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

Mascher, Martin; J. Schuenemann, Verena; Davidovich, Uri; Marom, Nimrod; Himmelbach, Axel; Hübner, Sariel

Published in:
Nature Genetics

DOI:
10.1038/ng.3611

Publication date:
2016

Document Version
Peer reviewed version

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

Mascher, Martin; J. Schuenemann, Verena; Davidovich, Uri; Marom, Nimrod; Himmelbach, Axel; Hübner, Sariel; Korol, Abraham; David, Michal; Reiter, Ella; Riehl, Simone; Schreiber, Mona; Vohr, Samuel H.; Green, Richard E.; Dawson, Ian K.; Russell, Joanne; Kilian, Benjamin; Muehlbauer, Gary J.; Waugh, Robert; Fahima, Tzion; Krause, Johannes; Weiss, Ehud; Stein, Nils

Published in:
Nature Genetics

Publication date:
2016

Document Version
Peer reviewed version

DOI: 10.1038/ng.3611

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[17,3,4], Ehud Weiss[11], Nils Stein[1]

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) University of Tübingen, Institute for Archaeological Sciences, Tübingen, Germany
4) University of Tübingen, Senckenberg Center for Human Evolution and Paleoenvironment, Tübingen, Germany
5) The Hebrew University, Institute of Archaeology, Jerusalem, Israel
6) University of Haifa, Laboratory of Archaeozoology, Zinman Institute of Archaeology, Haifa, Israel
7) University of British Columbia, Department of Botany, Vancouver, Canada
8) Tel Hai College, Department of Biotechnology, Upper Galilee, Israel
9) University of Haifa, Institute of Evolution, Haifa, Israel
10) University of Haifa, Department of Evolutionary and Environmental Biology, Haifa, Israel
11) Bar-Ilan University, The Martin (Szusz) Department of Land of Israel Studies and Archaeology, Ramat-Gan, Israel
12) University of California, Santa Cruz, Department of Biomolecular Engineering, Santa Cruz, USA
13) The James Hutton Institute, Cell and Molecular Sciences, Invergowrie, Dundee, United Kingdom
14) University of Minnesota, Department of Plant Biology, St. Paul, USA
15) University of Minnesota, Department of Agronomy and Plant Genetics, St. Paul, USA
16) University of Dundee, Division of Plant Sciences, Dundee United Kingdom

17) Max Planck Institute for the Science of Human History, Jena, Germany

18) present address: Bayer CropScience NV, BCS Breeding & Trait Development, Zwijnaarde (Gent), Belgium

19) These authors contributed equally to this work.

Correspondence should be addressed to Tzion Fahima (fahima@research.haifa.ac.il), Johannes Krause (krause@shh.mpg.de), Ehud Weiss (ehud.weiss@biu.ac.il) or Nils Stein (stein@ipk-gatersleben.de).

The cereal grass barley was domesticated about 10,000 years ago in the Fertile Crescent and became a founder crop of Neolithic agriculture[1]. Here, we report 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 revealed 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 in the past six millennia have not experienced a major lineage turnover although there is evidence for gene flow between cultivated and sympatric wild populations. We show the utility 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 datasets 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]. Our current knowledge of their domestication is largely derived from morphological analysis of archaeobotanical remains[1] and the 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 present (cal BP)[1], the oldest verified DNA sequences to date 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]. 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 genome sequences of five 6000-year old barley grains excavated at Yoram Cave in the Judean Desert, Israel. Yoram Cave is part of a complex of three difficult-to-access caves, located in the south-eastern cliff of the Masada Horst facing

the Dead Sea. High-resolution excavation (Online Methods, Fig. 1a, Supplementary Figures 1, 2) revealed 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 (Figure 1b), prompting us to attempt the retrieval of DNA sequences. We selected ancient barley grains (Figure 1c, Supplementary Figure 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, whose other halves were 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). Based on the fraction of reads that could be aligned to the barley reference genome, we estimated the content of endogenous DNA to range from 0.4 to 96.4 %. Sequence reads of eight samples showed fragment sizes and damage patterns characteristic of ancient DNA (Supplementary Table 1, Supplementary Figures 4, 5, 6), which demonstrates the authenticity of the samples[14,15]. Deamination-derived mismatches (C -> T, G -> A) occurred towards the ends of reads with frequencies between 1.9 and 21.8 %. We only used five samples with on average more than 12 % misincorporations[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 misincorporations 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 from 267 entries from *ex situ* collections representing extant populations of wild (*Hordeum vulgare* ssp. s*pontaneum)* and domesticated (*H. vulgare* ssp. *vulgare*) barley from across the range of the species[18] (Figure 2a). This dataset[18] comprised 1,688,807 single nucleotide polymorphisms (SNPs) (Table 1).

Principal component analysis[19] has revealed fundamental patterns of population structure across the present-day accessions[18]. The first principal component (PC1) clearly differentiated wild and domesticated barleys, while PC2 represented the variation in the wild barleys (Figure 2b). Least-square projection[20] of the ancient samples onto the PCA axes defined by the extant samples revealed 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 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 Figure 7), consistent with a 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 (Figure 2c), the extant accessions closest to the ancient samples were two-rowed domesticated barleys from the Southern Levant and Egypt (Figure 2b, d). A putative two-rowed phenotype of sample JK3014 can also be inferred from the allelic status of the *Six-rowed spike 1* gene[22] (Supplementary Figure 7).

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 since gene flow across larger geographic 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 between 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 barleys from Israel also shared ≥ 30 rare alleles with the ancient samples (Supplementary Table 2).

We measured relatedness of ancient samples to each of the wild barley accessions based on 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 geographic distance between Yoram Cave and each wild accession, but no significant correlation between IBS and geographic distance was found (R = –0.17, P = 0.108). However, when splitting the data into geographically proximal (< 250 km) and distant (> 250 km) samples, significant correlations between geographic distance and the relatedness score were detected for both subsets of the tested wild barley samples (proximal: R = 0.74, P = < 0.001; distant: R = –0.34, P = 0.006) (Figure 3a). The same analysis was conducted between wild barleys and extant landraces from the Fertile Crescent (Figure 3b, c) pinpointing the Upper Jordan Valley as a peak for genetic similarity with domesticated barley. Conducting this analysis using only data from the deeply sequenced JK3014 yielded similar results (Supplementary Figure 8). The comparison of modern cultivars and landraces from outside of the Fertile Crescent (Europe, North 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 (though not the only) center of origin of domesticated barley[8,26]. This hypothesis is supported by two archeological 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 between 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 a 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. The analysis with two ancestral populations confirmed the strong differentiation between wild and domesticated accessions observed in PCA, with a perfect correlation between the domestication status and the assignment to ancestral populations. Thus, all ancient samples were assigned to the domesticated group (Figure 4a). However, 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 wild

ancestry for European landraces was only 0.14 %. We also performed ADMIXTURE runs with the number of ancestral population (K) set to five because this K value had the lowest cross validation error (Supplementary Figure 9). In this analysis, domesticated barley was separated into three and wild barley into two clusters (Supplementary Figure 10). The deeply sequenced ancient sample JK3014 had 7.6 % ancestry from a cluster predominantly composed of wild accessions, while the average ancestry fraction in this cluster was only 1.7 % for domesticated barley. These observations 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, Lebanon), extant landraces from the Levant, extant European landraces and 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 (Figure 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, 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 a closer affinity to extant Levantine wild barley than present-day landraces from this region, as D(JK3014, extant Levantine landraces, Levantine wild barley, outgroup) was not significantly different from zero. These findings indicate that the genomes of both ancient and present-day cultivated barley 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 the allele frequencies in sympatric wild and domesticated barley through bi-directional gene flow may complicate inferences about the origin(s) and domestication history[34] at the fine genomic scale, while key domestication genes

(such as the *btr* genes) are resistant to wild introgression due to a strong selection against shattering spikes. Despit hybridization events between wild and domesticated barleys in the last six millennia, the overall picture is that the genomes of extant Levantine landraces have remained remarkably similar to how they were 6000 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. Whereas we found no indications of major lineage turnovers in the barley crop in the Southern Levant (as has, for example, been observed in Near Eastern pig populations[37,38]), the eventful history of this region makes it likely that the farmers that grew cereals there several millennia ago were not the ancestors of those that tended the present-day landraces[39]. One can speculate that conquerors and immigrants did not bring their 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 phylogeographic studies in exploring crop origins, and shows that 6,000 years BP domesticated barley 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

Accession Codes

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

Acknowledgments

We gratefully acknowledge Birgit Schäfer for providing photographs of barley spikes and Anne Fiebig for sequence data submission. This work was supported by a grant of the Israel Science Foundation (1179/13) to EW, 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 GJM and core funding of IPK Gatersleben to NS and MM. RW, JR and IKD were supported by Research Programme funding from the Scottish Government and the University of Dundee (RW). NM and UD acknowledge a grant from the Irene Levi Sela CARE Foundation.

