nature genetics

Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley

Martin Mascher[1,2,19], Verena J Schuenemann[3,4,19], Uri Davidovich[5], Nimrod Marom[6], Axel Himmelbach[1], Sariel Hübner[7,8], Abraham Korol[9,10], Michal David[11], Ella Reiter[3], Simone Riehl[3,4], Mona Schreiber[1], Samuel H Vohr[12], Richard E Green[12], Ian K Dawson[13], Joanne Russell[13], Benjamin Kilian[1,18], Gary J Muehlbauer[14,15], Robbie Waugh[13,16], Tzion Fahima[9,10], Johannes Krause[3,4,17], Ehud Weiss[11] & Nils Stein[1]

The cereal grass barley was domesticated about 10,000 years before the present in the Fertile Crescent and became a founder crop of Neolithic agriculture[1]. Here we report the genome sequences of five 6,000-year-old barley grains excavated at a cave in the Judean Desert close to the Dead Sea. Comparison to whole-exome sequence data from a diversity panel of present-day barley accessions showed the close affinity of ancient samples to extant landraces from the Southern Levant and Egypt, consistent with a proposed origin of domesticated barley in the Upper Jordan Valley. Our findings suggest that barley landraces grown in present-day Israel have not experienced major lineage turnover over the past six millennia, although there is evidence for gene flow between cultivated and sympatric wild populations. We demonstrate the usefulness of ancient genomes from desiccated archaeobotanical remains in informing research into the origin, early domestication and subsequent migration of crop species.

Genetic analyses of ancient DNA can greatly inform research into the origin, initial domestication and subsequent dispersal of crops and livestock, as evidenced by studies involving ancient DNA samples and genomic data sets of present-day populations of cattle[2], swine[3], dogs[4] and maize[5]. Wheat and barley, founder crops of agriculture in the ancient Near East and Europe, were domesticated in the Fertile Crescent, where their wild relatives still thrive today[1,6]. Current knowledge of their domestication is largely derived from morphological analysis of archaeobotanical remains[1] and population genetic analysis of present-day samples[7,8]. Although domesticated wheat and barley appear in the archaeological record by 10,000 calendar years before the present (cal BP)[1], the oldest verified DNA sequences thus far were retrieved from archaeobotanical specimens originating from Bronze Age China[9] and Ancient Egypt[10]. Claims about a small number of prehistoric wheat DNA molecules retrieved from Mesolithic paleosol[11] have remained contentious[12,13]. To our knowledge, there have been no studies where large quantities of ancient DNA sequences have been retrieved that could underpin the comparison of modern and ancient samples of Old World cereals at a genome-wide scale.

Here we report the genome sequences of five 6,000-year-old barley grains excavated at Yoram Cave in the Judean Desert in Israel. Yoram Cave is part of a complex of three difficult-to-access caves, located in the southeastern cliff of the Masada Horst facing the Dead Sea. High-resolution excavation (**Fig. 1a**, Online Methods and **Supplementary Figs. 1** and **2**) found a single undisturbed anthropogenic layer of Chalcolithic origin (ca. 6,200–5,800 cal BP). The rich plant assemblage of more than 100 taxa was well preserved (**Fig. 1b**), prompting us to attempt the retrieval of DNA sequences. We selected ancient barley grains (**Fig. 1c** and **Supplementary Fig. 3**) for DNA extraction because of barley's central role in ancient and modern agriculture and its remarkable adaptive features that make it a model plant in domestication genomics.

DNA extractions were performed from ten bisected grains and spikelet remains, with the other halves subjected to direct radiocarbon dating, confirming the Chalcolithic origin of the specimens (**Table 1**). Illumina sequencing of libraries yielded between 7.3 and 21.5 million paired-end reads (**Supplementary Table 1**). On the basis of the fraction of reads that could be aligned to the barley reference genome,

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany. [2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [3]Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany. [4]Senckenberg Center for Human Evolution and Paleoenvironment, University of Tübingen, Tübingen, Germany. [5]Institute of Archaeology, Hebrew University, Jerusalem, Israel. [6]Laboratory of Archaeozoology, Zinman Institute of Archaeology, University of Haifa, Haifa, Israel. [7]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada. [8]Department of Biotechnology, Tel Hai College, Upper Galilee, Israel. [9]Institute of Evolution, University of Haifa, Haifa, Israel. [10]Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel. [11]Martin (Szusz) Department of Land of Israel Studies and Archaeology, Bar-Ilan University, Ramat-Gan, Israel. [12]Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, California, USA. [13]Cell and Molecular Sciences, James Hutton Institute, Invergowrie, Dundee, UK. [14]Department of Plant Biology, University of Minnesota, St. Paul, Minnesota, USA. [15]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota, USA. [16]Division of Plant Sciences, University of Dundee, Dundee, UK. [17]Max Planck Institute for the Science of Human History, Jena, Germany. [18]Present address: Bayer CropScience, BCS Breeding and Trait Development, Zwijnaarde (Gent), Belgium. [19]These authors contributed equally to this work. Correspondence should be addressed to T.F. (fahima@research.haifa.ac.il), J.K. (krause@shh.mpg.de), E.W. (ehud.weiss@biu.ac.il) or N.S. (stein@ipk-gatersleben.de).

