HAL open science

Cross-species hybridization and the origin of North African date palms

Jonathan M Flowers, Khaled M Hazzouri, Muriel Gros-Balthazard, Ziyi Mo, Konstantina Koutroumpa, Andreas Perrakis, Sylvie Ferrand, Hussam S M Khierallah, Dorian Q Fuller, Frederique Aberlenc, et al.

To cite this version:

Jonathan M Flowers, Khaled M Hazzouri, Muriel Gros-Balthazard, Ziyi Mo, Konstantina Koutroumpa, et al.. Cross-species hybridization and the origin of North African date palms. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116 (5), pp.1651-1658. 10.1073/pnas.1817453116 . hal-03202324

HAL Id: hal-03202324

https://hal.science/hal-03202324v1

Submitted on 19 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Cross-species hybridization and the origin of North African date palms

Jonathan M. Flowers[a,b], Khaled M. Hazzouri[a,c], Muriel Gros-Balthazard[a], Ziyi Mo[a], Konstantina Koutroumpa[d], Andreas Perrakis[e], Sylvie Ferrand[a], Hussam S. M. Khierallah[f,1], Dorian Q. Fuller[g], Frederique Aberlenc[h], Christini Fournaraki[e], and Michael D. Purugganan[a,b,2]

[a]Center for Genomics and Systems Biology, New York University Abu Dhabi Research Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates; [b]Center for Genomics and Systems Biology, New York University, New York, NY 10003; [c]Khalifa Center for Genetic Engineering and Biotechnology, United Arab Emirates University, Al-Ain, United Arab Emirates; [d]Department of Systematic and Evolutionary Botany, University of Zurich, 8008 Zurich, Switzerland; [e]Mediterranean Plant Conservation Unit, International Centre for Advanced Mediterranean Agronomic Studies (CIHEAM) Mediterranean Agronomic Institute of Chania, 73100 Chania, Crete, Greece; [f]Date Palm Research Unit, College of Agriculture, University of Baghdad, Baghdad 10071, Iraq; [g]Institute of Archaeology, University College London, London WC1H 0PY, United Kingdom; and [h]Unité Mixte de Recherche (UMR) Diversity Adaptation and Development of Plants (DIADE), Institut de Recherche pour le Développement, 34394 Montpellier, France

Edited by John F. Doebley, University of Wisconsin-Madison, Madison, WI, and approved December 6, 2018 (received for review October 12, 2018)

Date palm (*Phoenix dactylifera* L.) is a major fruit crop of arid regions that were domesticated ∼7,000 y ago in the Near or Middle East. This species is cultivated widely in the Middle East and North Africa, and previous population genetic studies have shown genetic differentiation between these regions. We investigated the evolutionary history of *P. dactylifera* and its wild relatives by resequencing the genomes of date palm varieties and five of its closest relatives. Our results indicate that the North African population has mixed ancestry with components from Middle Eastern *P. dactylifera* and *Phoenix theophrasti*, a wild relative endemic to the Eastern Mediterranean. Introgressive hybridization is supported by tests of admixture, reduced subdivision between North African date palm and *P. theophrasti*, sharing of haplotypes in introgressed regions, and a population model that incorporates gene flow between these populations. Analysis of ancestry proportions indicates that as much as 18% of the genome of North African varieties can be traced to *P. theophrasti* and a large percentage of loci in this population are segregating for single-nucleotide polymorphisms (SNPs) that are fixed in *P. theophrasti* and absent from date palm in the Middle East. We present a survey of *Phoenix* remains in the archaeobotanical record which supports a late arrival of date palm to North Africa. Our results suggest that hybridization with *P. theophrasti* was of central importance in the diversification history of the cultivated date palm.

introgression | archaeobotany | domestication | crop wild relative | range expansion

Domesticated crops are among the most evolutionarily successful species in the world. From geographically restricted centers of origin, many domesticated species have dramatically and rapidly expanded their ranges, adapting to new environments and cultures within the span of a few hundred to thousands of generations. The precise genetic and evolutionary mechanisms that allow crop species to adapt to multiple environments remain unclear, but understanding these mechanisms provides insights into the nature of species range expansion and the dynamics of human cultural evolution since the Neolithic. Identifying how crop species range expansion has occurred in the past may also suggest new approaches for future crop breeding efforts, especially in light of climate change.

The date palm (*Phoenix dactylifera* L.) is a dioecious species in the Arecaceae (formerly Palmae) family and the most important fruit-bearing crop in arid regions of the Middle East and North Africa. Date palms grow primarily in hot arid habitats including desert oases or well-irrigated small farms or plantations where they are propagated via a mixed clonal–sexual system. The traditional range of date palm cultivation extends from Morocco to Egypt in North Africa; the Arabian Peninsula, Iraq, and Iran in the Middle East; and Pakistan in South Asia (1) (Fig. 1*A*). Population genetic analysis using microsatellites (2–6) and whole-genome sequences (6–8) indicates that Middle Eastern date palms are genetically differentiated from North African *P. dactylifera*, with a possible hybrid zone in Egypt and the Sudan (5, 7).

Date palms are one of the earliest domesticated tree crops. The oldest evidence of exploitation comes from seed remains excavated from Dalma Island, Abu Dhabi, and Kuwait that date to the Arabian Neolithic ∼7,000 y before present (yBP) (9). Evidence of date palm cultivation appears somewhat later in Mesopotamia at ∼6,700–6,000 yBP (10, 11) and in the Levant at ∼5,700–5,500 yBP (12). The early evidence of cultivation at the eastern edge of the Fertile Crescent and the Upper Arabian Gulf supports an ancient center of origin of date palm domestication in this region (1, 9, 12, 13). Evidence for morphological change in archaeological date stones suggests selection of larger-fruited domesticated forms took place between ∼5,000 yBP and 2,000 yBP in the Near East (14). The sugar date palm, *Phoenix sylvestris* (L.) Roxb., is the sister species of *P. dactylifera* (1, 15), but unlikely to be the wild progenitor of domesticated date palm (15, 16). A recent study identified wild *P. dactylifera* populations in Oman, which may represent a relictual population of the wild progenitor of cultivated date palms in the Middle East (6).

Significance

Date palm (*Phoenix dactylifera* L.) is one of the oldest tree crop species in the world and is a major fruit crop of arid regions of the Middle East and North Africa. We use whole-genome sequence data from a large sample of *P. dactylifera* and its wild relatives to show that hybridization between date palms and *Phoenix theophrasti* Grueter—a species endemic to the Eastern Mediterranean—is associated with the diversification of date palm.

Author contributions: J.M.F., D.Q.F., C.F., and M.D.P. designed research; J.M.F., K.M.H., K.K., A.P., S.F., H.S.M.K., D.Q.F., F.A., and C.F. performed research; J.M.F., K.M.H., M.G.-B., Z.M., and D.Q.F. analyzed data; and J.M.F. and M.D.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The sequences reported in this paper have been deposited in the Sequence Read Archive database (accession no. PRJNA495685). SNP data have been deposited in the Dryad Digital Repository (https://doi.org/10.5061/dryad.tm40gd8).

[1]Deceased August 15, 2018.

[2]To whom correspondence should be addressed. Email: mp132@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817453116/-/DCSupplemental.

