



**HAL**  
open science

## What lies behind a fruit crop variety name? A case study of the barnī date palm from al-‘Ulā oasis, Saudi Arabia

Muriel Gros-Balthazard, Vincent Battesti, Jonathan M Flowers, Sylvie Ferrand, Matthieu Breil, Sarah Ivorra, Jean-Frédéric Terral, Michael D. Purugganan, Rod A. Wing, Nahed Mohammed, et al.

### ► To cite this version:

Muriel Gros-Balthazard, Vincent Battesti, Jonathan M Flowers, Sylvie Ferrand, Matthieu Breil, et al.. What lies behind a fruit crop variety name? A case study of the barnī date palm from al-‘Ulā oasis, Saudi Arabia. *Plants, People, Planet*, 2023, 5 (1), pp.82-97. 10.1002/ppp3.10326 . hal-03806722v2

**HAL Id: hal-03806722**

**<https://hal.science/hal-03806722v2>**

Submitted on 2 Jan 2023

**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 - ShareAlike 4.0 International License

## RESEARCH ARTICLE

# What lies behind a fruit crop variety name? A case study of the barnī date palm from al-‘Ulā oasis, Saudi Arabia

Muriel Gros-Balthazard<sup>1,2</sup>  | Vincent Battesti<sup>3</sup>  | Jonathan M. Flowers<sup>2</sup>  |  
 Sylvie Ferrand<sup>2</sup>  | Matthieu Breil<sup>3,4</sup>  | Sarah Ivorra<sup>4</sup>  | Jean-Frédéric Terral<sup>4</sup>  |  
 Michael D. Purugganan<sup>2,5,6</sup>  | Rod A. Wing<sup>7</sup>  | Nahed Mohammed<sup>7</sup>  |  
 Yann Bourgeois<sup>8</sup> 

<sup>1</sup>DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France

<sup>2</sup>Center for Genomics and Systems Biology (CGSB), New York University Abu Dhabi, Abu Dhabi, UAE

<sup>3</sup>Eco-anthropologie, CNRS, Muséum national d'histoire naturelle, Université Paris Cité, at Musée de l'Homme, Paris, France

<sup>4</sup>ISEM, Institut des Sciences de l'Evolution-Montpellier, Université de Montpellier/CNRS/IRD/EPHE, Montpellier, France

<sup>5</sup>Center for Genomics and Systems Biology (CGSB), New York University, New York, USA

<sup>6</sup>Institute for the Study of the Ancient World, New York University, New York, USA

<sup>7</sup>Center for Desert Agriculture, Biological and Environmental Science and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

<sup>8</sup>School of Biological Sciences, University of Portsmouth, Portsmouth, UK

**Correspondence**

Muriel Gros-Balthazard, DIADE, University of Montpellier, CIRAD, IRD, 911 Avenue Agropolis, 34394/34830 Montpellier Cedex 5, France.

Email: [muriel.gros-balthazard@ird.fr](mailto:muriel.gros-balthazard@ird.fr)

Vincent Battesti, Eco-anthropologie, CNRS, Muséum national d'histoire naturelle, Université Paris Cité, at Musée de l'Homme, 17 place du Trocadéro, 75016 Paris, France. Email: [vincent.battesti@cnrs.fr](mailto:vincent.battesti@cnrs.fr)

**Funding information**

French Agency for AIUla Development (AFALULA)

**Societal Impact Statement**

The oasis of al-‘Ulā is subject to a vast development operation by the central government of the Saudi monarchy. Agriculture is not strictly speaking the first objective of this initiative, the emphasis being on tourism and thus on the vast historical heritage and landscape qualities of the region. Nevertheless, agriculture and, in particular, phoeniculture remain the main resource for the inhabitants. Characterizing the local date palm agrobiodiversity is key to the sustainable development of oases. In al-‘Ulā, documenting indigenous knowledge about the locally predominant barnī variety and characterizing its genetic integrity and mode of propagation represents the essential leverage needed by farm development project planners to develop local production.

**Summary**

- Understanding how farmers name and categorize their crops in relation to the way they are propagated is critical for a proper assessment of agrobiodiversity. Yet, indigenous knowledge is often overlooked in genetic studies, which may result in an underestimation of crop diversity, thereby preventing its conservation and mobilization for developing sustainable agroecosystems.

Muriel Gros-Balthazard and Vincent Battesti contributed equally to this work.

Transliteration system

DIN-31635 (except API x for ح [h]).

All qāf ح are pronounced gāf (IPA: g) in al-‘Ulā region.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Plants, People, Planet* published by John Wiley & Sons Ltd on behalf of New Phytologist Foundation.

Gros-Balthazard\*, Muriel, Vincent Battesti\*, Jonathan M. Flowers, Sylvie Ferrand, Matthieu Breil, Sarah Ivorra, Jean-Frédéric Terral, Michael D. Purugganan, Rod A. Wing, Nahed Mohammed & Yann Bourgeois, 2022 – « What lies behind a fruit crop variety name? A case study of the barnī date palm from al-'Ulā oasis, Saudi Arabia ». *Plants, People, Planet*, 5 (1), p. 82–97, doi:10.1002/ppp3.10326 – en ligne: <https://doi.org/10.1002/ppp3.10326> & <https://hal.archives-ouvertes.fr/hal-03806722>

- Here, we focus on the barnī date palm variety, a local elite variety of al-'Ulā oasis, Saudi Arabia. We conducted an ethnobotanical survey on local phoeniculture practices and generated whole-genome data to determine whether or not barnī palms are exclusively clonally (vegetatively) propagated. Further, we contrasted the genomes of barnī and two other palms from al-'Ulā with 112 *Phoenix* spp. to provide an initial insight into date palm diversity in this oasis.
- The survey reveals that the dates of the barnī palm bear distinct names, depending on their quality. Results show that barnī is a true-to-type cultivar, indicating clonal propagation by offshoots with name maintenance, even between distinct cultivating situations in al-'Ulā and a nearby oasis. Nonetheless, it is distinct from the prominent barnī cultivated in Oman. Its ancestry is comparable to other West Asian date palms, but another palm from this oasis shows influence from North Africa.
- What lies behind the cultivar name barnī in al-'Ulā and further afield in the Arabian Peninsula has been deciphered through the key disciplinary combination of social anthropology and genetics. Future studies will provide additional insights into the original genetic make-up of this millennia-old oasis.

#### KEYWORDS

agrobiodiversity, al-'Ulā oasis (Saudi Arabia), date palm (*Phoenix dactylifera* L.), ethnobotany, indigenous knowledge, intra-varietal genetic variation, local categorization, population genetics

## 1 | INTRODUCTION

Agrobiodiversity is a proven lever for the resilience and adaptation of our food production systems (Jarvis et al., 2016). It is the spectacular product of farmers' work over millennia, and as such, it encompasses not only the genetic diversity of crops and their wild relatives but also the local knowledge and practices associated to their cultivation (Bahuchet, 2017). Yet, the biological and social dimensions of this diversity are rarely examined jointly, even though it would promote a better understanding and evaluation of it, in turn fostering its conservation and mobilization to mitigate the adverse effects of global change (Caillon & Degeorges, 2007; Gros-Balthazard et al., 2020; Leclerc & Coppens d'Eeckenbrugge, 2011).

In predominantly vegetatively (clonally) propagated crops, farmers promote the maintenance of valuable phenotypes and adaptive potential of a lineage through the use of mixed sexual/clonal reproduction systems (McKey et al., 2010). On one hand, farmers propagate interesting genotypes vegetatively, in which case a name (of a variety or cultivar) is assumed to refer to a single genotype. On the other hand, sexual reproduction creates new combinations of genes, and the resulting seedlings may be incorporated by farmers to the cultivated pool. Farmers may voluntarily select a new seedling with attractive characteristics to become a new variety, under a new name (e.g., in cassava, Elias et al., 2000 or in yam, Scarcelli et al., 2006). By contrast, farmers may incorporate seedlings under an existing variety name, if they consider that the seedling has the same phenotypic traits, thus prompting intra-varietal diversity (e.g., in oca, Bonnavé

et al., 2014, or in date palm, Gros-Balthazard et al., 2020). This latter practice may lead to underestimation of agrobiodiversity because variety names may not correspond to single genotypes.

So far, very few studies have explored what lies behind a crop variety name and its connection to cultivation practices. In date palm, our aforementioned study, focusing on the oasis of Siwa in Egypt, remains unique in its integration of social and biological sciences (Battesti, 2013; Battesti et al., 2018; Gros-Balthazard et al., 2020). It appears crucial to develop this pluri-disciplinary approach in other oases to establish a comprehensive evaluation of date palm agrobiodiversity, in terms of both variety names and genetic diversity.

The date palm (*Phoenix dactylifera* L.) is the main crop of desert oasis agrosystems of North Africa and West Asia, first cultivated for its fruits: the dates. All parts of the plant, however, can be used—for food, architecture, handicrafts, and so forth—and the plant itself is the oasis system engineer (Battesti, 2005). Its importance can be seen in its denominations: In Arabic in particular, its generic name is always the same (*naxla*, نخلة), but the names of the varieties—each of which has its own use, conservation, taste, harvest period, pedoclimatic needs, and so forth—are prolific. Some are well known and refer to elite commercial varieties, such as medjool (or mejhoul [*majhūl*], see Zaid & Oihabi, 2022) or khalas [*xalāṣ*], while many names are rather found locally. Inventories of varieties have been carried out locally (e.g., in Tunisia; Rhouma, 1994, 2005), but their overall number remains difficult to evaluate and could exceed 3000 worldwide (Zaid & Arias-Jiménez, 1999).

It is usually assumed that date palm varieties are vegetatively propagated by farmers (as in Krueger, 2011, and discussed in Battesti, 2013). Indeed, although this dioecious species can reproduce sexually, its multiplication is mainly carried out vegetatively by farmers, who cut and replant the offshoots growing at its base. This technique of vegetative propagation maintains the selected features, particularly that of the fruit. By contrast, palms grown from seeds have the drawback for farmers of being half males, which do not produce dates, and if females, they produce dates solely after 6–7 years, and those are typically different from that of the mother plant. Therefore, seedlings found in palm groves “merely grew by accident” (Popenoe & Bennett, 1913) and are almost always regarded by farmers as of lower quality than those of the mother plant, thus resulting in their uprooting (but see Johnson et al., 2013; Newton et al., 2013).

According to both main local and scientific narratives, a name matches a variety that matches a genotype (a true-to-type cultivar). But numerous studies have highlighted the existence of genetic variability within a name (e.g., Al-Khalifah & Askari, 2003; Al-Ruqaishi et al., 2008; Elhoumaizi et al., 2006; Sabir et al., 2014). Several hypotheses were raised to explain this variability, such as the presence of somatic mutations (Elmeer et al., 2020; Gros-Balthazard et al., 2020) or the existence of homonyms, that is, when two different lines of clones are called the same. The latter is particularly relevant given that names may refer to general features (e.g., sukārī for sugary) or a common anthroponym (e.g., nabtat saīf, Saīf’s plant). Local practices of naming and categorizing palms may also promote diversity under a name. First, seedlings sharing some characteristics or qualities according to local standards, for example, fruit color or usage, may be called by a single name. The most obvious examples of such a so-called ‘local category’ (Battesti, 2013) are the categories of “males” or “seedlings” (the latter usually called *khalt* [xalt], Johnson et al., 2013). Second, a small number of clonal lines, all sharing, from farmers’ local point of view, the same phenotypic characteristics (in particular the fruit), and vegetatively propagated by farmers, may be called deliberately by a single name; we coined the result of this practice as ethnovariety (Battesti, 2013; Battesti et al., 2018; Gros-Balthazard et al., 2020). Consequently, the number of cultivated genotypes in the field of this clonally propagated crop can be very different from the number of named types, which can lead to an underestimation of local agrobiodiversity if the system of naming and classification of palms remains undocumented (Gros-Balthazard et al., 2020).

In this study, we explore which kind of identity lies behind the named type “barnī” (برني), the local elite date palm of the oasis of al-‘Ulā, province of Medina, in northwestern Saudi Arabia. “Barnī” is used today to designate a variety of date palm. In Saudi Arabia, it is grown in the provinces of Asir, Medina, Najran, Riyadh, and Tabuk (Al-Khayri et al., 2015), but whether they belong to a single clonal line or are cases of homonymy remains to be elucidated. According to Zhang et al. (2015), for instance, the “Barni Al Madinah” date is a medium to long date, cylindrical in shape, and of brown color. This description matches the barnī cultivated in al-‘Ulā, but it is unknown however if it is genetically the very same barnī. In addition, a “barnī” variety is grown in Oman where it is one of the top 10 producing varieties

(Al-Yahyai & Al-Khanjari, 2019), but Popenoe and Bennett noted that the “Oman variety apparently has no relation to the classical Birnī of Arabia and North Africa” (1913, p. 227). This possible homonymy may be based on the antiquity of the term barnī coupled with its reputation or connotation. Indeed, barnī is one of the few (e.g., ‘ajwa) date palm varieties mentioned in the collections of reports of the Sunnah of the Prophet Muhammad. In Sahih al-Bukhari, one of the collection of hadith, it is mentioned that Bilāl ibn Rabāh brought barnī to the Prophet (Muḥammad ibn Ismā‘īl al-Bukhārī, 2312, Book 40, Hadith 12). In this saying, it is quite clear that barnī stands for a date of superior quality. The etymology of the term “barnī” is confusing. Popenoe and Bennett (1913, pp. 226–227) tentatively ventured this for the barnī variety grown in Oman: What is called Burnī or Berni might derive from the name of a city named “Burn” or from the Persian “bir, fruit/drop” and “nik, good/heavy.” Before them, the lexicographer E. W. Lane (1863, p. 196) compiled these same etymologies (an Arabicized Persian word) from classical Arabic authors, adding a meaning of “clay vessel” also mentioned by these authors. We can cautiously hypothesize that a variety of date may have taken the name of the pottery that preserved it (not long ago in al-‘Ulā, for example, certain varieties of date were preserved and exported in paste form, see below, in goatskin called *šanna* or basketry called *mijlad*).

In this study, we explore the identity of the date palm barnī in al-‘Ulā and in the Arabian Peninsula. We assess whether the local identity of barnī, as given by the farmers, corresponds to a unique genetic identity (and therefore, strictly vegetatively propagated from offshoots locally) or if it refers to a multiplicity of genetic forms with an ethnobotanical survey and genetic analysis. The two approaches are essential and complementary. Indeed, an extensive ethnographical field survey will shed light on the local categorization processes and assist with designing an effective sampling strategy of the local diversity. The genetic data will permit an exploration of what lies behind the named type barnī in term of genetic variability and provide insight into how it relates to other date palms from the oasis and beyond.

## 2 | MATERIALS AND METHODS

### 2.1 | Ethnobotanic survey and sampling of the date palm named type barnī in al-‘Ulā

We carried out an ethnobotanical survey in the oasis of al-‘Ulā between April 2019 and November 2021, totaling more than 9 months divided into four fieldwork stays. It involved observations, non-directive and semi-structured interviews and theme-based group discussions with farmers. The purpose of the survey was to understand the local cognitive and practical relationships with plants and specifically the date palm of the named type barnī.

