bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1
2 **Title: Archaeogenomics of a ~2,100-year-old Egyptian leaf provides a new timestamp on**
3 **date palm domestication**
4
5 Oscar A. Pérez-Escobar[1*], Sidonie Bellot[1], Muriel Gros-Balthazard[2], Jonathan M. Flowers[2],
6 Mark Nesbitt[1], Philippa Ryan[1], Rafal M. Gutaker[2], Tom Wells[3], Rowan Schley[1], Diego
7 Bogarín[4], Natalia Przelomska[1,16], Steven Dodsworth[1,5], Rudy Diaz[1], Manuela Lehmann[6],
8 Peter Petoe[7], Wolf L. Eiserhardt[1,7], Michaela Preick[8], Michael Hofreiter[8], Irka Hajdas[9],
9 Alexandre Antonelli[1,10], Ilia J. Leitch[1], Barbara Gravendeel[11], Maria Fernanda Torres[12],
10 Guillaume Chomicki[13], Susanne S. Renner[14], Alexander S.T. Papadopulos[15], Michael
11 Purugganan[2], William J. Baker[1*]
12
13 **Affiliations**:
14 [1]Royal Botanic Gardens, Kew, Richmond TW9 3AE. London, UK.
15 [2]Center for Genomics and Systems Biology, New York University Abu Dhabi, United Arab
16 Emirates.
17 [3]University of Oxford, Oxford, OX1 3QU, UK.
18 [4]Lankester Botanical Garden, University of Costa Rica, San José, Costa Rica.
19 [5]University of Bedfordshire, Luton LU1 3JU, UK.
20 [6]British Museum, London, WC1B 3DG, UK
21 [7]Aarhus University, 8000 Aarhus C, Denmark.
22 [8]University of Potsdam, Potsdam, Germany.
23 [9]ETH Zurich, Switzerland.
24 [10]Gothenburg Global Biodiversity Centre and Department of Biological and Environmental
25 Sciences, University of Gothenburg, Sweden.
26 [11]Naturalis Biodiversity Center, Leiden, The Netherlands.
27 [12]University of Gothenburg, Gothenburg, Sweden
28 [13]Department of Biosciences, Stockton Rd, Durham University, Durham DH1 3LE, UK.
29 [14] Systematic Botany and Mycology, Department of Biology, University of Munich, 80638
30 Munich, Germany.
31 [15]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences,
32 University of Bangor, Bangor LL57 2UW, UK.
33 [16]Smithsonian Institution, Washington, D.C. USA
34
35 ***Correspondence**: o.perez-escobar@kew.org (OAPE) or w.baker@kew.org (WJB).
36
37
38
39
40
41
42
43
44
45

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

46
47 **Abstract [200 words – 200 max]:**

48 • The date palm (*Phoenix dactylifera*) has been a cornerstone of Middle Eastern and
49 North African agriculture for millennia. It is presumed that date palms were first
50 domesticated in the Persian Gulf and subsequently introduced into North Africa,
51 where their evolution in the latter region appears to have been influenced by gene flow
52 from the wild relative *P. theophrasti*, which is restricted to Crete and Turkey.
53 However, the timing of gene flow from *P. theophrasti* to *P. dactylifera* remains
54 unknown due to the limited archaeobotanical evidence of *P. theophrasti* and their
55 exclusion from population genomic studies.

56 • We addressed this issue by investigating the relatedness and ancestry of a ~2,100-
57 year-old *P. dactylifera* leaf from Saqqara (Egypt), combining genome sequencing of
58 this ancient specimen with a broad sample of date palm cultivars and closely related
59 species.

60 • The ancient Saqqara date palm shares close genetic ancestry with North African date
61 palm populations. We find clear genomic admixture between the Saqqara date palm,
62 *P. theophrasti* and the closest known relative *P. sylvestris*.

63 • Our study highlights that gene flow from *P. theophrasti* and *P. sylvestris* to North
64 African date palms had already occurred at least ~2,100 years ago, providing a
65 minimum timestamp for hybridisation between species.

66
67
68
69
70
71
72
73
74
75
76
77
78
79

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

80

81 **Introduction**

82

83 The start of plant crop domestication some 10,000-12,000 years ago was arguably one of the
84 most important events in human history (Diamond, 2002). The domestication of plant crop
85 and husbandry of animals allowed the sustained nutrition of large sedentary human
86 population settlements (Fuller et al., 2014; Larson et al., 2014; Richter et al., 2017; Arranz-
87 Otaegui et al., 2018). Elucidating the domestication history of major plant crops is thus an
88 important scientific challenge, which requires collaboration between scholars of archaeology,
89 anthropology, taxonomy, genetics and genomics. The widespread availability of high-
90 throughput DNA sequencing has revolutionized the study of plant crop domestication history,
91 leading to many unpreceded insights, such as the identification of crop progenitors (Ling et
92 al., 2013; Gros-Balthazard et al., 2017, Chomicki et al. 2020), hybridization and introgression
93 events following the origin of crops (e.g. Cornille et al., 2012; Hufford et al., 2012; Baute et
94 al., 2015; Muñoz-Rodrígez et al., 2018), refinement of the geographic origins of crops
95 (Besnard et al., 2017; Cubry et al., 2018), the identification of genes controlling key
96 domestication traits (Zhou et al., 2016; Stitzer and Ross-Ibarra, 2018) and more generally of
97 convergent evolutionary processes that have challenged orthodoxies on domestication
98 (reviewed by Purugganan, 2019). The application of genomic approaches to crop wild
99 relatives is also bringing critical new resources for crop improvement (reviewed by
100 Brozynska et al., 2016).

101 With more than 8 million tonnes of fruits produced yearly, the date palm (*Phoenix*
102 *dactylifera* L.) has been a cornerstone of Middle Eastern and North African agriculture for
103 millennia. Despite its economic importance, the date palm domestication history is far from
104 well understood (reviewed in Gros-Balthazard et al., 2018). Archaeological evidence, ancient
105 texts and iconographies all point to the use of date palms for millennia in North Africa, the
106 Middle East and as far as Pakistan (Tengberg, 2012). The first evidences of cultivation date to
107 the end of the 4th millennium B.C.E. in the Persian Gulf region (reviewed in Tengberg, 2012).
108 It is thus presumed that date palms were first domesticated in this region and subsequently
109 introduced into North Africa (reviewed by Gros-Balthazard et al., 2018). However, a study
110 that resequenced 62 date palm cultivars from the Middle East and North Africa found more
111 genetic diversity in North African date palm populations, challenging the simple Middle-
112 Eastern origin hypothesis (Hazzouri et al., 2015). Population genomic analyses of date palm
113 cultivars and other *Phoenix* species revealed extensive introgressive hybridization of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

114 North African date palm with *P. theophrasti* Greuter - up to 18% of the genome of North
115 African cultivars was shared with the Cretan date palm (Flowers et al., 2019). Clearly, date
116 palm evolution in North Africa has been influenced by gene flow from the wild relative *P.*
117 *theophrasti*. However, the timing of this introgressive event is unknown.
118 To address this important knowledge gap in date palm domestication, we sequenced
119 the genome of a ~2,100-year-old date palm (*P. dactylifera* L.) leaf found in Saqqara, Egypt,
120 radiocarbon-dated to the Late Period of ancient Egypt (357-118 B.C.E.). We found that the
121 genomic ancestry of the ancient Saqqara date palm can be traced to modern domesticated
122 North African *P. dactylifera* and the close wild relatives *P. sylvestris* and *P. theophrasti*.
123 Hybridisation and gene flow between North African *P. dactylifera, P. theophrasti* and *P.*
124 *sylvestris* had already taken place by ~2,100 years ago. Our study thus provides a minimum
125 bound on the timing for gene flow between date palms and their close living relatives.
126

127 **Materials and Methods**

128 ***Plant taxon sampling***

129 Our sampling builds upon recent population genomic studies of date palms by Flowers
130 et al. (2019) and Gros-Balthazard (2017) with a total of 36 individuals representing seven
131 species. We sampled seventeen individuals of wild and cultivated Asian and North African *P.*
132 *dactylifera* populations, as well as 18 accessions of five closely related species, namely *P.*
133 *atlantica*, *P. canariensis, P. reclinata, P. sylvestris* (the sister species of the date palm) and *P.*
134 *theophrasti* following the current accepted taxonomy of the genus *Phoenix* (Barrow, 1998;
135 Gros-Balthazard et al., 2020a). In addition, a shallow genomic representation of the New
136 Guinean palm *Licuala montana* was sequenced to produce a plastid genome assembly that
137 was subsequently used as the root for phylogenetic analyses. This species was chosen due to
138 the sister relationship between the palm tribes Phoeniceae (containing *Phoenix*) and
139 Trachycarpeae (containing *Licuala*) (Baker and Dransfield, 2016).
140 Whole genome sequence data from these taxa was obtained from the NCBI sequence
141 read archive repository. Twenty million reads were downloaded for each accession using the
142 tool *fastq-dump* of the SRA toolkit. Nearly all accessions sampled are linked to vouchers and
143 have known origins (Gros-Balthazard et al., 2017; Flowers et al., 2019); detailed information
144 on their provenance, average read length and number of bases downloaded are provided in
145 Tables S1 and S2.
146