Received 6 May; accepted 13 June; published online 18 July 2016; doi:10.1038/ng.3611

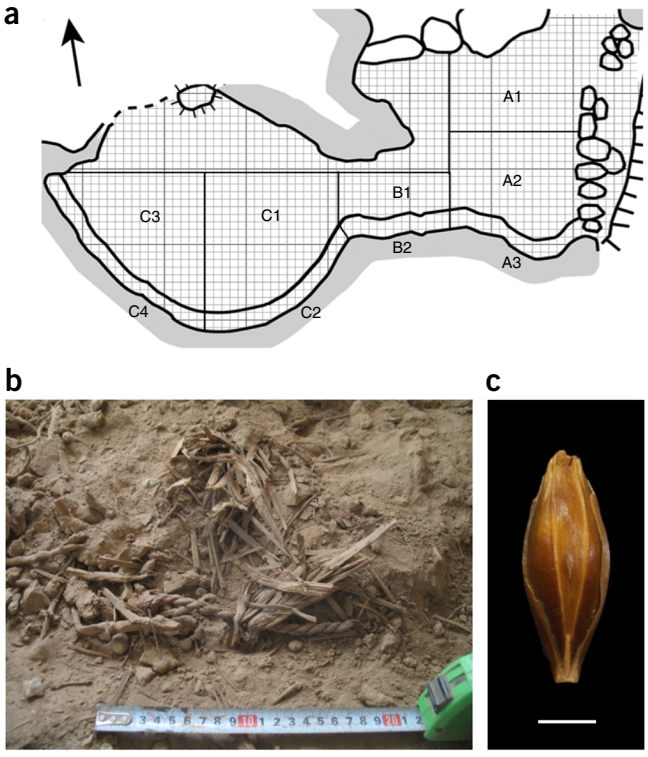

Figure 1 Ancient plant remains excavated at Yoram Cave. (**a**) Plan of the Southern Chamber of Yoram Cave showing the excavation grid and subunits. The arrow points north. (**b**) Photograph of locus 3 in square A2 during excavation. Note the excellent dry preservation of rope, reeds, seeds and pellets. (**c**) Photograph of a well-preserved desiccated barley grain found at Yoram Cave. Scale bar, 2 mm.

we estimated the content of endogenous DNA to range from 0.4 to 96.4%. Sequence reads for eight samples showed fragment sizes and damage patterns characteristic of ancient DNA (**Supplementary Figs. 4–6** and **Supplementary Table 1**), demonstrating the authenticity of the samples[14,15]. Deamination-derived mismatches (C•T and G•A) occurred toward the ends of reads, with frequencies between 1.9 and 21.8%. We only used five samples with on average more than 12% misincorporation[15] at the first base of sequence fragments for further experiments. Once the authenticity of these samples had been established, we treated five DNA extracts containing a large fraction of endogenous DNA with uracil-DNA glycosylase (UDG) to reduce nucleotide misincorporation caused by ancient DNA damage by removing deaminated cytosines[16]. Deep Illumina sequencing of the UDG-treated libraries yielded between 82.5 million and 5.1 billion reads (**Table 1**).

We compared the ancient barley genome sequences to present-day accessions on the basis of whole-exome capture[17] sequence data for 267 entries from *ex situ* collections representing extant populations of wild (*Hordeum vulgare* ssp. *spontaneum*) and domesticated (*Hordeum vulgare* ssp. *vulgare*) barley from across the range of the species[18] (**Fig. 2a**). This data set[18] comprised 1,688,807 SNPs (**Table 1**).

Principal-component analysis (PCA)[19] has shown fundamental patterns of population structure across present-day accessions[18]. The first principal component (PC1) clearly differentiated wild and domesticated barley accessions, and PC2 represented the variation in wild barley accessions (**Fig. 2b**). Least-square projection[20] of the ancient samples onto the PCA axes defined by the extant samples demonstrated the close affinity of ancient barley with present-day domesticated barley. The deep coverage of sample JK3014 allowed us to ascertain the allelic status of the domestication-related genes *NON-BRITTLE RACHIS 1* (*BTR1*) and *NON-BRITTLE RACHIS 2* (*BTR2*). In domesticated barley, one of *BTR1* and *BTR2* carry mutations that abolish the disarticulation of the spike at maturity[21]. JK3014 had a wild-type *BTR2* haplotype but carried the previously described 1-bp deletion in the coding sequence of *BTR1* (**Supplementary Fig. 7**), consistent with the high frequency of this mutation in barleys of the Southern Levant[21]. In agreement with the archaeobotanical classification of the ancient barley spike remains as being of the two-rowed type (**Fig. 2c**), the extant accessions closest to the ancient samples were two-rowed domesticated barleys from the Southern Levant and Egypt (**Fig. 2b**,**d**). A putative two-rowed phenotype for sample JK3014 can also be inferred from the allelic status of the *SIX-ROWED SPIKE 1* (*VRS1*) gene[22] (**Supplementary Fig. 7**).

Table 1 Summary of ancient barley samples used for genetic analyses

Sample name	Radiocarbon age[a]	Number of raw reads (in millions)	Percentage of mapped reads	Percentage of unique reads	Average read depth[b]	Number of called SNPs
JK2281	5,290 ± 27	256.1	31.4	26.0	0.54	162,110
JK3009	5,034 ± 36	89.2	61.7	67.3	0.46	133,365
JK3010	5,032 ± 37	94.7	62.3	57.9	0.96	278,505
JK3013	5,227 ± 37	82.5	49.4	39.8	0.19	18,949
JK3014	4,988 ± 36	5,131.2	86.4	28.3	20	1,283,396

[a]Uncalibrated radiocarbon years before the present (see **Supplementary Table 1** for calibrated dates). [b]Average read depth in targeted regions of the exome capture assay for JK2281–JK3013. The mode of the coverage distribution is given for JK3014.

Rare genetic variants can provide insights into the spatial structure of populations[23,24]. In inbreeding plants such as barley, isolation by distance[25] is common because gene flow across larger geographical distances is limited. We identified rare variants with minor allele counts of up to five across the ancient and extant barleys and determined the number of rare alleles shared by pairs of sequenced samples. Transitions were excluded from the analysis because deaminated cytosines cannot be repaired by UDG treatment if they are methylated[16] and thus can give rise to genotyping errors. The extant landraces that shared at least 30 rare alleles with the ancient samples were two-rowed accessions from Syria, Jordan, Egypt and Israel as well as six-rowed accessions from North Africa (**Supplementary Table 2**). Eight wild barley accessions from Israel also shared ≥30 rare alleles with the ancient samples (**Supplementary Table 2**).