Authors' contributions

EW, TF, NS, and JK conceived the study. EW, TF, NS, JK, VJS and MM designed experiments. NM, UD, MD, SR and EW performed excavations and archaeobotanical analyses. VJS, AH and ER performed the ancient DNA experiments. MM, SH, AK, MS, SHV and REG analyzed data. JR, MM, IKD, BK, GJM, NS and RW provided exome capture data. MM, VJS, AH, SR, TF, JK, EW and NS wrote the manuscript with input from all co-authors. All authors read and approved the manuscript.

Competing interests

The authors declare no competing financial interests.

References for the main text

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**, 1088-91 (2001).
3. Larson, G. *et al.* Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci U S A* **104**, 15276-81 (2007).
4. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalen, L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr Biol* **25**, 1515-9 (2015).
5. da Fonseca, R.R. *et al.* The origin and evolution of maize in the Southwestern United States. *Nature Plants* **1**(2015).
6. Harlan, J.R. & Zohary, D. Distribution of wild wheats and barley. *Science* **153**, 1074-80 (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. *Journal of Archaeological Science* **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.* Archaeology. 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., Prufer, 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**(2015).
13. Kistler, L. *et al.* Thermal age, cytosine deamination and the veracity of 8,000 year old wheat DNA from sediments. *bioRxiv* (2015).
14. Briggs, A.W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* **104**, 14616-21 (2007).
15. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, 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.* Adaptation of barley to different environments revealed in the exomes of a range- wide collection of landraces and wild relatives. *Nature Genetics* **in press**(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**, 489-94 (2009).

21. Pourkheirandish, M. *et al.* Evolution of the Grain Dispersal System in Barley. *Cell* **162**, 527-39 (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 U S A* **104**, 1424-9 (2007).
23. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *bioRxiv*, 022723 (2015).
24. Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet* **44**, 243-6 (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 U S A* **104**, 3289-94 (2007).
27. W. van Zeist & 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**, 588-95 (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**, 564-78 (2011).
31. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-64 (2009).
32. Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol Biol Evol* **28**, 2239-52 (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**, 1602-12 (2006).
34. Harlan, J.R. On the origin of barley. *USDA Agriculture Handbook* **338**, 9-31 (1968).
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**, 1785-91 (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 Catalhoyuk 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).

Figure legends for the main text

Figure 1: Ancient plant remains excavated at Yoram Cave. Plan of the Southern Chamber of Yoram Cave showing the excavation grid and sub-units **(a).** Photograph of Locus 3 in Square A2 during excavation – note the excellent dry preservation of rope, reeds, seeds and pellets **(b)**. Photograph of a well-preserved, desiccated barley grain found at Yoram Cave **(c)**.

Figure 2: Ancient barley samples are closely related to present-day landraces from the Levant. Ancient barley sequences were compared to exome sequence data of a present-day diversity panel. **(a)** The collection sites of landraces (black circles) and wild barleys (blue circles) are shown. Masada is marked with a red circle. **(b)** Principal component analysis showing ancient samples projected on the present-day diversity panel. The inset magnifies the PCA space around the ancient samples. ISR, SYR, LBN, EGY represent closely related landraces from Israel, 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 **(d).** Spike of 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.

Figure 3: Relationship between genetic similarity and geographic distance. Scatter plot of genetic similarity and geographic distance between 91 extant wild barley accessions sampled range-wide including the Fertile Crescent and **(a)** archaeological samples found at Yoram Cave, **(b)** two-row cultivated landrace from Israel, and **(c)** two-row cultivated landrace from Egypt. The geographic position attributed to each sample is: A (31.3141 N, 35.353 E), B (31.7156 N, 35.1871 E), C (31.193 N 29.904 E). Correlation coefficients and P-values for the geographically proximate and distant subsets are indicated 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 geographic regions and ancient barleys as determined by ADMIXTURE. The wild ancestry proportion is shown in black for extant samples and in red for ancient samples. **(b)** D statistics

for different quadruples of barley populations (P_1, P_2, P_3, O). Positive D values indicate that P_1 shares more derived allele with P_3 than P_2 does. Black bars indicate ±3 standard errors (SE), gray bars ±1 SE. Population names are abbreviated as follows: LevDom (Levantine landraces), LevWild (Levantine wild barley), Euro (European landraces), O (*H. pubiflorum,* outgroup). JK3014 is a deeply sequenced ancient sample.

Table 1: Summary of ancient barley samples used for genetic analyses.

Sample name	Radiocarbon age[1]	Number of raw reads	Percentage of mapped reads	Percentage of unique reads	Average read depth[2]	Number of called SNPs
JK2281	5290 ± 27	256.1 M	31.4%	26.0%	0.54	162,110
JK3009	5034 ± 36	89.2 M	61.7%	67.3%	0.46	133,365
JK3010	5032 ± 37	94.7 M	62.3%	57.9%	0.96	278,505
JK3013	5227 ± 37	82.5 M	49.4%	39.8%	0.19	18,949
JK3014	4988 ± 36	5,131.2 M	86.4%	28.3%	20	1,283,396

[1]Uncalibrated radiocarbon years before the present, see for Supplementary Table 1 for calibrated dates.

[2]Average read depth in target regions of the exome capture assay for JK2281-JK3013. The mode of the coverage distribution is given for JK3014.

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 has 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 south-eastern cliff of Masada Horst (Supplementary Figure 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 rock-fall. The Yoram Cave entrance is on an almost vertical cliff, some 4m above a goat trail at its base (Fig. 1B).

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

The cave entrance is 2.9 m wide with a fieldstone wall stretching along the entrance. The interior has two rooms (Figure 1a and Supplementary Figure 1). The northern room is approximately 7 m long and between 3.5 and 5 m wide, and 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. (Figure 1a, Supplementary Figures 1, 2). The heights of these 3 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 revealed three phases (from top to bottom):

1. 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;

2. An anthropogenic layer, representing the Chalcolithic period of human activity (Fig. S2); and

3. A pre-floor Chalcolithic layer.

Scant evidence of modern human activity was discovered on top of layer 1. 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 LB/IAI period.

Reeds (possibly *Typha*/*Phragmites* spp. see Schick et al.[42]) were found abundantly among the anthropogenic layer plant remains (Figure 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 a chance occupation. 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 sub-units (Figure 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 (e.g. 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 out of each excavated bucket from the "anthropogenic" loci (#1, 3, 4, 5, 7, 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, Supplementary Fig. 1) was performed by the excavation team headed by Uri Davidovich and Nimrod Marom 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 eight barley grains, two barley ear fragments, two wheat emmer grains and one 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 Tuebingen 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 sent for radiocarbon dating at Curt-Engelhorn-Zentrum Archaeometrie gGmbH Mannheim (Germany). The DNA extractions were 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 a first incubation for two hours at 37°C the plant remains were pelleted, the supernatants were taken off and stored at 4°C over night. Plant pellets were resuspended in extraction buffer a second time and incubated over night at 37°C. All extracts were purified simultaneously on the next day.