147 **Radiocarbon dating and ancient DNA extraction**

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

148 To determine with accuracy the age of the Saqqara date palm item (accession EBC
149 26796), one cm^2 of leaf removed from the edge of the sample was sent to the Laboratory of
150 Ion Beam Physics, ETH-Zurich. The leaf sample underwent a treatment with solvents and an
151 acid–base–acid washes (Hadjas et al., 2008) to remove potential contamination of waxes,
152 carbonates and humic acids. The dry, clean material (weighing 2.6 mg, equivalent to 1 mg of
153 carbon) was weighed into tin cups for combustion in the Elemental Analyser for subsequent
154 graphitization (Němec et al., 2010). The resulting graphite was pressed into aluminium
155 cathodes and the $^{14}C/^{12}C$ and $^{13}C/^{12}C$ ratios were measured using the Mini Carbon Dating
156 System dedicated accelerator mass spectrometry facility (Synal et al., 2007). The radiocarbon
157 age was calculated following the method described by Stuiver et al. (1977) using the
158 measured ^{14}C content after correction for standards, blank values and fractionation ($\delta^{13}C$
159 values were measured semi-simultaneously on graphite). The reported conventional age in
160 years BP (before 1950 AD or CE) was calibrated to a calendar age using OxCal version 4.2.4
161 (Reimer et al., 2013) and the IntCal13 atmospheric curve (Bronk-Ramsey, 2013) .
162 Ancient DNA (aDNA) was extracted by grinding a small piece of a leaflet (<1 cm^2,
163 5.8 mg) with a Retsch mill (MM 400). DNA extraction was performed following the modified
164 protocol of Wales et al. (2014) (Pedersen et al., 2014; Dabney et al., 2013). For the digestion
165 treatment, a lysation buffer containing 0.5 % (w/v) N-lauroylsarcosine (Sigma Aldrich
166 L9150-50G), 50 mM Tris-HCl (Thermo Fisher Scientific 15568025), 20 mM EDTA (VWR
167 E177-500MLDB) 150 mM NaCl (Thermo Fisher Scientific AM9760G), 3.3 % 2-
168 mercaptoethanol (Sigma Aldrich 63689-25ML-F), 50 mM DL-dithiothreitol (Sigma Aldrich
169 D9779-250MG) and 0.25 mg/mL Proteinase K (Promega V3021) was applied to the leaflet
170 powder as described in Wales et al. (2014). DNA purification was performed according to
171 Dabney et al. (2013) but with reduced centrifugation speed (450 x *g*), following Basler et al.
172 (2017).
173
174 **Ancient DNA library preparation and sequencing**
175 A genomic Illumina library was prepared from the extracted aDNA following the
176 single-stranded protocol of Korlević et al. (2015). The protocol included the treatment with
177 Uracil-DNA-Glycolase (New England Biolabs M0279) to remove uracil residues and
178 Endonuclease VIII (New England Biolabs M0299) to cleave DNA strands at abasic sites.
179 Circligase II (2.5 U/µl; Biozym 131406) was used for the fill-in reaction which was carried
180 out overnight. A quantitative PCR was performed on a PikoReal 96 Real-Time PCR machine
181 (Thermo Fisher Scientific TCR0096) using 0.2 % of the unamplified library and the following

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

182 thermal profile: 10 min initial denaturation step at 95 °C, followed by 40 cycles of: 15 s at 95
183 °C, 30 s at 60 °C, and 1 min at 72 °C. The quantitative PCR reaction mix contained a final
184 volume of 10 µL: 1 µL of diluted library, 1 x SYBR Green qPCR Master Mix (Applied
185 Biosystems 4309155), 0.5 µM of each primer IS7 and IS8. Three replicates of each library
186 were used. Indexing PCR was performed by the appropriate number of cycles according to
187 the results of the qPCR, with 8 bp indices added to the 5' and 3' adapters. The PCR and final
188 concentrations used were the same as described by Gansauge and Meyer (Korlević et al.,
189 2017), but with a final volume of 80 µL using 20 µL of template. DNA sequencing was
190 performed on an Illumina NextSeq 500 sequencing platform, using the 500/550 High Output
191 v2 kit (75 cycles, Illumina FC-404-2005), with a custom read-1 (Perdersen et al., 2014) and a
192 custom index-2 (Paijmans et al., 2017) sequencing primer. All extractions and library
193 preparations were performed in the ancient DNA facility of the University in Potsdam;
194 negative controls were included in all steps.
195

196 **Genome skimming of *Licuala montana***

197 We extracted genomic DNA from silica-dried leaf tissue of *L. montana* using the
198 Qiagen DNeasy Plant kit, following the manufacturer's protocol. A genomic Illumina paired-
199 end library was prepared using the NEBNext Ultra II library preparation kit, following the
200 manufacture's protocol and with an average insert size of 150 bp. Library sequencing was
201 performed by the company Genewiz (New Jersey, USA) on a HiSeq platform. A total of 4
202 million paired-end reads was produced.
203

204 **High-throughput read data processing**

205 The Illumina raw reads were quality filtered using Trim Galore v.0.4 (Krueger, 2015),
206 discarding sequences with an averaged phred33 score below 20. Pre- and post-trimming read
207 quality was assessed using FASTQC v.0.1 (Andrews et al., 2015). The proportion of
208 endogenous DNA sequence present in the ancient Saqqara date palm leaf extract was assessed
209 by blasting the trimmed read data against the nuclear and plastid genomes of *Phoenix*
210 *dactylifera* (Khalas variety, assembly GCA000413155.1 [Al-Mssallem et al., 2013]) using
211 BLAST+ v.2.8.1 (McGinnis and Madden, 2013), an e-value of 0.001 and a coverage
212 threshold of 80%. We then determined the proportion of nuclear/plastid ancient date palm
213 read data sequenced by filtering the number of hits mapped by blast onto nuclear, plastid and
214 mitochondrial scaffolds.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

215 Given the low proportion of nuclear genomic data recovered from the ancient Saqqara
216 date palm leaf (see *Results*), we mapped trimmed reads of both modern and ancient
217 accessions on nuclear and plastid targeted scaffolds (i.e. scaffolds with Saqqara date palm leaf
218 reads mapped). These represented 198 contigs, or 26.8% (149.01 Mb) of the *P. dactylifera*'
219 nuclear genome assembly. We investigated the proportion and position of mis-incorporated
220 nucleotides in the Saqqara date palm leaf DNA using aligned aDNA reads and the tool
221 mapDamage2 v.2.0.9 (Jónsson et al., 2013). We compared nucleotide mis-incorporation
222 patterns between the aligned aDNA reads of the Saqqara date palm leaf and DNA reads of
223 modern date palm accessions (SRR5120110). Finally, to reduce the fraction of mis-
224 incorporated nucleotides mapped onto the reference genome, we trimmed two bases at the 3'
225 and 5' end of the aDNA reads, using Trim Galore v.0.4. Read mapping, alignment and DNA
226 damage analyses were implemented through the pipeline PALEOMIX v.1.2.13 (Schubert et
227 al., 2014). The trimmed read data were mapped using the software bowtie v.2.3.4.1, followed
228 by a realigning step around indels and filtering of duplicated reads with the software GATK
229 v.3.8.1 (McKenna et al., 2010) and Picard-tools v.1.137 (Thomer et al., 2016). Mis-
230 incorporated nucleotides in DNA fragments are characteristic of sequence data derived from
231 historical and archaeobotanical specimens (Estrada et al., 2018). Read mapping, and average
232 coverage statistics for each accession sampled in this study are provided in Table S1.
233 To account for biases in the mapping of aDNA read data onto the reference genome
234 (Günther and Nettelblad, 2019) and test the robustness of our population genomic inferences
235 against missing data (Skotte et al., 2013), we also mapped the ancient and modern DNA reads
236 onto 18 highly contiguous scaffolds of a newly assembled nuclear genome of *P. dactylifera*
237 (four-generations backcross of a Bahree cultivar, assembly CA0009389715.1 [Hazzouri et al.,
238 2019]), representing 50% of the nuclear genome (~380 Mb). Read mapping and alignment
239 were conducted using the same procedure and tools as specified above.
240

241 **Plastid phylogenomic analyses of *Phoenix***

242 We produced consensus plastid genome sequences of modern date palm accessions
243 from the BAM files produced by PALEOMIX by following a modified statistical base-calling
244 approach of Li et al. (2008), i.e. minimum depth coverage of 10, and bases matching at least
245 50% of the reference sequence. Because the attained average coverage of the Saqqara date
246 palm leaf plastid genome was ~2x (Table S1), the consensus plastid sequence for this
247 accession was produced by using a minimum depth coverage of 2, bases matching at least
248 50% of the reference sequence and missing data represented as Ns whenever parts of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