We measured the relatedness of ancient samples to each of the wild barley accessions on the basis of the level of identity by state (IBS) calculated across all SNPs. The genetically closest wild accessions originated from a sampling site located in the Upper Jordan Valley. We then calculated the geographical distance between Yoram Cave and the sampling site for each wild accession, but no significant correlation between IBS and geographical distance was found ($R = -0.17$, $P = 0.108$). However, when splitting the data into geographically proximal (<250 km apart) and distant (>250 km apart) samples, significant correlations between geographical distance and the relatedness score were detected for both subsets of the tested wild barley samples (proximal: $R = 0.74$, $P \leq 0.001$; distant: $R = -0.34$, $P = 0.006$) (**Fig. 3a**). The same analysis was conducted comparing wild barleys and extant landraces from the Fertile Crescent (**Fig. 3b**,**c**), pinpointing the Upper Jordan Valley as a peak for genetic similarity with domesticated barley. Conducting this analysis using only data from

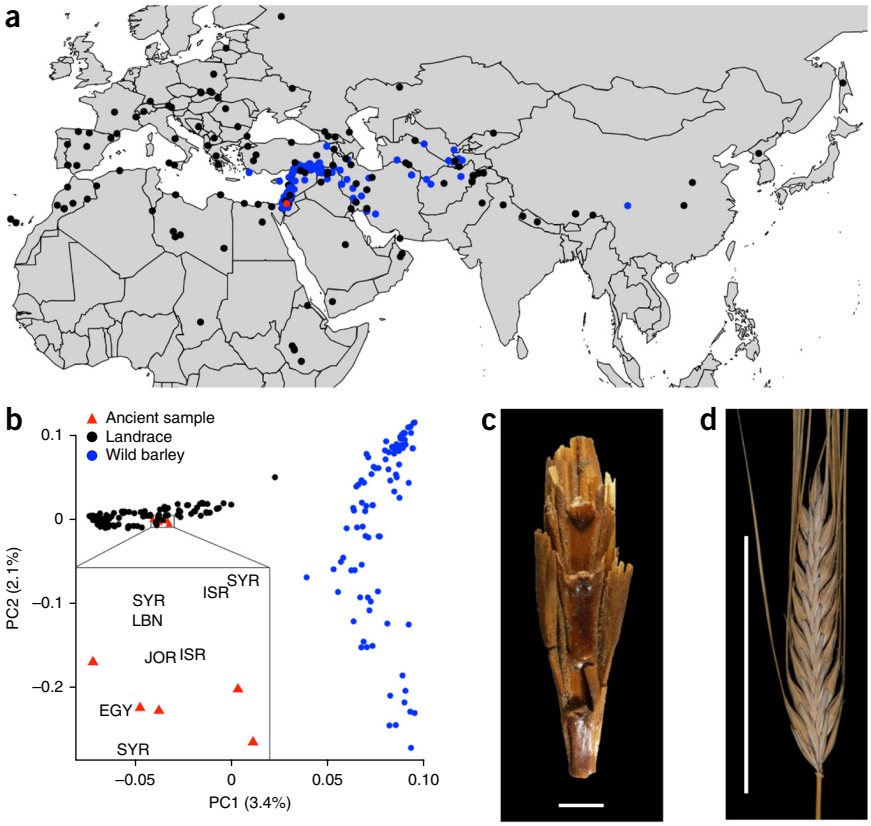

Figure 2 Ancient barley samples are closely related to present-day landraces from the Levant. Ancient barley sequences were compared to exome sequence data for a present-day diversity panel. (**a**) The collection sites of landraces (black circles) and wild barleys (blue circles) are shown. The location of Masada is marked with a red circle. (**b**) PCA showing ancient samples projected onto the PCA axes for the present-day diversity panel. The inset magnifies the PCA space around the ancient samples. ISR, JOR, SYR, LBN and EGY represent closely related landraces from Israel, Jordan, Syria, Lebanon and Egypt, respectively. The proportion of variance explained by each principal component is indicated in parentheses. (**c**) Well-preserved rachis of two-rowed domesticated barley from Yoram Cave. Scale bar, 2 mm. (**d**) Spike from an individual of a present-day two-rowed landrace barley (accession HOR8658) that is among the barleys most closely related to the ancient DNA sample from Yoram Cave (photo by B. Schäfer) Scale bar, 5 cm.

the deeply sequenced JK3014 sample yielded similar results (**Supplementary Fig. 8**). Comparison of modern cultivars and landraces from outside the Fertile Crescent (Europe and northern Asia) to our wild barley panel accessions pinpointed accessions from the Upper Jordan Valley as the most closely related (**Supplementary Table 3**). The Israel–Jordan area was proposed earlier as one (although not the only) center of origin for domesticated barley[8,26]. This hypothesis is supported by findings from two archaeological sites—Tel Aswad and Ohalo II—with the earliest traces of barley cultivation[27,28], which are within 80 km of the extant wild barley accessions in our panel that are genetically closest to the ancient samples.