We sampled in situ, in collaboration with local farmers, 23 barnī palms while conducting this ethnobotanical study (Table 1; Figure 1; Figure S1). Twenty-two were identified as such by and with their owner or otherwise by their permanent manager. The last palm,

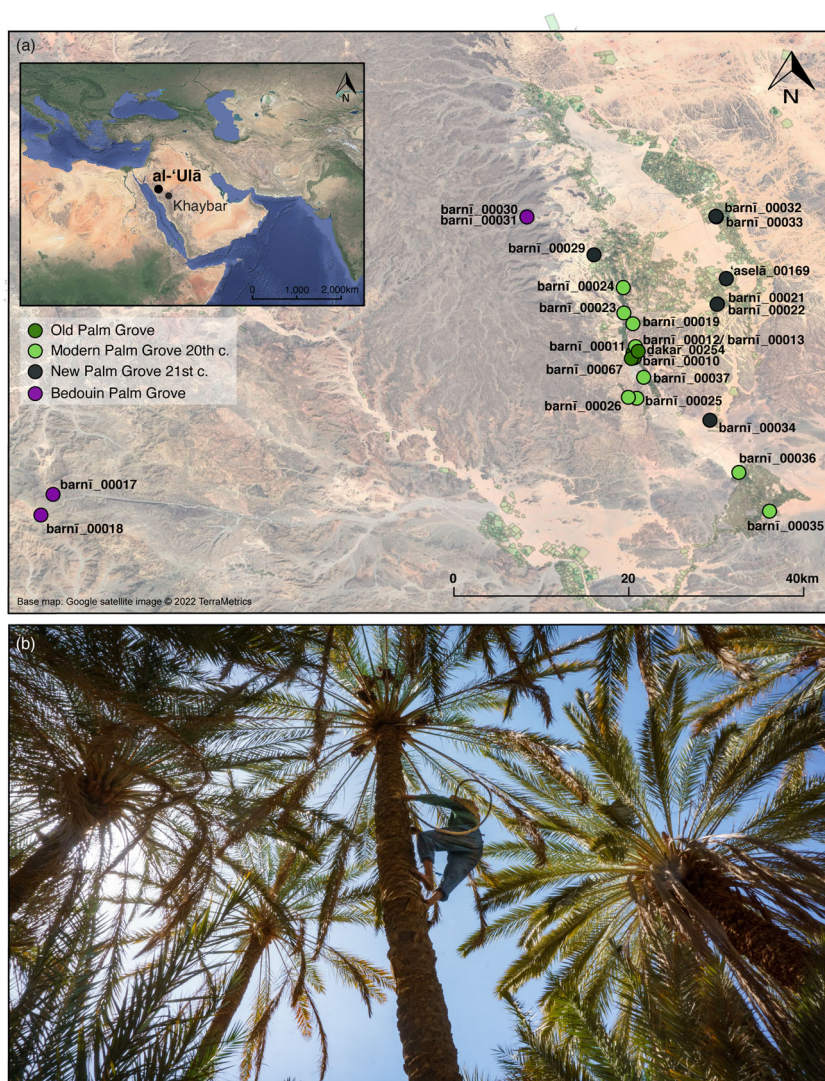
**TABLE 1** Sampling of barnī and other palm accessions for the intra-variety genetic study. We sampled 23 barnī from al-'Ulā oasis and surroundings along with three date palms (two barnī from Khaybar, Saudi Arabia and from al-Kamil, Oman, and one mabrouma from the United Arab Emirates)

ID	Name type	Place of sampling and type of palm grove where barnī palms were sampled in al-'Ulā
Samples of barnī from al-'Ulā region		
barnī_00010	barnī	al-'Ulā, Saudi Arabia (old palm grove)
barnī_00011	barnī	al-'Ulā, Saudi Arabia (old palm grove)
barnī_00012	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00013	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00017	barnī	al-'Ulā, Saudi Arabia (Bedouin palm grove)
barnī_00018	barnī	al-'Ulā, Saudi Arabia (Bedouin palm grove)
barnī_00019	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00021	barnī	al-'Ulā, Saudi Arabia (new 21st c. palm grove.)
barnī_00022	barnī	al-'Ulā, Saudi Arabia (new 21st c. palm grove)
barnī_00023	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00024	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00025	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00026	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00029	barnī	al-'Ulā, Saudi Arabia (new 21st c. palm grove.)
barnī_00030	barnī	al-'Ulā, Saudi Arabia (Bedouin palm grove)
barnī_00031	barnī	al-'Ulā, Saudi Arabia (Bedouin palm grove)
barnī_00032	barnī	al-'Ulā, Saudi Arabia (new 21st c. palm grove)
barnī_00033	barnī	al-'Ulā, Saudi Arabia (new 21st c. palm grove)
barnī_00034	barnī	al-'Ulā, Saudi Arabia (new 21st c. palm grove)
barnī_00035	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00036	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00037	barnī	al-'Ulā, Saudi Arabia (modern 20th c. palm grove)
barnī_00067	barnī	al-'Ulā, Saudi Arabia (old palm grove)
Samples of barnī and mabrouma collected outside of al-'Ulā		
barnī_00268	barnī	Khaybar, Saudi Arabia
barnī_Oman	barnī	al-Kamil, Oman
mabrouma	mabrouma	al-Shiwayb, United Arab Emirates
Clones used to calibrate the analyses		
barnī_00036A	barnī	al-'Ulā, Saudi Arabia
barnī_00036r1	barnī	al-'Ulā, Saudi Arabia
barnī_00036r2	barnī	al-'Ulā, Saudi Arabia
'asēla_00169A	'asēla	al-'Ulā, Saudi Arabia
'asēla_00169r1	'asēla	al-'Ulā, Saudi Arabia
'asēla_00169r2	'asēla	al-'Ulā, Saudi Arabia
ḍakar_00254	ḍakar	al-'Ulā, Saudi Arabia
ḍakar_00254r1	ḍakar	al-'Ulā, Saudi Arabia
ḍakar_00254r2	ḍakar	al-'Ulā, Saudi Arabia

Note: Ten additional accessions (three sets of clones) were used to calibrate the analyses.

so-called “potential barnī” (barnī\_00018), was sampled in a Bedouin palm grove where our Bedouin informant did not specifically point out the collected palm but stated that most of the palms there were supposed to be barnī; it was nonetheless identified as barnī by another informant who did not belong to the tribe owning the palm grove.

We voluntarily maximized both the diversity of the social criteria (sedentary and Bedouin, tribes, social groups, large and small landowners, and so forth) and the different types of palm groves (which include different farming conditions) (see Results; Figures 1 and 2) in order to verify whether some of these criteria could be explanatory.



**FIGURE 1** Date palm sampling. (a) Location of al-'Ulā and Khaybar oases, in Saudi Arabia, are presented in the top-left map, while the main map shows the sampling in al-'Ulā. (b) a farmer, workforce from Qena (Egypt), is climbing a barnī date palm for harvest, in al-Lutāt, an old palm grove of al-'Ulā. The same technique was used by us or the farmers to sample leaflets for analysis. October 25, 2021. Picture: Vincent Battesti

## 2.2 | Sampling of additional date palms and *Phoenix* spp.

In the first part of this study, we explored the intra-named type genetic variability of barnī through genome sequencing of 23 barnī palms sampled in al-'Ulā region, analyzed along with the genomes of 12 other date palms (Table 1; Figure 1; Dataset S1). Specifically, three accessions from other regions, having either the same name or a name that may also be used to refer to barnī, were included. The first one is a “barnī,” identified as such by the owner, sampled in the oasis of Khaybar located 200 km from al-'Ulā (Figure 1). The second one is also named “barnī” and is grown in al-Kamil, in Oman. The third is a GenBank accession (from Hazzouri et al., 2019) of a date palm cultivated in the UAE (al-Shiwayb) and called “mabrouma” by its owner, easily associated with the name given in al-'Ulā to a category of barnī dates, *mabrūm* (see Results).

Further, three sets of clonal accessions consisting of one quartet and two triplets were sampled to calibrate the analyses aiming at

identifying whether those palms represent a single clonal line (Table 1; Figure 1; Dataset S1; Figure S2). A quartet of barnī clones was constituted by sampling one accession a first time in 2019 (barnī\_00036) and re-sampling the same palm 2 years later (barnī\_00036A), along with two of its offshoots (barnī\_00036r1 and barnī\_00036r2). The two triplets were constituted by a male palm and a female variety (ḍakar\_00254 and 'asēla\_00169A, respectively) sampled along with two of their offshoots (ḍakar\_00254r1/ḍakar\_00254r2 and 'asēla\_00169r1/'asēla\_00169r2, respectively).

In the second part of the study, we assessed the diversity of date palms in al-'Ulā by analyzing three unique genomes from this region (barnī\_000268, ḍakar\_00254 and 'asēla\_00169A) along with 112 *Phoenix* spp. genomes from previous studies (Dataset S1) (Flowers et al., 2019; Gros-Balthazard et al., 2017, 2021; Hazzouri et al., 2015). This includes 88 date palms from 13 countries of North Africa and West Asia, 8 *Phoenix sylvestris*, 18 *Phoenix theophrasti*, and 1 *Phoenix reclinata*.



**FIGURE 2** Photographs of four different types of palm groves: (a) old palm grove, here a *bustān*, a garden nearby the old city of al-'Ulā oasis, April 15, 2019; (b) Bedouin palm grove in the bed of wādī Werd, about 100 km west of al-'Ulā oasis, November 8, 2019; (c) modern 20th century palm grove, here a farm with only lined up barnī date palms, in al-Khurayba sector, north of the old palm grove of al-'Ulā, September 23, 2021; (d) new 21st century palm grove, with a view of the agricultural farm *mazra'a* planted in al-'Oqeyb district, north of al-'Ulā, November 1, 2021. Pictures: Vincent Battesti

Overall, a total of 146 *Phoenix* spp. genomes were analyzed, of which 34 were new to this study and 112 were retrieved from GenBank SRA (Table 1; Figure 1; Dataset S1).

### 2.3 | DNA extraction, whole-genome sequencing, and bioinformatic processing

Genomic DNA was extracted from silica-dried leaf tissue using plant DNeasy mini kit (Qiagen, Venlo, Netherlands). Libraries ( $2 \times 100$  or 150 bp paired end) were constructed with either NEBNext Ultra II FS, Nextera DNA Flex, or Illumina truseq nano DNA library preparation kits, and sequenced on an Illumina NextSeq 550 or a NovaSeq 6000 system according to the manufacturer's protocols.

The detailed protocol for read processing, genome alignment, variant calling and filtering may be found in Methods S1. Briefly, we filtered reads based on quality and length before aligning them to the Barhee BC4 reference genome assembly (Hazzouri et al., 2019). We carried out low depth sequencing and performed population genetic analyses to identify the intra-varietal genetic variability by computing the genotype likelihoods from short read alignments. For the second part of our study, we called variants from higher coverage data using GATK v4.2.0.0 (McKenna et al., 2010) and filtered sites and genotypes based on several criteria detailed in Methods S1.

### 2.4 | Data analysis of intra-named type variability in the date palm barnī

Relatedness of the samples was quantified with the King-robust kinship estimator, given its robustness to SNP ascertainment bias and applicability to low-depth sequencing data (Waples et al., 2019), and calculated using NGSrelate v2 (Hanghøj et al., 2019). A principal component analysis was performed using PCAngsd v1.01 (Meisner & Albrechtsen, 2018). In both cases, genotype likelihoods were computed with ANGSD v0.933 (Korneliussen et al., 2014) using the GATK method (option -GL 2). Of note: only repeat masked annotated regions from the 18 linkage groups (Hazzouri et al., 2019) were used. Additionally, reads that did not map uniquely were discarded, and only those reads where the mate could be mapped were kept. We filtered out sites based on the following criteria: non-biallelic sites, minimum mapping quality and minimum base quality of 20, minimum number of individuals 12 (34%), minimum global depth 250 and max depth 415, minimum individual depth  $5\times$ , and SNPs with a  $p$  value  $< 1.10^{-6}$ .

Further, genetic distances among those samples were computed using ngsDist v1.0.10 (Vieira et al., 2016). ANGSD was used to compute genotype posterior probabilities with the same filtering options as above and downsampling the sites to obtain  $\sim 10,000$  sites. Bootstrap replicates ( $n = 100$ ) using blocks of 20 sites were generated, and fastME v2.1.5 (Guindon & Gascuel, 2003) was employed to compute the trees with default parameters. The *consensus* function from *ape R* package (Paradis & Schliep, 2019) was run to obtain a consensus tree where nodes found in 90% of the 100 bootstrap trees were represented. Finally, *TreeDyn* v198.3 (Chevenet et al., 2006) was used for plotting with mid-point rooting via the *phylogeny.fr* web interface (Dereeper et al., 2008).

## 2.5 | Data analysis of the genetic make-up of al-‘Ulā date palms

Three unique genotypes from al-‘Ulā (barnī, ‘aselā, and ḡakar) were compared with 112 *Phoenix* spp. genomes in order to have a first glimpse into the genetic diversity present within al-‘Ulā oasis. First, the structure of the genetic diversity was analyzed by estimating individual ancestries using ADMIXTURE (Alexander et al., 2009) v1.3.0 with a cross-validation of 100, and a principal component analysis (PCA) was run with the pcadapt (Luu et al., 2017) R package v4.3.3 filtering out SNPs with minor allele frequencies below 5%. A maximum likelihood tree was generated using RAXML-NG (Kozlov et al., 2019) v0.9.0. To do so, 20 maximum likelihood tree searches were performed using 10 random and 10 parsimony-based starting trees. The best scoring topology was picked and checked for robustness by performing 100 bootstrap replicates.

The fraction of heterozygote sites for each date palm accession was calculated using pixy (Korunes & Samuk, 2021). Finally, admixture tests were performed using the admixr R package v0.9.1 (Petr et al., 2019) (Methods S2).

To gain insight into the maternal origins of the date palms from al-‘Ulā, a bootstrapped chloroplast DNA tree was constructed using the Neighbor-joining method with the phangorn v2.8.1 package in R (Methods S1). The tree was rooted with *Phoenix reclinata* (PREC1).

Statistical analyses and plotting were conducted with the R Statistical Programming Language (R Core Team, 2022).

## 3 | RESULTS AND DISCUSSION

Our study focused on the date palm locally named barnī, the local elite date palm of al-‘Ulā oasis, Saudi Arabia. We first performed an ethnobotanic survey to both better understand folk categorization in conjunction with local date palm agrobiodiversity and set up an in situ sampling methodology with the essential cooperation of the local farmers. We then performed whole-genome sequencing of barnī date palms from al-‘Ulā region in order to assess whether this name refers to a unique clone, an ethnovariety or a local category with multiple genetic identities. By adding two barnī date palms from outside of al-‘Ulā (i.e., Khaybar, Saudi Arabia and al-Kamil, Oman) and one “mabrouma” (*mabrūm* is the name given in al-‘Ulā to a quality of barnī dates; see below) from the UAE, we further explored the genetic variability of barnī at the scale of the Arabian Peninsula. Finally, we studied three date palms from al-‘Ula, including a barnī, along with 112 *Phoenix* spp. to obtain a first glimpse into the genetic makeup of this millennia-old oasis.

### 3.1 | The cultivation of date palm in the oasis of al-‘Ulā

Our anthropological survey (a 10-month field survey in 2019–2022) of al-‘Ulā oasis and the region highlighted different types of palm

groves, each with its own social and spatial organization, space, and cultural practices, depending on their location, their history and the social group that exploit and own them (Notes S1; Figure 2). To summarize, two of them reflected historical growing situations, namely, the date palms grown in the subsistence-type gardens (*basāṭin*) of the palm grove near the old city of the oasis and those found in the Bedouin palm groves scattered in desert wadis outside the oasis (Figure 2a,b). The two other grove types were more recent and commercial: the modern palm groves established during the 20th century outside of the historical core area, but in its immediate vicinity and the more recent gardens established during the 21st century beyond the perimeter of the old and modern palm groves (Figure 2c,d).