249 reference plastid genome were not covered by aDNA reads. The whole plastid genome
250 consensus sequences were produced in Geneious v.8.0. Consensus plastid genome sequences
251 were aligned with Mauve using a progressive algorithm and assuming collinearity (Darling et
252 al., 2004). The resulting ~150,000 bp alignment was first trimmed to exclude mis-aligned
253 regions and positions with >90% missing data (final alignment length of 103,807 bp) and then
254 subjected to Maximum Likelihood (ML) tree inference in RAxML v8.0 (Stamatakis, 2014),
255 using the GTR substitution model, 25 gamma rate categories, and 1,000 bootstrap replicates.
256
257 **Population structure and nuclear phylogenetic position of the Saqqara date palm**
258 Given low-depth high-throughput sequencing data produced for the ancient Saqqara
259 date palm leaf, we relied on genotype likelihoods (GL) to place the aDNA nuclear genomic
260 data of the Saqqara date in range with genomic sequences of modern *Phoenix* samples. We
261 computed nuclear GL using the software ANGSD v.0.929 (Korneliussen et al., 2014), by
262 implementing the GATK GL model, inferring the minor and major alleles, and retaining
263 polymorphic sites with a minimum p-value of 1-e^{6}. To reveal the relationship of the ancient
264 Saqqara date leaf nuclear genome to the genetic diversity of modern samples of *Phoenix*, we
265 conducted principal component (PCA) and population structure analyses using nuclear GLs
266 and the tools PCangds (Meisner and Albrechtsen, 2018), and NGSadmix (Skotte et al., 2013)
267 of the software ANGDS, respectively. Because PCA can be particularly affected by the
268 proportion of overlapping sites between modern and ancient populations (Ausmees, 2019), we
269 computed covariance matrices by using only GLs derived from sites that were shared across
270 all the modern individuals and the ancient Saqqara date palm leave (i.e. option -minInd set to
271 35), and a maximum of 1000 iterations. Admixture analyses were conducted with number of
272 population (K) set from two to eight and a maximum of 20,000 iterations. The best K was
273 selected by comparing the resulting likelihood values derived from each K iteration.
274 To test for admixture between the Saqqara date palm and other lineages amongst date
275 palms or closely related species (*P. atlantica*, *P. sylvestris* and *P. theophrasti*), we used the D-
276 statistic framework as implemented in the software ANGSD. To account for the differences in
277 sequencing coverages obtained for modern individuals and the Saqqara leaf and to assess the
278 robustness of our introgression tests, two different approaches were followed, namely a)
279 between nuclear genomes of each individual (i.e. sampling one base from reads of one
280 individual per population [Korneliussen et al., 2014]) and b) between populations (i.e.
281 considering all reads from multiple individuals in each population [Soraggi et al., 2018]).
282 Nuclear GL were used as input and D-statistics for both tests. In (a) D-statistics were

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

283 calculated by sampling a random base at each analysed position in blocks of 5 million bp,
284 removing all transitions to rule out possible post-mortem base misincorporations in the
285 Saqqara sample together with reads with qualities lower than 30, and setting *P. reclinata* as a
286 fixed outgroup terminal following the same experimental design as in Flowers et al. (2019).
287 In (b) individuals were assigned to six populations defined according to their species identity,
288 e.g. all modern individuals of *P. dactylifera* were assigned to one population (Table S2); the
289 Saqqara leaf was assigned to its own population as well as the outgroup (*P. reclinata*),
290 following the recommendations provided by Soraggi et al. (2018). D-statistics were then
291 calculated by sampling reads from multiple individuals in each population. Moreover, to
292 account for the influence of polymorphisms in the outgroup taxon, we executed population
293 tests using polymorphic and non-polymorphic sites (Table S5) and only non-polymorphic
294 sites in the outgroup (i.e. "–enhance" flag in ANGSD; Table S6). The significance of the
295 analyses was assessed by executing a block-Jacknife test which derived standard errors and
296 Z–scores. For population level analyses, *p-values* were derived. The admixture of the Saqqara
297 date palm was discussed based solely on significant *p*- and D-statistic values (i.e. $|Z| > 3$). D-
298 statistic values, standard deviations, Z-scores and the number of evaluated sites for each
299 topological permutation are provided in Tables S3A-C. NGSadmix and D-statistic analyses
300 were conducted on filtered GLs derived from read data mapped on: a) 198 contigs
301 representing 26.8% of the *P. dactylifera* genome (assembly GCA000413155.1); and b) 18
302 scaffolds representing 50% of the nuclear genome of *P. dactylifera* (assembly
303 GCA0009389715.1, see *High-throughput read data processing* section of *Methods* above).
304

305 ## Results

306 ### DNA sequencing of an archaeological date palm leaf from Saqqara

307 Our archaeological sample is from an object made from date palm leaflets discovered
308 in the temple complex of the animal necropolis of Saqqara, an Egyptian UNESCO World
309 Heritage site located 20 km south of Cairo and adjacent to the Nile valley. The object,
310 currently held in the Economic Botany Collection at the Royal Botanic Gardens, Kew, was
311 recovered during the 1971-2 excavation season from the 'West Dump'. The site is a mixed
312 refuse deposit dating between 500-300 B.C.E. that also contained other objects such as
313 possible 'brushes' made from date palm, papyri, jar-stoppers, amulets and other debris
314 including seeds (Martin et al., 1981). The object consists of a plaited portion of a leaf
315 (including leaflets and rachis) and was originally considered as a 'head-pad' by excavators,
316 however there are no analogous finds from other sites supporting this interpretation. A

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

317 virtually identical object from the 'West Dump' at Saqqara is held in the British Museum
318 collections (accession EA68161) where it is identified as possibly 'part of the lid of a basket'.
319 We speculate that the object, hereafter referred to as "Saqqara leaf" (Fig. 1), may instead have
320 been part of the layering used to close and seal a vessel, similar to those found in the New
321 Kingdom (1570–1070 B.C.E) (Hope, 1977). We radiocarbon-dated the Saqqara leaf to 2,165
322 ± 23 BP (ETH-101122), or to a calibrated date of 357-118 B.C.E, thus confirming its burial
323 during the Late Period or the Ptolemaic Kingdom of ancient Egypt.
324 We sequenced ~400 million reads from the Saqqara leaf of which up to ~4% (i.e. *c.* 16
325 million reads) were identified to be from endogenous DNA of *P. dactylifera* (Table S1). As
326 expected from ssDNA libraries, nucleotide misincorporations (C to T), which are indicative of
327 DNA damage, predominantly occurred towards both ends of the reads, visible even after an
328 uracil reduction procedure. Read length distributions were centred on 35 bp (Fig. S1),
329 consistent with sequencing data from similarly-aged material (Scott et al., 2019; Ramos-
330 Madrigal et al., 2016). Although the average read depth was only ~2x, we obtained a near-
331 complete representation of the plastid genome of the Saqqara sample, covering 95% of the
332 plastome of modern date palms (see *Methods*). We also recovered up to ~755 million base
333 pairs of the *P. dactylifera* nuclear genome (Table S1).
334
335 **Phylogenetic placement of the Saqqara leaf and detection of introgression in its genome**
336 To identify the closest relatives of the Saqqara leaf, we used Illumina sequencing
337 reads available in the Sequence Read Archive to assemble the plastomes of 17 modern Asian
338 and African date palms (including possible wild-origin individuals from Oman, Gros-
339 Balthazard et al., 2017) and 17 individuals belonging to five closely related species (i.e. *P.*
340 *atlantica*, *P. canariensis, P. reclinata*, *P. sylvestris,* and *P. theophrasti*; Table S2). To
341 compare the outcome of our phylogenetic and population genomic analyses with results
342 obtained by previous studies, our taxon sampling strategy is virtually identical to the one
343 conducted by Flowers *et al.* (2019) and Gros-Balthazard *et al.* (2017). Maximum Likelihood
344 (ML) phylogenetic analyses on full plastome alignments revealed that the Saqqara leaf is
345 nested in a strongly supported clade (Likelihood Bootstrap Support [LBS]: 87) entirely
346 composed of North African cultivated date palms and two accessions of *P. atlantica* (Fig. 1;
347 Fig. S2), a disputed species currently restricted to Cape Verde (Gros-Balthazard et al., 2020a).
348 The North African clade is itself nested in a clade (LBS 99) of *P. sylvestris* samples, a species
349 found from Pakistan to Myanmar, long hypothesized to be the closest relative of *P.*
350 *dactylifera* (Barrow 1998; Pintaud et al., 2013; Gros-Balthazard et al., 2020a). Asian *P.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