Although self-fertilization is predominant in barley[29], wild barley is fully interfertile with the domesticated crop, and evidence for hybridization between the two has been reported[7,30]. To ascertain whether the genetic similarity of ancient and extant landraces is the outcome of shared ancestry or the result of later hybridization between local wild barley and domesticated forms, we performed model-based assignment of present-day and ancient samples to two ancestral groups corresponding to wild and domesticated barleys using ADMIXTURE[31], considering only transversion variants. Analysis with two ancestral populations confirmed the strong differentiation between wild and domesticated accessions observed in PCA, with a perfect correlation between domestication status and assignment to ancestral populations. Thus, all ancient samples were assigned to the domesticated group (**Fig. 4a**). However, the wild ancestry coefficients of two ancient samples were 4.2% and 8.7%. Present-day landraces from the Levant (Israel, Jordan, Lebanon and Syria) also showed an elevated fraction of wild ancestry (6.8%). By contrast, the average proportion of wild ancestry for European landraces was only 0.14%. We also performed ADMIXTURE runs with the number of ancestral populations (*K*) set to five because this *K* value had the lowest cross-validation error (**Supplementary Fig. 9**). In this analysis, domesticated barley was separated into three clusters and wild barley was separated into two clusters (**Supplementary Fig. 10**). The deeply sequenced ancient sample JK3014 had 7.6% ancestry from a cluster predominantly composed of wild accessions, whereas the average ancestry fraction in this cluster was only 1.7% for domesticated barley. These observations

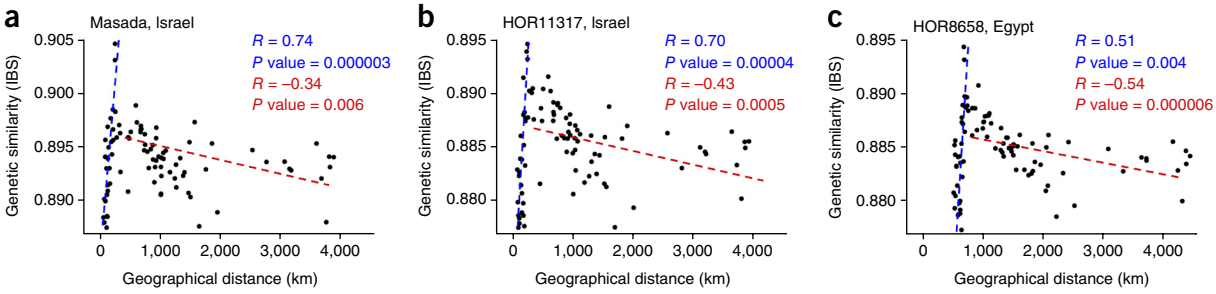

Figure 3 Relationship between genetic similarity and geographical distance. (**a**–**c**) Scatterplots of genetic similarity plotted against geographical distance in comparisons of 91 extant wild barley accessions sampled across the range, including the Fertile Crescent, with archaeological samples found at Yoram Cave (**a**), a two-rowed cultivated landrace from Israel (**b**) and a two-rowed cultivated landrace from Egypt (**c**). The geographical positions attributed to the compared samples are as follows: 31.3141° N, 35.353° E in **a**, 31.7156° N, 35.1871° E in **b** and 31.193° N, 29.904° E in **c**. Correlation coefficients (*R*) and *P* values for the geographically proximate and distant subsets appear in blue and red, respectively.

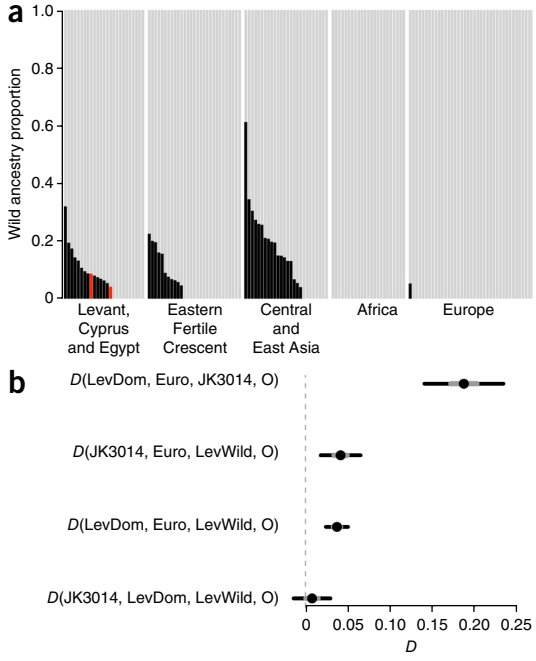

Figure 4 Gene flow between wild and domesticated barleys in the Levant. (**a**) Wild ancestry coefficients of landraces from different geographical regions and ancient barleys as determined by ADMIXTURE. The proportion of wild ancestry is shown in black for extant samples and in red for ancient samples; domesticated ancestry is shown in gray. (**b**) *D* statistics for different quadruples of barley populations (P_1, P_2, P_3 and O). Positive *D* values indicate that P_1 shares more derived alleles with P_3 than P_2 does. Black bars correspond to ±3 s.e.m., and gray bars correspond to ±1 s.e.m. Population names are abbreviated as follows: LevDom, Levantine landraces; LevWild, Levantine wild barleys; Euro, European landraces; O, *H. pubiflorum* outgroup). JK3014 is a deeply sequenced ancient sample.

suggest gene flow between wild and cultivated barley in regions of sympatry.