Our fieldwork was performed among all local social categories with several hundred farmers being interviewed in Arabic following the ethnographic methodology. This ethnobotanical survey suggests a very rich agrobiodiversity for date palm alone, drawing up a complex picture of variety names ( $n = 99$ , at this stage of our survey). Farmers classify the two million palms in the region into categories and assign names to them by consensus based on local criteria and shared features. We found that they typically propagate their palms by offshoots, as is usual in palm groves in the Sahara and Arabia (Munier, 1973), but we lack assurance that the clonal propagation has been applied consistently and systematically throughout al-‘Ulā region for all varieties. As a matter of fact, we witnessed reproduction by seed and analyzed emic discourses of palm biology that enable such practices, but we do not yet know the extent to which this technique is used (in practice and over time): it may have emerged in recent decades due to a less extensive knowledge of date palms by newcomers to phoeniculture (sedentarized Bedouins in particular, as declared by both the social group of oasis sedentaries and the Bedouins themselves).

### 3.2 | The barnī, a socialized date palm in al-‘Ulā

#### 3.2.1 | A local elite variety, but a recently increased supremacy over the local date agrobiodiversity

The ethnobotanical survey conducted on date agrobiodiversity in al-‘Ulā clearly revealed the special and shared status, today, of the named type barnī. Indeed, in al-‘Ulā, there is “the barnī” and “the rest,” *al-bāqī* (الباقية), or “the [other] varieties,” *al-aṣnāf* (الأصناف). It is by far the most cultivated variety and is found in all four agricultural contexts described above (Figure 2). It also is the most exported date variety of this oasis. It is considered by all farmers hardy and local and, as such, enjoys an elevated status among the inhabitants and is perceived to be a superior fruit and crop and to grow better. The alternate local elite variety, presently second in rank for all farmers, is the variety named ḡelwa (and when it is necessary to specify, the ḡelwa ḡamrā to distinguish it from another local variety, the ḡelwa beyḡā). Together, barnī and ḡelwa constitute the main varieties in the oasis and are usually the only two reported to be grown in Bedouin palm groves in the region.

The orientalist Jaussen and Savignac (1914, p. 40) noted in the early 20th century the elite status of the barnī and ḡelwa: “Dans



l'oasis d'el-'Ela on cultive principalement le palmier qui est la grande ressource du pays. On distingue ici deux principales espèces de dattes, les dattes douces, *ḥelweh* (حلوة) et les dattes suaves, *barnyeh* (برنية).” Similarly, Nasif (1988, p. 174) stated that “the best-known and best-liked of the various types of date produced is the ḥulwah or sweet date. (...) The barnī follows the ḥulwah in popularity.” This barnī variety seems to have been the reference variety for payment by farmers (in kind, with grain as well) for the work of the *mu'allim*, the person responsible for sharing irrigation water (Nasif, 1988, p. 249). The barnī hence played a central role in agricultural life.

Actually, a distinction must be made between two issues: a numerical supremacy (the case of barnī) and a shared preference (the case of ḥelwa). Local accounts collected during our ethnobotanical survey corroborate the population's shared preference for the taste of ḥelwa over the barnī, especially among elders. It was previously noted by Doughty (1921: p. 153) that the ḥelwa was the most valued, tasted like honey, and widely exported with the pilgrimage. He added that the barnī constituted the “cheaper household food” of al-'Ulā. We should probably understand that dates in general, and among them especially the barnī, were the staple food of the oasis inhabitants. The above-mentioned export of ḥelwa is no longer what it once was. In the past, according to local narratives, dates were exported in the form of paste, which suited the soft ḥelwa well. Today, individual fruits are preferred over paste in export markets thus making the semi-soft barnī a higher value commodity. Consumer preferences outside of al-'Ulā have changed and the export market has been more supportive of barnī than ḥelwa.

This explains the very broad numerical supremacy that barnī date palms have acquired today in al-'Ulā palm groves. Its election is undoubtedly not recent, but might have been largely amplified by the administration and the market from the second part of the 20th century. Government incentives—notably through the conditions for agricultural loans for “modern palm groves” which target export (particularly, it seems, towards Turkey's market)—have pragmatically favored a barnī monoculture. Meanwhile, the inhabitants have diversified their consumption and diet (more rice than dates). Our survey reveals in parallel that local people substituted (sometimes clearly uprooting and replacing) rare local varieties with the elite barnī. Local farmers of the sedentary group recounted their memories of ḥelwa date palms (for instance) that were purposefully uprooted to plant barnī offshoots (we even witnessed it), and clearly identified that as a consequence of a market appeal. The supremacy of barnī has been further increased with the considerable extension of the area cultivated in palm groves in the region over the last decades, and particularly in varietal monoculture of barnī.

### 3.2.2 | Naming a palm and naming its dates according to their quality

The elite status of the barnī is also reflected by the fact that the palm, the barnī, produces dates that are not named/sold under this name. According to our study, in al-'Ulā, three names are used to sort its

dates depending on date quality (*jawda*) for marketing (Figure 3): #1 *mabrūm*: the best quality; #2 *mašrūk*: almost good, but of lower quality (smaller, and has a *qešra* “skin,” i.e., a more wrinkled epidermis); #3 *'ādī*: good only for livestock, damaged with too much “skin” (a cracked epidermis; the term *'ādī* here refers to the notion of “ordinary,” “worthless”). The second quality or the merged second and third qualities are sometimes also referred to as the name *abū qešra* (because of that whitish skin, *qešra*, epidermis). In addition to the shape, the tastes differ with the qualities, according to most farmers. The *mabrūm* are said to taste better, be more presentable, and sell for much more (Notes S2). These different commercial



**FIGURE 3** Different quality of the dates harvested from a barnī date palm. (a) A date bunch where various qualities of dates can be spotted. Ṣadar palm grove, August 24, 2021. (b) The three different qualities in the hand of a farmer. On the right, the *mabrūm* (the best quality), in the middle, the *mašrūk* (almost good, but of lower quality), and the *'ādī* (good only for livestock, damaged with too much “skin”). Jabāna/Muḡeyra palm grove, August 17, 2021. (c) Team of workers busy sorting the harvest of barnī date palms, in al-Xaiḡib palm grove in al-'Ulā. In the foreground are boxes of the best dates, the so-called *mabrūm*. November 2, 2021. Pictures: Vincent Battesti

qualities of dates are present in the same bunch of the same date palm. It is said that the older the barnī palm, the greater the proportion of *mabrūm* in its yield. Besides, farmers say that this proportion varies across years, and since this variety now dominates the date economy (Notes S2), it is a sign of a bad or good phœnicultural harvest.

That barnī fruits go by different names seems to have already been noticed by Doughty a century ago (1921, p. 153), who stated that “there are many kinds.” This is best explained later by Nasif (1988, p. 174), who highlighted that “this date is sometimes divided into two kinds, when the better quality of this type is selected because it is without a skin; this class of the barnī is known as *mabrūm*.”

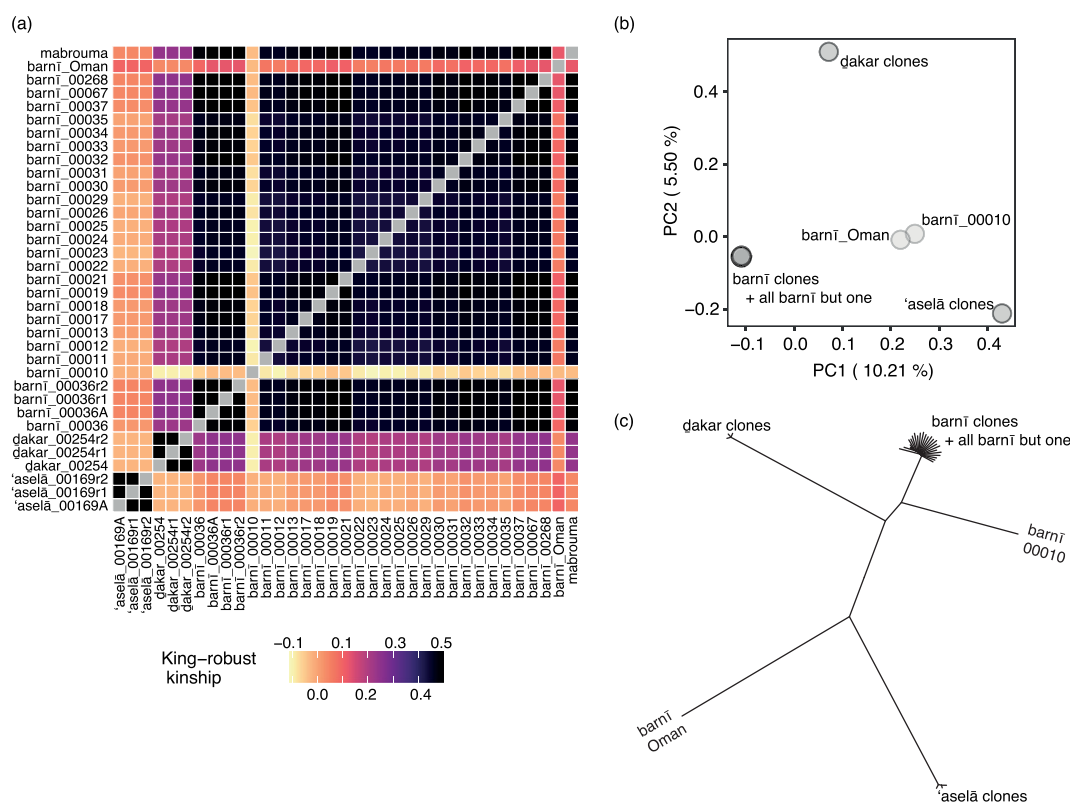
The case of naming a palm and its dates differently is unusual but not unique: One example is in Siwa (Egypt, a Berber-speaking oasis) of a variety, also elite, the *tasutet* palm whose dates are named *ṣacidi* (Battesti, 2013). Further, naming dates of the same variety according to their quality is also uncommon, although the case is not entirely unique either. For instance, in Siwa, the second elite cultivar (after the above-mentioned *ṣacidi*) is the true-to-type *alkak* date palm whose local name depends on fruit quality which varies according to growth conditions and age. The higher *alkak* date palms are called “*alkak n amles*,” meaning bearing smooth or wrinkle-free *alkak* dates, while the lower *alkak* date palms bearing smaller dates, and three times cheaper,

are called “*alkak nifūgen*” (Gros-Balthazard et al., 2020). The case is not entirely analogous to al-‘Ulā though: in Siwa, the palms bear either of these two names, and so does their production, while for the barnī, it is the dates from the same palm that are given different names.

### 3.3 | What lies behind the name barnī?

To study the genetic variability within the barnī, we sequenced the genomes of 35 date palms, including 23 barnī from al-‘Ulā collected in the four agricultural contexts (Figures 1 and 2), and obtained between 18 and 122 million of paired reads of size >76 bp (Table 1; Dataset S2). Sequencing reads from these runs and those retrieved from GenBank SRA were mapped to the Barhee BC4 genome assembly (Hazzouri et al., 2019), resulting mostly in low coverage alignments, ranging from 2.9× to 20.5× (9.3× on average; Dataset S2).

We assessed whether 26 barnī and *mabrouma* date palms from Saudi Arabia, Oman and the UAE are genetically identical using the King-robust kinship estimator (Waples et al., 2019) (Figure 4a; Dataset S3), a principal component analysis (Figure 4b), and a tree based on genetic distances (Figure 4c; Figure S3). In all three cases, 305,610,249 sites across the 18 linkage groups of the date palm genome were analyzed by ANGSD. After filtering, 1,508,939 sites



**FIGURE 4** Intra-named type variability of the barnī date palms. (a) Heatmap of the king-robust kinship estimator calculated across 1,508,939 sites in 35 date palms; (b) principal component analysis of 35 date palms (1,508,939 sites). Variance explained by each principal component (PC) is provided within parentheses; (c) tree of genetic distance calculated across 10,742 sites in 35 date palms. The consensus tree obtained from 100 bootstrap replicates may be found in Figure S3.

were retained for both the relatedness analysis and PCA analysis, and we further downsampled these sites to get 10,742 sites for computing genetic distances while limiting the effects of linkage.

Low coverage sequencing may yield kinship estimates that differ from the expectation of 0.5 for members of a single clone particularly for heterozygous samples such as date palms. We therefore sequenced three clones multiple times to assess the deviation from 0.5 that can be expected using our approach to kinship estimation (Figure S2). The barnī we sampled and sequenced twice revealed a King-robust kinship of 0.482, while kinship estimates for the mother palm and two offshoots ranged between 0.481 and 0.499 (Figure 4a; Dataset S3). Sequences of two clone triplets (ḍakar and ‘aselā) that each consisted of a palm tree and two of its offshoots each yielded a kinship estimate of 0.499. Since these samples are known to be clonal, we attribute any differences in kinship estimates from the expectation of 0.5 to be attributable to our low coverage approach. In the PCA, the accessions from each triplet/quartet of clones overlap (Figure 4b), while they group together in separate clades in the genetic distance tree (Figure 4c; Figure S3), as expected given their genetic identity.

### 3.3.1 | Barnī is a true-to-type cultivar in al-‘Ulā region

We found that the 23 barnī from al-‘Ulā are genetically identical, except barnī\_00010. Indeed, those 22 accessions cluster together in both the PCA (Figure 4b) and the tree (Figure 4c; Figure S3). Further, the King-robust kinship among them ranges from 0.424 to 0.491, nearing the theoretical 0.5 expectation and the 0.481–0.499 range observed among the known clones (Figure 4a; Dataset S3). On a technical note, we hypothesized that lower kinship estimates among the 22 barnī, compared with the known clones, may be due to a lower coverage in the former (Dataset S3). We tested the relationship between the fraction of sites with information for two individuals (used to calculate pairwise kinship) and the King-robust kinship estimate, and indeed found that they are highly positively correlated (Notes S3; Figure S4).

To understand why barnī\_00010 is genetically different, we returned to the field and found that it had been misidentified by the farm manager (a foreign worker) at the time of collection. The error is attributable to his lack of knowledge of the planting history of the sample as well as the youthfulness of the palm which made it difficult to identify using botanical characteristics. This demonstrates the need for meticulous sampling pre-informed by ethnobotanical and anthropological study.

The samples of the named type barnī were voluntarily collected from the different categories of palm groves of al-‘Ulā and these palm groves represent a great diversity of management, farming practices and social origins (Figure 2; Notes S1). Nevertheless, these differences did not lead to the selection of an ethnovariety. Indeed, all barnī whether from old, modern (20th c.) or new (21st c.) palm groves are identical. More intriguing, the four barnī from Bedouin palm grove (including the so-called “potential barnī” barnī\_00018) also are

identical to those found in the oasis palm groves. Those Bedouin palm groves found in the Balawī tribal territory about 100 km west of al-‘Ulā are of the picking palm grove type, i.e. characterized by very little labor investment, a monoculture of a few barely pollinated palms, in the bottom of the wādī, thus without necessary irrigation and without permanent habitation. This suggests a circulation, difficult over great distances, of palm offshoots between the apparently antagonistic Bedouin and sedentary worlds. The local saying that refers the Bedouin to “*ibel w ḡanem*” (dromedary and sheep) and the sedentary to *naxla* (the date palm) does suggest very different ways of life and mode of production, and consequently distinct domains of knowledge: Bedouins are less knowledgeable about date palms than oasian people, but apparently enough to maintain a clonal lineage in their picking palm groves (Notes S1).

This result—barnī is a true-to-type cultivar—was, in a sense, expected, since the mode of reproduction is supposedly vegetative in al-‘Ulā (according to our field survey, in spite of some discordant evidence of cases of seed propagation). Besides, barnī is a local elite variety and we have previously hypothesized that local elite date palms are likely to be true-to-type cultivars, despite their prevalence and therefore the mechanical probability of becoming an ethnovariety (Gros-Balthazard et al., 2020). The system of ethnovariety and local category would allow farmers to organize the diversity of palms with lower commercial value (but potentially high local value) while not multiplying the denominations for the same characteristics (i.e., not creating named types for palms that are locally considered the same); it offers a fairly flexible management of agrobiodiversity (Gros-Balthazard et al., 2020).