For library preparation, a well established protocol by 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 a quantitative PCR with the primers IS7 and IS8[44], the reagents of the DyNAmo Flash SYBR Green qPCR Kit (Biozym) and the Lightcycler 96 (Roche). Then, double indexed libraries were created by adding sample specific barcodes to both library adapters via amplification[45] followed by another quantification assay using the primers IS5 and IS6[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 shot-gun sequencing and enrichment additional libraries for the extracts JK2281E1, JK2281E2, JK3009E1, JK3010E1, JK3013E1 and JK3014E1 (Table 1, 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 uracil-DNA glycosylase (UDG) and endonuclease VIII during the 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 5µl library template, 4 units AccuPrime Taq DNA Polymerase High Fidelity (Invitrogen), 1 unit 10X AccuPrime buffer (containing dNTPs) and 0.3 µM IS5 and IS6 primers[44]. The following thermal profile was performed: 2-min initial denaturation at 94°C, followed by 4 to 17 cycles consisting of 30-sec denaturation at 94°C, a 30-sec annealing at 60°C and a 2-min elongation at 68°C with a final 5-min 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 10nM and pooled in equimolar amounts. Initial shotgun sequencing of libraries was undertaken with an 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 PE Cluster Kit v3-cBot-HS). Genome-wide shotgun sequencing of the UDG treated libraries was performed on an Illumina NextSeq 500 platform with 2 x 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 the extracts JK2281E1, JK2281E2 (Table 1, 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 quantitative PCR using the primers IS5 and IS6[44], the SYBR Green PCR Master Mix (Qiagen, Hilden) and the 7900 HT Fast Real-Time PCR system (Applied Biosystems).

After a dilution to 10 nM the 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 PE Cluster Kit v3-cBot-HS).

Four additional UDG-treated libraries of the JK3014E1 were produced for deeper sequencing as detailed previously, diluted to 10 nM and pooled in equimolar amounts together with the already sequenced JK3014E1 UDG-library. The sequencing of the pools was conducted on a HiSeq4000 platform with 2 x 75+8+8 cycles using the HiSeq 3000/4000 PE Cluster Kit, HiSeq 3000/4000 SBS Kit and the manufacturer's protocol for multiplex sequencing.

Raw sequence reads have been uploaded to 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[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 was done with novosort (Novocraft Technologies Sdn Bhd, Malaysia). Nucleotide misincorporation profiles were generated with mapDamage version 2.0[54] for the nuclear and the chloroplast genome. 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 the deeply sequenced sample JK3014, we also required the read depth to

be not larger than 30 (= x 3 the mode of the coverage distribution in exome capture target regions). Heterozygous calls were set to missing. In the 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 parental haplotypes of an individual is very low. Contamination with DNA of 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

Principal component analysis (PCA) was performed with EIGENSOFT 6.0.1[19] for five ancient barley samples and 228 extant barley *ex situ* accessions with clear domestication status and well-described geographic 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 was extracted from the filtered SNP table. Relatedness between the archeological 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 archeological samples was averaged using the geometrical mean to obtain one relatedness score (RS). We then calculated the geographic distance (GD) between the Yoram Cave location (latitude: 31.314 N, longitude: 35.353 E) and the sampling position of each wild accession based on its coordinates, converting the distances between coordinate positions to kilometres using the rough conversion metrics of 1 degree = 111 km. To capture the change in the correlation coefficient sign observed between geographic 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 break point.

Model-based ancestry estimation was performed with ADMIXTURE (ref. 31). For each K from 1 to 10, twenty replicate ADMIXTURE runs were performed on the genotype matrix of 233 samples (228 geo-referenced extant accessions of known domestication status[18] plus five ancient samples) using only transversion variants with a present genotype call for at least one ancient samples. Before running ADMIXTURE, linkage disequilibrium pruning was done with the R package SNPRelate[57] using the function snpgdsLDpruning() with the parameters "ld.threshold=0.4, slide.max.bp=100000, 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 the EIGENSOFT format with the SNPRelate[57] function snpgdsGDS2Eigen(). The barley relative *Hordeum pubiflorum* was used as an outgroup. We used exome sequencing reads of *H. pubiflorum* published by Mascher et al. (ref. 17) to call genotypes at variant positions with SAMtools.

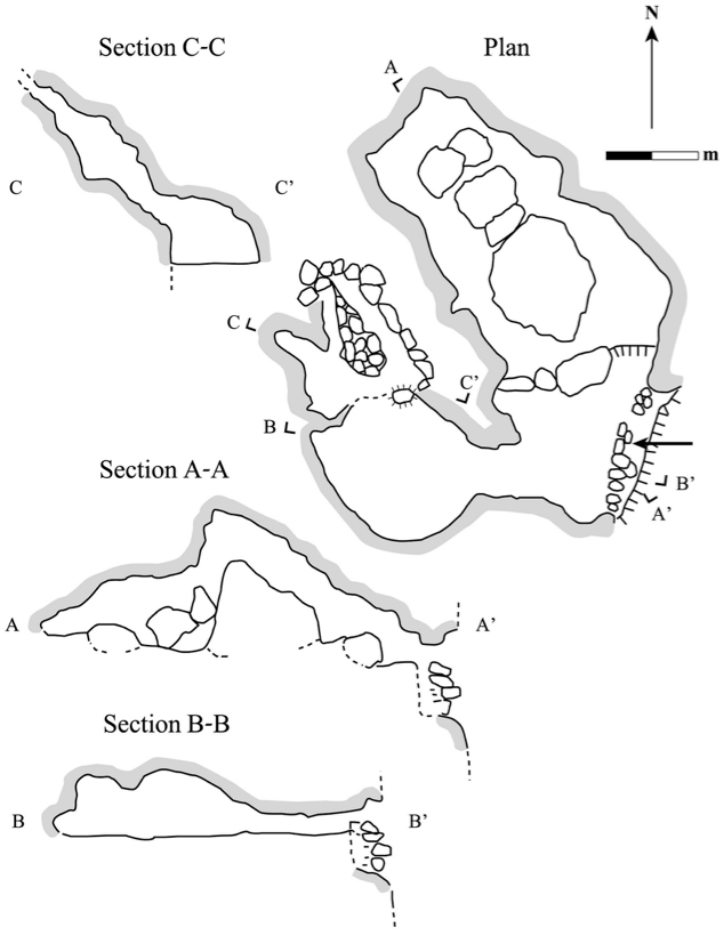
Read alignments at the *Btr1/2* and *Vrs1* loci were manually inspected with SAMtools "tview".

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

Additional references

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**, 71-9 (2012).
44. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb 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**, 820-2 (2015).
50. International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711-6 (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**, 571-90 (2007).
52. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
54. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P.L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-4 (2013).