We used *D* statistics[32] to corroborate the hypothesis of archaic admixture between wild and domesticated barley populations in the Levant. We considered five categories: ancient barley, extant wild barley from the Levant (Israel, Syria, Jordan and Lebanon), extant landraces from the Levant, extant European landraces and the outgroup *Hordeum pubiflorum*[33]. We calculated *D* for each ancient sample separately (**Supplementary Table 4**) and focus here on the results for the single deeply sequenced sample, JK3014 (**Fig. 4b**). *D*(extant Levantine landraces, extant European landraces, JK3014, outgroup) was significantly positive, confirming the close affinity of the ancient sample with the present-day Levantine landraces. The comparisons *D*(extant Levantine landraces, extant European landraces, Levantine wild barley, outgroup) and *D*(JK3014, extant European landraces, Levantine wild barley, outgroup) were also significantly positive ($z > 3$; **Supplementary Table 4**), indicating admixture between wild and domesticated barleys from Israel after the lineage leading to Levantine landraces split from the progenitors of European landraces. The ancient sample did not show closer affinity to extant Levantine wild barley than to present-day landraces from this region, as *D*(JK3014, extant Levantine landraces, Levantine wild barley, outgroup) was not significantly different from 0. These findings indicate that the genomes of both ancient and present-day cultivated barleys from the Levant show traces of archaic gene flow from sympatric wild accessions after the split between Levantine and European landraces, supporting the notion of hypothetical hybridization events between domesticated barley and sympatric wild stands[7,34]. As a consequence of this demographic scenario, the homogenization of allele frequencies in sympatric wild and domesticated barleys through bidirectional gene flow may complicate inferences about origin(s) and domestication history[34] on the fine genomic scale, whereas key domestication-related genes (such as *BTR1* and *BTR2*) are resistant to wild introgression because of strong selection against shattering spikes. Despite hybridization events between wild and domesticated barleys over the last six millennia, the overall picture is that the genomes of extant Levantine landraces have remained remarkably similar to how they were 6,000 years ago. This is despite climate change[35] and anthropogenic transformations of local flora and fauna, including changes in agricultural practices[36], which might have favored the introduction of landraces from other regions that were better adapted to the changing agricultural environment. Although we found no indications of major lineage turnovers in the barley crop from the Southern Levant (as have, for example, been observed in Near Eastern pig populations[37,38]), the eventful history of this region makes it likely that the farmers who grew cereals there several millennia ago are not the ancestors of those who tend the present-day landraces[39]. One can speculate that conquerors and immigrants did not bring crop seeds from their old homelands but favored locally adapted landraces.

Expanding on previous studies that reported the PCR amplification and sequencing of single genes from ancient wheat and barley samples[10,40], our results show that very ancient desiccated plant remains preserved under hot and arid conditions contain sufficient amounts of endogenous DNA to underpin genome-wide population genetic analyses in the context of diversity panels of extant individuals. Our analysis demonstrates the value of archaeogenomics in supporting contemporary genetic-based phylogeographical studies in exploring crop origins and shows that domesticated barley from 6,000 cal BP appeared remarkably similar to proximate extant landraces, indicating that the major domestication events had occurred by that time.

URLs. Novosort, http://www.novocraft.com/; R package mapdata, https://cran.r-project.org/web/packages/mapdata/index.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw read files for the ancient samples can be retrieved from the European Nucleotide Archive (ENA) under project PRJEB12197. The SNP genotype matrix is available at http://dx.doi.org/10.5447/IPK/2016/6. Passport information for the extant barley panel can be retrieved from http://dx.doi.org/10.5447/IPK/2016/3. The DOIs were registered with e!DAL[41].

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We gratefully acknowledge B. Schäfer for providing photographs of barley spikes and A. Fiebig for sequence data submission. This work was supported by a grant of the Israel Science Foundation (1179/13) to E.W., funding from the Endowed Chair in Molecular Genetics Applied to Crop Improvement at the University of Minnesota and the Triticeae Coordinated Agricultural Project, USDA-NIFA 2011-68002-30029 to G.J.M., and core funding of IPK Gatersleben to N.S. and M.M. R.W., J.R. and I.K.D. were supported by Research Programme funding from the Scottish government and the University of Dundee (R.W.). N.M. and U.D. acknowledge a grant from the Irene Levi Sela CARE Foundation.