On a practical note, processing and distributing (including identification and traceability) a heterogeneous or a homogeneous commercial product obviously differs. Our clarification of the status of this local elite variety in al-‘Ulā, covering thousands of hectares, is thus of great importance regarding the agronomic and economic development of the phoenicultural sector in this region. The barnī of al-‘Ulā is not only precisely identified by the farmers and multiplied strictly by offshoots, even by Bedouins, but can now also be easily identified through genetic fingerprints.

On a separate note, we found that the male accession used to calibrate the genetic identity analyses (ḍakar\_00254) is closely related to barnī, with a King-robust kinship estimate ranging from 0.194 to 0.284 (Figure 4; Dataset S3). Since barnī is by far the most cultivated named type in al-‘Ulā and that males mostly arise from accidental seedlings, this male is probably an offspring of barnī.

### 3.3.2 | The date palm barnī across the Arabian Peninsula

We compared the barnī of al-‘Ulā with two barnī date palms from outside this oasis (Khaybar, Saudi Arabia and Nizwa, Oman) and with a palm identified as mabrouma (name given in al-‘Ulā to the best date category of barnī, see above; Figure 3) in a private collection from the UAE.

We found that the barnī from Khaybar, a sedentary oasis palm grove 200 km away, is genetically identical to the barnī from al-‘Ulā, indicating that the named type barnī refers to the same entity not only at the scale of al-‘Ulā oasis, but possibly at the scale of the region. The mabrouma from a palm grove in the UAE is also identical to the barnī of al-‘Ulā. It originates from a collector's farm (the exotic character is often valued by collectors). Perhaps originating from al-‘Ulā, this palm was supplied to the owner under the name “mabrūma,” possibly in reference to the *mabrūm* quality of its dates.

On the other hand, the barnī from Oman turns out to be different from that of al-‘Ulā (Figure 4). This alternate barnī appears to be present in Oman, mainly in the Northern and Southern al-Sharqia regions and is the tenth most cultivated variety in the country (Al-Yahyai & Al-Khanjari, 2019; Al-Yahyai & Khan, 2015). Although Popenoe and Bennett (1913, pp. 226–227) recognized the probable difference between these barnī, our genetic study confirmed this example of homonymy.

### 3.4 | A first glimpse into the date palm agrobiodiversity in al-‘Ulā

In Northwestern Saudi Arabia, date palms have dominated the oasis agricultural system since at least the fourth century BCE (Bouchaud, 2013), and al-‘Ulā region is known to be home to the two-millennia-old Nabataean site of Hegra (Mada’in Šaleḥ). In this area, dates have been consumed since at least the end of the 2nd millennium BCE (Rohmer et al., in press) and they have a particular symbolic importance in the Nabatean period as attested by the date necklace excavated in a tomb of this period (Bouchaud et al., 2015). The region is positioned in a strategic location, on a critical trade route, namely, the incense road, connecting the South of the Arabian Peninsula with the Levantine region roughly during the 7th century BCE and the 2nd century CE. It is also at the crossroads of date palm diversity, between the distinct North African and West Asian gene pools (Flowers et al., 2019; Hazzouri et al., 2015).

So far, no studies have focused on the origin and extent of the diversity of Northwestern Saudi Arabia date palms, including al-‘Ulā oasis. A few studies focused on the genetic diversity of date palm varieties in Saudi Arabia (Aleid et al., 2015; Al-Khalifah & Askari, 2003; Al-Qurainy et al., 2011) but how those varieties, some potentially cultivated in al-‘Ulā, relate to cultivars from other regions remain to be elucidated.

Here, we analyzed the genome of the barnī of al-‘Ulā and of two other accessions from this oasis, one of the ‘asela variety (‘asela\_000169A) and a male (ḡakar\_00254), along with that of 112 *Phoenix* spp. (Dataset S1). Given the genetic identity of all barnī palms from al-‘Ulā, we picked the accession with the highest coverage, namely barnī\_000268, for downstream analyses. Aligning reads to the Barhee BC4 genome led to an average coverage of 17.3x across accessions (Dataset S2) and we identified 1,007,281 SNPs after quality filtering, which we used in subsequent analyses (Methods S1).

### 3.4.1 | Genetic relationships between date palms from al-‘Ulā and from North Africa and West Asia

The relationships and the genetic structure of the 3 date palms of al-‘Ulā and 88 date palms from North Africa and West Asia were determined by performing model-based genetic clustering (Figure 5a; Figures S5 and S6), reconstructing a phylogenetic tree (Figure 5b) and applying a PCA (Figure 5c; Dataset S4) using the SNP data. All three analyses (Figure 5) corroborated previous results on date palms, that is, that they can be split in two main clusters (North Africa and West Asia) with Egyptian accessions being a mix between the two (reviewed by Gros-Balthazard & Flowers, 2021).

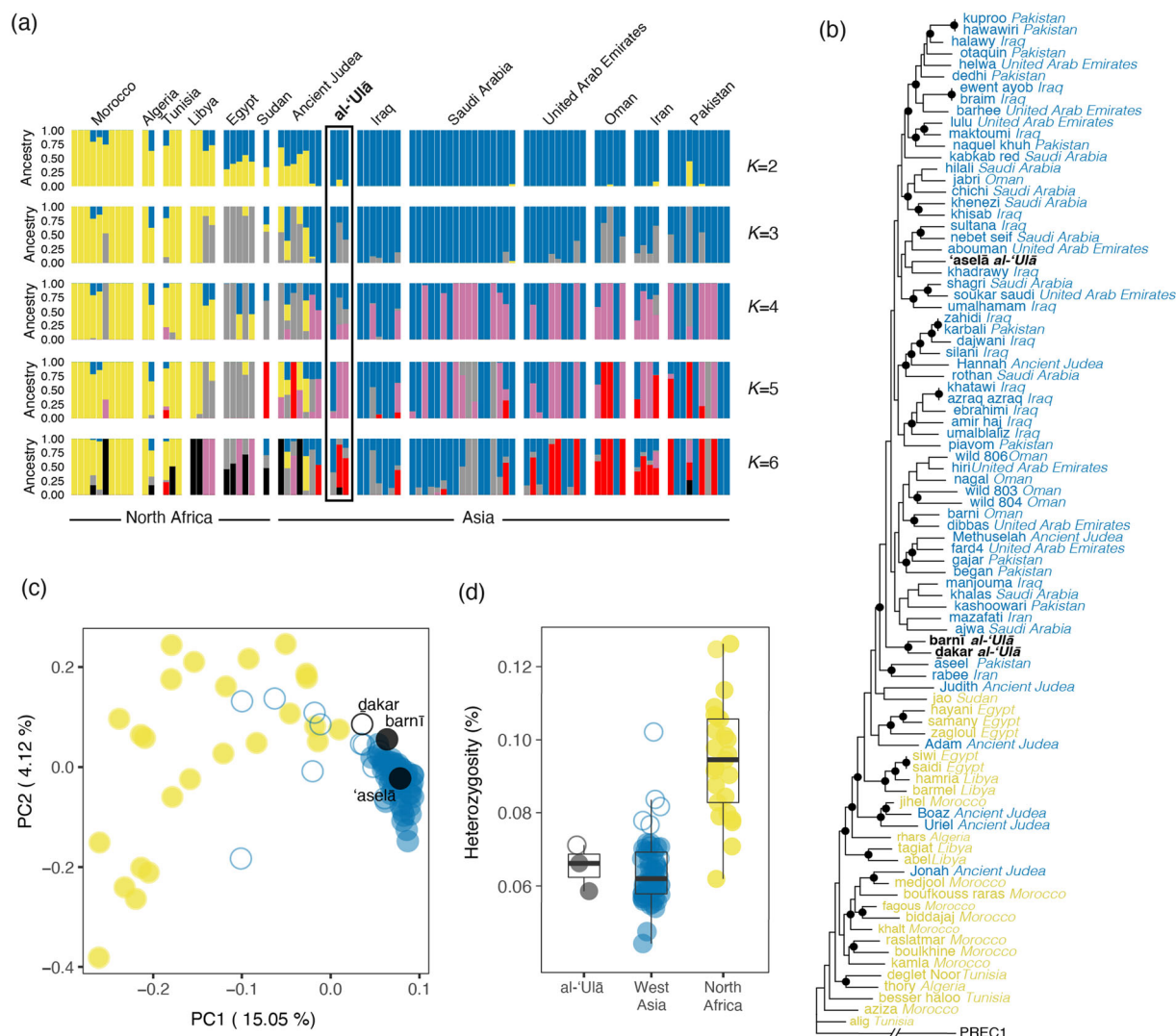
Regarding the diversity in al-‘Ulā, our results revealed that it may represent a unique mixing among West Asian date palms (Figure 5): The three al-‘Ulā palms indeed cluster with West Asian date palms, but the male ḡakar shows influence from North African diversity. It is found on the edge of the West Asian group in the PCA, close to North African date palms (Figure 5c) and shows mixed ancestry from the West Asian and the North African clusters in the clustering analysis (Figure 5a).

### 3.4.2 | Genetic diversity of al-‘Ulā date palms

To determine the genetic diversity of al-‘Ulā date palms, the proportion of heterozygous sites in each date palm accession was calculated using pixy (Korunes & Samuk, 2021), which confirmed results from previous reports (Flowers et al., 2019; Gros-Balthazard et al., 2017, 2021; Hazzouri et al., 2015): i.e., North African date palms display a higher diversity (mean heterozygosity  $0.094 \pm 0.016\%$ ) than cultivated West Asian date palms (mean heterozygosity  $0.062 \pm 0.0069\%$ ; one-sided Wilcoxon rank sum test,  $W = 1243$ ,  $P = 1.11 \times 10^{-11}$ ). Regarding the three al-‘Ulā date palms analyzed in this study, their diversity (mean heterozygosity  $0.065 \pm 0.0064\%$ ) was on average comparable to that of cultivated West Asian date palms (Wilcoxon rank sum test,  $W = 99$ ,  $P = 0.41$ ), and lower than that found in African date palms (one-sided Wilcoxon rank sum test,  $W = 3$ ,  $P = 0.0021$ ).

### 3.4.3 | Evidence of gene flow between al-‘Ulā date palms and the wild relative *P. theophrasti*

Interspecific introgression has shaped the diversity of North African and Levantine date palms since at least 2000 years (Flowers et al., 2019; Gros-Balthazard et al., 2021; Pérez-Escobar et al., 2021). The date palm was presumably domesticated in the Persian Gulf region during the fifth millennium BCE and its cultivation then spread across Arabia and further across North Africa. Modern North African date palms and ancient Levantine date palms show higher genetic diversity than that found in West Asia, which may at least partially be explained by gene flow from a wild relative species, that is, *Phoenix theophrasti*, whose present-day distribution includes Crete and the coast of



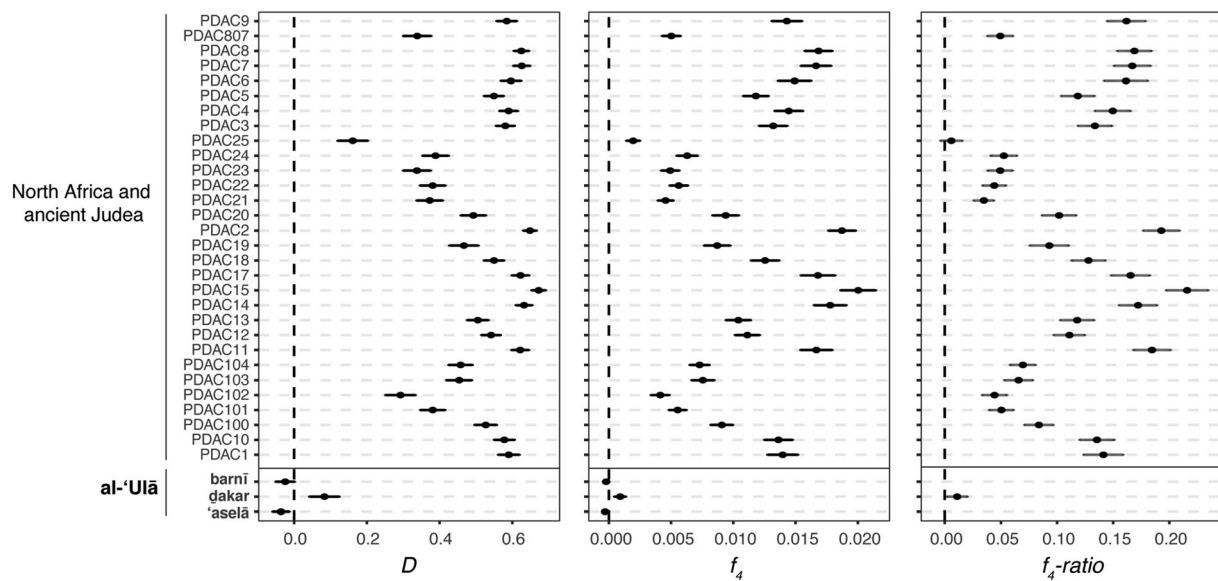
**FIGURE 5** Structure and diversity of 88 date palms from North Africa and West Asia, including al-'Ulā. (a) Ancestry coefficients from the West of the distribution to the East inferred from 28,406 single nucleotide polymorphisms (SNPs). The date palms from al-'Ulā are ordered as follow (from left to right): 'aselā, ḡakar, and barnī. Labels and additional  $K$  values may be found in Figure S4; (b) phylogenetic relationships inferred from 50,171 SNPs. The tree was rooted with *Phoenix reclinata*. Black circles indicate nodes with >80% bootstrap support. Labels in yellow correspond to North African date palms, in blue to West Asian date palms, and in black to those collected in al-'Ulā; (c) principal component analysis inferred from 14,137 SNPs. Black dots correspond to the date palms from al-'Ulā. Yellow dots correspond to North African date palms. Full blue dots correspond to West Asian date palms, the empty ones are those having a fraction of their ancestry from the African cluster in (a). The variance explained by each principal component (PC) is given in parentheses. Coordinates on PC1 and 2, and additional PCs may be found in Dataset S4; (d) fraction of heterozygote sites across 88 date palm genomes calculated from 105,236,083 sites including 1,007,281 SNPs. Same legend as panel (c)

Turkey, or a *theophrasti*-like species (Flowers et al., 2019; Gros-Balthazard & Flowers, 2021). The oasis of al-'Ulā has long been connected to the Levantine region and its ports on the Mediterranean Sea. Whether the genetic make-up of its date palms also displays ancestry from *P. theophrasti*, reflecting this possible bond, remains to be determined.

To evaluate whether the date palms from al-'Ulā showed evidence of introgression, both  $D$ - and  $f_4$ -statistics were computed (Figure 6; Methods S2; Datasets S5-S6). This revealed that both barnī and 'aselā do not show any evidence of introgression (Figure 6). On the other hand, North African date palms, and the ancient Judean date

palms, that were previously shown to be admixed, displayed significant positive  $D$ - and  $f_4$ -statistics (Figure 6; Flowers et al., 2019; Gros-Balthazard et al., 2021). Interestingly, this is also the case for the male ḡakar\_00254 from al-'Ulā (Figure 6).

We further estimated the fraction of *theophrasti* ancestry in the date palms showing evidence of admixture according to  $D$ - and  $f_4$ -statistics (Figure 6; Dataset S7), including the male ḡakar\_00254. We found that the latter displays about 1.1% of its genome from *Phoenix theophrasti*. This is in the lower range of what is observed in North African and Judean date palms (average 11.23%, ranging from 0.59 to 21.58%; Dataset S7).



