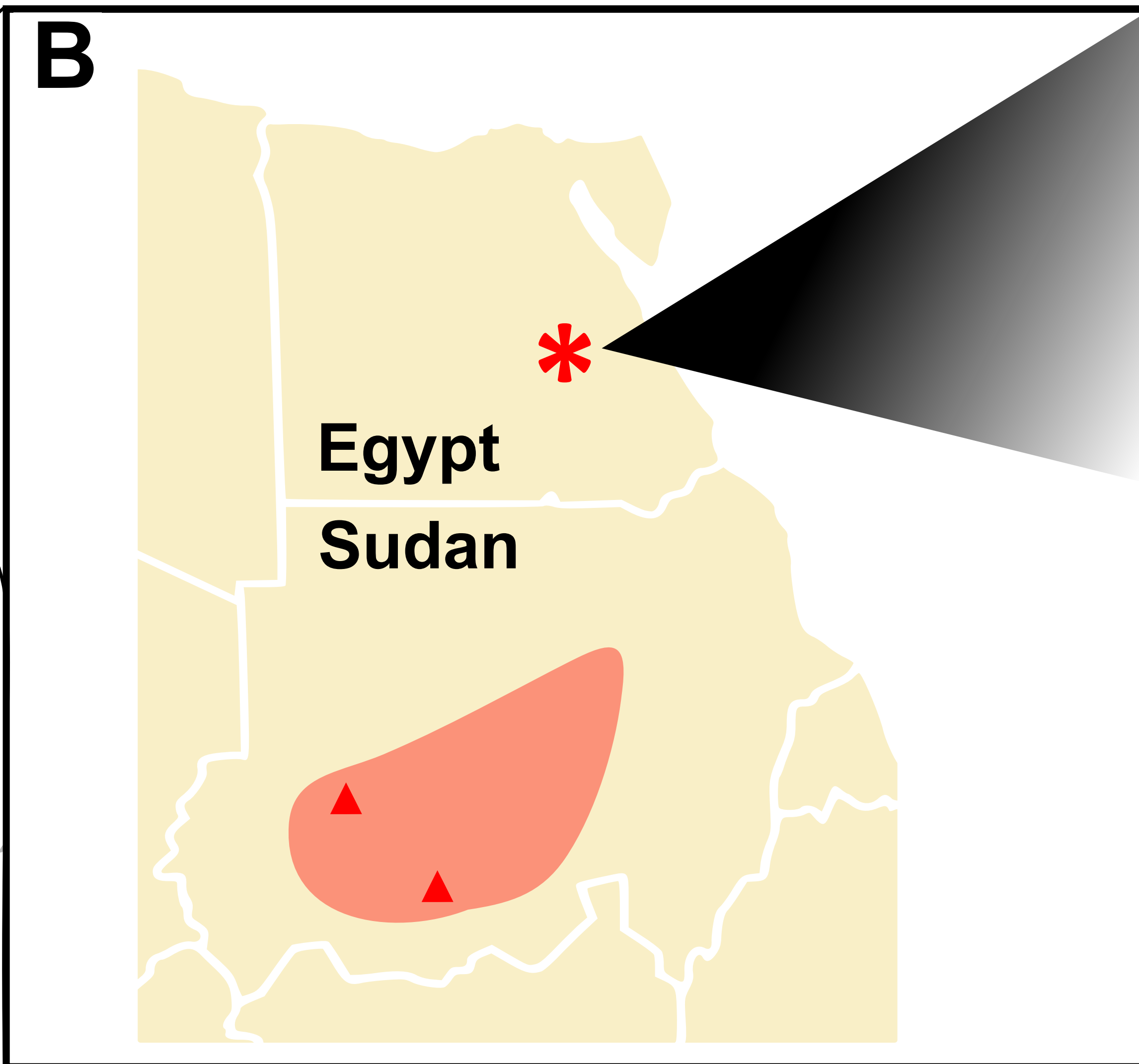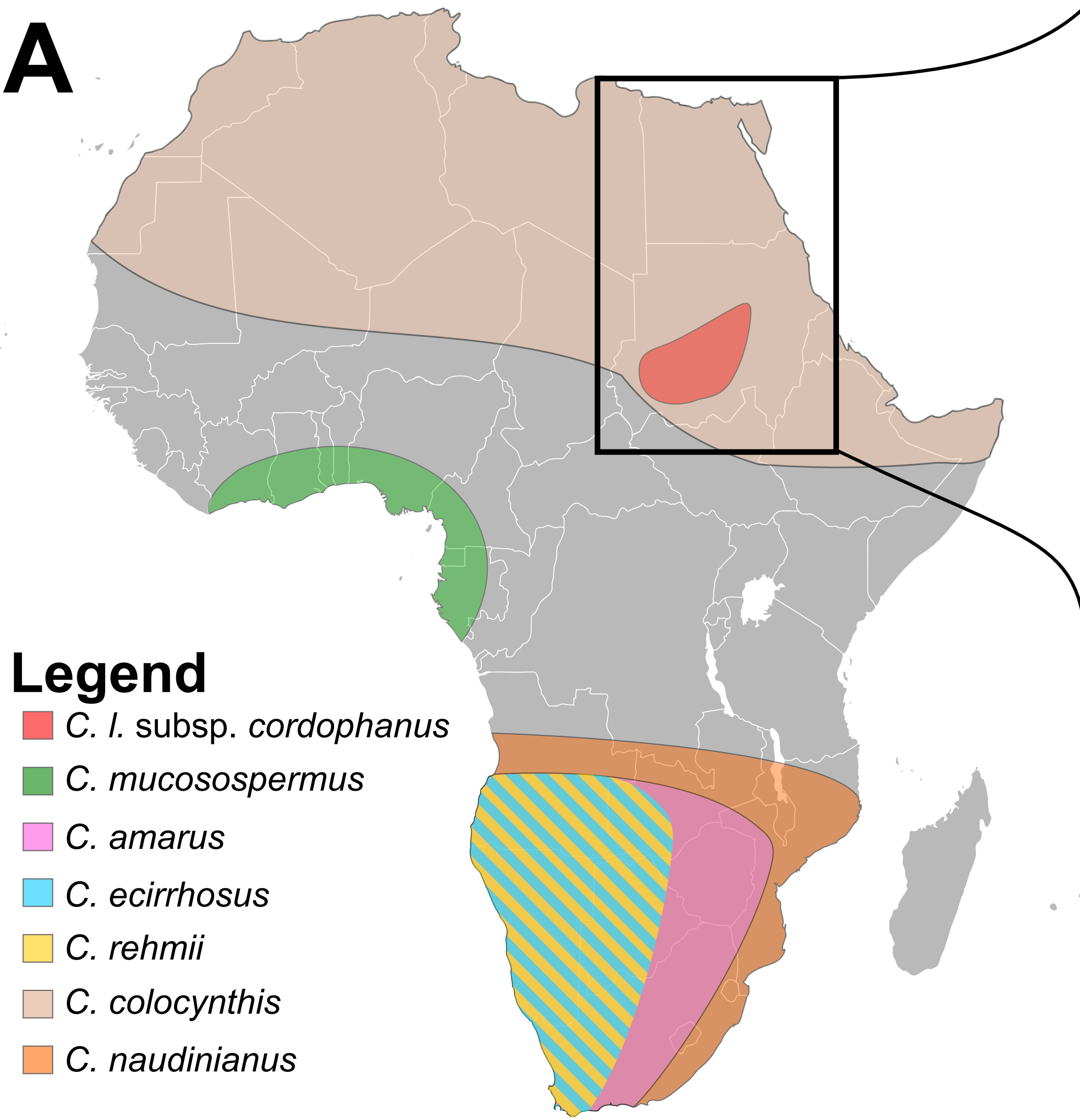450   *Bioinformatics* 29:1682-1684.

451
452

453 **Figure legends**

454 **Fig. 1 Distribution of *Citrullus* species, archaeological samples and illustrations of**
455 **watermelon in Ancient Egypt. (A)**, Map showing the distribution of *Citrullus* species in
456 Africa (*C. colocynthis* extends to West India). (**B**), An asterisk marks the location of the tomb
457 at Deir el-Bahari near Luxor, Egypt, in which a leaf fragment was found on "Unknown Man
458 C" (Cairo museum mummy CG61067) in the 19th century and deposited in the Economic
459 Plant Collection of Kew (*9-12*). Red triangles and patch mark the Kordofan region and
460 locations with wild watermelons in North and South Darfur. Top, sarcophagus of
461 Amenhotpou from Deir el-Bahari, a mummy from the same site (*19*). Bottom, the preserved
462 leaf and the sequenced fragment seen from below (*inset*, magnification 250x). (**C**), Papyrus de
463 Kamara (*30*), illustrating a *Citrullus* fruit (*inset*), interpreted as a wild watermelon by Keimer
464 (*20*). (**D-E**), Wall illustrations from Pharaonic tombs. (**D**), Tomb of Chnumhotep, Saqqara,
465 *ca*. 4450 BC. (**E),** Unspecified tomb from Meir, drawing by L. Manniche. Photo and drawing
466 courtesy of L. Manniche, June 2018.

