nature plants

LETTERS

PUBLISHED: 8 JANUARY 2015 | ARTICLE NUMBER: 14003 | DOI: 10.1038/NPLANTS.2014.3

The origin and evolution of maize in the Southwestern United States

Rute R. da Fonseca[1,2*], Bruce D. Smith[3], Nathan Wales[1], Enrico Cappellini[1], Pontus Skoglund[4], Matteo Fumagalli[5], José Alfredo Samaniego[1], Christian Carøe[1], María C. Ávila-Arcos[1,6], David E. Hufnagel[7], Thorfinn Sand Korneliussen[1], Filipe Garrett Vieira[1,5], Mattias Jakobsson[8,9], Bernardo Arriaza[10], Eske Willerslev[1], Rasmus Nielsen[1,11], Matthew B. Hufford[7], Anders Albrechtsen[2], Jeffrey Ross-Ibarra[12] and M. Thomas P. Gilbert[1,13*]

The origin of maize (*Zea mays mays*) in the US Southwest remains contentious, with conflicting archaeological data supporting either coastal[1–4] or highland[5,6] routes of diffusion of maize into the United States. Furthermore, the genetics of adaptation to the new environmental and cultural context of the Southwest is largely uncharacterized[7]. To address these issues, we compared nuclear DNA from 32 archaeological maize samples spanning 6,000 years of evolution to modern landraces. We found that the initial diffusion of maize into the Southwest about 4,000 years ago is likely to have occurred along a highland route, followed by gene flow from a lowland coastal maize beginning at least 2,000 years ago. Our population genetic analysis also enabled us to differentiate selection during domestication for adaptation to the climatic and cultural environment of the Southwest, identifying adaptation loci relevant to drought tolerance and sugar content.

Documenting ancient diffusion routes of domesticates and how they were modified when introduced into new regions has long been a challenge. For example, hybridization and gene flow have long confounded attempts to understand the origins of either indica rice[8] in the Indian subcontinent or maize in southern Mexico[9]. The origin and adaptation of maize in the US Southwest is a similarly difficult case. Following its initial domestication from the wild grass teosinte in southern Mexico[10,11], maize diffused throughout the Americas, spreading through much of the continental United States after its introduction to the Southwest around 4,100 calendar years before present (BP)[7]. There has been considerable debate about the arrival of maize into the Southwest, however, as early archaeological samples suggested a highland route[5,6], whereas more recent samples[1,2] and morphological similarity to extant Mexican maize support a lowland, Pacific coast route[3,4]. And while temporal variation in Southwest maize cob morphology has been described[2], the genetic changes responsible for adaptation to the Southwest environment during the last 4,000 years are still uncharacterized.

In order to resolve questions about the diffusion of maize into the Southwest as well as to track genetic changes in Southwest maize through time, we sampled DNA from archaeological specimens dating to around 4,000–3,000, 2,000 and 750 BP (SW3K, SW2K and SW750 hereafter), as well as four ancient Mexican samples dating to around 5,910 BP, 5,280 BP and 1,410 BP (Table 1) and a single modern open-pollinated highland Mexican maize accession (Supplementary Table 5). We generated sequence data from ancient samples using a hybridization target capture approach that was enriched for the exons of 348 genes (depth of covered sites ~10× on target and ~2× elsewhere; selection criteria are in Supplementary Tables 8, 9 and 11); our modern highland sample was sequenced using a whole-genome shotgun approach. To these data we added published sequence data from an additional ancient sample from Mexico[12] and modern samples of teosinte subspecies, *Zea mays parviglumis* and *Zea mays mexicana*, as well as Southwest and Mexican maize[13].

Comparison of shared derived alleles between ancient Southwest samples and the Mexican highland landrace Palomero de Jalisco or the Mexican lowland landrace Chapalote using D statistics[14] argues for a highland origin of the earliest Southwest maize (SW3K; Fig. 1a), consistent with low-density single nucleotide polymorphism data[15] from a sample of more than 2,000 modern maize landraces and teosinte (Supplementary Fig. 6). In contrast, values of D in SW2K support gene flow from Chapalote (Fig. 1a). TreeMix[16] also identifies introgression from lowland maize to the SW2K population (Fig 1b) and agrees with previous evidence for introgression from the teosinte *Z. mays. mexicana* into Mexican highland landraces[17]. Finally, admixture analysis (Fig. 1c, and Supplementary Fig. 5) reveals evidence of teosinte admixture in all ancient Southwest maize. As there is no history of teosinte in the Southwest, this is consistent with a highland origin. Assignment to the group that includes the lowland samples Chapalote and Reventador, however, increases in the SW2K and SW750 samples; we interpret the lack of observed admixture with teosinte or Mexican maize in the extant Southwest Santo Domingo landrace (USA17) to be a result of recent extensive genetic exchange with other American landraces (Supplementary Fig. 5). Together, these results argue for a complex origin of Southwest maize, originally