**FIGURE 6** Admixture between date palms from North Africa, ancient Judea, al-'Ulā and *P. theophrasti*.  $D$ -statistic and  $f_4$ -statistic, both testing whether modern North African, ancient Judean, and al-'Ulā date palms show an excess of shared alleles with *P. theophrasti*, in which case they are significantly positive. We estimated the  $D$ -statistics and  $f_4$ -statistics using the following tree: (((test sample, West Asian date palms), *P. theophrasti*), *P. reclinata*). The  $f_4$ -ratio statistic, indicates the fraction of *P. theophrasti* genomes found in the test sample and calculated as  $f_4(A,O; X,C)/f_4(A,O; B,C)$ , where X is the test sample; A is a sister species, namely *P. sylvestris*; B and C are the mixing populations, namely, West Asian date palms and *P. theophrasti*, respectively; and O is the outgroup, that is, *P. reclinata*. It was calculated only for samples showing evidence of an excess of shared alleles with *P. theophrasti* as evidenced by  $D$ - and  $f_4$ -statistics, that is all samples except barnī and 'aselā. A negative  $D$  and  $f_4$  in this context imply a greater degree of allele sharing between West Asian samples and *theophrasti*, whereas positive values would imply greater sharing between the test sample and *theophrasti*. More details on the methods may be found in Methods S2 while the data are in Datasets S5, S6, and S7.

The evidence of gene flow from *P. theophrasti* found in date palms of the Levant has been hypothesized to be related to the growing control of the Roman Empire in the region 2000 years ago, favoring exchange of goods, including dates, with North Africa (Gros-Balthazard et al., 2021). Indeed, a changeover from absence or low *P. theophrasti* ancestry to ~10% of *P. theophrasti* ancestry coincided with a shift in imperial control of the region in favor of the Romans. The region of al-'Ulā, with the famous site of Hegra, was part of the Nabatean kingdom, and a stop on the trading routes connecting southern and eastern Arabia to Petra and the Mediterranean Sea. The site is also located, since 106 CE, on the border of the newly created Roman province of Arabia on the ruins of this kingdom. To observe evidence of ancestry that is rather characteristic of North Africa and the Levant is therefore not unlikely. Whether this reflects ancient (i.e., from the Nabateo-Roman period) or other gene flow remains to be elucidated.

### 3.4.4 | Maternal origin of al-'Ulā date palms

In date palm, two deeply diverged chlorotypes (so-called western and eastern) have been reported (Pintaud et al., 2013). The eastern (or oriental) chlorotype is found in most West Asian date palms, while in North Africa, the so-called occidental is prominent. Previous analysis of date palm chlorotypes showed a gradient in frequency of the

oriental chlorotype from low in Northwestern Africa to ~50% in Egypt, which suggest strong gene flow from West Asia (Gros-Balthazard et al., 2017; Zehdi-Azouzi et al., 2015). Our analysis of chloroplast sequences from three al-'Ulā date palms revealed that all bear the oriental chlorotypes indicating a maternal origin from West Asia (Figure S7).

## 4 | CONCLUSION AND PROSPECTS

What lies behind variety names in clonally propagated crops has been insufficiently explored, although it is a key element for assessing agrobiodiversity (e.g., in oca, Bonnavé et al., 2014 or in date palm, Gros-Balthazard et al., 2020). Many studies have highlighted variation within variety names; in date palms (Al-Khalifah & Askari, 2003; Al-Ruqaishi et al., 2008; Elhoumaizi et al., 2006; Sabir et al., 2014), and in other clonally propagated fruit crops as well (e.g., in grapevine, Meneghetti et al., 2012 or in olive, Lazović et al., 2018), although most do not reference deliberate cultivation practices as a source of this variation (but see Battesti et al., 2018; Gros-Balthazard et al., 2020).

By engaging with local farmers, we have established that barnī cultivated in al-'Ula and its surroundings, is a true-to-type cultivar, that is, its local identity given by the farmers corresponds to a unique genetic identity. It implies that, locally, barnī date palms have always been

strictly vegetatively propagated, even by Bedouins in remote desert areas and in different palm grove farming systems. But geographic scale matters: barnī from al-'Ulā oasis is distinct from that of Oman, and we therefore confirmed the homonymy foreseen by Popenoe and Bennett (1913).

In al-'Ulā alone, to date, we estimate that about 99 varieties are cultivated, some of which have been given names close to barnī: for example, barniyat al-'aiṣ, barniyat banāt sa'ad, and barniyat bader. The two latter are probably old local rare named types, while barniyat al-'aiṣ (the "barnī from al-'Aiṣ") clearly designates a variety said from the oasis of al-'Aiṣ (KSA), which is referred to in al-'Aiṣ as barnī. This stresses the importance of understanding the categorization and naming systems used by farmers in relation to the way they propagate palms for a proper understanding and assessment of crop biodiversity.

In addition, our joint analyses of three date palms from al-'Ulā and other *Phoenix* specimens, have revealed an intriguing diversity patterns, where, although clustering with West Asian date palms, one of them shows influence from North Africa. Further analyses, comprising all date palm diversity from al-'Ulā area, will provide further insights into the diversity and history of the keystone species of this antique oasis.

## ACKNOWLEDGEMENTS

We are grateful to the farmers, who opened their gardens and farms to us, shared their knowledge and know-how, and offered us unflinching hospitality.

We thank Claire Newton for providing a sample; Marc Arnoux, Nizar Drou, Michael Dhar, and the New York University Abu Dhabi Bioinformatics Core for assistance with DNA sequencing and bioinformatic analyses; Luis Rivera Serna from the Center for Desert Agriculture at KAUST for assistance with bioinformatic analyses. This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. We are particularly grateful to Shenglong Wang for his unwavering support while using NYU cluster.

This work was funded by the French Agency for AIUla Development (AFALULA) with his Saudi partner the Royal Commission for AIUla (RCU) through a grant awarded to Vincent Battesti and Muriel Gros-Balthazard (project al-'Ulā DPA: Ethnographic, genetic, and morphometric analyses of the date palm agrobiodiversity in al-'Ulā oasis). We wish to thank them for their financial and logistic support, and in particular Elisabeth Dodinet and Stéphane Forman. Legally, the Royal Commission for AIUla is the owner and provider of the material.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

MGB and VB conceived and designed the analysis. VB led the ethnobotanical survey accompanied by MGB. MGB, VB, SF, JMF, MDP, RW and NM contributed to data. MGB, VB, MB, SI, JFT, JMF, MDP, YB contributed to the analyses. MGB and VB wrote the paper.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supporting information of this article. Sources for all downloaded genomic data are stated in Dataset S1. Sequencing reads for the newly sequenced accessions can be found in the public sequence database GenBank under the BioProject ID PRJNA817028 and BioSample IDs can be found in Dataset S1.

## ORCID

Muriel Gros-Balthazard  <https://orcid.org/0000-0002-2587-3946>

Vincent Battesti  <https://orcid.org/0000-0002-5793-1098>

Jonathan M. Flowers  <https://orcid.org/0000-0002-8752-205X>

Sylvie Ferrand  <https://orcid.org/0000-0003-1088-5941>

Matthieu Breil  <https://orcid.org/0000-0002-7981-7317>

Sarah Ivorra  <https://orcid.org/0000-0003-0314-8054>

Jean-Frédéric Terral  <https://orcid.org/0000-0003-1921-2161>

Michael D. Purugganan  <https://orcid.org/0000-0002-9197-4112>

Rod A. Wing  <https://orcid.org/0000-0001-6633-6226>

Nahed Mohammed  <https://orcid.org/0000-0002-8857-3246>

Yann Bourgeois  <https://orcid.org/0000-0002-1809-387X>

## REFERENCES

- Aleid, S. M., Al-Khayri, J. M., & Al-Bahrany, A. M. (2015). Date palm status and perspective in Saudi Arabia. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *Date palm genetic resources and utilization* (pp. 49–95). Springer Netherlands. [https://doi.org/10.1007/978-94-017-9707-8\\_3](https://doi.org/10.1007/978-94-017-9707-8_3)
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Al-Khalifah, N. S., & Askari, E. (2003). Molecular phylogeny of date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia by DNA fingerprinting. *Theoretical and Applied Genetics*, 107(7), 1266–1270. <https://doi.org/10.1007/s00122-003-1369-y>
- Al-Khayri, J. M., Jain, S. M., & Johnson, D. V. (Eds.) (2015). *Date palm genetic resources and utilization*. Springer Netherlands. <https://doi.org/10.1007/978-94-017-9707-8>
- Al-Qurainy, F., Khan, S., Al-Hemaid, F. M., Ali, M. A., Tarroum, M., & Ashraf, M. (2011). Assessing molecular signature for some potential date (*Phoenix dactylifera* L.) cultivars from Saudi Arabia, based on chloroplast DNA sequences rpoB and psbA-trnH. *International Journal of Molecular Sciences*, 12(10), 6871–6880. <https://doi.org/10.3390/ijms12106871>
- Al-Ruqaishi, I. A., Davey, M., Alderson, P., & Mayes, S. (2008). Genetic relationships and genotype tracing in date palms (*Phoenix dactylifera* L.) in Oman, based on microsatellite markers. *Plant Genetic Resources*, 6(1), 70–72. <https://doi.org/10.1017/S1479262108923820>
- Al-Yahyai, R., & Al-Khanjari, S. (2019). Biodiversity of date palm in the Sultanate of Oman. *African Journal of Crop Science*, 7(10), 1–7.
- Al-Yahyai, R., & Khan, M. M. (2015). Date palm status and perspective in Oman. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *Date palm genetic resources and utilization* (pp. 207–240). Springer Netherlands. [https://doi.org/10.1007/978-94-017-9707-8\\_6](https://doi.org/10.1007/978-94-017-9707-8_6)
- Bahuchet, S. (2017). *Les jardiniers de la nature*. Odile Jacob.
- Battesti, V. (2005). *Jardins au désert, Évolution des pratiques et savoirs oasiens: Jérid tunisien*. Éditions IRD. <https://doi.org/10.4000/books.irdeditions.10160>
- Battesti, V. (2013). L'agrobiodiversité du dattier (*Phoenix dactylifera* L.) dans l'oasis de Siwa (Égypte): Entre ce qui se dit, s'écrit et s'oublie. *Revue d'ethnoécologie*, 4. <https://doi.org/10.4000/ethnoecologie.1538>

- Battesti, V., Gros-Balthazard, M., Ogéron, C., Ivorra, S., Terral, J.-F., & Newton, C. (2018). Date palm agrobiodiversity (*Phoenix dactylifera* L.) in Siwa Oasis, Egypt: Combining ethnography, morphometry, and genetics. *Human Ecology*, 46(4), 529–546. <https://doi.org/10.1007/s10745-018-0006-y>
- Bonnaive, M., Bleeckx, G., Rojas Beltrán, J., Maughan, P., Flamand, M.-C., Terrazas, F., & Bertin, P. (2014). Farmers' unconscious incorporation of sexually-produced genotypes into the germplasm of a vegetatively-propagated crop (*Oxalis tuberosa* Mol.). *Genetic Resources and Crop Evolution*, 61(4), 721–740. <https://doi.org/10.1007/s10722-013-0068-z>
- Bouchaud, C. (2013). Exploitation végétale des oasis d'Arabie: Production, commerce et utilisation des plantes. L'exemple de Madâ'in Sâlih (Arabie Saoudite) entre le IV<sup>e</sup> siècle av. J.-C. et le VII<sup>e</sup> siècle apr. J.-C. *Revue d'ethnoécologie*, 4. <https://doi.org/10.4000/ethnoecologie.1217>
- Bouchaud, C., Sachet, I., Dal Prà, P., Delhospital, N., Douaud, R., & Leguilloux, M. (2015). New discoveries in a Nabataean tomb. Burial practices and 'plant jewellery' in ancient Hegra (Madâ'in Sâlih, Saudi Arabia). *Arabian Archaeology and Epigraphy*, 26(1), 28–42. <https://doi.org/10.1111/aae.12047>
- Caillon, S., & Degeorges, P. (2007). Biodiversity: Negotiating the border between nature and culture. *Biodiversity and Conservation*, 16(10), 2919–2931. <https://doi.org/10.1007/s10531-007-9149-7>
- Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., & Christen, R. (2006). Tree-Dyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, 7, 439. <https://doi.org/10.1186/1471-2105-7-439>
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., & Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36(Web Server), W465–W469. <https://doi.org/10.1093/nar/gkn180>
- Doughty, C. M. (1921). *Travels in Arabia Deserta* (Edition (Vol. 1)). P.L. Warner and J. Cape. <http://archive.org/details/travelsinarabiad01doug>
- Elhoumaizi, M. A., Devanand, P. S., Fang, J., & Chao, C.-C. T. (2006). Confirmation of 'Medjool' date as a landrace variety through genetic analysis of 'Medjool' accessions in Morocco. *Journal of the American Society for Horticultural Science*, 131(3), 403–407. <https://doi.org/10.21273/JASHS.131.3.403>
- Elias, M., Rival, L. M., & Mckey, D. (2000). Perception and management of cassava (*Manihot esculenta* Crantz) diversity among Makushi Amerindians of Guyana (South America). *Journal of Ethnobiology*, 20, 239–265.
- Elmeer, K., Mattat, I., Al Malki, A., Al-Mamari, A.-G., Al-Jabri, A., Buhendi, A., Alkhabaz, S., Abu-Idrees, A., Abdulkareem, A., Udupa, S. M., & Baum, M. (2020). Intra-cultivar variability at microsatellite loci in date palm cultivars across the GCC countries. *QScience Connect*, 2020(1), 3. <https://doi.org/10.5339/connect.2020.3>
- Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., Ferrand, S., Khierallah, H. S. M., Fuller, D. Q., Aberlenc, F., Fournarakis, C., & Purugganan, M. D. (2019). Cross-species hybridization and the origin of North African date palms. *Proceedings of the National Academy of Sciences*, 116(5), 1651–1658. <https://doi.org/10.1073/pnas.1817453116>
- Gros-Balthazard, M., Battesti, V., Ivorra, S., Paradis, L., Aberlenc, F., Zango, O., Zehdi-Azouzi, S., Moussouni, S., Naqvi, S. A., Newton, C., & Terral, J. (2020). 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(8), 1818–1840. <https://doi.org/10.1111/eva.12930>
- Gros-Balthazard, M., & Flowers, J. M. (2021). A brief history of the origin of domesticated date palms. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *The date palm genome* (Vol. 1). Phylogeny, biodiversity and mapping. (pp. 55–74). Springer International Publishing. [https://doi.org/10.1007/978-3-030-73746-7\\_3](https://doi.org/10.1007/978-3-030-73746-7_3)
- Gros-Balthazard, M., Flowers, J. M., Hazzouri, K. M., Ferrand, S., Aberlenc, F., Sallon, S., & Purugganan, M. D. (2021). The genomes of ancient date palms germinated from 2,000 y old seeds. *Proceedings of the National Academy of Sciences*, 118(19), e2025337118. <https://doi.org/10.1073/pnas.2025337118>
- Gros-Balthazard, M., Galimberti, M., Kousathanas, A., Newton, C., Ivorra, S., Paradis, L., Vigouroux, Y., Carter, R., Tengberg, M., Battesti, V., Santoni, S., Falquet, L., Pintaud, J.-C., Terral, J.-F., & Wegmann, D. (2017). The discovery of wild date palms in Oman reveals a complex domestication history involving centers in the Middle East and Africa. *Current Biology*, 27(14), 2211–2218.e8. <https://doi.org/10.1016/j.cub.2017.06.045>
- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5), 696–704. <https://doi.org/10.1080/10635150390235520>
- Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience*, 8(5), giz034. <https://doi.org/10.1093/gigascience/giz034>
- Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S. M., Rosas, U., Pham, G. M., Meyer, R. S., Johansen, C. K., Fresquez, Z. A., Masmoudi, K., Haider, N., El Kadri, N., Idaghdour, Y., Malek, J. A., Thirkhill, D., Markhand, G. S., Krueger, R. R., Zaid, A., & Purugganan, M. D. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications*, 6(1), 8824. <https://doi.org/10.1038/ncomms9824>
- 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., Rosas, U., Zhang, J., Talag, J., Lee, S., Kudrna, D., Powell, R. F., Leitch, I. J., Krueger, R. R., Wing, R. A., ... Purugganan, M. D. (2019). Genome-wide association mapping of date palm fruit traits. *Nature Communications*, 10(1), 4680. <https://doi.org/10.1038/s41467-019-12604-9>
- Jarvis, D. I., Hodgkin, T., Brown, A. H. D., Tuxill, J. D., López Noriega, I., Smale, M., Sthapit, B. R., & Samper, C. (2016). *Crop genetic diversity in the field and on the farm: Principles and applications in research practices*. YALE University Press.
- Jaussen, A., & Savignac, R. (1914). *Mission archéologique en Arabie* (mars-mai 1907) 2. El-'Ela, d'Hégra à Teima, Harrah de Tebouk (Ernest Leroux, Éditeur). Publications de la Société des fouilles archéologiques. <https://archive.org/details/missionarcheolog21jaus>
- Johnson, D. V., Jain, S. M., & Al-Khayri, J. M. (2013). Seedling date palms (*Phoenix dactylifera* L.) as genetic resources. *Emirates Journal of Food and Agriculture*, 25(11), 809. <https://doi.org/10.9755/ejfa.v25i11.16497>
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 13, 356. <https://doi.org/10.1186/s12859-014-0356-4>
- Korunes, K. L., & Samuk, K. (2021). PIXY: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, 21(4), 1359–1368. <https://doi.org/10.1111/1755-0998.13326>
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAXML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
- Krueger, R. R. (2011). Date palm germplasm. In S. M. Jain, J. M. Al-Khayri, & D. V. Johnson (Eds.), *Date palm biotechnology* (pp. 313–336). Springer Netherlands. [https://doi.org/10.1007/978-94-007-1318-5\\_16](https://doi.org/10.1007/978-94-007-1318-5_16)
- Lane, E. W. (1863). *An Arabic-English lexicon* (Vol. 1-8). Williams and Norgate.
- Lazović, B., Klepo, T., Adakalić, M., Šatović, Z., Arbeiter, A. B., Hladnik, M., Strikić, F., Liber, Z., & Bandelj, D. (2018). Intra-varietal variability and genetic relationships among the homonymic East Adriatic olive (*Olea europaea* L.) varieties. *Scientia Horticulturae*, 236, 175–185. <https://doi.org/10.1016/j.scienta.2018.02.053>