Published online January 14, 2019.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

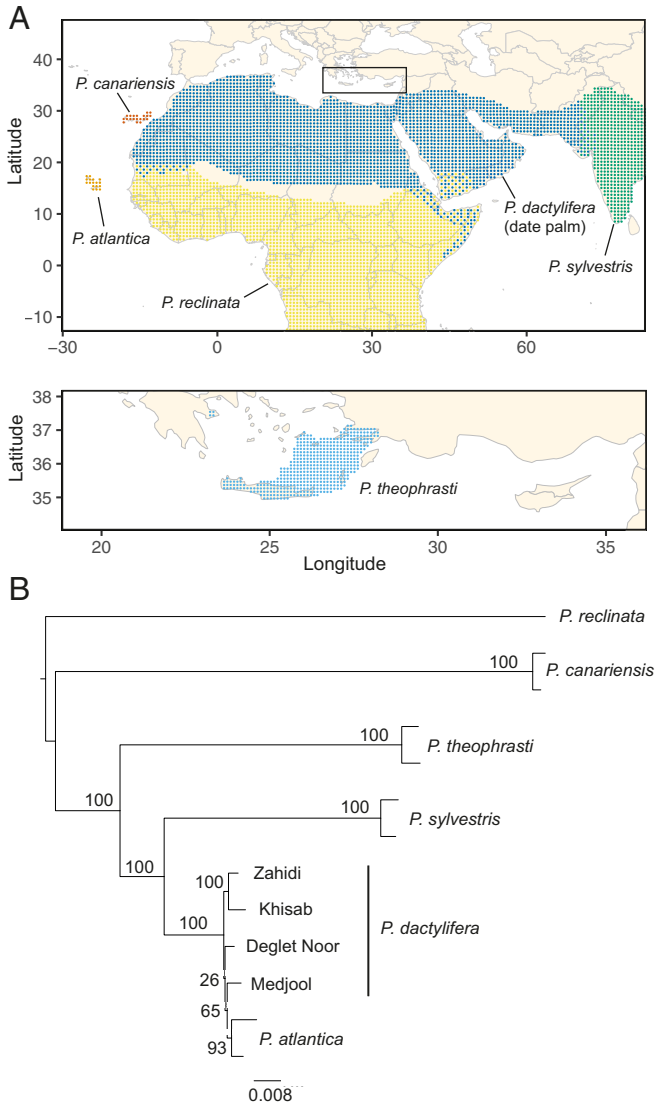

Fig. 1. The geographic distribution and phylogeny of cultivated date palm (*P. dactylifera*) and its wild relatives. (*A*) Approximate geographic distribution of *Phoenix* species included in this study (27). (*B*) Maximum-likelihood phylogeny of *Phoenix* species based on a subset of SNPs from whole-genome sequencing data. Support values are the percentage of RAxML bootstrap replicates that support each node. The tree is midpoint rooted.

Alternate origin hypotheses propose a domestication center of date palm in North Africa (1, 17, 18). These hypotheses are poorly supported by the archaeobotanical record, however, as evidence of date palm cultivation appears ∼3,000 y later in the records of this region compared with the Middle East. *P. dactylifera* remains in the Nile Valley are found as early as the predynastic period, more than 5,000 yBP (19, 20), but cultivation may not have begun until the Middle Kingdom, ∼4,000 yBP (14, 21). In Libya, cultivation began ∼2,800–2,400 yBP (18, 22). Date palm is unknown from elsewhere in the Saharan Maghreb and the sub-Saharan Sahel until much later (23–25).

The archaeobotanical record therefore is consistent with a late arrival of date palm to North Africa and suggests a model where date palm was domesticated in the Near or Middle East and later expanded to the African continent (26). Paradoxically, however, nucleotide diversity in North African date palm is at least 20% higher than that in the Middle East (6, 7) and diversity at microsatellite loci is comparable between the two populations (5). These observations are inconsistent with a bottleneck associated with founder-effect colonization in North Africa. Evidence of an unknown ancestry component in North African cultivated date palm further implies a more complex history in this region (6). The genetic distinctiveness of North African date palms, their absence from early archaeological sites in North Africa, and elevated levels of nucleotide diversity perpetuate the enigma of the origins of date palms in the western part of their range.

To examine the origin of North African date palms, we resequenced a large sample of cultivars from Morocco to Pakistan and five wild *Phoenix* relatives that occur either peripatrically or allopatrically to cultivated date palm (Fig. 1*A*). Here we present evidence that the North African population is the product of introgressive hybridization between cultivated date palm and the Cretan date palm *Phoenix theophrasti* Greuter, a species endemic to Crete and the Eastern Mediterranean. We demonstrate that introgression has been central in shaping patterns of diversity genome-wide, which supports introgression as being an important factor that shaped the domestication history of date palm. The growing list of examples of interspecific hybridization associated with domestication suggests that hybridization is a common evolutionary genetic mechanism for the adaptation of both annual and perennial crops.

Results

Plant Samples, Nucleotide Diversity, and the Phylogeny of *Phoenix* Species. We resequenced the genomes of 71 cultivated date palm varieties and multiple genomes from each of its 5 closest wild relatives in the genus *Phoenix* to address questions about the origin of the domesticated date palm (*P. dactylifera*). Our sample set included common varieties of dates from its traditional range of cultivation (7) and 2–18 individuals from wild relatives that include *P. sylvestris*; *P. theophrasti*; *Phoenix atlantica* A. Chev.; *Phoenix canariensis* hort. ex Chabaud; and a single member of a putative outgroup species from sub-Saharan Africa, *Phoenix reclinata* Jacq. (*SI Appendix*, Table S1).

Samples were sequenced to moderate to deep coverage (5–56×, mean = 22×; *SI Appendix*, Table S2) using 2 ×100 bp paired-end Illumina sequencing. The close relationships of the wild relatives to date palm allowed us to align >96% of short reads of each sequenced genome (*SI Appendix*, Table S2) to the draft assembly of the date palm genome (28). We identified 14,402,469 single-nucleotide polymorphisms (SNPs) across *Phoenix* species which we used in population and phylogenomic analysis.

Reconstruction of the phylogeny of date palm and its wild relatives using the whole-genome SNP data supported the close relationship of *P. sylvestris*, *P. theophrasti*, *P. atlantica*, and domesticated date palm *P. dactylifera*, (the "date palm group") (1). *P. canariensis* and *P. reclinata* are more distant relatives (Fig. 1), consistent with previous analyses (1, 15) and the Cape Verde Islands endemic, *P. atlantica*, is a member of a well-supported *P. dactylifera* clade consistent with the samples in our analysis being feral date palms (1, 6). Within the date palm group, *P. sylvestris* is the sister species of date palm, and *P. theophrasti* is a more diverged species (Fig. 1). Estimates of nucleotide diversity in the *Phoenix* wild relatives suggest *P. canariensis* ($\pi = 0.0117$) > *P. sylvestris* ($\pi = 0.0105$) > *P. theophrasti* ($\pi = 0.0072$), while estimates in the two populations of date palm indicate higher diversity in North Africa compared with the Middle East (*SI Appendix*, Table S3) as previously reported (6, 7).