AUTHOR CONTRIBUTIONS

E.W., T.F., N.S. and J.K. conceived the study. E.W., T.F., N.S., J.K., V.J.S. and M.M. designed experiments. N.M., U.D., M.D., S.R. and E.W. performed excavations and archaeobotanical analyses. V.J.S., A.H. and E.R. performed the ancient DNA experiments. M.M., S.H., A.K., M.S., S.H.V. and R.E.G. analyzed data. J.R., M.M., I.K.D., B.K., G.J.M., N.S. and R.W. provided exome capture data. M.M., V.J.S., A.H., S.R., T.F., J.K., E.W. and N.S. wrote the manuscript with input from all co-authors. All authors read and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Zohary, D., Hopf, M. & Weiss, E. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford University Press on Demand, 2012).
2. Troy, C.S. *et al.* Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088–1091 (2001).
3. Larson, G. *et al.* Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. USA* **104**, 15276–15281 (2007).
4. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* **25**, 1515–1519 (2015).
5. da Fonseca, R.R. *et al.* The origin and evolution of maize in the Southwestern United States. *Nat. Plants* **1**, 14003 (2015).
6. Harlan, J.R. & Zohary, D. Distribution of wild wheats and barley. *Science* **153**, 1074–1080 (1966).
7. Poets, A.M., Fang, Z., Clegg, M.T. & Morrell, P.L. Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biol.* **16**, 173 (2015).
8. Badr, A. *et al.* On the origin and domestication history of barley (*Hordeum vulgare*). *Mol. Biol. Evol.* **17**, 499–510 (2000).
9. Li, C. *et al.* Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. *J. Archaeol. Sci.* **38**, 115–119 (2011).
10. Palmer, S.A., Moore, J.D., Clapham, A.J., Rose, P. & Allaby, R.G. Archaeogenetic evidence of ancient Nubian barley evolution from six to two-row indicates local adaptation. *PLoS One* **4**, e6301 (2009).
11. Smith, O. *et al.* Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* **347**, 998–1001 (2015).
12. Weiß, C.L., Dannemann, M., Prüfer, K. & Burbano, H.A. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *eLife* **4**, e10005 (2015).
13. Kistler, L. *et al.* Thermal age, cytosine deamination and the veracity of 8,000 year old wheat DNA from sediments. Preprint at *bioRxiv* http://dx.doi.org/10.1101/032060 (2015).
14. Briggs, A.W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104**, 14616–14621 (2007).
15. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).
16. Briggs, A.W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
17. Mascher, M. *et al.* Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**, 494–505 (2013).
18. Russell, J. *et al.* Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat. Genet.* http://dx.doi.org/10.1038/ng.3612 (2016).
19. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
20. Reich, D., Thangaraj, K., Patterson, N., Price, A.L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
21. Pourkheirandish, M. *et al.* Evolution of the grain dispersal system in barley. *Cell* **162**, 527–539 (2015).
22. Komatsuda, T. *et al.* Six-rowed barley originated from a mutation in a homeodomain–leucine zipper I–class homeobox gene. *Proc. Natl. Acad. Sci. USA* **104**, 1424–1429 (2007).
23. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history *Nat. Commun.* **7**, 10408 (2016).
24. Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nat. Genet.* **44**, 243–246 (2012).
25. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
26. Morrell, P.L. & Clegg, M.T. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc. Natl. Acad. Sci. USA* **104**, 3289–3294 (2007).
27. van Zeist, W. & Bakker-Heeres, J.A.H. Archaeological studies in the Levant. 1. Neolithic sites in the Damascus basin: Aswad, Ghoraifé, Ramad. *Palaeohistoria* **24**, 165–256 (1985).
28. Snir, A. *et al.* The origin of cultivation and proto-weeds, long before Neolithic farming. *PLoS One* **10**, e0131422 (2015).
29. Abdel-Ghani, A.H., Parzies, H.K., Omary, A. & Geiger, H.H. Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. *Theor. Appl. Genet.* **109**, 588–595 (2004).
30. Russell, J. *et al.* Analysis of >1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. *New Phytol.* **191**, 564–578 (2011).
31. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
32. Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
33. Jakob, S.S. & Blattner, F.R. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Mol. Biol. Evol.* **23**, 1602–1612 (2006).
34. Dickson, A.D. *et al. Barley: Origin, Botany, Culture, Winter Hardiness, Genetics, Utilization, Pests* (US Science and Education Administration, US Department of Agriculture, 1978).
35. Russell, J. *et al.* Genetic diversity and ecological niche modelling of wild barley: refugia, large-scale post-LGM range expansion and limited mid-future climate threats? *PLoS One* **9**, e86021 (2014).
36. Frumin, S., Maeir, A.M., Kolska Horwitz, L. & Weiss, E. Studying ancient anthropogenic impacts on current floral biodiversity in the Southern Levant as reflected by the Philistine migration. *Sci. Rep.* **5**, 13308 (2015).
37. Meiri, M. *et al.* Ancient DNA and population turnover in Southern Levantine pigs—signature of the Sea Peoples migration? *Sci. Rep.* **3**, 3035 (2013).
38. Giuffra, E. *et al.* The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785–1791 (2000).
39. Haber, M. *et al.* Genome-wide diversity in the Levant reveals recent structuring by culture. *PLoS Genet.* **9**, e1003316 (2013).
40. Bilgic, H., Hakki, E.E., Pandey, A., Khan, M.K. & Akkaya, M.S. Ancient DNA from 8400 year-old Çatalhöyük wheat: implications for the origin of Neolithic agriculture. *PLoS One* **11**, e0151974 (2016).
41. Arend, D. *et al.* e!DAL—a framework to store, share and publish research data. *BMC Bioinformatics* **15**, 214 (2014).

ONLINE METHODS

Archaeology. *Yoram Cave.* Yoram Cave is archaeologically significant as one of the rare cave sites with a single layer of human occupation according to current radiocarbon dating findings. Unlike most other Judean Desert caves, there are no findings from the later Roman and Byzantine Periods. In addition, it is one of the rare cave sites that have not suffered from modern looting or hyena burrowing. It is the only Chalcolithic cave site in the Judean Desert that has been excavated by high-resolution sampling methods. The cave's plant assemblage has been preserved by drying, supporting use in possible DNA-based studies.

Location and description of the cave. Yoram Cave is part of a cave complex (Masada caves–South), with three caves located in the southeastern cliff of Masada Horst (**Supplementary Fig. 2**), facing the Dead Sea. Access to the cave complex is relatively difficult as it requires walking along narrow goat paths on a sharp incline rockfall. The Yoram Cave entrance is on an almost vertical cliff, some 4 m above a goat trail at its base (**Fig. 1b**).

Water sources are scarce. Some small rock depressions, holding floodwaters for a few months, are found about 150 m southwest of the cave complex. The nearest permanent springs are in Tze'elim canyon, some 5 km in walking distance to the north.

The cave entrance is 2.9 m wide with a fieldstone wall stretching along the entrance. The interior has two rooms (**Fig. 1a** and **Supplementary Fig. 1**). The northern room is approximately 7 m long and between 3.5 and 5 m wide, and it contains large boulders. The southern room consists of three areas: an entrance (A), a short corridor (B) and a small inner cubicle (C). The latter room's maximum length is 6.5 m, and its width is between 2 and 2.5 m. (**Fig. 1a** and **Supplementary Figs. 1** and **2**). The heights of these three areas range between 0.2 and 1.8 m. Most human activities and related plant remains were found in the southern room.

Stratigraphy. Excavations in the southern room found three phases (from top to bottom): (i) a biogenic layer, mostly the result of nesting activities of large birds of prey (possibly bearded vulture, *Gypaetus barbatus*, or members of the eagle family), including large and small twigs, bones and droppings; (ii) an anthropogenic layer, representing the Chalcolithic period of human activity (**Supplementary Fig. 2**); and (iii) a pre-floor Chalcolithic layer. Scant evidence of modern human activity was discovered on top of the first layer. An initial round of radiocarbon dating validated the excavators' field observation of a Chalcolithic origin of the anthropogenic layer (ca. 6,200–5,800 cal BP). The biogenic layer was dated to the Late Bronze Age/Iron Age I period.