- Leclerc, C., & Coppens d'Eeckenbrugge, G. (2011). Social organization of crop genetic diversity. The  $G \times E \times S$  interaction model. *Diversity*, 4(1), 1–32. <https://doi.org/10.3390/d4010001>
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). *pcadapt*: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17(1), 67–77. <https://doi.org/10.1111/1755-0998.12592>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McKey, D., Elias, M., Pujol, B., & Duputié, A. (2010). The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist*, 186(2), 318–332. <https://doi.org/10.1111/j.1469-8137.2010.03210.x>
- Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, 210(2), 719–731. <https://doi.org/10.1534/genetics.118.301336>
- Meneghetti, S., Poljuha, D., Frare, E., Costacurta, A., Morreale, G., Bavaresco, L., & Calò, A. (2012). Inter- and intra-varietal genetic variability in Malvasia cultivars. *Molecular Biotechnology*, 50(3), 189–199. <https://doi.org/10.1007/s12033-011-9423-5>
- Munier, P. (1973). *Le palmier-dattier*. G.-P. Maisonneuve&Larose.
- Nasif, A. A. (1988). *Al-'Ulā: An historical and archaeological survey with special reference to its irrigation system*. Riyadh: King Saud University. (Ph.D. thesis in Victoria University of Manchester, 1981).
- Newton, C., Gros-Balthazard, M., Ivorra, S., Paradis, L., Pintaud, J.-C., & Terral, J.-F. (2013). *Phoenix dactylifera* and *P. sylvestris* in northwestern India: A glimpse of their complex relationships. *Palms*, 57(1), 37–50.
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pérez-Escobar, O. A., Bellot, S., Przelomska, N. A. S., Flowers, J. M., Nesbitt, M., Ryan, P., Gutaker, R. M., Gros-Balthazard, M., Wells, T., Kuhnhäuser, B. G., Schley, R., Bogarín, D., Dodsworth, S., Diaz, R., Lehmann, M., Petoe, P., Eiserhardt, W. L., Preick, M., Hofreiter, M., ... Baker, W. J. (2021). Molecular clocks and archeogenomics of a late period Egyptian date palm leaf reveal introgression from wild relatives and add timestamps on the domestication. *Molecular Biology and Evolution*, 38(10), 4475–4492. <https://doi.org/10.1093/molbev/msab188>
- Petr, M., Vernot, B., & Kelso, J. (2019). admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*, 35(17), 3194–3195. <https://doi.org/10.1093/bioinformatics/btz030>
- Pintaud, J.-C., Ludeña, B., Aberlenc-Bertossi, F., Zehdi, S., Gros-Balthazard, M., Ivorra, S., Terral, J.-F., Newton, C., Tengberg, M., Abdoukader, S., Daher, A., Nabil, M., Saro Hernández, I., González-Pérez, M. A., Sosa, P., Santoni, S., Moussouni, S., Si-Dehbi, F., & Bouguedoura, N. (2013). Biogeography of the date palm (*Phoenix dactylifera* L., arecaceae): Insights on the origin and on the structure of modern diversity. *Acta Horticulturae*, 994, 19–38. <https://doi.org/10.17660/ActaHortic.2013.994.1>
- Popenoe, P. B., & Bennett, C. L. (1913). *Date growing in the old world and the new/by Paul B. Popenoe (with a chapter on the food value of the date)*. West India Gardens. <https://doi.org/10.5962/bhl.title.32190>
- R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rhouma, A. (1994). *Le palmier dattier en Tunisie, I. Le patrimoine génétique* (Vol. 1). Arabesques, INRA Tunisie, GRIDAO France, PNUD/FAO.
- Rhouma, A. (2005). *Le palmier dattier en Tunisie, I. Le patrimoine génétique* (Vol. 2). IPGRI, UNDP, GEF/FEM, INRAT.
- Rohmer, J., Lesguer, F., Bouchaud, C., Purdue, L., Alsuhaibani, A., Tourtet, F., Monchot, H., Dabrowski, V., Decaix, A., & Desormeau, X. (in press). New clues to the development of the oasis of Dadan. Results from a test excavation at Tall al-Salimiyah (al-'Ulā, Saudi Arabia). *Proceedings of the Seminar for Arabian Studies*.
- Sabir, J. S. M., Arasappan, D., Bahieldin, A., Abo-Aba, S., Bafeel, S., Zari, T. A., Edris, S., Shokry, A. M., Gadalla, N. O., Ramadan, A. M., Atef, A., Al-Kordy, M. A., El-Domyati, F. M., & Jansen, R. K. (2014). Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and close phylogenetic relationships among cultivars. *PLoS ONE*, 9(4), e94158. <https://doi.org/10.1371/journal.pone.0094158>
- Scarcelli, N., Tostain, S., Vigouroux, Y., Agbangla, C., Dainou, O., & Pham, J.-L. (2006). Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. *Molecular Ecology*, 15(9), 2421–2431. <https://doi.org/10.1111/j.1365-294X.2006.02958.x>
- Vieira, F. G., Lassalle, F., Korneliusson, T. S., & Fumagalli, M. (2016). Improving the estimation of genetic distances from next-generation sequencing data: Genetic distances from NGS data. *Biological Journal of the Linnean Society*, 117(1), 139–149. <https://doi.org/10.1111/bij.12511>
- Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, 28(1), 35–48. <https://doi.org/10.1111/mec.14954>
- Zaid, A., & Arias-Jiménez, E. J. (Eds.). (1999). *Date palm cultivation*. FAO, Plant Production and Protection Paper.
- Zaid, A., & Oihabi, A. (Eds.). (2022). *Mejhoul Variety, The Jewel of Dates. Origin, distribution and international market*. Khalifa International Award for Date Palm and Agricultural Innovation.
- Zehdi-Azouzi, S., Cherif, E., Moussouni, S., Gros-Balthazard, M., Abbas Naqvi, S., Ludeña, B., Castillo, K., Chabrilange, N., Bouguedoura, N., Bennaceur, M., Si-Dehbi, F., Abdoukader, S., Daher, A., Terral, J.-F., Santoni, S., Ballardini, M., Mercuri, A., Ben Salah, M., Kadri, K., ... Aberlenc-Bertossi, F. (2015). Genetic structure of the date palm (*Phoenix dactylifera*) in the Old World reveals a strong differentiation between eastern and western populations. *Annals of Botany*, 116(1), 101–112. <https://doi.org/10.1093/aob/mcv068>
- Zhang, C.-R., Aldosari, S. A., & Nair, M. G. (2015). Determination of the variability of sugars in date fruit varieties. *Journal of Plantation Crops*, 43(1), 53–61.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Gros-Balthazard, M., Battesti, V., Flowers, J. M., Ferrand, S., Breil, M., Ivorra, S., Terral, J.-F., Purugganan, M. D., Wing, R. A., Mohammed, N., & Bourgeois, Y. (2023). What lies behind a fruit crop variety name? A case study of the barni date palm from al-'Ulā oasis, Saudi Arabia. *Plants, People, Planet*, 5(1), 82–97. <https://doi.org/10.1002/ppp3.10326>

## ***Plants, People, Planet* Supporting Information**

**Article title:** What lies behind a fruit crop variety name? A case study of the barnī date palm from al-‘Ulā oasis, Saudi Arabia

**Authors:** Muriel Gros-Balthazard, Vincent Battesti, Jonathan M. Flowers, Sylvie Ferrand, Matthieu Breil, Sarah Ivorra, Jean-Frédéric Terral, Michael Purugganan, Rod A. Wing, Nahed Mohammed, Yann Bourgeois.

The following Supporting Information is available for this article:

**Fig. S1** Pictures of barnī date palms.

**Fig. S2** Pictures of the date palms groups of clones used to calibrate the relatedness analyses.

**Fig. S3** Consensus tree of 35 date palms based on genetic distance calculated using genotype likelihoods across 10,742 sites.

**Fig. S4** Relationship between the fraction of sites covered by one or more reads in both sequenced genomes and the King-robust kinship estimator.

**Fig. S5** Ancestry coefficients inferred in 88 date palms using a randomly downsampled set of 28,406 Single Nucleotide Polymorphisms (SNPs).

**Fig. S6** Cross-validation error of the admixture model.

**Fig. S7** Chloroplast DNA tree with uncorrected distances reconstructed using the Neighbor-joining method.

**Dataset S1** List of *Phoenix* spp. accessions, their geographic origin and source.

**Dataset S2** Sequencing and read alignment metrics.

**Dataset S3** Kinship among 35 date palms assessed using 1,515,793 sites across the genome.

**Dataset S4** Coordinates in the Principal Component Analysis of 88 date palms including three from al-‘Ulā.

**Dataset S5** Summary of *D*-tests of admixture in date palm and *Phoenix theophrasti*.

**Dataset S6** Summary of  $f_4$ -statistics for testing of admixture in date palm and wild *Phoenix* species.

**Dataset S7** Summary of  $f_4$ -ratio statistics for inferring the proportion of *Phoenix theophrasti* ancestry in the North African, ancient Judean and ḡakar (al-'Ulā) date palm genomes.

**Methods S1** Detailed protocol for sequencing read processing, genome alignment, variant calling and Single Nucleotide Polymorphisms (SNPs) filtering.

**Methods S2** Admixture tests and ancestry proportions.

**Notes S1** Date palms growing situations in al-'Ulā region.

**Notes S2** Categorizing the production of the barnī date palm.

**Notes S3** On the relationship between coverage and the King-robust kinship estimator.

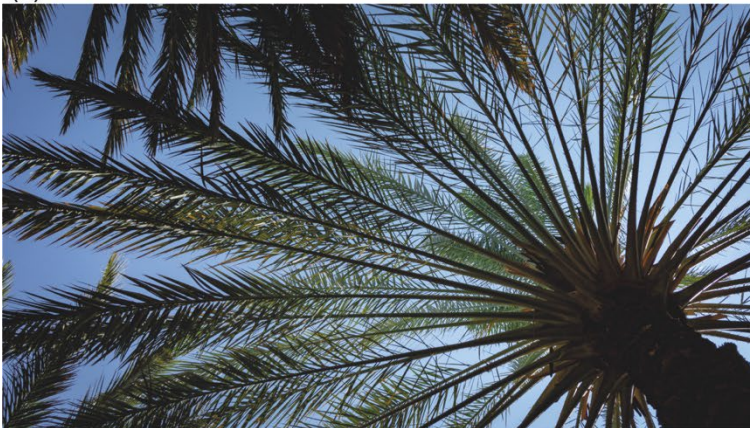
## References present in supporting information

Gros-Balthazard\*, Muriel, Vincent Battesti\*, Jonathan M. Flowers, Sylvie Ferrand, Matthieu Breil, Sarah Ivorra, Jean-Frédéric Terral, Michael D. Purugganan, Rod A. Wing, Nahed Mohammed & Yann Bourgeois, 2022 – « What lies behind a fruit crop variety name? A case study of the barnī date palm from al-'Ulā oasis, Saudi Arabia ». *Plants, People, Planet* (Fruits in Focus), p. 1-16, doi:10.1002/ppp3.10326 – online: <https://hal.archives-ouvertes.fr/hal-03806722>

(a)



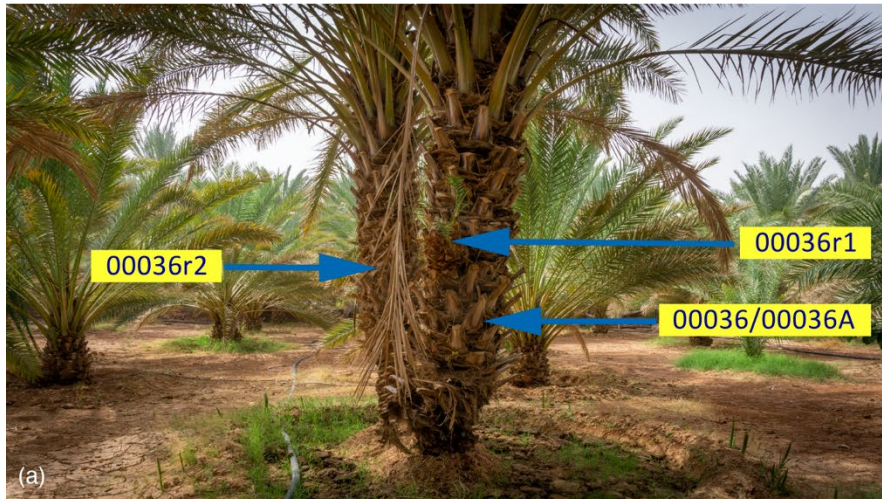
(b)