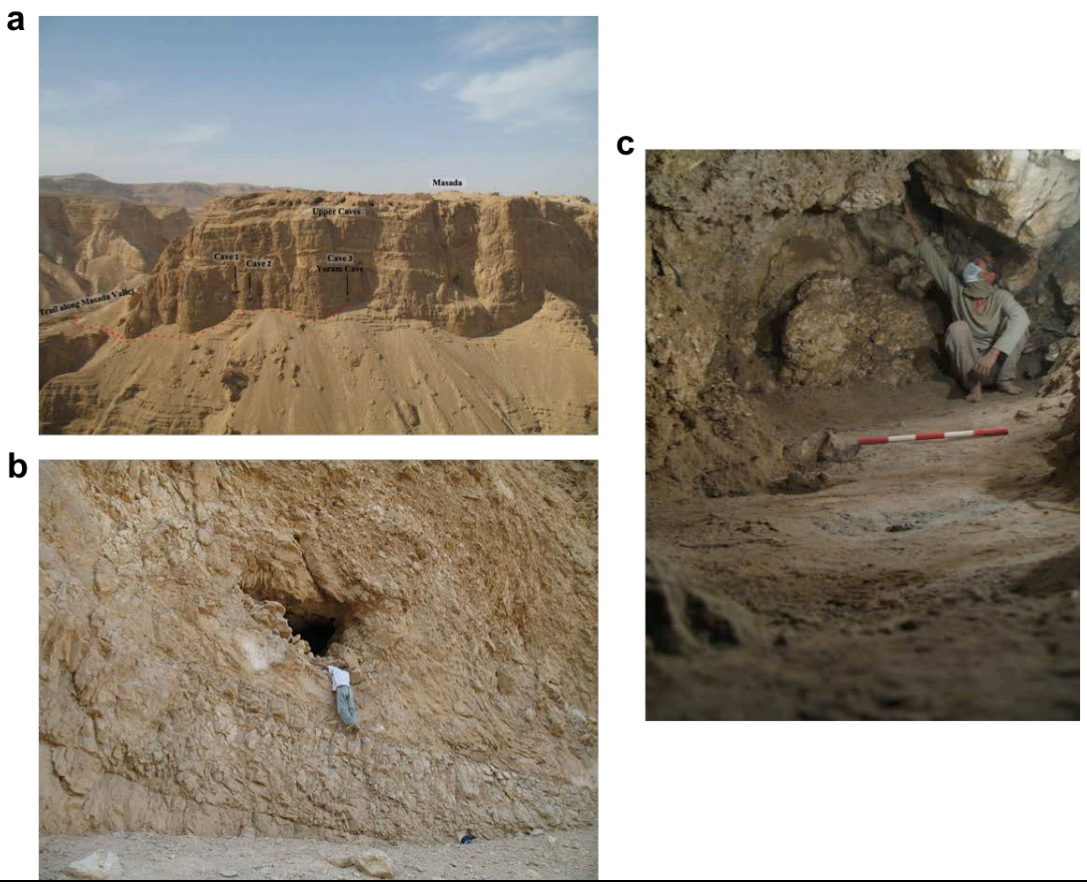
55. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
56. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
57. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-8 (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**, 1801-6 (2007).
59. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-93 (2012).



Supplementary Figure 1

Plan of Yoram Cave – top plan and sections.

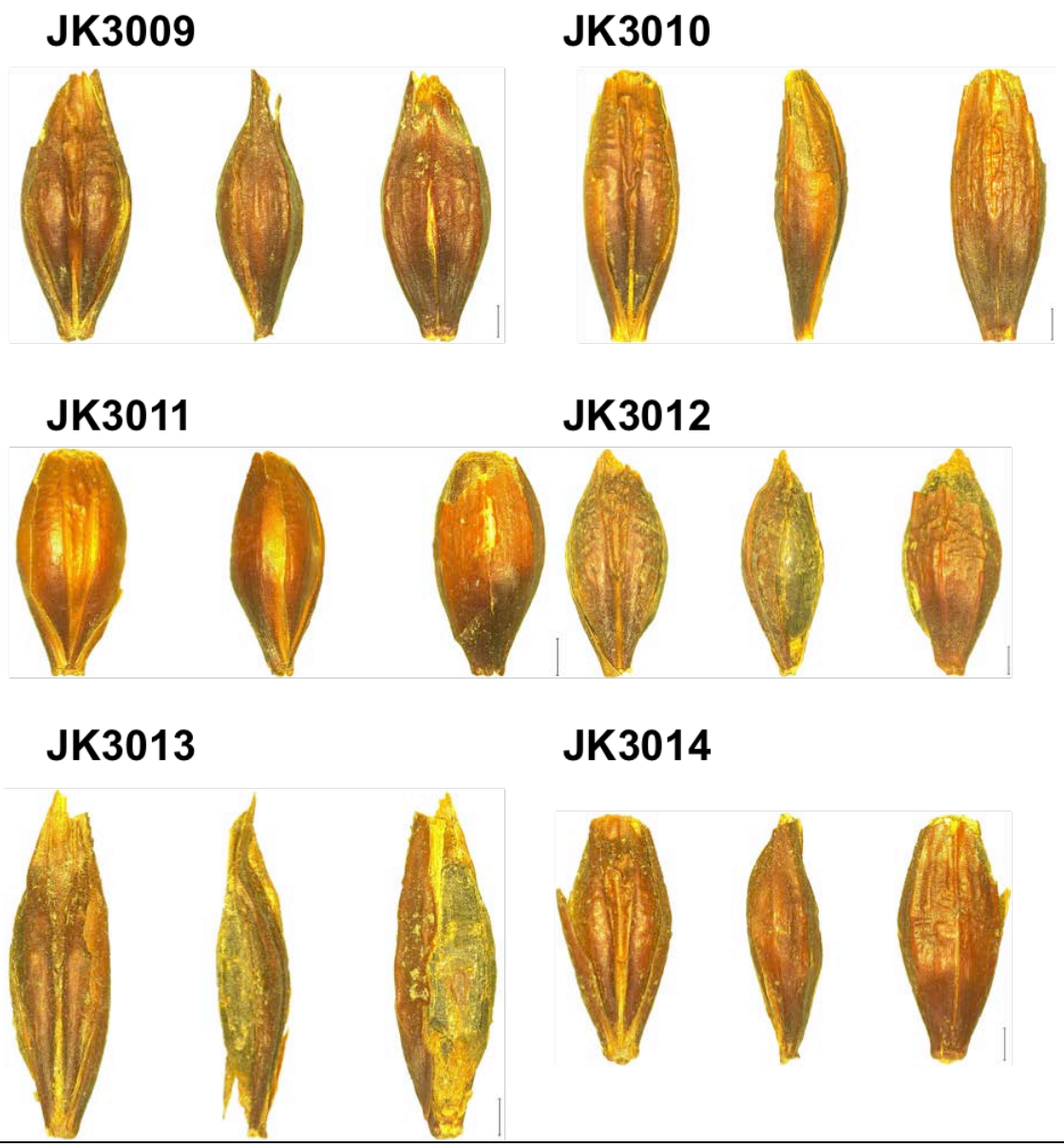
Arrows indicates the entrance to the cave and the human-made wall across it. Note the boulders in the northern room and the relatively horizontal surface in the southern room. Gray shaded borders indicate that the edge of the cave is cut into rock.



Supplementary Figure 2

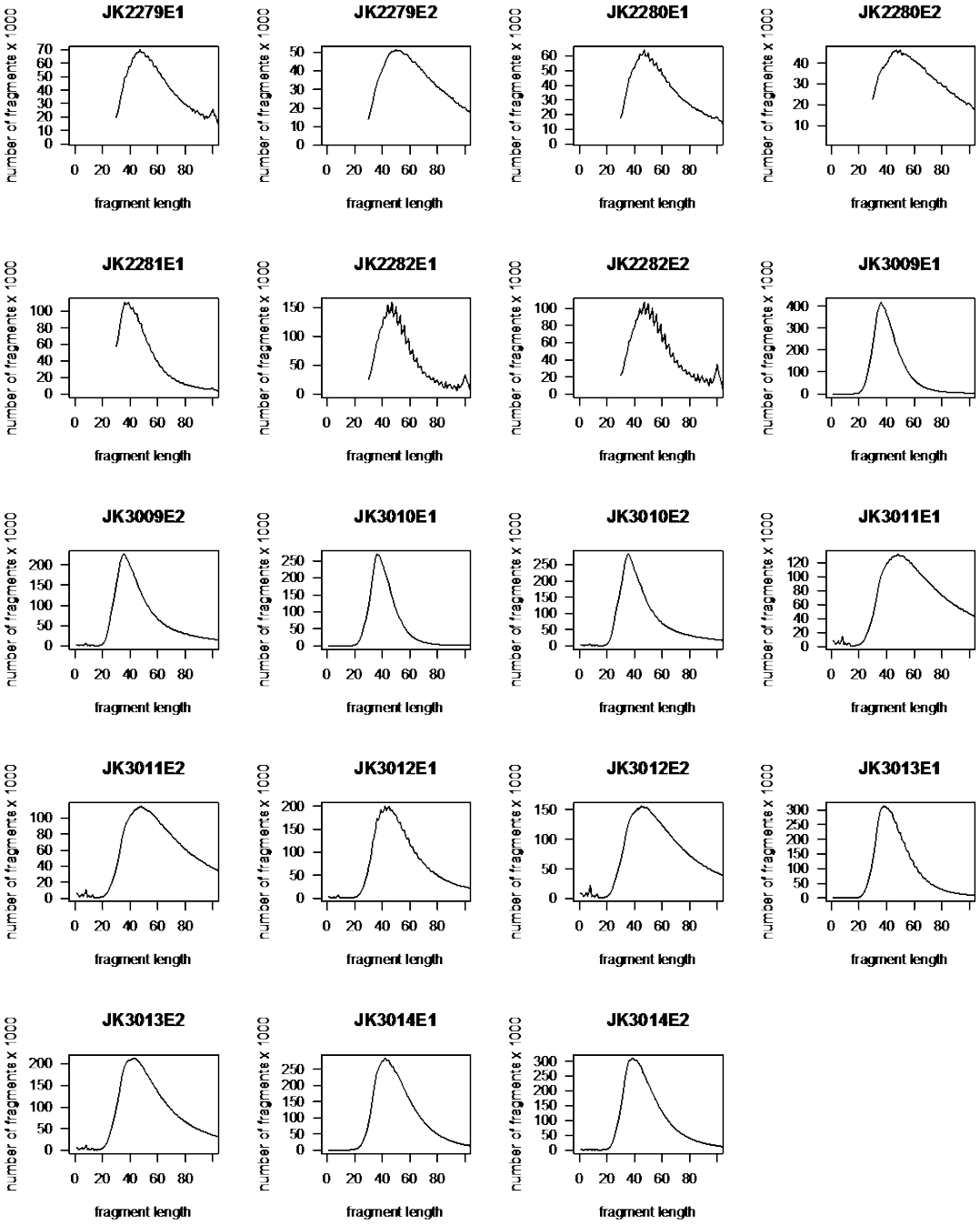
Photographs from the excavation.