467

468 **Fig. 2 Phylogenomics of *Citrullus* reveals the watermelon progenitor and confirm that**
469 **New Kingdom leaves belong to *Citrullus lanatus*. (A)** Phylogeny of all *Citrullus* species
470 based on 143 low copy nuclear genes, rooted on the relevant outgroups (*17*) (see also SI
471 Appendix, **Figs. S1, S2**). Values on branches are bootstrap support from 1,000 replicates
472 under the same ML model applied to nuclear (left) or plastid DNA sequences (right). (**B-I**),
473 Illustrations of all *Citrullus* species. (**B**) *Citrullus lanatus* (watermelon). (**C**) *C. lanatus* subsp.
474 *cordophanus* (Kordofan melon). (**D**) *C. mucosospermus* (egusi melon). (**E**), *C. amarus* (citron
475 melon [wild], preserving melon [domesticated]). (**F**) *C. ecirrhosus* (tendril-less melon). (**G**),
476 *C. rehmii*. (**H**) *C. colocynthis* (colocynth). (**I**) *C. naudinianus*. (**J-K**) Key changes in genetic
477 pathways underlying bitterness and lycopene synthesis in *Citrullus*. (**J**) Cucurbitacin pathway
478 in *Citrullus* highlighting key changes in copy number and functionality of the fruit-specific
479 transcription factor *ClBt* that controls fruit bitterness. (**K**) Lycopene pathway in *Citrullus*, and