Population Clustering Suggests That North African Date Palms Are Admixed. To identify genetic clusters in *P. dactylifera* and its wild relatives, we conducted model-based clustering with the program STRUCTURE (29). STRUCTURE runs with the independent allele-frequency model produced clusters at K = 2 to K = 4 that showed evidence of mixed ancestry of North African date palm.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

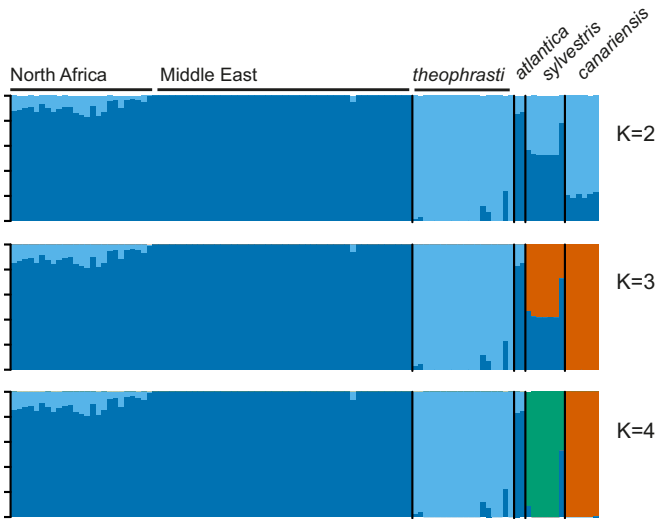

Fig. 2. North African date palm has mixed ancestry from Middle Eastern date palm and a wild relative, *P. theophrasti*. Shown is a STRUCTURE (29) diagram for K = 2 to K = 4 using the independent allele-frequency model.

Approximately 5–15% of the North African sample genomes were shared with *P. theophrasti* and the remainder with Middle Eastern date palm (Fig. 2 and *SI Appendix*, Table S4). Higher K values did not assign samples to any additional meaningful sources.

Running STRUCTURE with the independent model may underestimate the number of clusters. We therefore repeated the analysis with the correlated allele-frequency model (*SI Appendix*, Fig. S1 and Table S5). K = 2 to K = 4 produced qualitatively similar results as above. However, At K = 5 and higher, a third ancestry component appears in the North African population which does not trace to any population in our analysis (*SI Appendix*, Fig. S1). This could represent an additional source of variation in North Africa or could be attributable to known artifacts of the correlated frequencies model, which is subject to oversplitting at higher K values (29, 30).

Finally, we conducted a set of analyses restricted to species pairs ("hierarchical" analysis) (31). Pairwise STRUCTURE runs limited to date palm samples and those from a single wild relative species yielded results qualitatively similar to the corresponding correlated and independent allele-frequency analyses that included all samples. The analysis restricted to *P. theophrasti* and date palm also supported the mixed ancestry of North African samples and this wild relative (*SI Appendix*, Fig. S2 and Table S6). STRUCTURE analyses are therefore consistent with the North African population being admixed between cultivated date palm and this wild relative or a *P. theophrasti*-like population.

Tests of Admixture Indicate North African Date Palms Are Products of Interspecific Hybridization. Evidence of mixed ancestry of date palms in the STRUCTURE analysis is suggestive of hybridization, but interpretation must be treated with caution (30). We therefore conducted explicit tests of admixture to establish whether North African date palms are indeed the product of interspecific hybridization between Middle Eastern *P. dactylifera* and the wild Cretan species, *P. theophrasti*. First, we performed ABBA–BABA, or *D* tests (32, 33), to test for excesses of shared derived polymorphism that can be attributed to gene flow between species. We conducted all combinations of tests among date palm and its wild relatives by assuming the population configuration *D*(P1,P2,P3,O), where P1 and P2 are Middle Eastern and North African populations of domesticated date palms, respectively; P3 is a wild relative; and *P. reclinata* is the outgroup (O). A positive *D* statistic in this configuration indicates gene flow between North African date palm (P2) and a wild relative (P3), while negative tests indicate gene flow between Middle Eastern date palm (P1) and a wild species.

We observe a positive *D* statistic for the test with *P. theophrasti* as the wild relative ($D = 0.58$) (SE $\pm$ 0.02; $Z = 37.24$; Fig 3*A* and *SI Appendix*, Table S7), which indicates an excess of shared derived alleles between *P. theophrasti* and North African date palm and suggests interspecific hybridization between this wild relative and date palm. This result is robust to the choice of outgroup, as replacing *P. reclinata* with *P. canariensis* produces

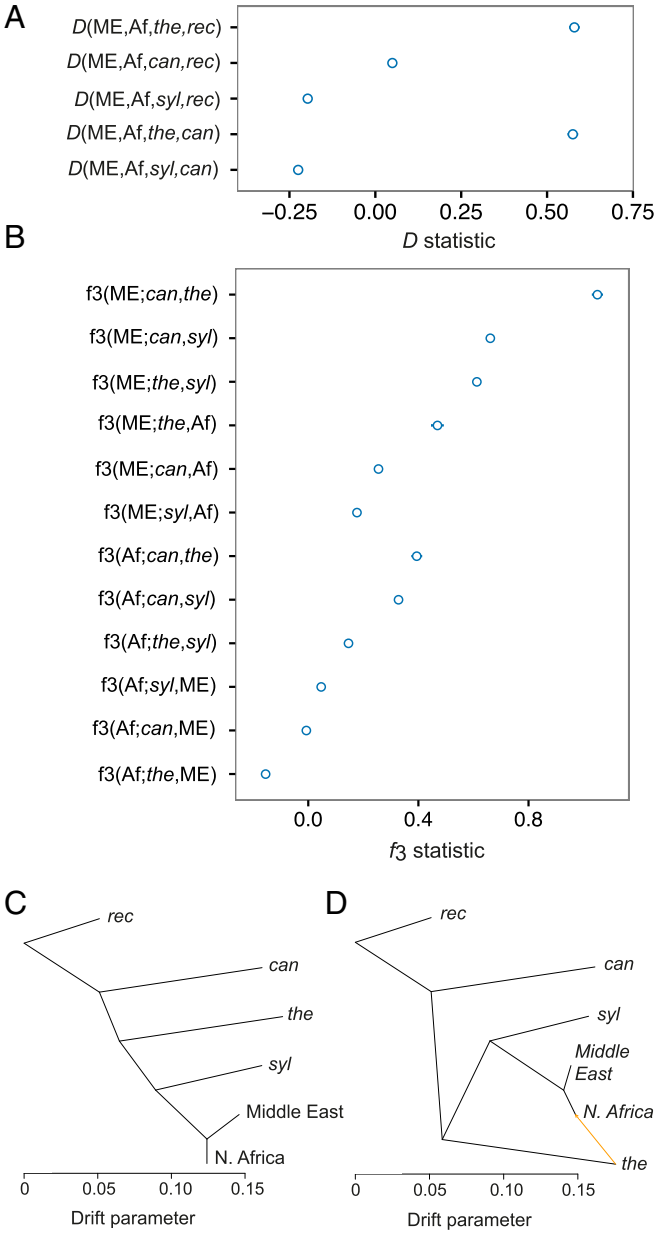

Fig. 3. Admixture between North African date palm and *P. theophrasti*. (*A*) *D* statistics $\pm$ SE. (*B*) f_3 statistics $\pm$ SE. (*C* and *D*) Maximum-likelihood tree based on *TreeMix* with block size of 3,000 SNPs for (*C*) zero ($m = 0$) and (*D*) one migration ($m = 1$) event. These models explain 98.6% and 99.9% of the variance in relatedness among populations. The migration edge pointing from *P. theophrasti* to North African date palm has a mixture weight of 15.7%. Abbreviations: Af, North Africa; *can*, *P. canariensis*; ME, Middle East; *rec*, *P. reclinata*; *syl*, *P. sylvestris*; and *the*, *P. theophrasti*.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

a similar test outcome (Fig. 3*A* and *SI Appendix*, Table S7). Tests where samples from the Maghreb countries are replaced with cultivars from Egypt also show an excess of derived allele sharing with *P. theophrasti* although the degree of sharing is less than in cultivars sampled from elsewhere in North Africa (*SI Appendix*, Table S7). *D* tests including *P. sylvestris* and *P. canariensis* are negative and positive, respectively, possibly supporting additional gene flow with date palm and other wild relatives. However, these are minor contributions compared with those of *P. theophrasti* and may represent hybridization in anthropogenic contexts (Fig. 3*A* and *SI Appendix*, Table S7).