Masada Southern Cave complex, with three caves located in the south-eastern cliff of the Masada Horst **(a)**. The dotted red line marks the easiest trail giving access to the caves. The entrance of Yoram Cave, facing south-east, in an almost vertical cliff, some 4 meters above the trail leading to the cave **(b)**. The south room during excavation of locus 3 **(c)**.



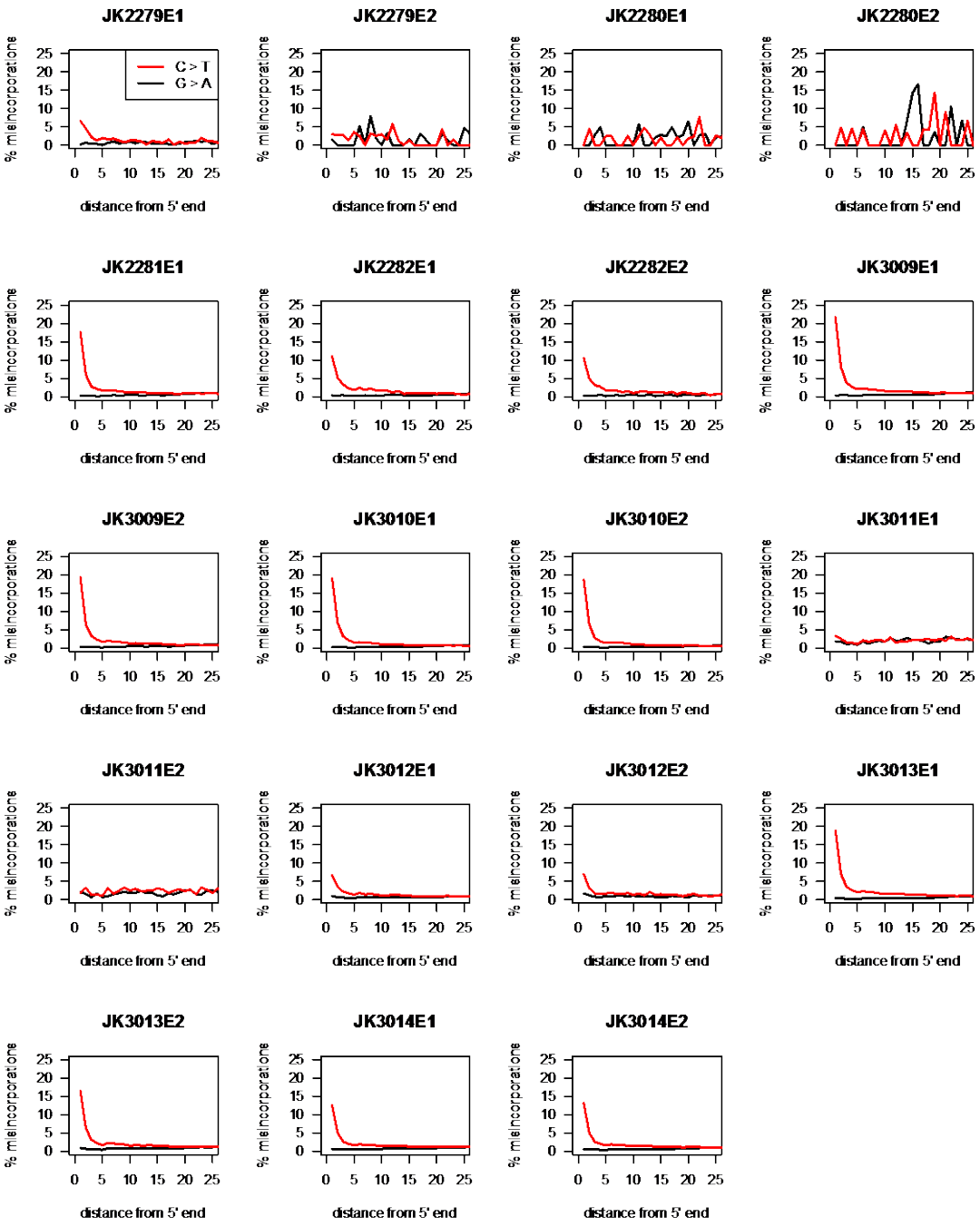
Supplementary Figure 3

Photographs of ancient barley grains used for DNA extraction.