[1]Centre for GeoGenetics, University of Copenhagen, 1350 Copenhagen, Denmark. [2]The Bioinformatics Centre, University of Copenhagen, 2200 Copenhagen, Denmark. [3]Program in Human Ecology and Archaeobiology, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA. [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. [6]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA. [7]Department of Ecology, Evolution, & Organismal Biology, Iowa State University, 50011, USA. [8]Department of Evolutionary Biology, Uppsala University, Uppsala 752 36, Sweden. [9]Science for Life Laboratory, Uppsala University, Uppsala 752 36, Sweden. [10]Instituto de Alta Investigación, Universidad de Tarapacá, 15101 Arica, Chile. [11]Department of Integrative Biology and Statistics, University of California, Berkeley, California 94720-3140, USA. [12]Department of Plant Sciences, Center for Population Biology and Genome Center, University of California, Davis, California 95616, USA. [13]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia, 6102, Australia. *e-mail: rute.r.da.fonseca@gmail.com; mtpgilbert@gmail.com

Table 1 | Identification and summary statistics of the ancient samples sequenced in this study.

Age group	Type of analyses*	IDS	Intercept of radiocarbon age with calibration curve years BP†	Cob morphology shape (pineapple, P; cylinder, C), row number, cob diameter	Site	Retained nucleotides	Average depth (targets)
SW3K	a,d	SW443	2,780		McEuen Cave, USA	8,230,593	13
	t,a,d	SW4Ba	3,390		Bat Cave, USA	17,621,611	9
SW2K	a,d,s	SW207	1,860	P, 12 row, 2.0 cm	Tularosa Cave, USA	5,870,362	11
	s	SW256	~1,850–1,750	P, 12 row, 2.5 cm		3,768,546	9
	s	SW261	~1,850–1,750	P, 10 row, 1.9 cm		4,613,152	9
	a,d,s	SW264	1,820	P, 12 row 1.8 cm		15,134,398	14
	s	SW278	~1,850–1,750	P, 12 row, 2.2 cm		3,431,137	10
	a,d,s	SW280	~1,850–1,750	P, 10 row, 2.1 cm		5,209,183	6
	s	SW283	1,860; 1,850; 1,830	P, 12 row, 2.2 cm		5,642,954	3
	s	SW288	~1,850–1,750	P, 12 row, 2.3 cm		148,791	1
	s	SW296	~1,850–1,750	P, 10 row, 1.9 cm		2,072,254	5
	t,a,d,s	SW298	1,770; 1,760; 1,740	P, 12 row, 2.0 cm		80,568,726	10
SW750	s	SW105	670	C, 10 row, 1.5 cm	Tularosa Cave, USA	2,058,626	4
	a,d,s	SW107	~700–900	C, 8 row, 1.3 cm		34,929,483	20
	a,d,s	SW109	~700–900	C, 8 row, 1.3 cm		12,364,145	15
	a,d,s	SW110	~700–900	C, 8 row, 1.6 cm		35,088,565	17
	a,d,s	SW111	~700–900	C, 8 row, 1.5 cm		29,640,515	19
	a,d,s	SW112	~700–900	C, 8 row, 1.4 cm		22,887,209	16
	s	SW118	~700–900	C, 8 row, 1.2 cm		3,855,808	4
	a,d,s	SW121	790	C, 10 row, 1.5 cm		29,736,402	7
	a,d,s	SW124	~700–900	C, 8 row, 1.3 cm		33,518,448	18
	a,d,s	SW132	740	C, 8 row, 1.4 cm		17,131,288	18
	t,a,d,s	SW146	690	C, 8 row, 1.3 cm		111,329,149	12
	s	SW1b9	740	C, 8 row, 1.5 cm		68,634	2
	a	SW1AX	670		Turkey House Ruin, USA	59,526,622	25
	a	TH563	5,910	4 ranks, 8 rows, 1.2 cm	Tehuacan Caves, Mexico	9,544,881	3
	t,a	TH564	5,280; 5,160; 5,140; 5,100	4 ranks, 8 rows, 1.1 cm		10,791,297	5
	a	TH157	1,410	8 rows, 1.5 cm		18,126,654	2
	a	AR14B			Arica, Chile	5,328,366	16
	a	AR1A9				11,261,584	11
	a	AR1A8				286,639,854	24
	a	AR171				159,400,189	21

* t, TreeMix (Fig. 1a); a, NGSadmix (Fig. 1b, and Supplementary Fig. 5); d, D-statistics (Fig. 1c, Supplementary Fig. 12); s, selection tests (Figs 2 and 3, and Supplementary Fig. 10).
†INTCAL09 calibration curve.

entering the United States via a highland route by 4,000 BP and subsequently receiving gene flow from lowland maize via the Pacific coastal corridor starting around 2,000 BP.