351 *dactylifera* (LBS 100) form a clade that is sister to the above-described clade of African *P.*
352 *dactylifera* plus *P. sylvestris* (LBS 100).
353 To determine the genomic affiliation of the nuclear genome of the Saqqara sample to
354 either North African or Asian modern *P. dactylifera* populations, we conducted a model-free
355 principal component (PCA) and a model-based clustering analysis using genotype likelihoods
356 derived from the nuclear genomes of all accessions, the latter assuming two to eight ancestral
357 populations. We conducted genomic clustering analyses using two genomes of *P. dactylifera*
358 as reference with different levels of completeness and contiguity to account for read mapping
359 and missing data biases (Skotte et al., 2013; Günter and Nettelblad, 2019; see *Methods*).
360 Regardless of the reference genome used, with four populations the Saqqara genome grouped
361 with populations of *P. dactylifera* with Asian ancestry. North African individuals of *P.*
362 *dactylifera* and *P. atlantica* shared most of their genome with Asian modern date palm
363 populations albeit with a relatively small proportion of their genome admixed with *P.*
364 *theophrasti* (Fig. 2; Fig. S4), thus supporting previous findings (Flowers et al., 2019). When
365 assuming five ancestral populations, *P. dactylifera* segregated into two populations with
366 African and Asian ancestry, respectively (Fig. 2; Fig. S4). Here, the clustering indicated that
367 the majority (90-98%) of the analysed nuclear sequences from the Saqqara genome displayed
368 components of North African domesticated *P. dactylifera* individuals and Cape Verde's *P.*
369 *atlantica*, whilst the remaining 1-10% could be traced to both domesticated and wild Asian *P.*
370 *dactylifera* individuals (Fig. 2; Fig. S4). Allele sharing between *P. dactylifera* and *P.*
371 *sylvestris* was also evident in clustering analyses with four and five populations, also
372 supporting previous findings by Flowers et al. (2019). The PCA revealed similar results to
373 those obtained by the model-based clustering analyses. Regardless of the reference genome
374 used, the covariance matrices inferred from ~27,000 to ~35,000 filtered sites placed the
375 Saqqara date palm genome closest to modern North African date palm individuals in a cluster
376 made of accessions of *P. dactylifera* (Fig. 2c,d).
377 We conducted introgression tests (i.e. D-statistics) to trace gene exchange between the
378 Saqqara leaf in relation to the modern date palms and the closely related *P. atlantica*, *P.*
379 *sylvestris* and *P. theophrasti*, using *P. reclinata* as an outgroup based on previous studies
380 (Flowers et al., 2019; Fig. 3; Fig. S3). To account for the differences in sequencing coverage
381 obtained from modern individuals and the ancient Saqqara genome, these topological tests
382 were conducted using two approaches tailored to separately evaluate individuals (i.e. by
383 sampling one base from reads of one individual per population) and populations (i.e. by
384 considering all reads from all individuals in each population; Soraggi et al., 2018). Both

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

385 approaches were also implemented using two reference genomes to account for potential
386 sequence biases (Günther and Nettelblad, 2019; see *Methods*). Analyses considering
387 individuals separately and populations (regardless of the genome of reference) gave virtually
388 identical results regarding the relatedness of the Saqqara leaf.

389 Introgression tests between modern individuals and the Saqqara leaf involved the
390 evaluation of 39,280 and 503,386 nuclear bases, with an average of 81 and 586 bases per
391 analysis using highly fragmented and contiguous genome assemblies as reference,
392 respectively (Tabs. S3, S4). We found no signal of introgression from *P. sylvestris* in the
393 Saqqara leaf nuclear genome as inferred from the analysis conducted on a highly fragmented
394 genome. However, when computing D (Saqqara, *X*, *P. sylvestris*; *P. reclinata*, where *X* is *P.*
395 *atlantica* or *P. dactylifera*) using as a reference the contiguous genome assembly, the Saqqara
396 sample shared more derived alleles with *P. sylvestris* than with *X* (Table S4). Signal of
397 introgression from *P. theophrasti* in the Saqqara sample was evident in analyses conducted on
398 both highly fragmented and contiguous reference genomes. Here, when computing D
399 (Saqqara, *P. sylvestris*; *P. theophrasti*, *P. reclinata*), the Saqqara sample shared more derived
400 alleles with *P. theophrasti* than with *P. sylvestris* ($Z < -3.1$ and $Z < -4.8$ for highly fragmented
401 and contiguous reference genome, respectively; Fig. 3A & Tabs. S3, S4).

402 Population tests using the highly fragmented genome of reference evaluated 3,472 to
403 9,469 bases, with an average of 192.93 and 527.6 bases per analysis considering polymorphic
404 and non-polymorphic sites in the outgroup, respectively (Tabs. S5, S6). In contrast, analyses
405 based on the continuous reference genome assessed 10,400 to 16,565 bases, with an average
406 of 577.82 and 920.3 bases per analysis considering polymorphic and non-polymorphic sites in
407 the outgroup, respectively (Tabs. S5, S6). Altogether, the population test analyses revealed
408 results consistent with introgression analyses conducted at the individual level, regardless of
409 the genome of references employed, thus providing support for the past occurrence of gene
410 flow between the Saqqara leaf and *P. theophrasti*. Here, when computing D (Saqqara, *X*; *P.*
411 *theophrasti*, *P. reclinata*, where *X* refers to either *P. sylvestris*, *P. atlantica*, or *P. dactylifera*),
412 the Saqqara leaf genome shared more derived alleles with *P. theophrasti* than the latter with
413 any of the other closely related taxa evaluated ($Z < -3.93 < -6.9$ & $Z < -6.73 < -17.08$ for
414 highly fragmented and contiguous reference genomes, respectively; Fig. 3B; Tabs. S5, S6).
415 As for individuals-based tests, only the D-statistic analysis conducted on the contiguous
416 genome revealed gene flow between the Saqqara leaf and *P. sylvestris* (with D[Saqqara, *P.*
417 *atlantica*; *P. sylvestris*, *P. reclinata*] = $Z < -5.7$; Table S5).

418

419 **Discussion**

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

420 Generating genomic data from plant archaeological remains of known origin and
421 unequivocal age provides a unique window into the timing and sequence of plant crop
422 domestication and diffusion processes (Estrada et al., 2018; Gutaker et al., 2017; Swarts et al.,
423 2017; Scott et al., 2019). Archaeological evidence thus far has not informed us about the
424 occurrence and timing of genetic exchanges between date palms and their wild relatives, and
425 whether their distribution overlapped in the past (Flowers et al., 2019). Our study is the first
426 to address this gap by applying archaeogenomic approaches to shed light on date palm
427 evolutionary history. Though low in overall proportion, we retrieve sufficient genetic
428 information from the endogenous aDNA, highlighting the potential for further genomic
429 analysis using additional archaeological remains, from other species, places and times.
430 Comparisons of our plastid and nuclear topologies and population structure analysis
431 conducted on two reference genomes representing contrasting levels of contiguity and
432 completeness provide robust evidence for the genomic affiliation of the Saqqara leaf with
433 modern North African *P. dactylifera* populations, as well as the occurrence of ancient gene
434 flow between *P. dactylifera, P. theophrasti,* and *P. sylvestris*. The clustering of *P. sylvestris*
435 with selected North African date palm cultivars in plastid phylogenies has been previously
436 reported by several studies (Pintaud et al. 2013; Chaluvadi et al., 2019; Flowers et al., 2019;
437 Mohamoud et al., 2019), thus opening the question of whether gene flow or ancestral
438 polymorphisms are responsible for such patterns (Flowers et al., 2019). *Phoenix dactylifera*
439 and *P. sylvestris* overlap their distribution ranges in north western India and Pakistan, they are
440 interfertile and known to produce fertile hybrids (Newton et al., 2013), suggesting that gene
441 flow between both species is plausible.
442 Recently, extensive sequencing of over 200 organellar genomes of *P. dactylifera*
443 revealed that date palm cultivars contain four haplotypes that are tightly linked to the
444 geographical origin of the cultivar (Mohamoud et al., 2019), but the time of their
445 diversification is largely unknown. In particular, one major haplotype (NA1, see Fig. 2 in
446 Mohamoud et al., 2019) reported for North Africa is thought to be highly divergent from the
447 remaining three haplotypes and is shared with *P. sylvestris* (Mohamoud et al., 2019). The
448 trace of gene flow between the Saqqara leaf and *P. sylvestris* detected by our introgression
449 analyses suggests that the recurrent clustering patterns of individuals from both species in
450 plastid phylogenies could be derived from one or several chloroplast-capture processes
451 mediated by hybridisation. In addition, by confidently placing the Saqqara leaf plastid
452 genome in the NA1 haplotype, we set for the first time a minimum age for the origin of this
453 plastid subpopulation to *c.* 2,100 yrs BP.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

454 In the Nile valley, date stones are recovered from at least seven archaeological sites in
455 Egypt and three in Sudan spanning the Middle Kingdom (2,500-1,650 BC) to 2nd intermediate
456 period, but they are only commonly present from the New Kingdom (1,570-1,070 BC)
457 onwards (see reviews in Murray 2000; Zohary et al., 2015; and database in Flowers et al.,
458 2019). More limited evidence from the Old Kingdom (2700-2100 BC) includes findings of
459 two date stones, and occasional fragments of other plant parts, from Giza (Malleson and
460 Miracle 2018). Textual evidence from the Old Kingdom also refers to imported dates (Tallet
461 2017). As such, some potential early finds may represent either imports or cultivation (Gros-
462 Balthazard et al 2020). Examples of date stones recovered from earlier predynastic sites,
463 notably from El Omari and Hierakonpolis, might be intrusive and lack reliable context (R.
464 Friedman, pers. com. 1 March 2020) (Flowers et al., 2019). There are also occasional
465 (potential) examples of date palm leaves and fibre from various sites, mostly funeral contexts,
466 from around 3,800 B.C.E. onwards (Vartavan et al., 2000).
467 Flowers et al. (2019) suggested date cultivation in Egypt was established between the
468 Middle to New Kingdom periods based on presence/absence of date stones recovered from
469 archaeological sites within their database. Alternatively, Gros-Balthazard et al (2020) argue
470 that cultivation and cultural importance is only clear from the New Kingdom onwards. The
471 new importance of date culture during the New Kingdom is also reflected artistically, for
472 instance in garden scenes within tomb wall-paintings (Parkinson 2008). Further west of the
473 Nile valley, evidence for date cultivation at Zinkekra in Libya is clear from the early first
474 millennium B.C.E, and also provides an early example of oasis agriculture in North Africa
475 (Pelling 2013, Van der, Veen, & Westley, 2010). Additionally, a datestone dating to c.1400-
476 1300 B.C.E. from the Wadi Tanzzuft, some 400 km further south-west, suggests earlier
477 evidence for mid to later 2nd millennium date cultivation in the central Sahara (Mattingly and
478 Wilson 2010, di Lernia and Manzi 2002). It may be that small scale cultivation was a
479 precursor to a more widespread practice. This might reflect the time and expertise it takes to
480 establish date palms effectively via clonal propagation rather than from seed, or to develop an
481 adequate ratio of male/female trees and the practice of deliberate fertilisation (Bacon 1948).
482 Thirdly, improved water management techniques were developed by the New Kingdom
483 (Murray 2000).
484 Evidence demonstrating ancient date palm use and exploitation is recorded by reliable
485 finds of date seeds around 5,500 B.C.E. from the Arabian Peninsula (Beech et al., 2003) and
486 around 4,000 B.C.E. from Mesopotamia (Gillet, 1981; Tengberg, 2012). Date palms also
487 occurred in Jordan from between 4,800-4,250 B.C.E. and in Israel from around 3,500-3,200