Supplementary Figure 4

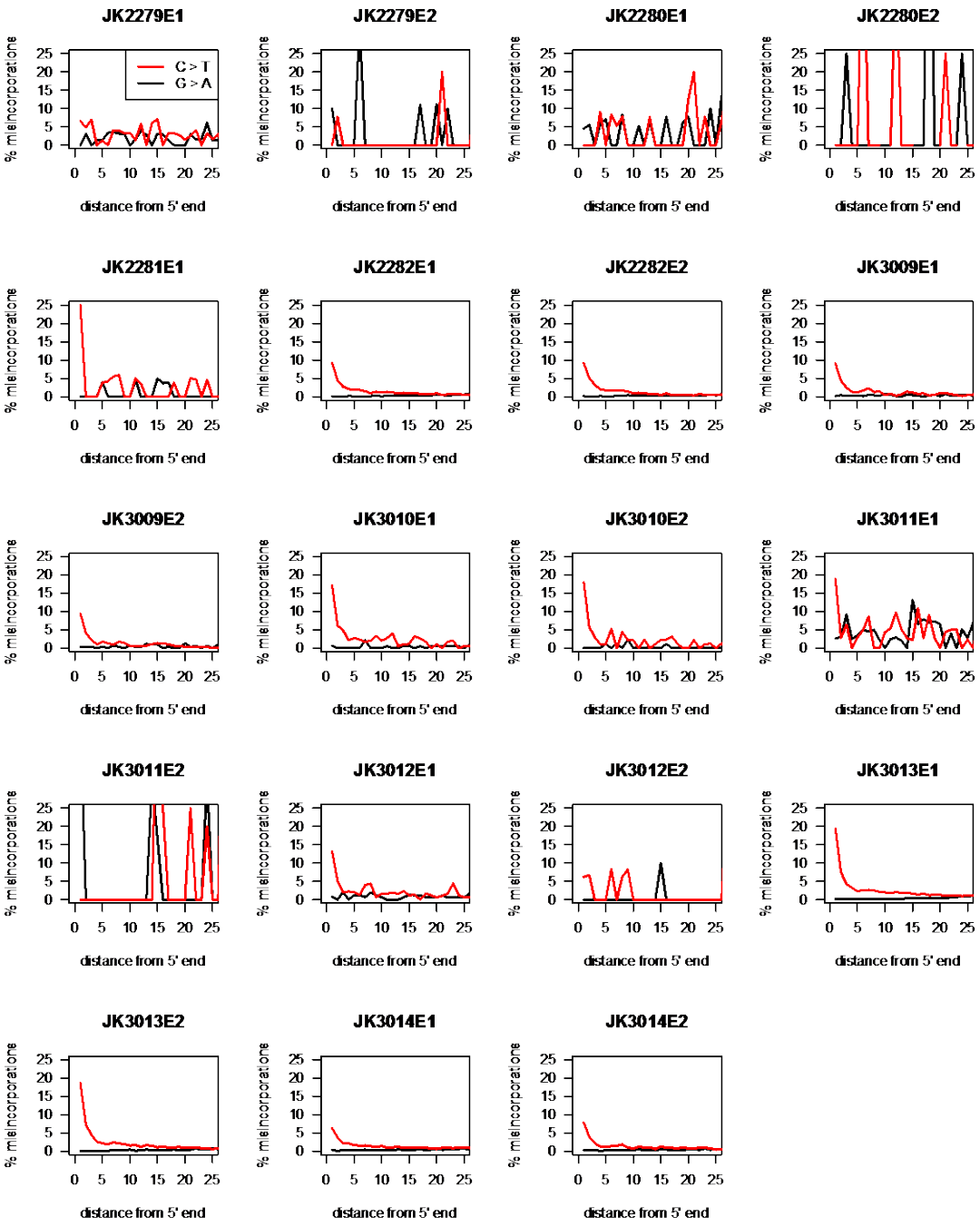
Length distribution of sequence fragments.



Supplementary Figure 5

Nucleotide misincorporation profiles in reads mapped to the whole-genome shotgun assembly of barley cv. Morex.

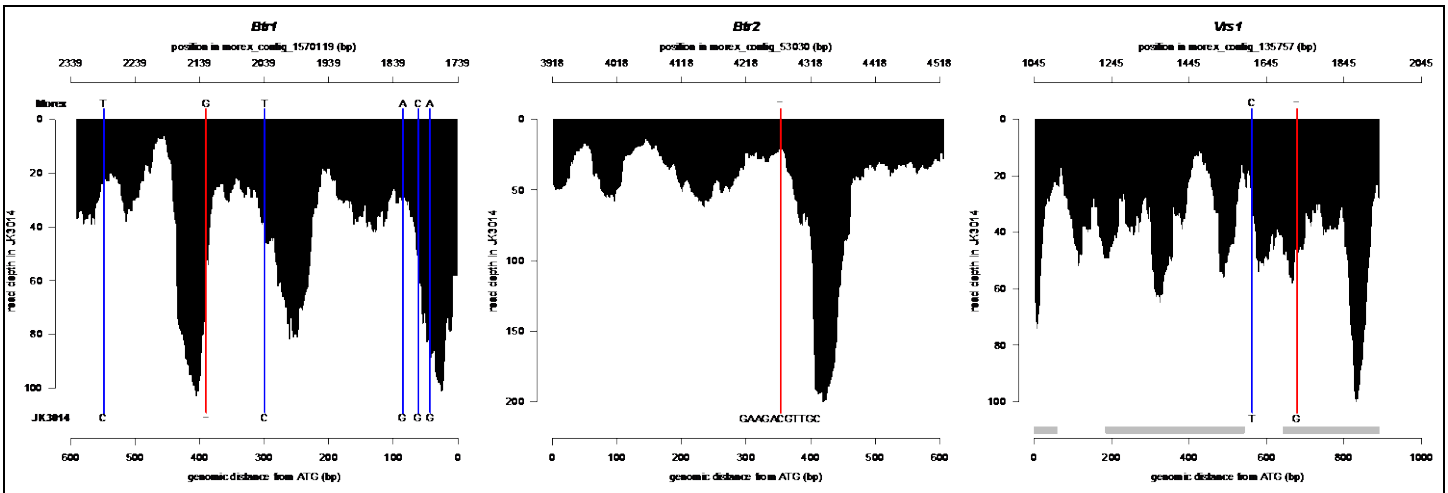
The proportion of C > T misincorporations (red) is compared to the G > A baseline (black).



Supplementary Figure 6

Nucleotide misincorporation profiles in reads mapped to the barley chloroplast genome assembly.

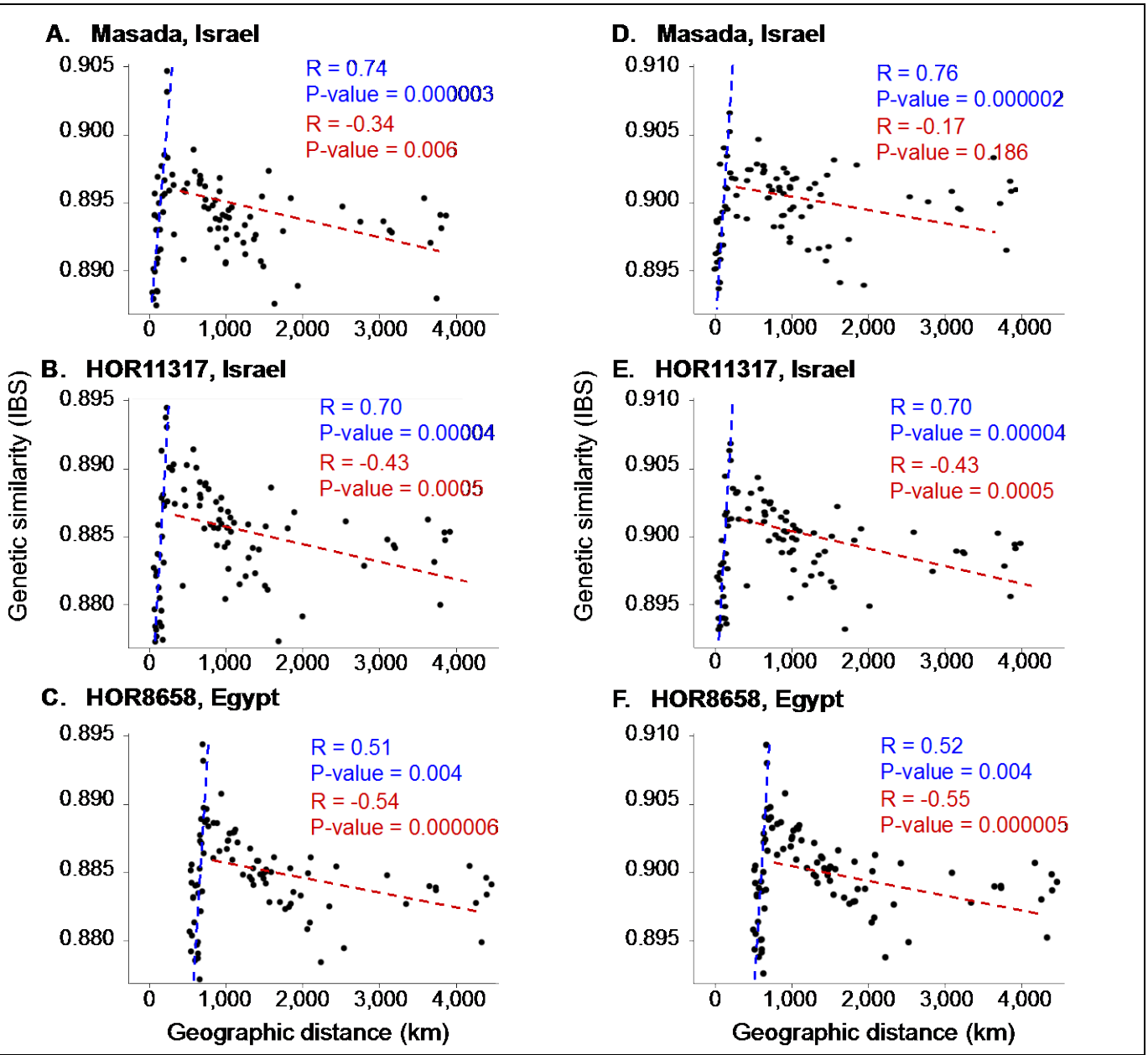
The proportion of C > T substitutions (red) is compared to the G > A baseline (black).