Reeds (possibly *Typha* or *Phragmites* species; see ref. 42) were found abundantly among the anthropogenic layer plant remains (**Fig. 1b**). The appearance of reeds alongside various rope segments (plants still to be identified) and a small mat section in one of the excavated samples hint that simple mats probably covered the cave floor. Such mats indicate preparations for prolonged stay in the cave, rather than chance occupancy. The human-built wall in the entrance is further evidence for the prudent use of the cave.

Excavation and sampling. Excavation was conducted by a high-resolution excavation method, with the excavated space being divided into subunits (**Fig. 1a**). These were meticulously sampled, with each sample going through a sorting procedure, using 1-mm and 100-µm mesh sieves. During sifting, various categories of finds were separated (for example, archaeological artifacts, macro- and microecofacts—archaeozoological and archaeobotanical remains), which were packed separately. Additional separation was undertaken on plant material, with 1 liter of sediment from each excavated bucket from the 'anthropogenic' loci (1, 3, 4, 5, 7 and 9) was kept for archaeobotanical analysis in the laboratory. Larger samples were also retained when plant remains were visible to the naked eye during excavation.

Mapping of Yoram Cave (**Fig. 1a** and **Supplementary Fig. 1**) was performed by the excavation team headed by U.D. and N.M. in 2007 using standard cave mapping equipment, including a Leica Disto D3 laser inclinometer and a Silva Ranger 3 prismatic compass; the grade of mapping was 5C. Field maps were later graphically edited using Limelight software.

DNA extraction and library preparation. A panel of 13 samples was initially selected for this study consisting of 8 barley grains, 2 barley ear fragments, 2 wheat emmer grains and 1 emmer ear fragment. All subsequent sampling procedures, DNA extractions and library preparations were carried out in clean-room facilities dedicated to ancient DNA research at Tübingen University. During the sampling process, all samples weighing more than 15 mg were divided into two parts, of which one part was used for subsequent DNA extraction and the other one was sent for radiocarbon dating at Curt Engelhorn Zentrum Archaeometrie (Mannheim, Germany). DNA extraction was conducted on 5 to 30 mg of plant material using the PTB extraction protocol detailed by Kistler[43] with the following modifications: all samples were extracted twice (E1 and E2). After an initial incubation for 2 h at 37 °C, the plant remains were pelleted, and the supernatants were taken off and stored at 4 °C overnight. Plant pellets were resuspended in extraction buffer a second time and incubated overnight at 37 °C. All extracts were purified simultaneously on the next day.

For library preparation, a well-established protocol from Meyer and Kircher[44] was used to convert a 20-µl aliquot of each DNA extract into double-stranded Illumina libraries. Adaptor ligation to the fragments was quantified using qPCR with primers IS7 and IS8 (ref. 44), the reagents of the DyNAmo Flash SYBR Green qPCR kit (Biozym) and a LightCycler 96 instrument (Roche). Then, double-indexed libraries were created by adding sample-specific barcodes to both library adaptors via amplification[45] followed by another quantification assay using primers IS5 and IS6 (ref. 44) to estimate the efficiency of the indexing PCR. All extraction and library blanks were treated accordingly. These libraries were used subsequently for initial shotgun sequencing.

For genome-wide shotgun sequencing and enrichment, additional libraries for extracts JK2281E1, JK2281E2, JK3009E1, JK3010E1, JK3013E1 and JK3014E1 (**Table 1** and **Supplementary Table 1**) were prepared from 50-µl aliquots of all DNA extracts following the methods described above[44,45] with one modification: all extracts and blanks were treated with UDG and endonuclease VIII during library preparation to avoid potential sequencing artifacts that are caused by the characteristic ancient DNA damage pattern due to the deamination of cytosine to uracil over time[16].

For all indexed libraries, a second amplification was carried out in 100-µl reactions using 5 µl of library template, 4 U AccuPrime Taq DNA Polymerase High Fidelity (Invitrogen), 1 U 10× AccuPrime buffer (containing dNTPs) and 0.3 µM IS5 and IS6 primers[44]. The following thermal profile was performed: 2 min of initial denaturation at 94 °C followed by 4 to 17 cycles consisting of 30 s of denaturation at 94 °C, 30 s of annealing at 60 °C and 2 min of elongation at 68 °C with a final 5 min of elongation at 68 °C. After amplification, the products were purified using MinElute spin columns (Qiagen) according to the manufacturer's protocol and quantified using an Agilent Bioanalyzer DNA 1000 Chip.

All libraries for initial and genome-wide shotgun sequencing were then diluted to a 10 nM concentration and pooled in equimolar amounts. Initial shotgun sequencing of libraries was undertaken on the Illumina HiSeq 2500 platform, using a paired-end dual-index run with 2 × 100 + 7 + 7 cycles and the manufacturer's protocols for multiplex sequencing (TruSeq Paired-End Cluster Kit v3-cBot-HS). Genome-wide shotgun sequencing of the UDG-treated libraries was performed on the Illumina NextSeq 500 platform with 2 × 150 + 8 + 8 cycles using the NextSeq High-Output reagent kit v1 and the manufacturer's protocol for multiplex sequencing.

The UDG-treated libraries from extracts JK2281E1 and JK2281E2 (**Table 1** and **Supplementary Table 1**) were treated separately: after the second amplification, the libraries were enriched using a sequence capture assay for the barley exome[17], as described by Himmelbach *et al.*[46] with one modification: the concentration of the DNA fragments recovered from the capture was determined by qPCR using primers IS5 and IS6 (ref. 44), SYBR Green PCR Master Mix (Qiagen, Hilden) and a 7900 HT Fast Real-Time PCR system (Applied Biosystems).

After dilution to 10 nM, sequencing was carried out on the Illumina HiSeq 2500 platform as described by Mascher *et al.*[17] with a paired-end single-index run using 101 + 6 + 100 cycles and the manufacturer's protocols for multiplex sequencing (TruSeq Paired-End Cluster Kit v3-cBot-HS).