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

488 B.C.E. (Zohary et a., 2015). Taken together, these findings suggest an origin of date palm
489 domestication around the Persian Gulf or Mesopotamia followed by the subsequent spread of
490 the crop into North Africa. Several studies show two genetically distinct populations of
491 domesticated *P. dactylifera* in North Africa and the Middle East (Gros-Balthazard et al.,
492 2017; Gros-Balthazard et al., 2020b). The diversity of African date palms is, however, higher
493 than expected following such a founder event, raising the possibility of an independent origin
494 of date palm domestication in Africa and/or further introgression of genetic material from
495 other *Phoenix* species (Hazzouri et al. 2015; Gros-Balthazard et al., 2017).

496 The study by Flowers et al. (2019) provided some explanation for the origin of the
497 elevated genetic diversity in North African *P. dactylifera* by showing that parts of their
498 genome were most closely related to *P. theophrasti*, a species currently restricted to the
499 coastal areas of Crete, the Aegean islands and Turkey (Boydack, 2019). Following the only
500 known archaeobotanical record of *P. theophrasti* found in northern Israel, the timing of such
501 genetic exchange was hypothesised to have occurred ~7,500 yrs ago when the geographic
502 ranges of the two species may have overlapped (Flowers et al., 2019). However, the
503 authenticity of such an archaeological macrofossil is questionable because its identification
504 relied solely on morphological comparisons of seed size, a character proven to be labile in
505 wild and domesticated date palms (Gros-Balthazard et al., 2017; J. Dransfield, pers. com. 20
506 Feb 2020) and informative only when studied using statistical frameworks (Terral et al., 2012;
507 Gros-Balthazard et al., 2017). Thus, before our study there was no concrete evidence
508 supporting a minimum age for the genetic exchange between North African populations of *P.*
509 *dactylifera* and *P. theophrasti*. Modern hybrid zones between *P. dactylifera* and *P.*
510 *theophrasti* are known, and the species are known to hybridise in botanical gardens and
511 plantations (Gros-Balthazard et al., 2013). The small but consistent proportion of derived
512 alleles shared between the Saqqara date palm and *P. theophrasti* identified here (Figure 3)
513 provides evidence that hybridisation between these species had already occurred before
514 ~2,100 yrs BP. Nevertheless, caution is required when interpreting our D-statistics because
515 the number of homologous bases analysed (ranging from hundreds to thousands; Tabs. S3-S6)
516 represents only a small fraction of the nuclear genome. A better representation of the Saqqara
517 nuclear genome, ideally attained through target capture to increase the proportion of
518 endogenous nuclear DNA, would help to further define the proportion of the Saqqara genome
519 that has been inherited from *P. theophrasti* or any other wild relative. In spite of the ancient
520 timing of this introgression, the regions inherited from *P. theophrasti* comprise as much as
521 18% of the modern nuclear genome of North African domesticated date palm (Flowers et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

522 2019). This high proportion suggests that such alleles may confer an advantage over their date
523 palm homologs, enabling them to persist in the date palm genome despite the absence of a
524 current contact zone between both species. Alternatively, the early implementation of clonal
525 propagation of offshoots derived from hybrid individuals could have contributed to the
526 survival of admixed genotypes. Archaeological evidence supporting such agricultural
527 practice, however, is lacking.
528

529 **Conclusion**

530 As the first documented instance of successful retrieval of ancient DNA from date
531 palm, our study provides a timestamp for the occurrence of a key introgression process in the
532 evolutionary history of this culturally and economically important crop. Our plastid and
533 nuclear topological frameworks together with the genomic composition analysis involving
534 different reference genomes, levels of contiguity and genomic representations, consistently
535 indicate that the Saqqara palm belongs to the same clade as the modern North African *P.*
536 *dactylifera*, and that by c. 2,000 yrs ago, this lineage had already undergone considerable
537 genetic differentiation from Asian date palms, including from wild date palm populations.
538 Our results also suggest a minimum date for ancient interspecific gene flow between North
539 African *P. dactylifera, P. sylvestris* and *P. theophrasti*.
540 Our study highlights how an integrated approach comprising genomics,
541 phylogenomics and archaeobotany can yield new insights on the timing and processes
542 involved in the domestication and diffusion of one of the oldest fruit crops. The research
543 showcases the importance of safeguarding and curating biological collections, which are now
544 providing a multitude of uses in evolutionary research that could never have been dreamt of at
545 the time samples were collected. Our results provide the temporal framework of date palm
546 introgression with *P. theophrasti*, and a sound base for further biogeographical and molecular
547 dating studies that could include a larger sampling of modern and ancient populations to
548 clarify a) whether gene flow with other wild relatives also occurred and b) where and in what
549 environmental conditions these processes took place.
550

551 **Supplementary Information**
552 Supplemental Information includes four figures and six tables.
553

554 **Author Contributions**
555 O.A.P.E., S.B., M.N., M.G.-B. and W.B. conceived the study. M.G.-B., J.F., O.A.P.E.,
556 S.B., M.H., M.Pu., A.S.T.P., and R.G., designed in-silico analyses. M.N., P.R., and M.L.
557 conducted archaeobotanical research. P.P., B.G., M.H., M.Pr. and I.H conducted lab work.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

558 O.A.P.E., T.W., R.S., R.D., and D.B. conducted in-silico analyses. O.A.P.E., M.N., P.R.,
559 S.B., S.D., I.L., M.Pr., S.S.R. G.C., and W.R. wrote the manuscript, with contributions from
560 all authors.
561

562 **Acknowledgements**
563 We would like to thank Valentina Gasperini (The British Museum) for her advice on
564 the Saqqara excavation contexts, and Renée Friedman (Director of excavations at the
565 Predynastic site of Hierakonpolis) for her advice on the reliability of date stone archaeological
566 findings and the need for re-evaluation of date leaf finds.
567

568 **Funding**
569 The authors acknowledge funding from the Lady Sainsbury Orchid Trust and the Swiss
570 Orchid Foundation to O.A.P.E.; a Garfield Weston Foundation postdoctoral fellowship to
571 S.B., the Swedish Research Council, the Swedish Foundation for Strategic Research, the Knut
572 and Alice Wallenberg Foundation and the Royal Botanic Gardens, Kew to A.A.
573

574 **Competing interests**
575 The authors declare no competing financial interests.
576

577 **References**

578 Andrews S, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. (2015). FastQC: A quality
579 control tool for high throughput sequence data. Babraham Bioinformatics. Babraham
580 Inst.
581 Al-Mssallem I.S., Hu S., Zhang X., Lin Q., Liu W., Tan J., Yu X., Liu J., Pan L., Zhang T., *et*
582 *al.* (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nature*
583 *Communications* **4**: 1–9. Available at: http://dx.doi.org/10.1038/ncomms3274.
584 Arranz-Otaegui A., Carretero L.G., Ramsey M.N., Fuller D.Q., Richter T. (2018).
585 Archaeobotanical evidence reveals the origins of bread 14,400 years ago in
586 northeastern Jordan. *Proceedings of the National Academy of Sciences, USA* **115**:
587 7925-7930.
588 Ausmees K. (2019). Evaluation of methods handling missing data in PCA on genotype data:
589 applications for ancient DNA. Uppsala University, 10 p.
590 Bacon S.H. (1948) Crops of the Sudan. In J.D. Tothill (Ed) *Agriculture in the Sudan*, Oxford
591 University Press, 302-400.
592 Baker W.J., Dransfield J. (2016). Beyond Genera Palmarum: progress and prospects in palm
593 systematics. *Botanical Journal of the Linnaean Society* **182**: 207–233.
594 Barrow S.C. (1998). A monograph of *Phoenix* L. (Palmae: Coryphoideae). *Kew Bulletin* **53:**
595 513–575.
596 Basler N., Xenikoudakis G., Westbury M. V., Song L., Sheng G., Barlow A. (2017).
597 Reduction of the contaminant fraction of DNA obtained from an ancient giant panda
598 bone. *BMC Research Notes* **10:** 754.
599 Baute G.J., Kane N.C., Grassa C.J., Lai Z., Rieseberg L.H. (2015). Genome scans reveal
600 candidate domestication and improvement genes in cultivated sunflower, as well as
601 post-domestication introgression with wild relatives. *New Phytologist* **206:** 830-838.
602 Beech M. (2003). Archaeobotanical evidence for early date consumption in the Arabian gulf.
603 In the date palm – from traditional resource to green wealth, The Emirates Center for
604 Strategic Studies and Research, ed. (Abu Dhabi: Bloomsbury Academic), pp. 11–31.
605 Besnard G., Terral J. F., Cornille, A. (2017). On the origins and domestication of the olive: a
606 review and perspectives. *Annals of Botany* **121**: 385-403.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