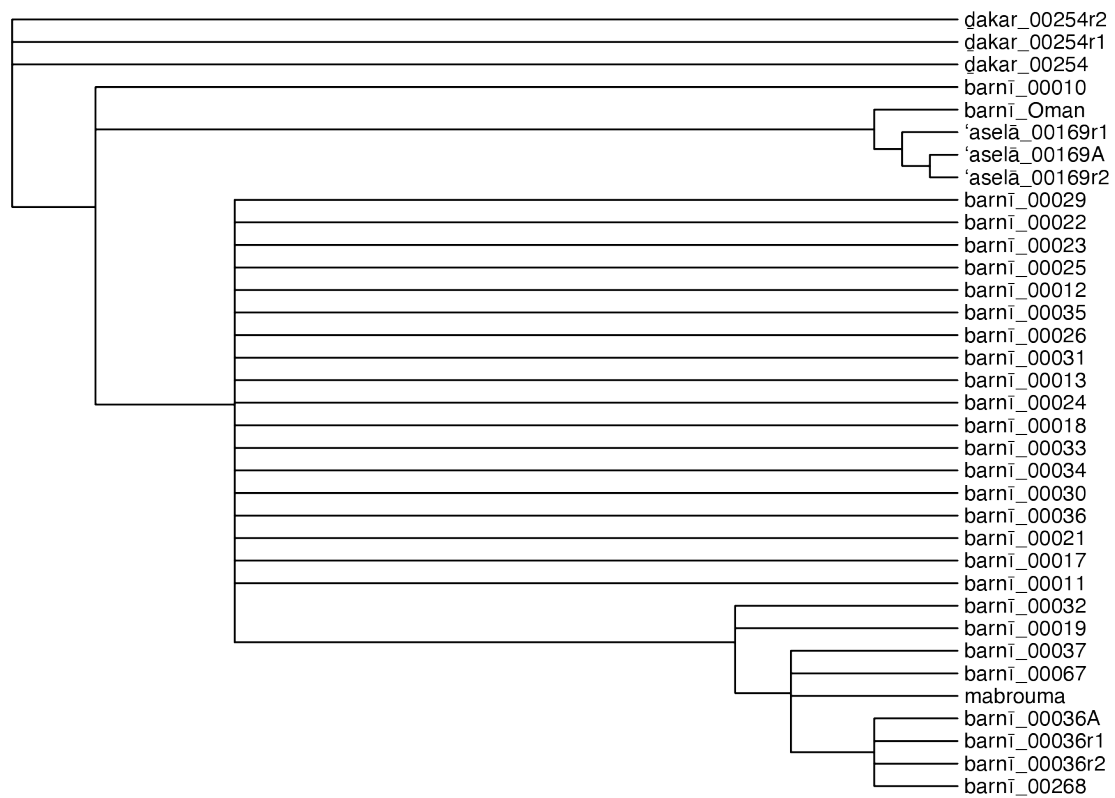
(c)



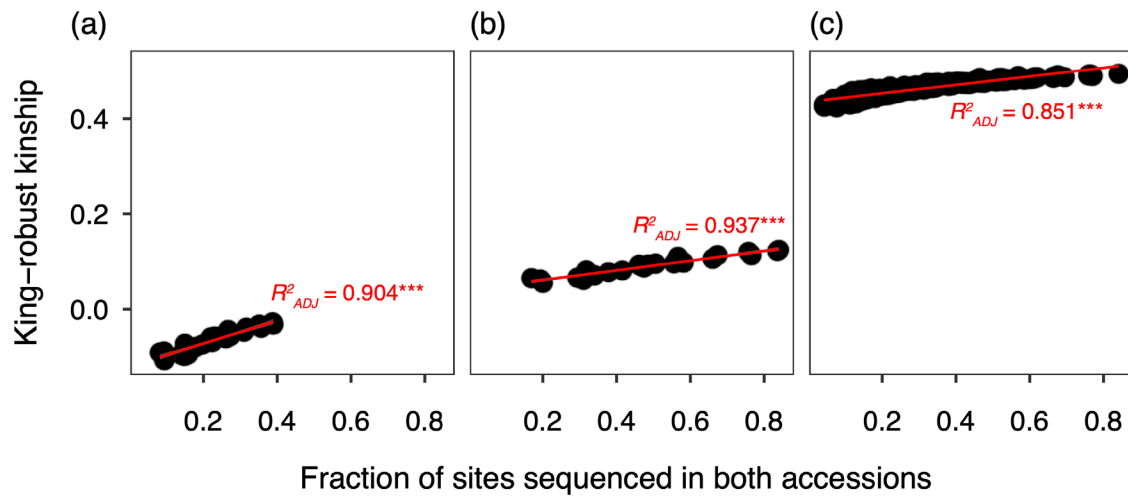
**Fig. S1** Pictures of barnī date palms. (a) A young barnī date palm, with two growing offshoots still attached to its base (other offshoots have had their leaves cut off). To their left are two offshoots that are in their first year of planting (wrapped with dry palms to protect these young plants). Photo taken at a Qarāqer farm, north of al-‘Ulā. Nov. 11th 2019; (b) The date palm sampled as barnī\_00024, photographed in low angle, showing the arrangement of its leaves around its terminal bud and its leaflets on its palms. In the palm grove of al-Oziyāt, north of al-‘Ulā. Nov. 10th 2019; (c) A barnī date palm in a new 21st c. palm grove, northeast of al-‘Ulā, displaying young inflorescences (future date bunches) some of which are still attached by a link: the farmer has slipped into the inflorescences male spikelets still visible (top left) for pollination. April 6th, 2019. Pictures: Vincent Battesti.



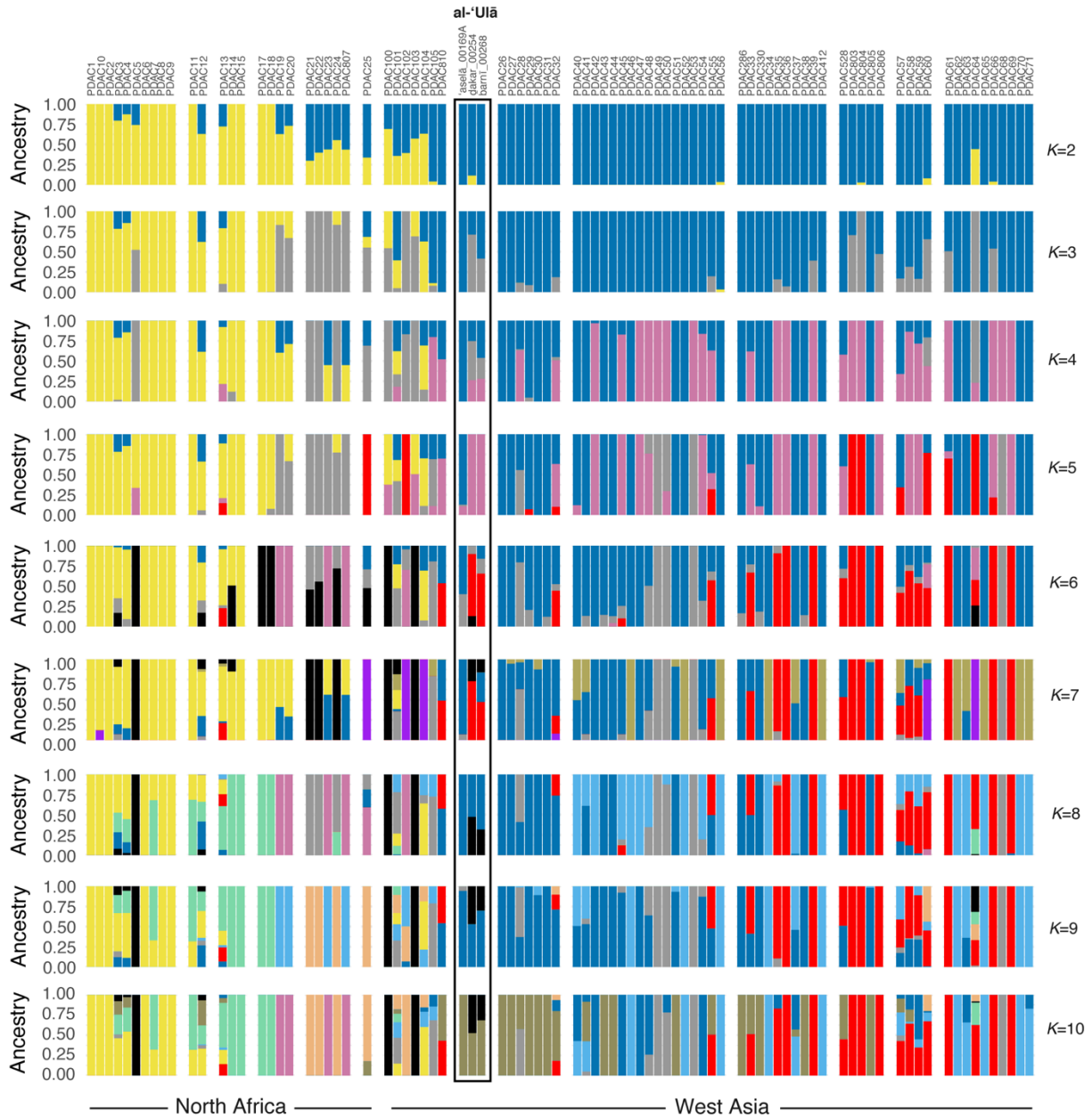
**Fig. S2** Pictures of the date palms groups of clones used to calibrate the relatedness analyses. (a) The mother plant barnī 00036, sampled twice (barnī\_00036 and barnī\_00036A), and its offshoots (00036r1 and 00036r2), (b) The male ḍakar (ḍakar\_00254) and its two sampled offshoots (00254r1 and 00254r2), (c) The offshoots of ‘aselā, of which two were sampled (00169r1 and r2) in front of the mother palm (‘aselā\_00169A). Pictures: Vincent Battesti.



**Fig. S3** Consensus tree of 35 date palms based on genetic distance calculated using genotype likelihoods across 10,742 sites. It was obtained by computing a consensus tree where only clades found in 90% of the 100 bootstrap replicates are displayed. The non-consensus tree where genetic distances are visible can be found in Fig. 4c.

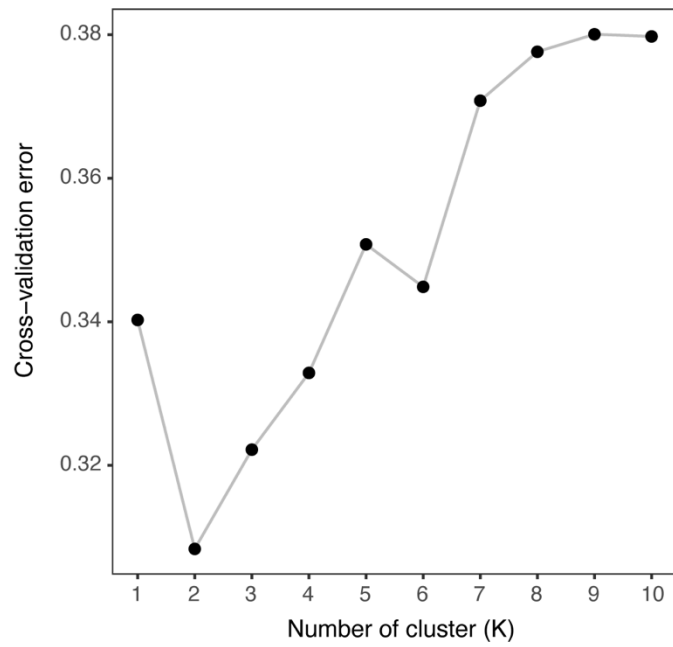


**Fig. S4** Relationship between the fraction of sites covered by one or more reads in both sequenced genomes and the King-robust kinship estimator. (a) Kinship between the sample barni\_00010 and the other 22 barni; b) Kinship between barni\_Oman and the other 22 barni; c) Kinship among the 22 barni accessions. Correlations (adjusted  $R^2$ ) were calculated using a linear mixed model and asterisks indicate p value < 0.01.

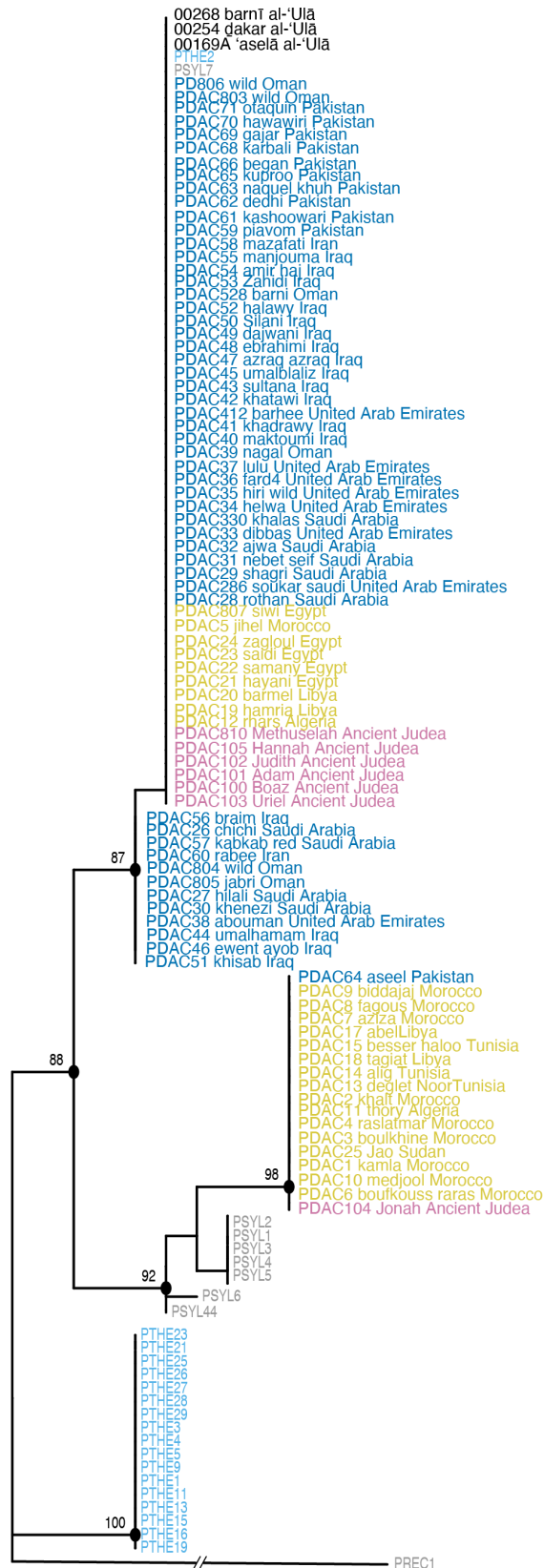


**Fig. S5** Ancestry coefficients inferred in 88 date palms using a randomly downsampled set of 28,406 Single Nucleotide Polymorphisms (SNPs). Cross-validation error plot may be found in Fig. S6.





**Fig. S6** Cross-validation error of the admixture model. The corresponding ancestry plot may be found in Fig. 5a.



**Fig. S7** Chloroplast DNA tree with uncorrected distances reconstructed using the Neighbor-joining method. Support values represent the percent of replicates supporting each node from 100 bootstrap replicates. The tree was rooted with *Phoenix reclinata* (PREC1). All three date palms from al-'Ulā (in black) are found within a clade comprising the West Asian date palms (in blue) and bear the so-called oriental chlorotype. North African date palms are written in yellow, ancient Judean date palms in pink, *Phoenix theophrasti* in light blue, and *Phoenix sylvestris* in grey.

**Methods S1** Detailed protocol for sequencing read processing, genome alignment, variant calling and Single Nucleotide Polymorphisms (SNPs) filtering.

**Read processing and genome alignment.** Reads were demultiplexed and those passing Illumina quality control filters were processed with Trimmomatic (Bolger et al., 2014) v. 0.39 to remove contaminating adapter sequences. For adapter removal, we used the adapter and Nextera transposase sequence database included with Trimmomatic (v. 0.39) downloaded with the following setting ILLUMINACLIP:<adapter library>:2:30:10 and only reads pairs where both reads in a pair were 76 bp or longer following trimming were retained for subsequent steps.

We aligned reads to the reference genome from Hazzouri *et al.*, 2019. This genome is a PacBio/Illumina genome assembly from a male date palm from a fourth-generation backcross (BC4) with a female of the Barhee cultivar. Its size is ~772 Mb and it contains 18 primary contigs (49.9% of total length) along with 2,371 unplaced contigs. The nuclear and chloroplast genomes were combined to form a single modified reference sequence that was used for variant calling. Site filtering was nonetheless carried out separately.