Four additional UDG-treated libraries of JK3014E1 were produced for deeper sequencing as detailed previously, diluted to a 10 nM concentration and pooled in equimolar amounts together with the already sequenced JK3014E1 UDG-treated library. Sequencing of the pools was conducted on the HiSeq 4000 platform with 2 × 75 + 8 + 8 cycles using the HiSeq 3000/4000 Paired-End

doi:10.1038/ng.3611

Cluster Kit, the HiSeq 3000/4000 SBS Kit and the manufacturer's protocol for multiplex sequencing.

Raw sequence reads have been uploaded to the EMBL ENA short-read archive (accession PRJEB12197).

Processing and mapping of sequence reads. Overlapping paired-end reads were merged using scripts provided by Kircher[47] for samples JK2279–JK2284 or with leeHom[48] (using the parameter --ancientdna) for the other samples. Length distribution of the merged reads was calculated using AWK and the Unix tools 'sort' and 'uniq' as described in Supplement S4 of Gallego Llorente *et al.*[49]. Merged reads were aligned to (i) the whole-genome shotgun assembly of barley cv. Morex[50] and (ii) the chloroplast genome assembly of cv. Morex[51] with BWA-MEM version 0.7.12 (ref. 52) using default parameters. Conversion to BAM format and calculation of mapping statistics were performed with SAMtools[53]. Sorting of BAM files and duplicate removal were performed with Novosort (Novocraft Technologies). Nucleotide misincorporation profiles were generated with mapDamage version 2.0 (ref. 54) for the nuclear and chloroplast genomes. Genotypes of five UDG-treated ancient samples at 1,688,807 known SNP positions[18] were called using single-sample variant calling with SAMtools (version 0.1.19, commands SAMtools mpileup and bcftools view)[55] using only reads with mapping quality above Q30 and considering bases with quality above Q20. SNP positions were retained if their quality score was at least 30 and they were covered by at least two reads. In the case of deeply sequenced sample JK3014, we also required the read depth to be not greater than 30 (equal to three times the mode of the coverage distribution in exome capture target regions). Heterozygous calls were set to missing. In deeply sequenced ancient DNA sample JK3014, 0.7% of variants were called heterozygous (compared to 1.7% in the extant samples). In an inbreeding crop, the divergence between the parental haplotypes of an individual is very low. Contamination with DNA from extant barley would thus become evident in an elevated fraction of heterozygous calls. The absence of such a pattern lends further support to the authenticity of the ancient samples. Coverage statistics were calculated with SAMtools[53] and BEDTools[56].

Population genetic analysis. PCA was performed with EIGENSOFT 6.0.1 (ref. 19) for 5 ancient barley samples and 228 extant barley *ex situ* accessions with clear domestication status and well-described geographical origins[18]. Least-square projection as implemented in the smartPCA program of EIGENSOFT was used to project the ancient samples onto the PCA axes defined by the extant samples.

To investigate the relatedness between the Yoram Cave samples (and extant landraces) and wild barley accessions representing the entire natural distribution range, the corresponding data were extracted from the filtered SNP table. Relatedness between the archaeological samples and each wild accession was measured by the level of IBS calculated across all SNPs using the SNPRelate package[57] in R. For each wild accession, relatedness to the five archaeological samples was averaged using the geometrical mean to obtain one relatedness score (RS). We then calculated the geographical distance (GD) between the Yoram Cave location (31.314° N: 35.353° E) and the sampling position of each wild accession on the basis of its coordinates, converting the distances between coordinate positions to kilometers using the rough conversion metrics of 1° = 111 km. To capture the change in the correlation coefficient sign observed between geographical distance and genetic relatedness, the data were split into geographically proximate and distant categories using the most related wild accession coordinates rounded up to the nearest 50 km as a breakpoint.

Model-based ancestry estimation was performed with ADMIXTURE[31]. For each *K* value from 1 to 10, 20 replicate ADMIXTURE runs were performed on the genotype matrix of 233 samples (228 geo-referenced extant accessions of known domestication status[18] plus 5 ancient samples) using only transversion variants with a present genotype call for at least one ancient sample. Before running ADMIXTURE, linkage disequilibrium pruning was performed with the R package SNPRelate[57] using the function snpgdsLDpruning() with parameters ld.threshold = 0.4, slide.max.bp = 100000 and slide.max.n = 50. Replicate ADMIXTURE runs were combined with CLUMPP[58].

D statistics were calculated using ADMIXtools[59] after the SNP matrix had been converted to EIGENSOFT format with the SNPRelate[57] function snpgdsGDS2Eigen(). The barley relative *H. pubiflorum* was used as an outgroup. We used exome sequencing reads of *H. pubiflorum* published by Mascher *et al.*[17] to call genotypes at variant positions with SAMtools.

Read alignments at the *Btr1*, *Btr2* and *Vrs1* loci were manually inspected with SAMtools 'tview'.

The map in **Figure 2a** was generated with the R package 'mapdata'.

42. Schick, T. The early basketry and textiles from caves in the northern Judean desert. *Part II'Atiqot* **41**, 223–239 (2002).
43. Kistler, L. Ancient DNA extraction from plants. *Methods Mol. Biol.* **840**, 71–79 (2012).
44. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, prot5448 (2010).
45. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
46. Himmelbach, A., Knauft, M. & Stein, N. Plant sequence capture optimised for Illumina sequencing. *Bio Protoc.* **4**, e1166 (2014).
47. Kircher, M. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol. Biol.* **840**, 197–228 (2012).
48. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
49. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015).
50. International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716 (2012).
51. Saski, C. *et al.* Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor. Appl. Genet.* **115**, 571–590 (2007).
52. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
55. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
56. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
57. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
58. Jakobsson, M. & Rosenberg, N.A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806 (2007).
59. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).