607 Boydack M. (2019). A new subspecies of *Phoenix theophrasti* Greuter (*Phoenix theophrasti*
608 Greuter subsp. *golkoyana* Boydak) from Turkey. *Forestist* **69:** 133-144.
609 Bronk-Ramsey C. (2013). Recent and planned developments of the program OxCal.
610 *Radiocarbon* **55:** 720–730.
611 Brozynska M., Furtado A., Henry, R.J. (2016). Genomics of crop wild relatives: expanding
612 the gene pool for crop improvement. *Plant Biotechnology Journal* **14**: 1070-1085.
613 Chaluvadi S.R., Young P., Thompson K., Bahri B.A., Gajera B., Narayanan S., Krueger R.,
614 Bennetzen J.L. (2019). *Phoenix* phylogeny, and analysis of genetic variation in a
615 diverse collection of date palm (*Phoenix dactylifera*) and related species. *Plant*
616 *Diversity* **41:** 330–339. Available at: https://doi.org/10.1016/j.pld.2018.11.005.
617 Chomicki, G., Schaefer, H. and Renner, S.S., 2020. Origin and domestication of
618 Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. *New*
619 *Phytologist* **226:** 1240-1255.
620 Cornille A., Gladieux P., Smulders M. J., Roldán-Ruiz I., Laurens F., Le Cam B., Nerseyan
621 A., Clavel J., Olonova M., Feguey L., Gabrielyan I., Zang X-G., Tenaillon M., Giraud
622 T. (2012). New insight into the history of domesticated apple: secondary contribution
623 of the european wild apple to the genome of cultivated varieties. *PloS Genetics* **8:**
624 e1002703
625 Cubry P., Tranchant-Dubreuil C., Thuillet A.-C., Monat C., Ndjiondjop M.-N., Labadie K.,
626 Cruaud C., Engelen S., Scarcelli N., Rhoné B., *et al*. (2018). The rise and fall of african
627 rice cultivation revealed by analysis of 246 new genomes. *Current Biology* **28:** 2274–
628 2282
629 Dabney J., Knapp M., Glocke I., Gansauge M.T., Weihmann A., Nickel B., Valdiosera C.,
630 García N., Pääbo S., Arsuaga J.L., *et al.* (2013). Complete mitochondrial genome
631 sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA
632 fragments. *Proceedings of the National Academy of Sciences, USA* **110:** 15758–15763.
633 Darling A.C.E., Mau B., Blattner F.R., Perna N.T. (2004). Mauve: Multiple alignment of
634 conserved genomic sequence with rearrangements. *Genome Research* **14:** 1394–1403.
635 Diamond J. (2002). Evolution, consequences and future of plant and animal domestication.
636 *Nature* **418**: 700-707.
637 di Lernia, S. and Manzi, G. (eds). 2002. Sand, Stones and Bones. The archaeology of death
638 in the Wadi TanezzuJt valley (5000-2000 BP). Firenze: All'Insegna dell Giglio.
639 Estrada O., Breen J., Richards S.M., Cooper A. (2018). Ancient plant DNA in the genomic
640 era. *Nature Plants* **4:** 394–396. Available at: http://dx.doi.org/10.1038/s41477-018-
641 0187-9.
642 Flowers J.M., Hazzouri K.M., Gros-Balthazard M., Mo Z., Koutroumpa K., Perrakis A.,
643 Ferrand S., Khierallah H.S.M., Fuller D.Q., Aberlenc F., *et al.* (2019). Cross-species
644 hybridization and the origin of North African date palms. *Proceedings of the National*
645 *Academy of Sciences, USA* **116:** 1651–1658.
646 Fuller D.Q., Denham T., Arroyo-Kalin M., Lucas L., Stevens C.J., Qin L., Allaby R.G.,
647 Purugganan M.D. (2014). Convergent evolution and parallelism in plant domestication
648 revealed by an expanding archaeological record. *Proceedings of the National Academy*
649 *of Sciences, USA* **111**: 6147-6152.
650 Gillet J.B. (1981). Section C: Botanical samples. In Eridu, F. Safar and A. M. Mustafa, eds.
651 (Baghdad: State Organisation of Antiquities and Heritage), pp. 317–318.
652 Gros-Balthazard M. (2013). Hybridization in the genus *Phoenix*: a review. *Emirates J. Food*
653 *Agric.* **25:** 831–842.
654 Gros-Balthazard M., Galimberti M., Kousathanas A., Newton C., Ivorra S., Paradis L.,
655 Vigouroux Y., Carter R., Tengberg M., Battesti V., *et al.* (2017). The discovery of wild
656 date palms in Oman reveals a complex domestication history involving centers in the
657 Middle East and Africa. *Current Biology* **27:** 2211–2218

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gros-Balthazard M., Hazzouri K.M., Flowers J.M. (2018). Genomic insights into date palm origins. *Genes* **9**: 502–523.

Gros-Balthazard M., Baker W.J., Leitch I., Pellicer K., Powell R.F., Bellot S. (2020a). Systematics and evolution of the genus *Phoenix*: towards understanding the date palm's origins. *ResearchGate* **10.13140/RG.2.2.14666.59847**

Gros-Balthazard M., Battesti V., Ivorra S., Paradis L., Aberlenc F., Zango O., Zehdi S., Moussouni S., Naqvi S.A., Newton C., *et al.* (2020b). On the necessity of combining ethnobotany and genetics to assess agrobiodiversity and its evolution in crops: a case study on date palms (*Phoenix dactylifera* L.) in Siwa Oasis, Egypt. *Evolutionary Applications* **13:** 1818-1840.

Gutaker R.M., Burbano H.A. (2017). Reinforcing plant evolutionary genomics using ancient DNA. *Current Opinion in Plant Biology* **36:** 38–45.

Günter T., Nettelblad C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics* **15:** e1008302.

Hajdas I. (2008). The radiocarbon dating and its applications in Quaternary studies. *Quat. Sci. J.* **57:** 2–24.

Hazzouri K.M., Gros-Balthazard M., Flowers J.M., Copetti D., Lemansour A., Lebrun M., Masmoudi K., Ferrand S., Dhar M.I., Fresquez Z.A., *et al.* (2019). Genome-wide association mapping of date palm fruit traits. *Nature Communications* **10;** 1–14. Available at: http://dx.doi.org/10.1038/s41467-019-12604-9.

Hope C. (1977). Excavations at Malkata and the Birket Habu 1971-1974: jar sealings and amphorae of the 18th dynasty (Warminster: Aris and Phillips).

Hufford M.B., Xu X., Van Heerwaarden J., Pyhäjärvi T., Chia J.M., Cartwright R.A., Elshire R.J., Glaubitz J.C., Guill K.E., Kaeppler S.M., Lai, J. (2012). Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44:** 808-811.

Jónsson H., Ginolhac A., Schubert M., Johnson P.L.F., Orlando L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29:** 1682–1684.

Korlević P., Gerber T., Gansauge M.T., Hajdinjak M., Nagel S., Aximu-Petri A., Meyer M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59:** 87–93.

Korneliussen T.S., Albrechtsen A., Nielsen R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**: 1.

Krueger F. (2015). Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Inst.

Larson G., Piperno D.R., Allaby R.G., Purugganan M.D., Andersson L., Arroyo-Kalin M., Barton L., Vigueira C.C., Denham T., Dobney K., Doust A.N. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences, USA* **11**: 6139-6146.

Li H., Ruan J., Durbin R. (2008). Mapping short DNA sequencing reads and calling variants using mapping. *Genome Research* **18**: 1851–1858. Available at: http://pixfunlobdot.59.to/content/18/11/1851.full.

Ling H.Q., Zhao S., Liu D., Wang J., Sun H., Zhang C., Fan H., Li D., Dong L., Tao Y., Gao C. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496:** 87-90.

Malleson, C., & Miracle, R. (2018). Giza Botanical Database [WWW Document]. https ://doi.org/10.6078/M7JH3J99

Martin G.T., Ward H., Emery W. (1981). The sacred animal necropolis at North Saqqa□ra. The southern dependencies of the main temple complex (Michigan: Egypt Exploration Society).