Second, we tested for admixture between date palm and its wild relatives with f_3 tests (34, 35). The f_3 statistic tests whether a population is the product of admixture between two reference, or source, populations with a significantly negative f_3 statistic supporting admixture and all other outcomes being uninformative about the admixture history (34). The test with North African date palm as the test population and *P. theophrasti* and Middle Eastern date palm as reference populations was negative and suggests that North African date palms have a mixed ancestry from these two populations [f_3(North Africa; Middle East, *P. theophrasti*) = −0.15; $|Z|$ = 19.8; Fig. 3*B* and *SI Appendix*, Table S8). All other f_3 tests either did not differ significantly from zero or were positive (Fig. 3*B* and *SI Appendix*, Table S8). The negative f_3(North Africa; Middle East, *P. theophrasti*) test supports admixture between date palm and a wild species, and also suggests that the Middle Eastern population or a Middle Eastern-like population is a source of alleles segregating in North Africa.

We modeled the population history of date palm and its relatives with the maximum-likelihood method *TreeMix* (36). This modeling approach infers the population tree based on the covariance in allele frequencies among populations and incorporates admixture through the addition of migration edges between populations that are poor fits to the strict tree model. To determine whether gene flow between date palm and *P. theophrasti* is supported by the data, we first inferred the maximum-likelihood tree without migration (Fig. 3*C*). A model rooted with *P. reclinata* yielded the same relationships among *Phoenix* species as the maximum-likelihood phylogeny reported above (Fig. 1*B*) and explained 98.6% of the variance in relatedness between populations. Addition of a single migration event to the model predicted gene flow between *P. theophrasti* and the lineage leading to the North African population of date palm (Fig. 3*D*) and increased the percentage of variance in relatedness explained to 99.9%. This migration edge was stable in independent runs of *TreeMix* with both different species included in the model and with different block sizes to account for linkage disequilibrium (LD) (*SI Appendix*, Table S9). Together, the admixture tests and *TreeMix* results suggest that the ancestry of North African date palm can be traced to Middle Eastern date palm and *P. theophrasti*.

Mixed Ancestry of North African Date Palm. We estimated the percentage of ancestry of North African date palms derived from *P. theophrasti* and Middle Eastern date palm using f_4 statistics (35) and *TreeMix* mixture weights (36). First, the ratio of appropriate f_4 statistics can provide an estimate of the proportion of North African ancestry that is derived from the Middle East and *P. theophrasti* by assuming the phylogenetic relationships in Fig. 1*B*. We estimated the percentage of ancestry (α) derived from Middle Eastern date palm at 82% (SE ± 0.01; *SI Appendix*, Table S10) and the *P. theophrasti* fraction as $1 - \alpha$ = 18%. Repeating the estimation of α by replacing samples from the Maghreb with Egyptian samples as the test population yielded a smaller *P. theophrasti* and larger Middle Eastern component in these samples (α = 95%) (SE ± 0.007; *SI Appendix*, Table S10). Second, *TreeMix* models with gene flow provide mixture weights estimates for each migration edge, which approximate ancestry proportions (36). A mixture weight of 15.7% on the migration edge from *P. theophrasti* to North African date palm with m = 1 (Fig. 3*D*) is comparable to our estimate from the f_4-ratio approach. These results suggest that 5–18% of the North African date palm genome is derived from the Cretan date palm with varieties from countries west of Egypt sharing greater ancestry with this wild relative.

Introgressed Regions in the North African Date Palm Genome. We identified introgressed genomic segments in the North African population using a combination of approaches. We traced individual alleles in this population to their most likely source population by characterizing variant sites that are fixed for alternate alleles in the Middle Eastern population and *P. theophrasti*. This analysis yielded 1,556,435 nucleotide fixations, of which 90.2% are polymorphic in North African *P. dactylifera*. Of the sites segregating for a Middle Eastern-like and *P. theophrasti*-like allele in this population, the latter is typically the minor allele with most observed at low to moderate frequency. A total of 1,252,999 of 1,404,273, or 89.2% of these polymorphic sites in North Africa, have a *P. theophrasti*-like allele frequency of 30% or less. Only 9.8% of sites that are fixed between Middle Eastern date palm and *P. theophrasti* are fixed for the Middle Eastern-like allele in North Africa and none are fixed for the *P. theophrasti*-like variant.

To characterize regions with introgressed haplotypes that have risen to moderate-to-high frequency, we estimated the introgression fraction (f_d) for the population configuration *D*(Middle East, North Africa, *P. theophrasti*, *P. reclinata*) (above and *SI Appendix*, *SI Materials and Methods* and ref. 37). A positive f_d implies introgression from *P. theophrasti* into North African date palm. We identified blocks of two or more consecutive 5-kb intervals with an f_d in the upper 10th percentile of the genome-wide distribution of this statistic. Applying this criterion, we identified 1,281 introgressed segments of 10 kb or larger, which totaled 24.6 Mb of the draft assembly. The median length of tracts with outlier f_d was 15 kb and the largest region was 105 kb on scaffold NW_008246809.1. This approach yields an underestimate of tract lengths owing, in part, to the current fragmented state of the draft assembly. Moreover, the signature of introgression is not limited to these outlier regions as evidenced by an elevated introgression fraction and altered patterns of diversity throughout much of the genome (below). Fig. 4*A* shows f_d in sliding windows across the longest genome scaffolds. Fig. 4*B* shows an example 3-Mb region with both elevated f_d and Tajima's *D* in North Africa and the incidence of high-frequency *theophrasti*-like alleles in this region. A gene tree reconstructed from phased genotypes confirmed the existence of shared haplotypes between *P. theophrasti* and North African date palm cultivars in this region (Fig. 4*C*).