Processed reads were aligned to this unmasked genome using bwa mem (Li, 2013) v. 0.7.15-r1140. The bwa mem aligner was run with the -M option to mark supplementary reads (0x800 bitwise flag) as secondary (0x100). Sample alignments were processed with SamSort from Picard-tools (<https://github.com/broadinstitute/picard>) v. 2.8.2 to coordinate-sort the alignments. We used MarkDuplicates (Picard-tools) to flag duplicate read pairs. Processed alignments were summarized with samtools v. 1.9 (Li et al., 2009) using the stats option.

**Variant calling and Single Nucleotide Polymorphisms (SNPs) filtering.** SNP-calling and genotyping was performed with the Genome Analysis Toolkit (GATK) v. 4.2.0 HaplotypeCaller (McKenna et al., 2010) run in Genomic Variant Call Format (GVCF) mode followed by joint-genotyping with GenotypeGVCFs. We used the option *-all-sites* to include not only variant sites but also loci found to be non-variant after genotyping.

The sites were filtered using GATK v4.2.0, Picard-tools v.2.23.8, bcftools v.1.14 (Li, 2011) and vcftools v.0.1.16 (Danecek et al., 2011). For the nuclear genomes, for both variant and invariant sites, we restricted analysis to the non-repetitive fraction of the genome assembly by excluding

SNPs in repetitive regions identified during the annotation of the genome assembly and to the regions that are not associated to sex determination (Gros-Balthazard et al., 2021; Hazzouri et al., 2019). We removed indels, excluded multi-allelic SNPs and sites having an average depth below 10. We set as missing genotypes showing a depth below 10 or above 50. We also excluded SNPs meeting the following conditions: within 6bp of indel polymorphisms, strand bias estimated using Fisher's exact test (FS) > 60.0, strand bias estimated by the symmetric odds ratio test (SOR) > 3.0, quality by depth (QD) < 8.0, mapping quality (MQ) < 40.0, mapping quality rank sum (MQRankSum) < -3.0, read position rank sum (ReadPosRankSum) < -1.5, base quality rank sum (BaseQRankSum) < -8.0, quality (QUAL) < 30. We filtered out genotypes having genotype quality (GQ) < 20.0 and SNPs with a genotype call rate < 80%. This procedure yielded a filtered site set of 105,236,083 bases including 1,007,281 SNPs that served as the basis for subsequent analyses.

### **Reconstruction of a chloroplast DNA tree**

We called Single Nucleotide Polymorphisms (SNPs) and genotypes for the chloroplast genome (cpDNA) (Hazzouri et al. 2019) using the same workflow as the nuclear genome, but applied a unique set of filters to obtain the final call set. The large inverted repeat (IR) regions characteristic of the date palm cpDNA (Yang et al. 2010) were identified by BLAST and excluded from further consideration. A separate BLAST analysis was conducted to identify regions of the cpDNA that share a high degree of similarity with the mitochondrial genome (mtDNA) (Fang et al. 2012). Any region of the cpDNA that BLAST to the mtDNA (Fang et al. 2012) with significance of  $1e-6$  or less was excluded from the analysis. Hard filters were then applied such that SNPs meeting any of following conditions were removed quality by depth (QD) < 2.0, strand bias estimated using Fisher's exact test (FS) > 60.0, mapping quality (MQ) < 40.0, mapping quality rank sum (MQRankSum) < 12.5, ReadPosRankSum < -8.0 (see GATK website for tag definitions). SNPs within 10 bp of an indel in the raw (unfiltered) indel call set were also removed. Finally, we excluded any site with missing data or with one or more heterozygote genotype calls as they may represent instances of heteroplasmy or genotyping artifacts. Remaining homozygous genotypes were converted to haploid to produce a sequence alignment with 52 SNPs. This alignment was used to construct a cpDNA tree with uncorrected distances using the Neighbor-joining method with the phangorn (v. 2.8.1) package in R.

**Methods S2** We tested for the presence and extent of admixture between date palms and *Phoenix theophrasti* using the R package *admixr* (Petr et al., 2019) v.0.9.1, which provides an implementation of ADMIXTOOLS (Patterson et al., 2012). We used a script available online to generate the input in EIGENSTRAT format from our VCF (<https://github.com/joanam/scripts/blob/master/convertVCFtoEigenstrat.sh>). We ran the tool using the full Single Nucleotide Polymorphisms (SNPs) set.

We focused on determining whether al-'Ulā date palms show evidence of gene flow with *P. theophrasti* as evidenced in North African cultivars, a few West Asian date palms, and most of the ancient Judean date palms (Flowers et al., 2019; Gros-Balthazard et al., 2021). We use the same pattern of test, fitting the phylogenetic relationships among *Phoenix*, as in Gros-Balthazard et al., 2021.

We calculated Patterson's *D* also known as the ABBA-BABA test (Green et al., 2010; Patterson et al., 2012). The implementation of the test in *admixr* requires a rooted and asymmetric four population tree in the form (((W, X), Y), Z), where Z is an outgroup to a clade formed by W, X and Y, and Y is an outgroup to a clade formed by W and X. The test is used to evaluate if the data is inconsistent with the null hypothesis that the tree is correct and that there is no gene flow between Y and either W or X. It is based on comparing the proportions of BABA and ABBA sites patterns observed in the data. *D* is calculated as follow:  $D = (n_{BABA} - n_{ABBA}) / (n_{BABA} + n_{ABBA})$ . We also calculated the  $f_4$ -statistic, which is very similar to *D* except that the denominator is the total number of sites (Patterson et al., 2012). Gene flow between populations W and Y leads to an increase of shared alleles between populations resulting in an elevated number of BABA sites and thus a positive *D*-statistic and a positive  $f_4$ -statistic. Here, we tested for gene flow between North African, ancient Judean and al-'Ulā date palms individually (test sample) and *P. theophrasti*. We conducted the tests separately including *P. reclinata* as the outgroup, following a tree reconstruction (Gros-Balthazard et al., 2021). We therefore estimated the *D*-statistics using the following tree: (((test sample, West Asian date palms), *P. theophrasti*), *P. reclinata*). We used all samples together for each population/species and after removing admixed samples, as identified in Gros-Balthazard et al. 2021 (PDAC64, PTHE2, PTHE3, PTHE4, PTHE26, PTHE27, PSYL7 and PSYL44). Significance was tested with block

jackknife with each linkage groups and unplaced contigs treated as a separate block. The standard error of the statistics was used to calculate a standard score (Z-score = statistics / standard error), and we used an absolute Z-score >2 to assess statistically significant results. Significance was tested with block jackknife with each linkage groups and unplaced contigs treated as a separate block. The standard error of the statistics was used to calculate a standard score (Z-score = statistics / standard error), and we used an absolute Z-score >2 to assess statistically significant results.

To infer the proportion of ancestry in the date palms displaying evidence of admixture based on the  $D$  and  $f_4$ -statistic, we calculated the  $f_4$ -ratio statistics, as described in Patterson *et al.* 2012. More precisely, we defined the following tree (((A,B), C), O). The  $f_4$ -ratio is  $f_4(A,O;X,C)/f_4(A,O;B,C)$ , where X is the test sample, A is a sister species, namely *P. sylvestris*, B and C are the mixing populations, namely West Asian date palms and *P. theophrasti*, respectively, and O is the outgroup, that is *P. reclinata*. We used all samples together for each population/species and after removing admixed samples (see above; PDAC64, PTHE2, PTHE3, PTHE4, PTHE26, PTHE27, PSYL7 and PSYL44). The resulting alpha value corresponds to the ancestry proportion of B in X, and by extension 1-alpha provides the proportion of C in X, therefore the proportion of *P. theophrasti* in the test sample X. Negative values are uninformative. Significance was tested with block jackknife with each linkage groups and unplaced contigs treated as a separate block. The standard error of the statistics was used to calculate a standard score (Z-score = statistics / standard error), and we used an absolute Z-score >2 to assess statistically significant results.

**Notes S1** Date palms growing situations in al-'Ulā region.

We consider four distinct date palms growing situations in al-'Ulā region (Battesti & Marty, *in prep.*), of which a quick summary is given below:

- The old oasis palm grove of al-'Ulā (the former “core area” according to the terminology of the Royal Commission for AlUla (RCU) administration, and now the palm grove of the so-called “Cultural Oasis District” according to the new terminology of the RCU administration): it is the historical center of oasian production of dates and agrobiodiversity of date palm (dozens of varieties) in the region. Date palms were cultivated by sedentary (*ḥaḍāraī*) social groups (mainly al-'Alawī) in palm grove small enclosed gardens (*basātīn*). They do still exist and produce, but a part of them is in bad shape and the other part reorganized. The classic form of farming that is carried out there is layered polyculture with an elaborate soil design work.
- The old Bedouin palm groves: it is the historical way for Bedouin social groups, with a pastoral way of life, to produce a (small) part of their food needs in dates (other historical means are the submission of oasis people, purchase/exchange and predation). These small palm groves are located in the pastoral tribal lands outside of the area of the sedentary oases, scattered in desert wadis. They have a range of labor investment levels, from nearly none in the picking palm groves (few date palms cultivated with neither fruit trees nor annual crops growing beneath) to small gardens around a spring with some fruit trees. The date palm agrobiodiversity is, moreover, restricted to the two local elite varieties, barnī and ḥelwa (*ḥamrā'*). These Bedouin palm groves do still exist.
- The modern 20<sup>th</sup> century palm groves: these palm groves were the first to be established outside the historical area of the old al-'Ulā palm grove, but in its immediate (north and south) vicinity. These palm groves were mostly created by sedentary oasis people, or by Bedouin groups already settled in al-'Ulā. The gardens or farms (*mazāra'*) no longer have organic forms, but are linear properties, with date palms lined up planted at the same time, thus having the same age and height. Besides a few fruit trees, date monoculture is the rule, and these farms depend heavily on the only variety for export, the barnī. The purpose is clearly no longer local food but export income.

- The new 21<sup>st</sup> century palm groves: these palm groves are the most recent and have been established well beyond the first perimeter of old and modern palm groves. The spirit that governs the exploitation of these farms (*mazāra'*) is generally that of agricultural entrepreneurship, particularly Bedouin, which does not negate forms of small varietal collections of date palms (especially exotic, sometimes also local especially when the owner is 'Alawī) and a crop diversification into citrus fruit or moringa, for instance.



**Notes S2** Categorizing the production of the barnī date palm.

In al-‘Ulā, the barnī variety is the only one that really benefits from such precision of quality categories (*mabrūm* for the best dates, *mašrūk* for the second grade, and *‘ādī* for the remains), which some still refine, with a #1bis quality, the “(*mabrūm*) *aşfar*” (which look like *mabrūm*, but whose seeds are small and poorly formed), or more recently the *mabrūm jumbo* (for their superior dimensions), term coined by wholesalers and commercial largescale farmers. The other date varieties, which we currently estimate at  $n \approx 99$  in al-‘Ulā (work in progress), are locally valued for their differences, but their respective productions are just sorted into good and bad dates with no specific names. Some local farmers logically argue that this refinement in the categorization of barnī qualities is simply a result of the market.

The data regarding the relative proportions of the different qualities of dates on a barnī palm fluctuates widely, depending on the type of palm grove and farm. This example can be considered correctly in the average: a producer in al-Oziyāt (northern part of al-‘Ulā) gave the year 2018 as bad with 40% *mabrūm* and 60% *mašrūk* and *‘ādī*, while he deemed the year 2019 as good with an inverse ratio of 60% and 40%.

The commitment to a high production of *mabrūm* is easily explained from a cash crop perspective: the market values are very different. For example (in 2019), *mabrūm* was priced at 24 riyals/kg, *mašrūk* at 10 riyals/kg, and *‘ādī* was left for 2.5 riyals/kg. For years 2020 and 2021, prices were greatly impacted downward with the Covid-19 pandemic: less export, because fewer foreign wholesale buyers came to buy (visa and vaccination certificate problem) and absence of the millions of pilgrims (the pilgrimage was suspended by the kingdom) who usually buy in the holy city of Medina, before their return home, large quantities of “the date of al-Madīna al-munawara”, which is for some part the barnī from al-‘Ulā). As a matter of fact, if our focus area is expanded to the entire province of Medina (Saudi Arabia), there seem to be three varieties of barnī (Aleid et al., 2015): from Medina, from al-‘Aīş—as we mentioned, putatively barniyat al-‘aīş cultivated also in al-‘Ulā—, and from al-‘Ulā. Whether they are homonyms or genetically the same variety cannot be certified at this point and will be assessed in the second phase of our Al-‘Ulā DPA (Date Palm Agrobiodiversity) Project (al-‘Ulā DPA).

**Notes S3** On the relationship between coverage and the King-robust kinship estimator.

We investigated why the King-robust kinship estimator is lower within the barnī accessions we believe to be clones compared to that found within each quartet/triplets of known clones. The latter being sequenced at a higher coverage, we hypothesized that it could affect the estimator, although we used genotype likelihoods rather than calls and a relatedness estimator presumably robust to Single Nucleotide Polymorphisms (SNPs) ascertainment biases linked to low coverage (Waples et al., 2019). Based on the estimated kinships (Fig. 4ab; Dataset S3), we found three types of relationships among the barnī samples: 1) barnī\_00010 is unrelated to other barnī, 2) barnī\_Oman is unrelated to other barnī, and 3) all the other barnī are clones. We therefore visualized the relationship between the pairwise coverage (fraction of sites covered by 1 or more reads in both sequenced genomes used for the calculation) and the resulting King-robust kinship estimator. We calculated the correlations using a linear mixed model (*lm* function in *stats R* package) and indeed found that pairwise coverage affects the estimator in the three cases (adjusted  $R^2$  ranged from 0.851 to 0.937, p-values < 0.01).

## References present in supporting information

- Aleid, S. M., Al-Khayri, J. M., & Al-Bahrany, A. M. (2015). Date Palm Status and Perspective in Saudi Arabia. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Éds.), *Date Palm Genetic Resources and Utilization* (p. 49-95). Springer Netherlands. [https://doi.org/10.1007/978-94-017-9707-8\\_3](https://doi.org/10.1007/978-94-017-9707-8_3)
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic : A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., Ferrand, S., Khierallah, H. S. M., Fuller, D. Q., Aberlenc, F., Fournaraki, C., & Purugganan, M. D. (2019). Cross-species hybridization and the origin of North African date palms. *Proceedings of the National Academy of Sciences*, *116*(5), 1651-1658. <https://doi.org/10.1073/pnas.1817453116>
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspina, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., ... Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710-722. <https://doi.org/10.1126/science.1188021>
- Gros-Balthazard, M., Flowers, J. M., Hazzouri, K. M., Ferrand, S., Aberlenc, F., Sallon, S., & Purugganan, M. D. (2021). The genomes of ancient date palms germinated from 2,000 y old seeds. *Proceedings of the National Academy of Sciences*, *118*(19), e2025337118. <https://doi.org/10.1073/pnas.2025337118>
- 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., Rosas, U., Zhang, J., Talag, J., Lee, S., Kudrna, D., Powell, R. F., Leitch, I. J., Krueger, R. R., Wing, R. A., ... Purugganan, M. D. (2019). Genome-wide association mapping of date palm fruit traits. *Nature Communications*, *10*(1), 4680. <https://doi.org/10.1038/s41467-019-12604-9>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. <https://arxiv.org/abs/1303.3997v2>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence

- Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.  
<https://doi.org/10.1093/bioinformatics/btp352>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. <https://doi.org/10.1101/gr.107524.110>
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, 192(3), 1065-1093. <https://doi.org/10.1534/genetics.112.145037>
- Petr, M., Vernot, B., & Kelso, J. (2019). admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*, 35(17), 3194-3195.  
<https://doi.org/10.1093/bioinformatics/btz030>
- Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, 28(1), 35-48. <https://doi.org/10.1111/mec.14954>