McGinnis S., Madden T.L. (2004). BLAST: At the core of a powerful and diverse set of

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

709 sequence analysis tools. *Nucleic Acids Research* **32**: W20.
710 McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K.,
711 Altshuler D., Gabriel S., Daly M., *et al.* (2010). The genome analysis toolkit: A
712 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
713 *Research* **20:** 1297–1303.
714 Mattingly and Wilson (2010) Concluding thoughts, made in Fazzan? In D.J. Mattingly (Ed)
715 Archaeology of Fazzan, Vol 3: Excavations of C.M. Daniels, Society for Libyan
716 Studies Monograph, London, 523-530.
717 Meisner J., Albrechtsen A. (2018). Inferring population structure and admixture proportions
718 in low-depth NGS data. *Genetics* **210:** 719-731.
719 Mohamoud Y.A., Mathew L.S., Torres M.F., Younuskunju S., Krueger R., Suhre K., Malek
720 J.A. (2019). Novel subpopulations in date palm (*Phoenix dactylifera*) identified by
721 population-wide organellar genome sequencing. *BMC Genomics* **20:** 1–7.
722 Murray M.A. (2000). Fruits, vegetables, pulses and condiments. In Ancient Egyptian
723 materials and technology, N. P.T. and I. Shaw, eds. (Cambridge: Cambridge University
724 Press), pp. 609–655.
725 Němec M., Wacker L., Gäggeler H. (2010). Optimization of the graphitization process AT
726 AGE-1. *Radiocarbon* **52:** 1380–1393.
727 Newton C., Gros-Balthazard M., Ivorra S., Paradis L., Pintaud J.-C., Terral J.-F. (2013).
728 *Phoenix dactylifera* and *P. sylvestris* in Northwestern India: a glimpse of their complex
729 relationships. *Palms* **57:** 37-50.
730 Parkinson R. (2008) The Painted Tomb-Chapel of Nebamun, The British Museum, London.
731 Paijmans J.L.., Baleka S., Henneberger K., Barlow A. (2017). Sequencing single-stranded
732 libraries on the Illumina NextSeq 500 platform. *arXiv* **1711.11004**.
733 Pedersen J.S., Valen E., Velazquez A.M.V., Parker B.J., Rasmussen M., Lindgreen S., Lilje
734 B., Tobin D.J., Kelly T.K., Vang S., *et al.* (2014). Genome-wide nucleosome map and
735 cytosine methylation levels of an ancient human genome. *Genome Research* **24:** 454–
736 466.
737 Pelling R (2013). 'The archaeobotanical remains' in Mattingly, D J, Daniels, C M, Dore, J N,
738 Edwards, D, Leone, A, & Thomas, D C (eds) The Archaeology of Fazzān. Volume 4,
739 Survey and Excavations at Old Jarma (Ancient Garama) carried out by C. M. Daniels
740 (1962-69) and the Fazzān Project (1997-2001). London: Society for Libyan Studies. 4:
741 473-94.
742 Pintaud J., Zehdi S., Couvreur T., Barrow S., Henderson S., Aberlenc-Bertossi F., Tregear J.,
743 Billotte, N. (2010). Species delimitation in the genus *Phoenix* (Arecaceae) based on
744 SSR markers, with emphasis on the identity of the date palm (*Phoenix dactylifera*). In
745 Diversity, phylogeny, and evolution in the monocotyledons, O. Seberg, G. Petersen, B.
746 Anders, and D. Jerrold, eds. (Aarhus: Aarhus University Press), pp. 267–286.
747 Pintaud J.C., Ludeña B., Aberlenc-Bertossi F., Gros-Balthazard M., Ivorra S., Terral J.F.,
748 Tengberg M., Hernández I.S., González-Pérez M.A., Sosa P., *et al.* (2013).
749 Biogeography of the date palm (*Phoenix dactylifera* L., Arecaceae): insights on the
750 origin and on the structure of modern diversity. *Acta Hortic.* **994:** 19–38.
751 Purugganan M.D. (2019). Evolutionary insights into the nature of plant domestication.
752 *Current Biology* **29:** 705-714.
753 Ramos-Madrigal J., Smith B.D., Moreno-Mayar J.V., Gopalakrishnan S., Ross-Ibarra J.,
754 Gilbert M.T.P., Wales N. (2016). Genome sequence of a 5,310-year-old maize cob
755 provides insights into the early stages of maize domestication. *Current Biology* **26:**
756 3195–3201.
757 Reimer P.J., Bard E., Bayliss A., Beck J.W., Blackwell P.G., Ramsey C.B., Buck C.E., Cheng
758 H., Edwards R.L., Friedrich M., *et al.* (2013). IntCal13 and Marine13 radiocarbon age
759 calibration curves 0–50,000 years cal BP. *Radiocarbon* **55:** 1869-1887.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

760 Richter T., Arranz-Otaegui A., Yeomans L., Boaretto, E. (2017). High resolution AMS dates
761 from Shubayqa 1, northeast Jordan reveal complex origins of Late Epipalaeolithic
762 Natufian in the Levant. *Scientific Reports* **7:** 1-10.
763 Schubert M., Ermini L., Sarkissian C. Der, Jónsson H., Ginolhac A., Schaefer R., Martin
764 M.D., Fernández R., Kircher M., McCue M., *et al.* (2014). Characterization of ancient
765 and modern genomes by SNP detection and phylogenomic and metagenomic analysis
766 using PALEOMIX. *Nature Protocols* **9:** 1056–1082.
767 Scott M.F., Botigué L.R., Brace S., Stevens C.J., Mullin V.E., Stevenson A., Thomas M.G.,
768 Fuller D.Q., Mott R. (2019). A 3,000-year-old Egyptian emmer wheat genome reveals
769 dispersal and domestication history. *Nature Plants* **5:** 1120–1128.
770 Skotte L., Korneliussen T.S., Albrechtsen A. (2013). Estimating individual admixture
771 proportions from next generation sequencing data. *Genetics* **195:** 693–702.
772 Soraggi S., Wiuf C., Albrechtsen, A. (2018). Powerful inference with the D-statistic on low-
773 coverage whole-genome data. *G3 Genes, Genomes, Genetics* **8:** 551–566.
774 Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis
775 of large phylogenies. *Bioinformatics* **30:** 1312–1313.
776 Stitzer M.C., Ross-Ibarra J., 2018. Maize domestication and gene interaction. *New Phytologist*
777 **220:** 395-408.
778 Stuiver M., Polach H.A. (1977). Reporting of C-14 data - Discussion. *Radiocarbon* ***19*:** 355–
779 363.
780 Swarts K., Gutaker R.M., Benz B., Blake M., Bukowski R., Holland J., Kruse-Peeples M.,
781 Lepak N., Prim L., Romay M.C., *et al.* (2017). Genomic estimation of complex traits
782 reveals ancient maize adaptation to temperate North America. *Science* **357:** 512–515.
783 Synal H.A., Stocker, M., and Suter, M. (2007). MICADAS: A new compact radiocarbon
784 AMS system. *Nuclear Instruments and Methods in Physical Research Section B Beam*
785 *Interaction with Materials and Atoms* **259:** 7–13.
786 Tallet P. (2017). Du pain et des céréales pour les équipes royales: Le grand papyrus
787 comptable du ouadi el-Jarf (papyrus H). *Nehet* **5:** 99–117.
788 Tengberg M. (2012). Beginnings and early history of date palm garden cultivation in the
789 Middle East. J. Arid Environ. *86*, 139–147. Available at:
790 http://dx.doi.org/10.1016/j.jaridenv.2011.11.022.
791 Terral J.F., Newton C., Ivorra S., Gros-Balthazard M., de Morais C.T., Picq S., Tengberg M.,
792 Pintaud J.C. (2012). Insights into the historical biogeography of the date palm (*Phoenix*
793 *dactylifera* L.) using geometric morphometry of modern and ancient seeds. *Journal of*
794 *Biogeography* **39:** 929–941.
795 Thomer, A.K., Twidale, M.B., Guo, J., and Yoder, M.J. (2016). Picard Tools. In Conference
796 on Human Factors in Computing Systems - Proceedings.
797 Van der Veen M., Westley B. (2010). 'Palaeoeconomic studies' in Mattingly, D.J., Daniels,
798 C.M., Dore, J.N., Edwards, D. and Hawthorne, J. (eds) The Archaeology of Fazzan Vol
799 3, Excavations carried out by C. M. Daniels. London: Society for Libyan Studies,
800 Departments of Antiquities, 488-522p.
801 Vartavan C., Amoros V.A., Arakelyan, A. (1997). Codex of ancient Egyptian plant remains
802 (London: SAIS).
803 Wales N., Andersen K., Cappellini E., Ávila-Arcos M.C., Gilbert M.T.P. (2014).
804 Optimization of DNA recovery and amplification from non-carbonized
805 archaeobotanical remains. *PLoS One* **9:** e86827.
806 Zhou Y., Ma Y., Zeng J., Duan L., Xue X., Wang H., Lin T., Liu Z., Zeng K., Zhong Y.,
807 Zhang S. (2016). Convergence and divergence of bitterness biosynthesis and regulation
808 in Cucurbitaceae. *Nature Plants* **2:** 1-8.
809 Zohary, D., Hopf, M., and Weiss, E. (2015). Domestication of plants in the Old World
810 (Oxford: Oxford University Press).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400408; this version posted November 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