Supplementary Figure 7

Haplotype of *Btr1*, *Btr2* and *Vrs1*

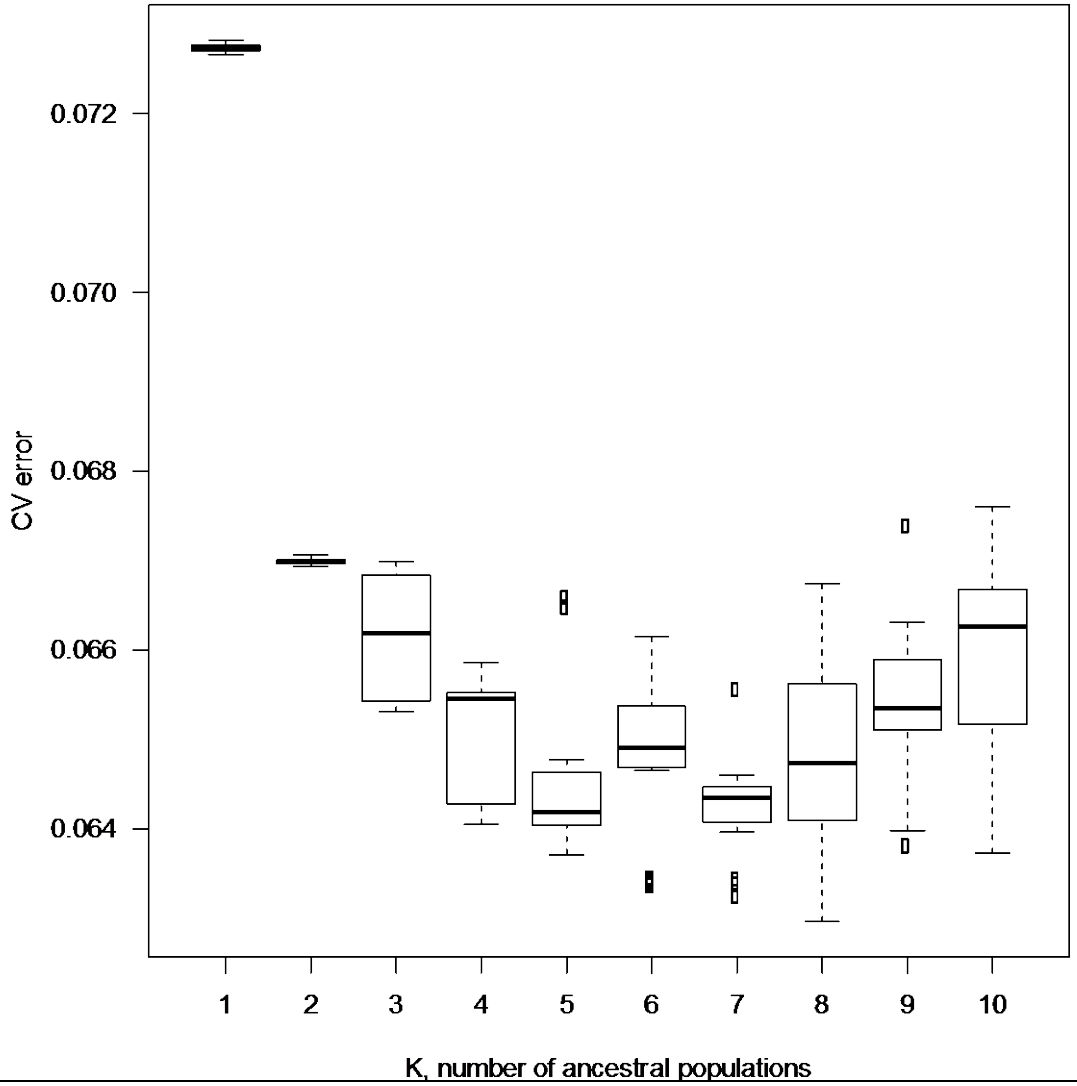
Contigs of the Morex WGS assembly harboring these genes were identified by BLAST searches. The read depth in the deeply sequenced sample JK3014 is shown in black. The position in the respective WGS contigs is indicated in the upper axis. The distance (in the genomic sequence) from the start codon of each gene is shown in the bottom axis. *Btr1* and *Btr2* are single-exon genes. Gray bars indicate the positions of the three exons of *Vrs1*. The positions of SNPs are highlighted by vertical lines. The Morex allele is shown above, and the JK3014 allele below the lines. SNPs with a previously reported functional effect are shown in red. JK3014 carries a loss-of-function allele of *Btr1*, while *Btr2* and *Vrs1* are wild-type alleles. The coding sequence of *Btr1* and *Btr2* is identical to the haplotype of cv. Haruna Nijo as reported by Komatsuda et al., 2015 (NCBI GenBank accession KR813337.1). The sequence of *Vrs1* matches the *Vrs1.b2* allele as designated by Komatsuda et al., 2007.



Supplementary Figure 8

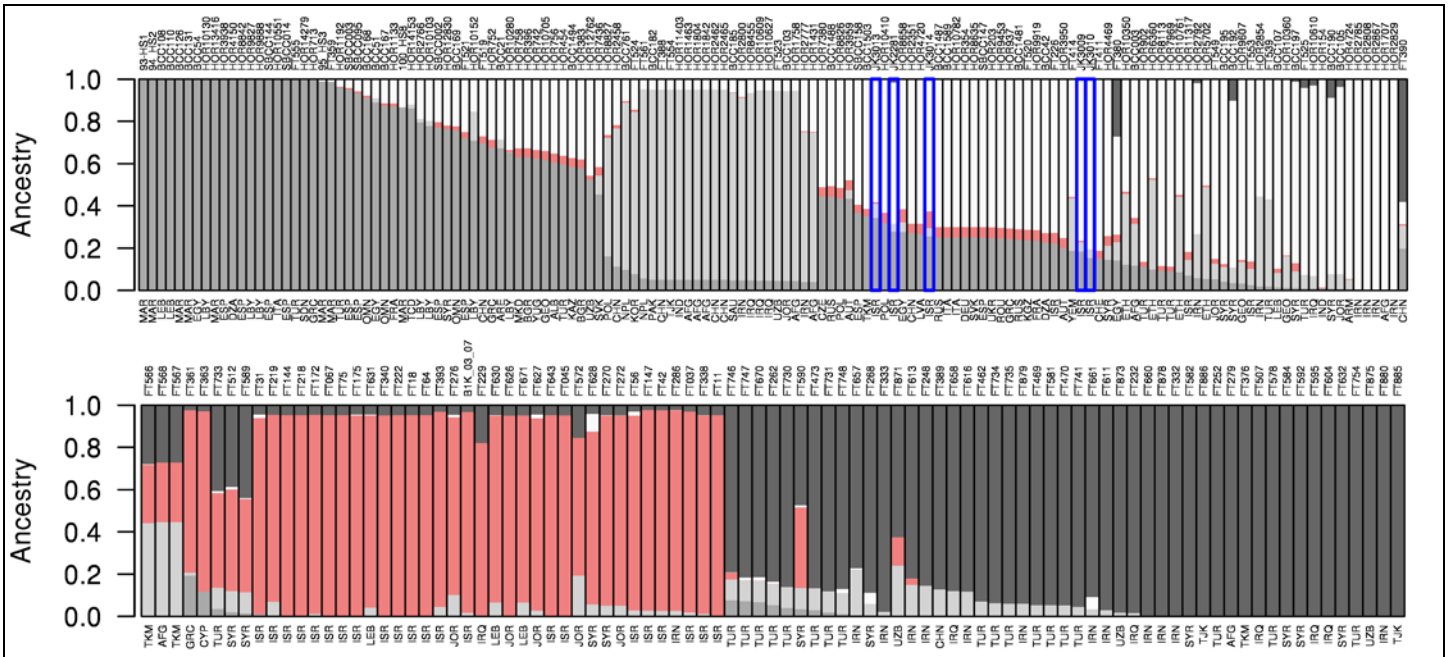
Relationship between genetic similarity and geographic distance