480 phylogenetic distribution of a key substation in LYCB (lycopene β-cyclase) gene that is
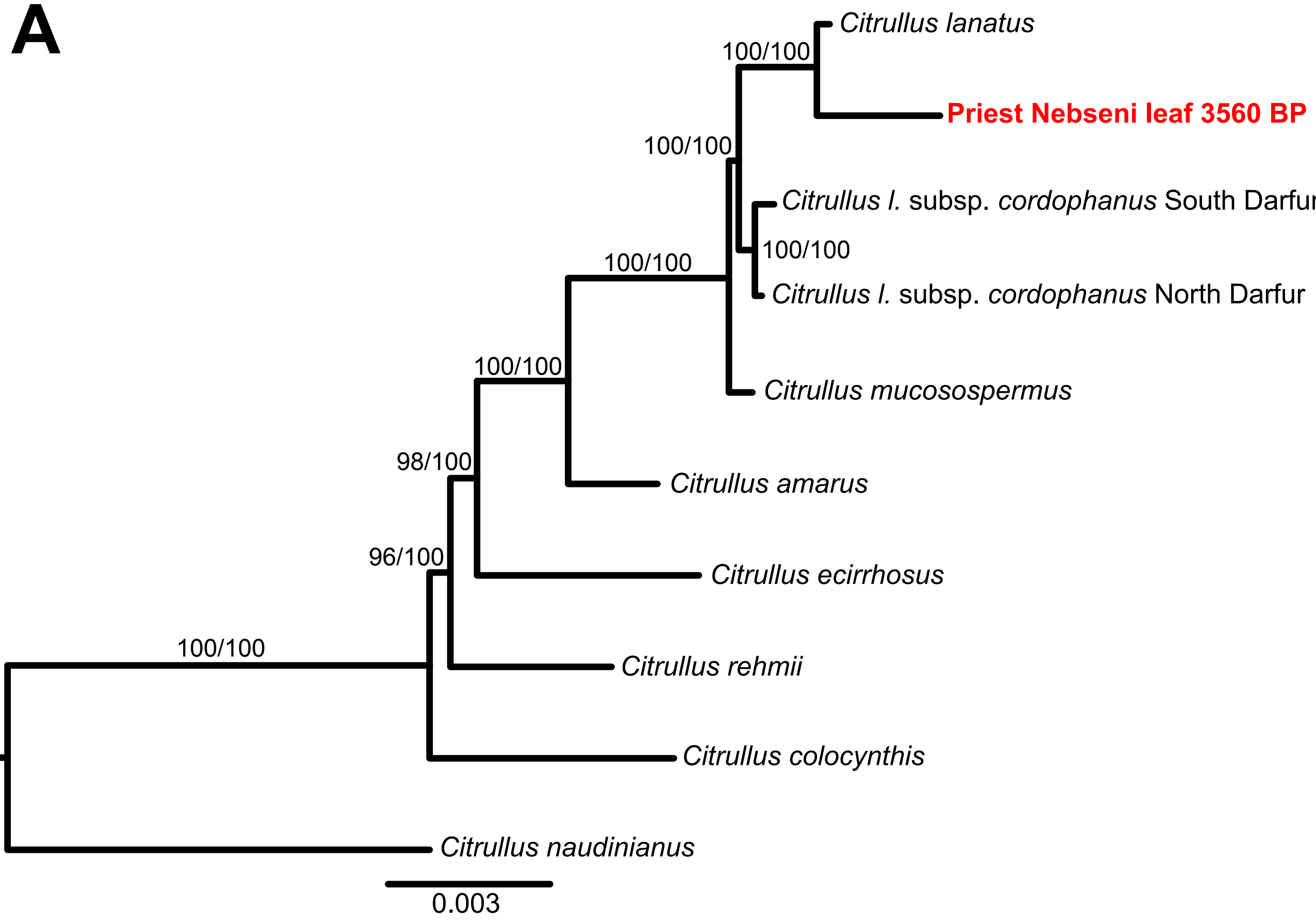481 linked to lycopene accumulation and controls fruit color.

482

483 **SI Appendix consists of SI text, Table S1, SI figures 1-7, and supplementary datasets 1,**
484 **2, and 3**. All DNA sequence alignments generated for the phylogenomic analyses as well as
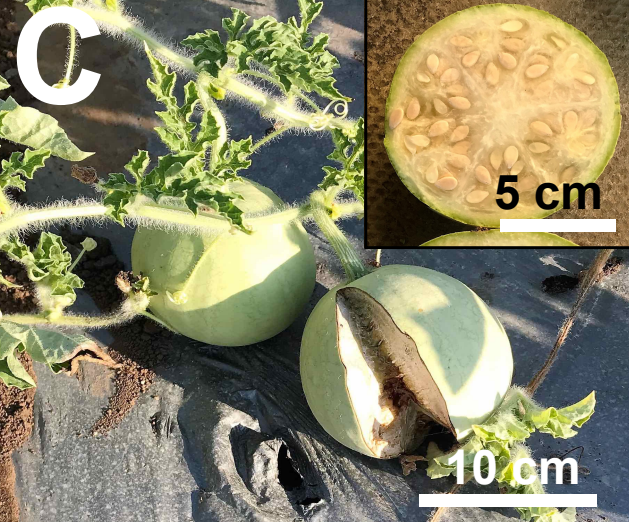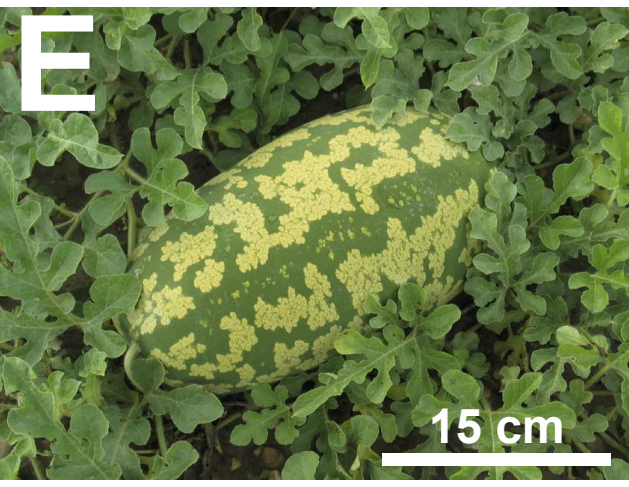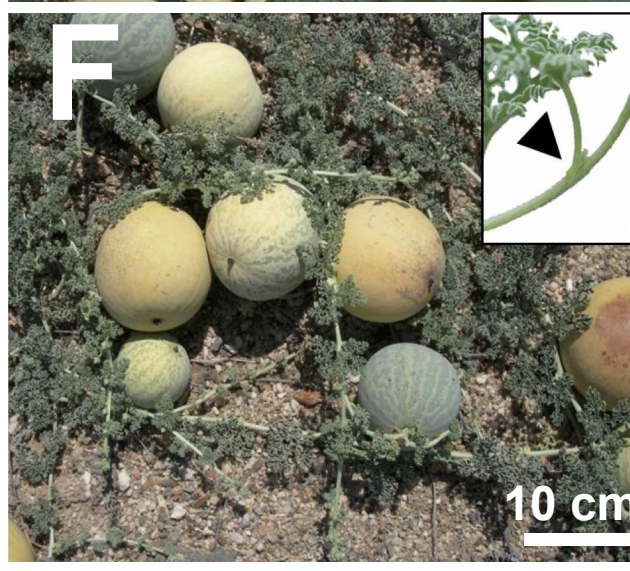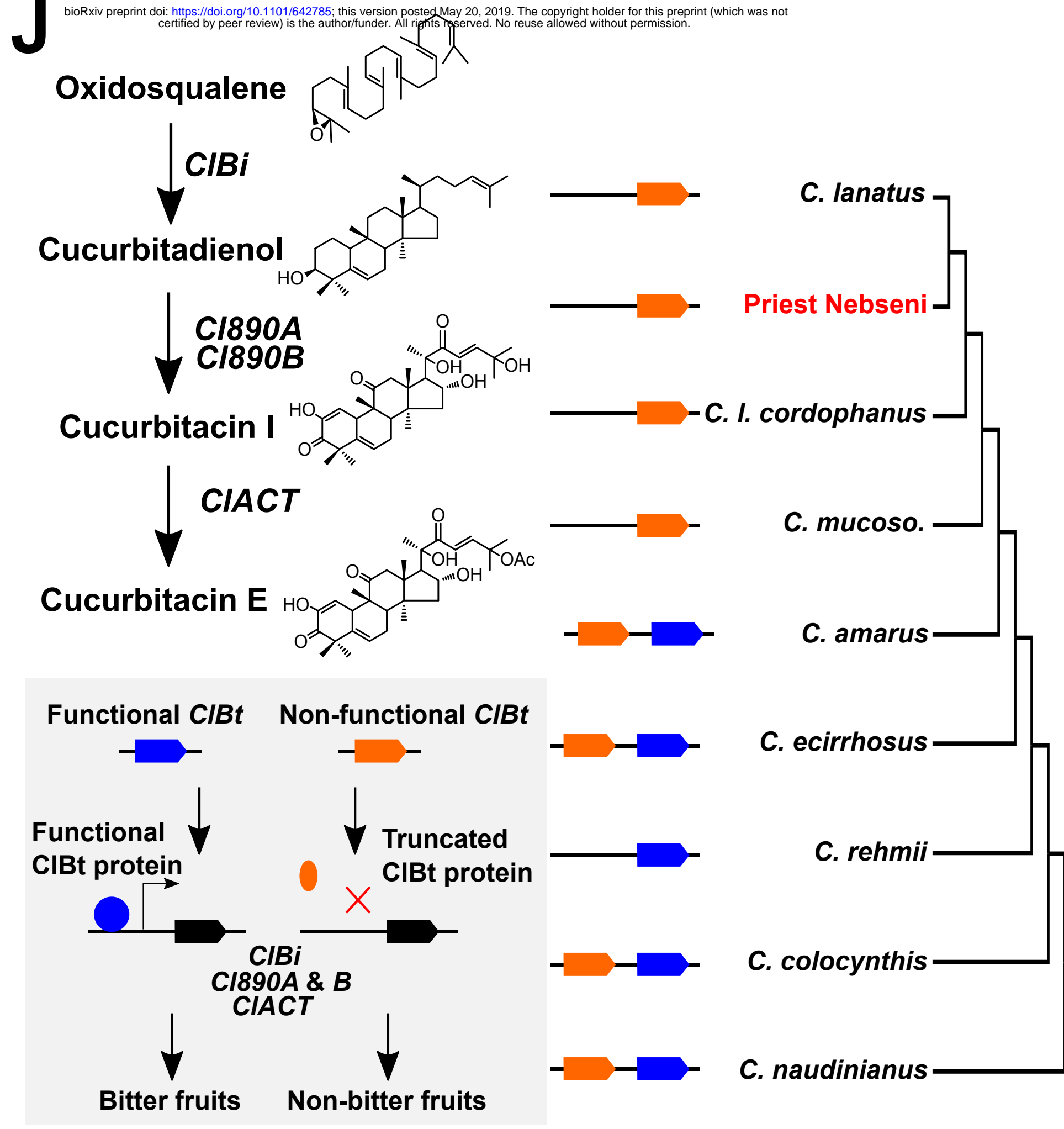485 for the genes of interests. Dryad accession number forthcoming.