We conducted a phylogenetic analysis of chloroplast (cpDNA) and mitochondrial (mtDNA) genomes to examine patterns of introgression in organellar genomes and evaluate whether cultivated date palm *P. dactylifera* or *P. theophrasti* served as the maternal progenitor in the interspecific cross. Both neighbor-joining and maximum-likelihood analyses of cpDNA genome sequences largely supported the species phylogeny recovered from the whole-genome analysis (*SI Appendix*, Fig. S3). *P. theophrasti* samples formed a distinct clade indicating that cpDNA from this species is not introgressed in North Africa. Middle Eastern and North African date palms appear as distinct and well-supported lineages, although the Middle Eastern haplotype is also found in some North African samples, a result consistent with a previous report (5). While the North African clade is distinct and well supported, it shares a most recent common ancestor with *P. sylvestris*. Single *P. theophrasti* (Gölköy, Turkey) and *P. sylvestris* (Faisalabad, Pakistan) samples that were admixed with date palm in the STRUCTURE analysis based on nuclear genotypes (Fig. 2) possess *P. dactylifera* cpDNA

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

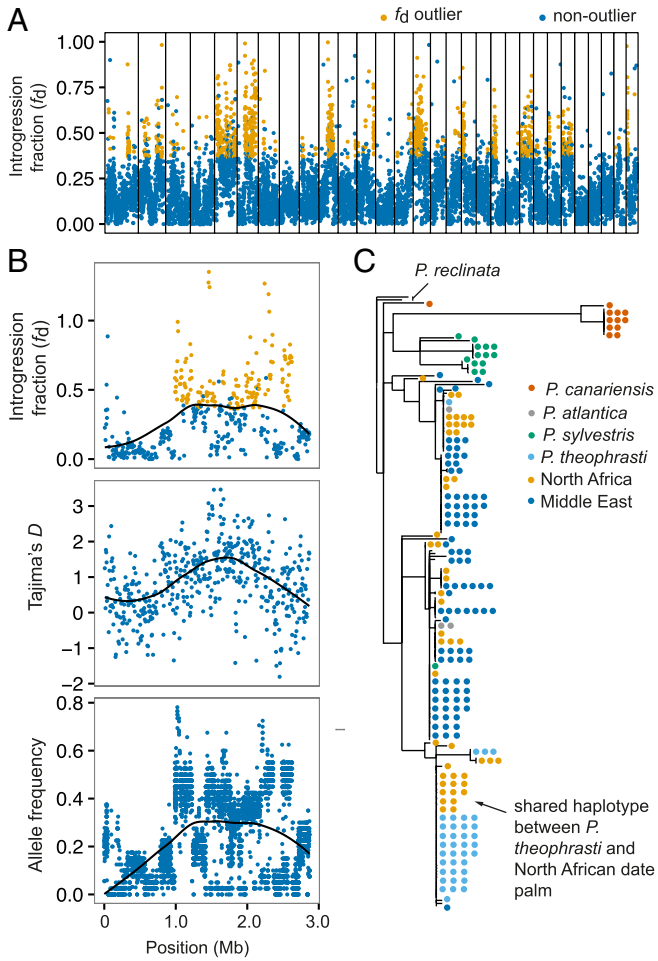

Fig. 4. Introgressed regions in North African date palm. (*A*) The introgression fraction (f_d) in 5-kb nonoverlapping windows from the 30 longest scaffolds in the draft assembly. (*B*) Population genetic summary statistics of North African date palm for scaffold NW_008246512.1 in 5-kb windows. *B*, *Bottom* shows the frequency of *P. theophrasti*-like alleles (main text) segregating in North Africa. Each circle represents a SNP. Curves were fitted to the data using local polynomial regression (LOESS). (*C*) Neighbor-joining 50% majority rule tree based on phased haplotypes in a 5-kb interval at position 1,100 kb of the scaffold in *B*. Colored circles represent the population origin of phased sequences. The single haplotype from the Middle East in the *P. theophrasti*/North Africa clade bearing the shared haplotype is from a Pakistani variety (Aseel) with a mixed Middle Eastern/North African ancestry (7). Single haplotypes from each of two probable hybrids (a *P. theophrasti* sample from Almyros, Crete and a *P. sylvestris* sample from Faisalabad, Pakistan) are nested within *P. dactylifera* clades as are haplotypes from two *P. atlantica* samples.

haplotypes (*SI Appendix*, Fig. S3) consistent with their being hybrid samples. Similar phylogenetic relationships were apparent in the analysis of the mtDNA genome (*SI Appendix*, Fig. S4). The distinctness of the *P. theophrasti* cpDNA and mtDNA haplotypes and their absence from North African date palm samples suggest an asymmetry in the direction of the interspecific cross and indicate that *P. theophrasti* was the paternal (pollen) contributor.

Impact of Introgression on Genetic Diversity in North Africa. Introgression has impacted genome-wide patterns of genetic diversity in North African date palm. Higher nucleotide diversity in the North African population ($P < 2.2 \times 10^{-16}$; two-tailed Wilcoxon signed-rank test, Fig. 5*B* and ref. 7) could be explained by introgression. To determine whether introgression has contributed to elevated diversity in North Africa, we rank ordered the genome based on f_d and found that highly introgressed regions show higher levels of nucleotide diversity in North Africa (*SI Appendix*, Fig. S5). However, regions with little or no evidence of introgression on average also show higher diversity in North Africa, an observation that suggests that hybridization alone may be insufficient to account for elevated diversity in North Africa. An effect of introgression on Tajima's *D* is also apparent. At the whole-genome level, Tajima's *D* is higher in North Africa than in the Middle East ($P < 2.2 \times 10^{-16}$, *SI Appendix*, Table S3). This appears to be driven at least in part by introgression as elevated Tajima's *D* is most pronounced in regions with highest f_d (*SI Appendix*, Fig. S5).

The impact of introgression is also apparent in measures of population differentiation and LD. Using F_{st} as a measure of population subdivision, we observe that the two regional populations of date palm are moderately diverged ($F_{st} = 0.085$). However, the North African population is less diverged from *P. theophrasti* ($F_{st} = 0.369$) than is the Middle Eastern population ($F_{st} = 0.403$; $P < 2.2 \times 10^{-16}$, two-tailed Wilcoxon signed-rank test, Fig. 5*A*). When we compare F_{st} across genomic intervals ranked by f_d, we observe that the reduction in F_{st} between North African date palm and *P. theophrasti* is most pronounced in regions with a high introgression fraction (*SI Appendix*, Fig. S5). Moreover, F_{st} between Middle Eastern and North African populations is elevated in these same regions (*SI Appendix*, Fig. S5), thus suggesting that population structure observed between these populations is at least partially due to introgression from *P. theophrasti* in North Africa.

If these patterns are the result of recent introgression from a wild relative, we expect the North African population to show evidence of admixture LD. At equilibrium, a population with larger effective population size (N_e) should have a faster rate of LD decay. However, although the North African population has higher nucleotide diversity and therefore larger N_e at equilibrium (7), LD decays at a slower rate in this population (Fig. 5*C*). LD in North Africa reaches half its maximum at ∼30 kb, while in the Middle East the half-decay distance is ∼20 kb, a pattern consistent with recent admixture of North African date palm with a distant population. These observations suggest that the mixed ancestry of North African date palm has profoundly impacted genome diversity in this population.