Scatter plot of genetic similarity and geographic distance between 91 extant wild barley accessions sampled across the Fertile Crescent and (A) archaeological sample JK3014 found at Yoram Cave and sequenced to higher depth using all SNPs, (B) two-row cultivated landrace from Israel, (C) a two-rowed cultivated landrace from Egypt, (D) the ancient sample JK3014 found at Yoram Cave and sequenced to higher depth excluding transition SNPs, (E) two-rowed cultivated landraces from Israel excluding transition SNPs, and (F) two-rowed cultivated landraces from Egypt excluding transition SNPs. The geographic position attributed to each sample is: A, D (31.3141 N, 35.353 E); B,E (31.7156 N, 35.1871 E); C,F (31.193 N 29.904 E). Correlation coefficients and P-values for the geographically proximate and distant subsets are indicated in blue and red, respectively.



Supplementary Figure 9

Cross-validation error of ADMIXTURE analysis. Box plots were calculated from 20 replicate runs for each K.



Supplementary Figure 10

ADMIXTURE analysis for K=5 for domesticated (top) and wild (bottom) samples.

Colors in both panels correspond to the same ancestral populations. Sample names and countries of origin are indicated above and below the plots, respectively. Ancient samples are highlighted by blue borders.

Supplementary Table 1. Sequencing statistics of ancient barley DNA without UDG treatment. Read statistics after UDG treatment are provided in Table 1.

Sample	C14 age[1]	Cal age[2]	Extraction	Raw reads	Merged reads	% mapped	% unique	% damaged
JK2279	5065 ± 27	3942-3802	1	8,208,520	3,548,936	31.5%	18.0%	6.6%
JK2279	5065 ± 27	3942-3802	2	7,809,034	3,240,281	2.8%	27.2%	3.0%
JK2280	NA	NA	1	7,410,426	3,192,474	18.6%	8.4%	4.4%
JK2280	NA	NA	2	8,537,472	3,150,021	4.7%	15.0%	4.8%
JK2281	5290 ± 27	4226-4047	1	7,257,940	3,210,243	56.9%	55.8%	17.7%
JK2282	5275 ± 27	4225-4006	1	9,601,504	4,513,912	95.9%	9.8%	11.1%
JK2282	5275 ± 27	4225-4006	2	8,401,768	3,772,270	75.7%	10.4%	10.5%
JK3009	5034 ± 36	3939-3775	1	18,769,156	8,976,322	67.5%	77.5%	21.8%
JK3009	5034 ± 36	3939-3775	2	15,308,126	7,037,725	33.2%	68.6%	19.5%
JK3010	5032 ± 37	3940-3773	1	11,237,658	5,427,102	71.6%	78.1%	19.2%
JK3010	5032 ± 37	3940-3773	2	17,609,772	8,071,918	37.8%	58.3%	18.8%
JK3011	5115 ± 37	3968-3812	1	19,100,026	8,175,478	0.7%	54.4%	3.4%
JK3011	5115 ± 37	3968-3812	2	15,648,476	6,838,637	0.4%	58.9%	1.9%
JK3012	4987 ± 37	3886-3707	1	17,736,280	7,985,422	42.6%	12.3%	6.6%
JK3012	4987 ± 37	3886-3707	2	20,000,020	8,788,418	2.8%	24.4%	7.0%
JK3013	5227 ± 37	4048-3976	1	19,482,948	9,219,005	60.7%	50.2%	18.9%
JK3013	5227 ± 37	4048-3976	2	21,479,860	9,742,769	11.7%	48.9%	16.5%
JK3014	4988 ± 36	3886-3708	1	20,069,866	9,681,200	87.8%	81.9%	12.5%
JK3014	4988 ± 36	3886-3708	2	20,901,314	9,939,663	73.3%	74.4%	13.1%

[1]Uncalibrated radiocarbon years before the present; [2]Calibrated age, calendar years before Christ

Supplementary Table 2: Rare allele sharing between ancient and modern samples. The table lists extant samples that share at least 30 rare alleles (minor allele count ≤ 5) with the ancient samples. Only transversion variants were considered for this analysis.

sample	country of origin	domestication status	number of shared alleles
HOR4469	Syria	landrace 2-row	89
HOR2766	Libya	landrace 6-row	63
FT549	Jordan	landrace 2-row	56
HOR8658	Egypt	landrace 2-row	56
FT75	Israel	spontaneum	45
HOR10280	Libya	landrace 2-row	42
BCC107	Lebanon	landrace 2-row	40
FT037	Israel	spontaneum	38
FT31	Israel	spontaneum	38
FT340	Israel	spontaneum	38
100_HS8	Morocco	landrace 6-row	37
HOR10103	Libya	landrace 6-row	36
FT218	Israel	spontaneum	35
HOR11317	Israel	landrace 2-row	35
HOR14153	Chad	landrace 6-row	35
FT276	Jordan	spontaneum	34
FT631	Lebanon	spontaneum	32
FT393	Israel	spontaneum	30

Supplementary Table 4: D statistics for different comparisons between wild and domesticated barleys from the Levant, European landraces and ancient samples. Standard errors and Z scores (D/SE) are given in columns SE and Z.

				all SNPs			transversions only		
P1	**P2**	**P3**	**Outgroup**	**D**	**SE**	**Z**	**D**	**SE**	**Z**
Levantine landraces	European landraces	Levant wild	H. pubiflorum	0.037	0.004328	8.546	0.0342	0.005323	6.418
Levantine landraces	European landraces	JK2281	H. pubiflorum	0.1076	0.019226	5.594	0.1007	0.021775	4.623
Levantine landraces	European landraces	JK3009	H. pubiflorum	0.0898	0.015975	5.62	0.0499	0.01801	2.768
Levantine landraces	European landraces	JK3010	H. pubiflorum	0.1671	0.016772	9.963	0.1534	0.018811	8.155
Levantine landraces	European landraces	JK3013	H. pubiflorum	0.0955	0.027177	3.516	0.0986	0.037957	2.598
Levantine landraces	European landraces	JK3014	H. pubiflorum	0.1864	0.015474	12.045	0.1854	0.016067	11.538
JK2281	European landraces	Levant wild	H. pubiflorum	0.061	0.01632	3.737	0.0116	0.021853	0.531
JK2281	Levantine landraces	Levant wild	H. pubiflorum	0.0265	0.01447	1.83	0.0018	0.020225	0.088
JK3009	European landraces	Levant wild	H. pubiflorum	0.0562	0.0166	3.386	0.0662	0.023755	2.789
JK3009	Levantine landraces	Levant wild	H. pubiflorum	0.019	0.014787	1.288	0.0617	0.022587	2.731
JK3010	European landraces	Levant wild	H. pubiflorum	0.0366	0.010696	3.424	0.0227	0.015691	1.445
JK3010	Levantine landraces	Levant wild	H. pubiflorum	0.0031	0.009611	0.325	-0.0068	0.014167	-0.48
JK3013	European landraces	Levant wild	H. pubiflorum	0.0883	0.028889	3.058	0.1282	0.056598	2.266
JK3013	Levantine landraces	Levant wild	H. pubiflorum	0.0809	0.029933	2.704	0.1068	0.055261	1.933
JK3014	European landraces	Levant wild	H. pubiflorum	0.0413	0.007807	5.288	0.0385	0.010813	3.558
JK3014	Levantine landraces	Levant wild	H. pubiflorum	0.0077	0.007189	1.066	0.0068	0.009839	0.696