811
812
813 **Figure legends**
814 **Figure 1**. Phylogenetic placement of the Saqqara specimen amongst *Phoenix* species. (**a**)
815 Maximum Likelihood analysis of whole plastome sequences showing the placement of the
816 Saqqara specimen amongst *Phoenix* species (accession numbers for each terminal are
817 provided in Fig. S2). (**b**) Individual of *Phoenix dactylifera* bearing fruits. (**c**) Individual of the
818 sugar date tree (*P. sylvestris*), the closest known leaving relative of *P. dactylifera*. (**d**) Male
819 inflorescence of *P. theophrasti*. (**e**) Individuals of *P. atlantica*. (**f**) Saqqara 26796, a jar-
820 stopper made of date palm leaflets (excavation inventory number 102, Kew Economic Botany
821 Collection number 26796). (**g**) A similar object to the Saqqara specimen number 26796, also
822 made of date palm leaflets thought to be a basket-lid and found in Saqqara. Photos: Penelope
823 Dawson (**b**), Sasha Barrow (**c**), John Dransfield (**d**), William J. Baker (**e**), Mark Nesbitt (**f**), ©
824 The Trustees of the British Museum CC BY-NC-SA 4.0 (**g**).
825
826 **Figure 2**. Genome ancestry of the Saqqara specimen. Population structure and principal
827 component analyses (PCA) based on estimated nuclear GLs derived from a highly fragmented
828 (**[a, c]**, GCA000413155.1) and a highly contiguous reference genome (**[b, d]**,
829 GCA0009389715.1). Structure analyses with population number (K) from 2-6 (**a, b**) show
830 admixture amongst wild and cultivated date palm populations, including the Saqqara leaf, and
831 closely related *Phoenix* species. The geographical origin of modern individuals of *P.*
832 *dactylifera* is provided at the bottom of the plot. Detailed cluster and delta likelihood values
833 from K 1-8 are provided on Fig. S4. Covariance matrices derived from PCA in (**b, d**) reveal a
834 close affinity of the Saqqara specimen with modern individuals of North African *P.*
835 *dactylifera* and the Cape Verde's *P. dactylifera*. The remaining individuals of *P. dactylifera*
836 not labelled in the plots belong to Asian populations.
837
838 **Figure 3**. Introgression of the Saqqara leaf with modern individuals of *P. theophrasti* inferred
839 from nuclear bases. (**a**) Results of D-statistic analyses derived from nuclear genotype
840 likelihoods (GLs) for the Saqqara date leaf amongst date palm populations and closely related
841 species (*P. atlantica*, *P. sylvestris* and *P. theophrasti*), with *P. reclinata* fixed as the outgroup,
842 using as a reference a highly fragment reference genome. The outcome of all possible
843 permutations conducted during the D-statistic test between all individuals sampled in this
844 study are provided in Table S3 and Figure S3. Table S4 provides the outcome of D-statitstic
845 conducted on all possible combinations between all individuals and using a contiguous
846 reference genome. (**b**) Three instances of D-statistic analyses for the Saqqara date leaf
847 conducted amongst populations of date palms and closely related species and a highly
848 fragmented reference genome supporting gene flow between *P. theophrasti* and the Saqqara
849 date leaf. The outcome of all possible permutations between populations and of analyses
850 conducted using a contiguous reference genome is provided in Tables S5 and S6.
851
852

(a)

Licuala montana

P. reclinata

100

85

P. theophrasti

100

P. canariensis

9.0e-4

African varieties

Asian varieties

P. sylvestris

100

99

99

Saqqara-26796

P. atlantica

87

Khalte

Medjool

Jabri

98

Khalte

99

Deglet

Wild individual

98

Siwi

P. sylvestris

Khalas

Wild individual

Rabia

99

Moshwaq

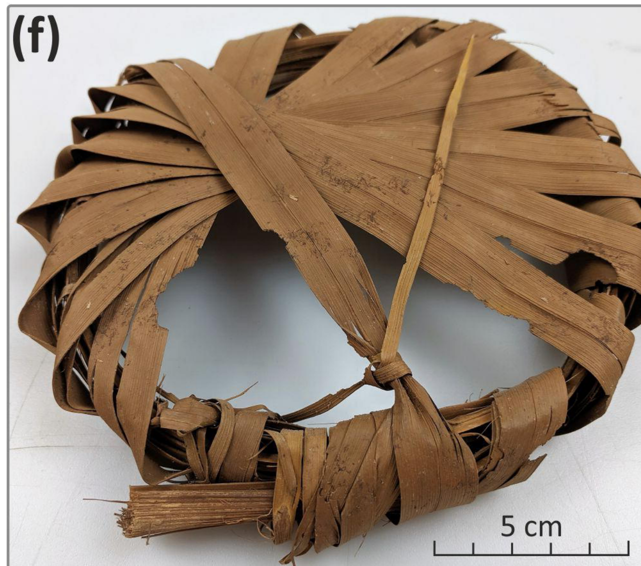
Sukkariat

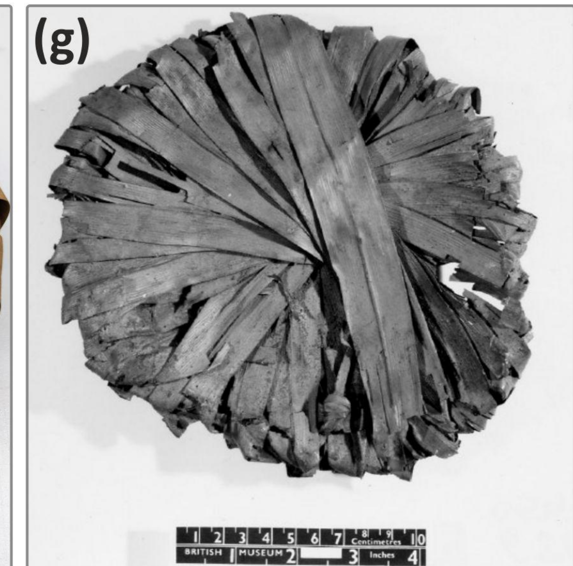
Dekhaini

Perny

Shalaby

P. dactylifera





(a)

Saqqara

P. theophrasti *P. atlantica* *P. dactylifera* *P. sylvestris* *P. canariensis*

Admixture proportion

K = 2
K = 3
K = 4
K = 5
K = 6

African Asian

(b)

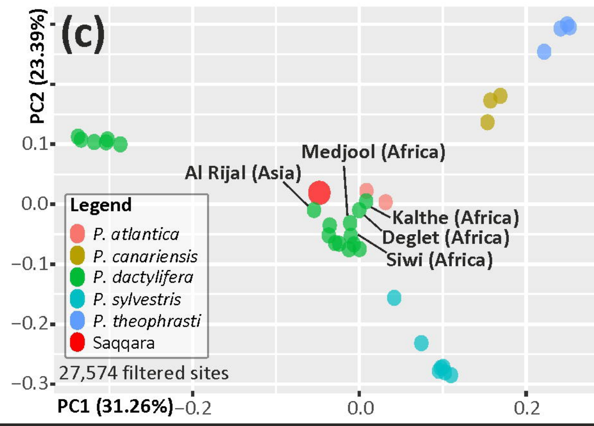
Saqqara

P. theophrasti *P. atlantica* *P. dactylifera* *P. sylvestris* *P. canariensis*

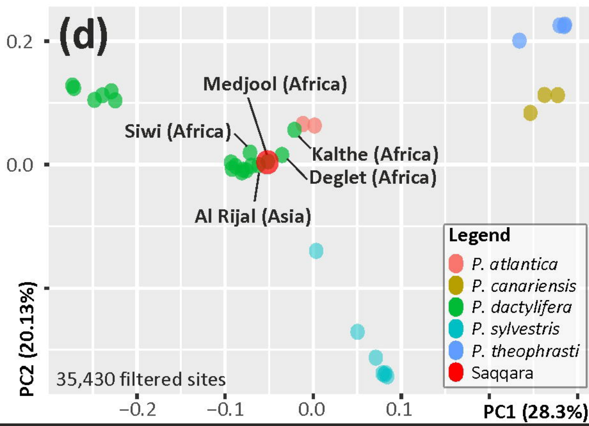
K = 2
K = 3
K = 4
K = 5
K = 6

African Asian

Khalas highly fragmented reference genome
198 scaffolds - 149 Mbases



Barhee BC4 contiguous reference genome
18 scaffolds - 380 Mbases



(a)

H2 H3 Out
D < 0

H2 H3 Out
D = 0

H2 H3 Out
D > 0

Topologies
H2: P. atlan. - H3: P. dact.
H2: P. sylv. - H3: P. atlan.
H2: P. sylv. - H3: P. dact.
H2: P. sylv. - H3: P. theo.
H2: P. theo. - H3: P. atlan.
H2: P. theo. - H3: P. dact.
H2: P. theo - H3: P. sylv.

Legend
Saqqara ancient date
Number of BABA patterns
Number of ABBA patterns

−0.75 −0.50 −0.25 0 0.25 0.50 0.75
D-stat

(b)

Outgroup
P. theophrasti (H3)
P. dactylifera (H2)
H1

D-stat: -0.085
Z: -3.936
***p*-value:** 0.000083

0 50 100 150 200 250 300 350

Outgroup
P. theophrasti (H3)
P. atlantica (H2)
H1

D-stat: -0.156
Z: -6.225
***p*-value:** 0

0 50 100 150 200 250 300 350

Outgroup
P. theophrasti (H3)
P. sylvestris (H2)
H1

D-stat: -0.141
Z: -6.433
***p*-value:** 0

0 50 100 150 200 250 300 350
Number of sites