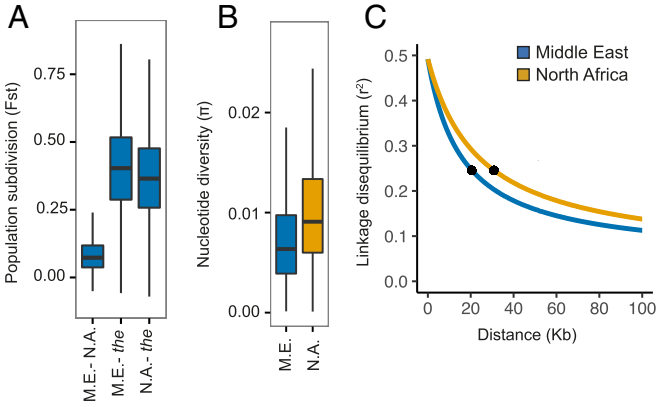

Fig. 5. Patterns of genetic diversity in North African and Middle Eastern date palm populations. (*A*) Boxplot distributions of F_{st} based on 5-kb nonoverlapping intervals between pairs of populations (M.E., Middle East; N.A., North Africa; *the*, *P. theophrasti*). (*B*) Distribution of nucleotide diversity in Middle Eastern and North African date palm populations. (*C*) Decay of LD calculated separately for each population. Symbols on decay curves represent the half-decay distance at 20.5 kb and 30.8 kb for Middle Eastern and North African populations.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

Private Alleles in Date Palm Populations. Our analysis raises questions about the possibility of an unsampled population that may have been a source of ancestry in the North African population. We examined the distribution of private polymorphisms among populations and *Phoenix* wild relatives. When measured as a fraction of total SNPs within each group, we find that the North African population does not have a large class of private polymorphisms (5.7%) relative to Middle Eastern date palm (13.4%) (*SI Appendix*, Table S11). This suggests any additional unsampled source of variation in the North African population is unlikely to be a divergent lineage.

Archaeobotanical Records Are Consistent with a Late Appearance of Date Palm in North Africa. We assembled archaeobotanical records from the published literature to address questions about the historical distribution of date palm (*SI Appendix*, *SI Text* and Dataset S1). Inspection of these records indicates that date palm appears in the archaeobotanical record of North Africa later than in the Near and Middle East (Dataset S1 and *SI Appendix*, Figs. S6–S8). A single pair of date stones recovered from Takarkori, Southwest Libya, might be as old as 9,000 yBP (38), but is here rejected as plausibly recent and intrusive. The earliest records on the African continent date to predynastic Egypt (ca. 5,500 yBP) and Sudan (ca. 4,000 yBP). In the Maghreb, the oldest reliable records are from Libya (ca. 3,000 yBP). West of Libya, there is one record of date palm at Volubilis, Morocco dating to the Classical Period (ca. 2,000 yBP), but otherwise no archaeobotanical records before the Islamic Period (ca. 700 yBP) when seed remains appear at the medieval sites of Igîlîz, Volubilis, and Sijilmasa, Morocco and Essouk and Gao Saney, Mali. These observations are consistent with the view that date palm may have only recently expanded to the African continent.

Discussion

We have presented evidence that the North African date palm population has a mixed ancestry that can be traced to hybridization between date palm and a wild relative *P. theophrasti*, the Cretan date palm. *P. theophrasti* is a distinct species (Figs. 1*B* and 3*C* and ref. 15) with a present-day distribution that is restricted primarily to geographically isolated locations on the island of Crete, the East Aegean Islands, and the Anatolian coast of Turkey (Fig. 1*A* and refs. 39–42). Evidence that the North African population has mixed ancestry between this Aegean Sea endemic and cultivated date palm suggests that *P. dactylifera* on the African continent has a hybrid origin and that *P. theophrasti* may be the unknown source of variation in North Africa identified by Gros-Balthazard et al. (6).

The geographic context of hybridization between *P. theophrasti* and *P. dactylifera* is obscured by the fact that the present-day distribution of *P. theophrasti* does not overlap with the traditional range of date palm cultivation (Fig. 1). One model is that hybridization occurred in the Levant or elsewhere in the Eastern Mediterranean where the ranges of the two species may have once overlapped. *Phoenix* stones identified as *P. theophrasti* have been found at a site in northern Israel from ∼9,000 yBP, suggesting that this species had a wider distribution in the past and may have been exploited (38, 43). *P. dactylifera* was also distributed historically in the Levant including the Jordan River valley and the Dead Sea until the Middle Ages when the Judean date palm population went extinct (44). Whether the two species historically occurred sympatrically in the Eastern Mediterranean or elsewhere is unclear.

The late appearance of date palm in the archaeobotanical record of North Africa suggests a model where Middle Eastern date palm expanded its range to the African continent. This model predicts a bottleneck associated with founder event colonization that was not supported by previous genome-wide analysis showing elevated genetic diversity in North Africa (6, 7). Evidence of hybridization, while appearing to explain much of the elevated diversity in this population, also does not appear to fully account for differences in nucleotide diversity between populations. These observations suggest that a simple expansion plus hybridization model may be insufficient to account for patterns in the genomic data. More complex expansion models such as those including postexpansion inbreeding (7) or additional bottlenecking in the Middle East could account for our observations.

Alternatives to the expansion model include those that propose an additional source of variation in addition to *P. theophrasti* and Middle Eastern date palm in the North African gene pool. For example, it is possible that the archaeobotanical record provides an incomplete picture of *Phoenix* in North Africa and that a wild population of *P. dactylifera* similar to the relictual population recently discovered on the Arabian Peninsula (6) once existed on the African continent. This population either could have been domesticated independently or was not domesticated, but served as an additional source of variation for an extant cultivated population with *P. theophrasti* and Middle Eastern date palm ancestry. A distinct haplotype found at high frequency in the organellar genomes of the North African population (*SI Appendix*, Figs. S3 and S4) (45) provides some support for such a lineage. However, we note that the relatively small numbers of private alleles in North Africa (5.7% of SNPs in North Africa vs. 13.4% in the Middle East) suggest that if such a population exists, it is unlikely to be a divergent lineage or has not contributed significantly to the genomic constitution of North African date palms.

Phoenix species are known to hybridize and produce viable offspring (16, 27, 46, 47) and either natural hybridization or horticultural practices could account for the introgression of interspecific alleles into *P. dactylifera*. For example, the practice of hand pollination of date palm was widely known throughout the ancient world (48) and existing practices of fertilization with interspecific pollen (49) and seedling propagation (50) suggest a possible mechanism for hybrid origins. Alternatively, putative natural hybrids between *P. theophrasti* and date palm have been reported (40) and are also apparent in our analysis. Although *P. theophrasti* samples appear as a distinct cluster in our STRUCTURE results (Fig. 2), individuals with some *P. dactylifera* ancestry include a sample from a reported hybrid population in Gölköy, Turkey (40); two samples from Almyros, Crete; and two samples from Epidaurus, Peloponesse. These samples may represent instances of crop-to-wild gene flow (51) or be relicts from an earlier hybridization event.

Crop species frequently retain the capacity to hybridize with their wild relatives and hybrid genotypes are sources of novelty, superior quality, and adaptive traits. Evidence of introgressive hybridization in annual crops such as rice (52) and maize (53) and perennial tree crops including citrus (54), almonds (55), and apricots (56) supports a role for hybridization in the domestication process. The adaptive significance of hybridization and the traits subject to selection are often unknown except in exceptional cases (53). However, it is becoming increasingly clear that introgressive hybridization is an important source of diversity (53) that frequently accompanies expansion of domesticated species such as apples (57) and grapevine (58). Discovery that a regional population of domesticated date palm is also the product of hybridization helps resolve long-standing questions about the origins of date palm and indicates that introgression may be an important factor that dramatically alters the genome of domesticated species, thus providing novel diversity for adaptation during domestication.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

Materials and Methods

A detailed description of materials and methods is provided in *SI Appendix, SI Materials and Methods*. Briefly, we obtained varieties of *P. dactylifera* and *Phoenix* wild relatives from various sources, including collections from wild populations, germplasm collections, research facilities, and ornamental gardens (*SI Appendix*, Table S1). Whole-genome 2 × 100 paired-end sequencing was conducted on an Illumina HiSeq 2500 Sequencer with one-half to one lane per sample. Data were aligned to the date palm draft assembly (28) available at the National Center for Biotechnology Information (NCBI) with Burrows–Wheeler aligner (59). SNPs were called following standard protocols (60) and filtered accordingly.

Phylogenomic analysis was conducted on a subset of SNPs separately for nuclear, cpDNA, and mtDNA genomes using a combination of randomized axelerated maximum likelihood (RAxML) (61) and R packages phangorn (62) and ape (63). Model-based clustering of population genomic data was conducted with STRUCTURE (29). A model of population splits and mixtures was fitted to the data using *TreeMix* (36). Admixture tests with f_3 and *D* statistics were conducted with Popstats (64). Estimates of ancestry proportions were obtained from *TreeMix* mixture weights and with the f_4-ratio approach as implemented in Popstats. Population and introgression summary statistics were estimated directly from sample alignments with analysis of next generation sequencing data (ANGSD) (65) or from SNPs with vcftools (66) and a script from Martin et al. (37). Statistical analysis was conducted with the R Statistical Programing Language (67).

Archaeobotanical records of date palm were gathered from published reports, based on finding monograph chapters based on previous regional reviews (e.g., refs. 68 and 69), a database search of "literature on archaeological remains of cultivated plants 1981–2004" (70), and a Google scholar search for additional recent journal articles. It includes all reports, most of which are dated based on associated artifactal and archaeological evidence (indicated as "ass."), in some cases by associated radiocarbon dates (C14) or direct accelerator mass spectrometry (AMS) dates on crop remains. Few of these are directly dated *Phoenix* stones themselves, but those available are indicated as "AMS*." Calibrated radiocarbon dates have been summed and the 1-sigma range has been taken to represent the phase of the sample, from a which a median has been calculated. In the case of associated dating evidence a standard period age range has been taken from recent archaeological literature (based on the expertise of D.Q.F.) and a median taken from this range. Where there are concerns over the true antiquity of remains, such as uncarbonized *Phoenix* stones of large size from Takrakori, Libya, attributed to a 8,000-yBP context (38), these are flagged in our database (Dataset S1) and excluded from the synthesis of date palm spread. Additional details are provided in *SI Appendix, SI Text*.

DNA sequence data new to this study have been deposited in the Sequence Read Archive (71). SNP data have been deposited in the Dryad Digital Repository (72).

ACKNOWLEDGMENTS. We thank Marc Arnoux and Nizar Drou in the New York University Abu Dhabi (NYUAD) Center for Genomics and Systems Biology for technical assistance. We thank Robert R. Krueger (US Department of Agriculture), Claudio Littardi [Centro Studi e Ricerche per le Palme - Sanremo (CSRP), Italy], José Plumed (Botanical Garden of Valencia, Spain), Emmanuel Spick (Jardin des Plantes de Montpellier, France), William J. Baker (Kew Gardens, United Kingdom), Syed Summar Abbas Naqvi (Institute of Horticultural Sciences, Pakistan), Joel A. Malek (Weill Cornell Medical College in Qatar), Hendrik J. Visser [Date Palm Research and Development Unit (DPRU), United Arab Emirates University (UAEU)], Abdelouahhab Zaid (DPRU, UAEU) Khaled Masmoudi (International Center for Biosaline Agriculture), Nadia Haider (Atomic Energy Commission of Syria), Nabila El Kadri (Technical Center of Dates, Ministry of Agriculture, Kebili, Tunisia), Youssef Idaghdour (NYUAD), Deborah Thirkhill (Arizona State University Date Palm Collection), and Ghulam S. Markhand (Date Palm Research Institute, Abdul Latif University, Pakistan) for providing samples. We also thank Jeffrey Ross-Ibarra and two anonymous reviewers who helped improve the manuscript. We thank members of the M.D.P. laboratory and Jessica Molina Abdala for helpful discussions. This work was made possible by Jean-Christophe Pintaud. This research was funded in part by an NYUAD Institute grant, as well as by grants from the US National Science Foundation Plant Genome Research Program and the Zegar Family Foundation (to M.D.P.).

1. Barrow SC (1998) A monograph of Phoenix L. (Palmae: Coryphoideae). *Kew Bull* 53:513–575.
2. Zehdi S, et al. (2012) Molecular polymorphism and genetic relationships in date palm (Phoenix dactylifera L.): The utility of nuclear microsatellite markers. *Sci Hort* 148:255–263.
3. Arabnejad H, Bahar M, Reza Mohammadi H, Latifian M (2012) Development, characterization and use of microsatellite markers for germplasm analysis in date palm (Phoenix dactylifera L.). *Sci Hort* 134:150–156.
4. Cherif E, et al. (2013) Male-specific DNA markers provide genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. *New Phytol* 197:409–415.
5. Zehdi-Azouzi S, et al. (2015) Genetic structure of the date palm (Phoenix dactylifera) in the Old World reveals a strong differentiation between eastern and western populations. *Ann Bot* 116:101–112.
6. Gros-Balthazard M, et al. (2017) The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa. *Curr Biol* 27:2211–2218.e8.
7. Hazzouri K, et al. (2015) Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nat Commun* 6:8824.
8. Mathew LS, et al. (2015) A genome-wide survey of date palm cultivars supports two major subpopulations in Phoenix dactylifera. *G3* 5:1429–1438.
9. Beech M (2003) Archaeobotanical evidence for early date consumption in the Arabian Gulf. *The Date Palm–From Traditional Resource to Green Wealth*, ed The Emirates Center for Strategic Studies and Research (The Emirates Center for Strategic Studies and Research, Abu Dhabi, United Arab Emirates), pp 11–32.
10. Safar F, Mustafa MA, Lloyd S (1981) *Eridu* (Ministry of Culture and Information, State Organization of Antiquities and Heritage, Baghdad).
11. Neef R (1991) Plant remains from archaeological sites in lowland Iraq: Tell el'Oueil. *Oueili: Travaux de 1985* [Works of 1985], ed Huot JL (Éditions Recherches sur les Civilisations, Paris), pp 321–329.
12. Zohary D, Hopf M, Weiss E (2012) *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in South-west Asia, Europe, and the Mediterranean Basin* (Oxford Univ Press, New York), 4th Ed.
13. Tengberg M (2012) Beginnings and early history of date palm garden cultivation in the Middle East. *J Arid Environ* 86:139–147.
14. Fuller DQ (2018) Long and attenuated: Comparative trends in the domestication of tree fruits. *Veg Hist Archaeobot* 27:165–176.
15. Pintaud JC, et al. (2010) Species delimitation in the genus Phoenix (Arecaceae) based on SSR markers, with emphasis on the identity of the date palm (Phoenix dactylifera). *Diversity, Phylogeny, and Evolution in the Monocotyledons*, eds Seberg O, Petersen G, Barfod AS, Davis JI (Aarhus Univ Press, Aarhus, Denmark), pp 267–286.
16. Newton C, et al. (2013) Phoenix dactylifera and P. sylvestris in Northwestern India: A glimpse of their complex relationships. *Palms* 57:37–50.
17. Munier P (1981) Origin of date palm cultivation and propagation in Africa: Historical notes on the principal African palm growing areas. *Fruits* 36:437–450.
18. Van der Veen M (1995) Ancient agriculture in Libya: A review of the evidence. *Acta Palaeobot* 35:85–98.
19. El-Hadidi MN (1982) The predynastic flora of the Heirakonpolis region. *The Predynastic of Hierakonpolis: An Interim Report*, ed Hoffman MA (Cairo Univ Herbarium, Faculty of Science, Giza, Egypt), pp 102–115.
20. Fahmy AG, Khodary S, Fadl M, El-Garf I (2008) Plant macroremains from an elite cemetery at predynastic Hierakonpolis, Upper Egypt. *Int J Bot* 4:205–212.
21. van Zeist W (1983) Fruits in foundation deposits of two temples. *J Archaeol Sci* 10:351–354.
22. Mercuri AM, Bosi G, Buldrini F (2013) Seeds, fruits and charcoal from the Fewet compound. *Arid Zone Archaeol Monogr* 6:177–190.
23. Lézine AM, Zheng W, Braconnot P, Krinner G (2011) Late Holocene plant and climate evolution at Lake Yoa, northern Chad: Pollen data and climate simulations. *Clim Past* 7:1351–1362.
24. Nixon S, Murray MA, Fuller DQ (2011) Plant use at an early Islamic merchant town in the West African Sahel: The archaeobotany of Essouk-Tadmakka (Mali). *Veg Hist Archaeobot* 20:223–239.
25. Ruas MP, Tengberg M, Ettahiri AS, Fili A, Van Staëvel JP (2011) Archaeobotanical research at the medieval fortified site of Igiliz (Anti-Atlas, Morocco) with particular reference to the exploitation of the argan tree. *Veg Hist Archaeobot* 20:419–433.
26. Munier P (1973) *Le Palmier-Dattier* [The Date Palm] (Maisonneuve et Larose, Paris).
27. Gros-Balthazard M (2013) Hybridization in the genus Phoenix: A review. *Emir J Food Agric* 25:831–842.
28. Al-Mssallem IS, et al. (2013) Genome sequence of the date palm Phoenix dactylifera L. *Nat Commun* 4:2274.
29. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
30. Lawson DJ, van Dorp L, Falush D (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat Commun* 9:3258.
31. Janes JK, et al. (2017) The K = 2 conundrum. *Mol Ecol* 26:3594–3602.
32. Green R, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722.
33. Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28:2239–2252.
34. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461:489–494.
35. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.
36. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967.
37. Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol* 32:244–257.
38. Rivera D, et al. (2014) Carpological analysis of Phoenix (Arecaceae): Contributions to the taxonomy and evolutionary history of the genus. *Bot J Linn Soc* 175:74–122.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021

39. Boydak M (1985) The distribution of Phoenix theophrasti in the Datça Peninsula, Turkey. *Biol Conserv* 32:129–135.
40. Boydak M, Barrow S (1995) A new locality for Phoenix in Turkey: Gölköy-Bödrum. *Principes* 39:117–122.
41. Greuter W (1995) Phoenix theophrasti. *The Red Book of Rare and Threatened Plants of Greece*, eds Phitos D, Strid A, Snogerup S, Greuter W (World Wide Fund for Nature Greece: K. Michalas S.A., Athens, Greece), pp 412–413.
42. Tsakiri M, Kougioumoutzis K, Iatrou G (2016) Contribution to the vascular flora of Chalki Islands (East Aegean, Greece) and bio-monitoring of a local endemic taxon. *Willdenowia* 46:175–190.
43. Kislev M, Hartmann A, Galili E (2004) Archaeobotanical and archaeoentomological evidence from a well at Atlit-Yam indicates colder, more humid climate on the Israeli coast during the PPNC period. *J Archaeol Sci* 31:1301–1310.
44. Goor A (1967) The history of the date through the ages in the Holy Land. *Econ Bot* 21:320–340.
45. Zehdi-Azouzi S, et al. (2016) Endemic insular and coastal Tunisian date palm genetic diversity. *Genetica* 144:181–190.
46. Wrigley G (1995) Date palm, (Phoenix dactylifera). *Evolution of Crop Plants*, eds Smartt J, Simmonds NW (Longman Scientific and Technical, Harlow, UK), pp 399–403.
47. González-Pérez MA, Sosa PA (2009) Hybridization and introgression between the endemic Phoenix canariensis and the introduced P. dactylifera in the Canary Islands. *Open For Sci J* 2:78–85.
48. Zohary D, Spiegel-Roy P (1975) Beginnings of fruit growing in the Old World. *Science* 187:319–327.
49. Popenoe PB (1913) *Date Growing in the Old World and the New* (West India Gardens, Altadena, CA).
50. Johnson DV, Al-Khayri JM, Jain SM (2013) Seedling date palms (Phoenix dactylifera L.) as genetic resources. *Emir J Food Agric* 25:809–830.
51. Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annu Rev Ecol Syst* 30:539–563.
52. Choi JY, et al. (2017) The rice paradox: Multiple origins but single domestication in Asian rice. *Mol Biol Evol* 34:969–979.
53. Hufford MB, et al. (2013) The genomic signature of crop-wild introgression in maize. *PLoS Genet* 9:e1003477.
54. Wu GA, et al. (2014) Sequencing of diverse Mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotech* 32: 656–662.
55. Delplancke M, et al. (2012) Gene flow among wild and domesticated almond species: Insights from chloroplast and nuclear markers. *Evol Appl* 5:317–329.
56. Zhang Q, et al. (2018) The genetic architecture of floral traits in the woody plant Prunus mume. *Nat Commun* 9:1702.
57. Cornille A, et al. (2012) New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet* 8:e1002703.
58. Myles S, et al. (2011) Genetic structure and domestication history of the grape. *Proc Natl Acad Sci USA* 108:3530–3535.
59. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]. Preprint, posted May 26, 2013.
60. Poplin R, et al. (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv:10.1101/201178. Preprint, posted July 24, 2018.
61. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
62. Schliep KP (2011) phangorn: Phylogenetic analysis in R. *Bioinformatics* 27:592–593.
63. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289–290.
64. Skoglund P et al. (2015) Genetic evidence for two founding populations of the Americas. *Nature* 525:104–108.
65. Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* 15:356.
66. Danecek P, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
67. R Core Team (2015) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna), Version 3.2.2.
68. Pelling R (2014) Patterns in the archaeobotany of Africa: Developing a database for North Africa, the Sahara, and the Sahel. *The Archaeology of African Plant Use*, eds Stevens CJ, Nixon S, Murray MA, Fuller DQ (Left Coast Press, Walnut Creek, CA), pp 205–224.
69. Fuller DQ (2015) The economic basis of the Oustul splinter state: Cash crops, subsistence shifts and labour demands in the Post-Meroitic transition. The Kushite World. Beiträge zur Sudanforschung Beiheft 9 [Contributions to Sudan Research Supplement]. *Proceedings of the 11th International Conference for Meroitic Studies, Vienna, 1-4 September 2008*, ed Zach MH (Verein der Förderer der Sudanforschung, Vienna), pp 33–60.
70. Kroll H (2005) Literature on archaeological remains of cultivated plants 1981-2004. Available at archaeobotany.de/database.html. Accessed January 2, 2016.
71. Leinonen R, Sugawara H, Shumway M (2010) The Sequence Read Archive. *Nucleic Acids Res* 39:D19–D21.
72. Flowers JM, et al. (2019) Data from "Cross-species hybridization and the origin of North African date palms." Dryad Digital Repository. Available at datadryad.org. Deposited December 26, 2018.

Downloaded at UNIVERSITE MONTPELLIER 11 on April 19, 2021