Maize was faced with a number of environmental challenges upon arrival in the Southwest, from extreme aridity to new dietary preferences[7]. Our population-level samples corresponding to temporally distinct occupations of the same cave site (Tularosa cave: SW2K, $n = 10$; SW750, $n = 12$), combined with published genomic data of the maize progenitor *Z. m. parviglumis* (Supplementary Table 4), allow us to distinguish evidence for these more recent adaptations from selection that occurred during maize domestication. We first used the population branch statistic PBS[18] to identify genes with the highest dissimilarity between teosinte and our ancient Southwest landraces (Fig. 2a). These genes were likely to be early targets of maize domestication that preceded arrival in the Southwest. Many of these genes also show a very negative Tajima's D, consistent with the effects of strong selection (Fig. 2a), and seven of the top ten genes (Supplementary Table 1) are located in previously identified selected regions[19]. The top gene, *zagl1*, corresponds to a MADS-box transcription factor associated with shattering, a key domestication feature strongly selected for by human harvesting[20]. Several other genes are also well known for their roles in domestication: *ba1* has a major role in the architecture of maize[21], *zcn1* and *gi* are associated with the regulation of flowering[20,22] and *tga1* controls the change from encased to exposed kernels[23].

Comparison of the ancient maize population samples from Tularosa cave then let us assess changes between 2,000 and 750 years BP, a period of ongoing adaptation to the Southwest. Median values of Tajima's D in the SW750 population are higher than in the SW2K (Supplementary Fig. 8 and Supplementary Table 2), consistent with model-based estimates suggesting a smaller effective population (Supplementary Fig. 9). Nonetheless, we find several genes showing evidence of selection. The top PBS outlier in the SW750 population is a dehydration-responsive element-binding protein shown to be upregulated as much as 50-fold in maize roots under drought conditions[24], perhaps a signature of adaptation to arid Southwest conditions (Supplementary Fig. 10). Analysis of genes in the starch biosynthesis pathway provides perhaps the best example of the power of our population-sampling approach. While the reduction of diversity at *ae1* is seen in all Southwest maize, consistent with selection during domestication, diversity at *sugary1* (*su1*) is reduced more than 60% between the SW2K and SW750 populations (Fig. 3). *su1* also shows an elevated PBS and a negative Tajima's D (Fig. 2) consistent with strong selection. The timing of selection on *su1* appears to correlate with a shift towards larger cobs and floury kernel endosperm in archaeological maize around 800–1000 AD[2]. Both *ae1* and *su1* affect the structure of amylopectin[25], which is involved in the pasting properties of maize tortillas and porridge[26]. Furthermore, it has been shown that storing non-structural carbohydrates can be beneficial in a drought scenario, consistent with adaptation to the Southwest climate[27]. The *su1* mutation with the highest allele frequency difference between SW2K and modern individuals (Supplementary Fig. 3) is known to cause the partial replacement of starch by sugar in sweetcorn[28]. Several Native American tribes grew sweetcorn

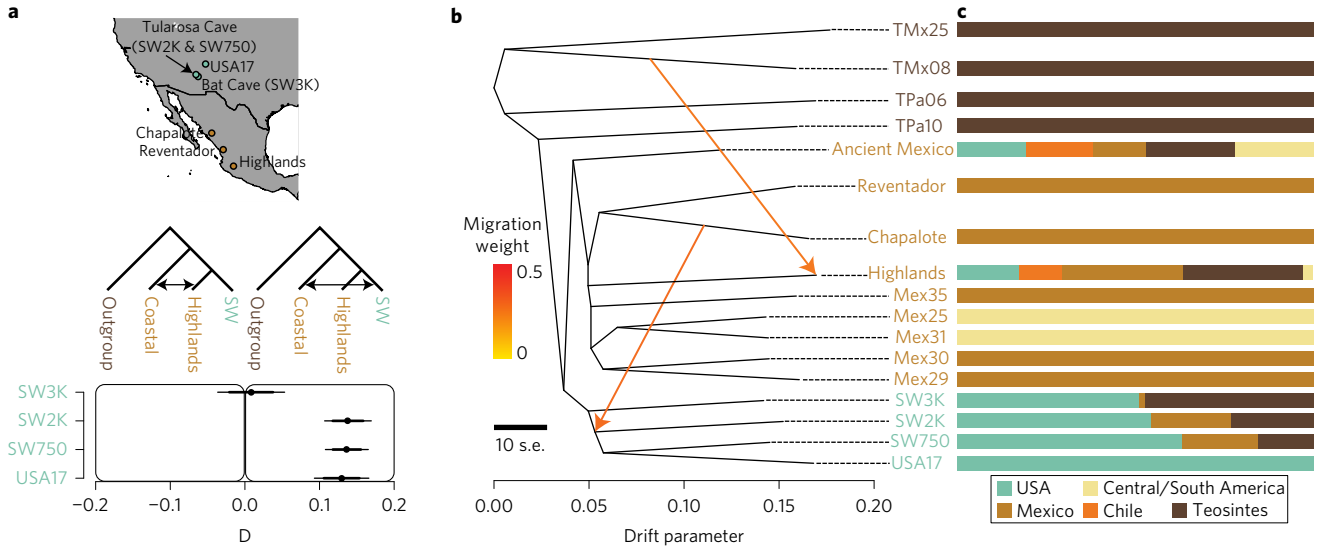

Figure 1 | Origins of the Southwest ancient maize samples. a, SW3K, SW2K and SW750 correspond to Southwest maize from ~3,000, ~2,000 and ~750 BP. The ancient Mexican sample dates to 5,100 BP (TH564). The Mex prefix indicates modern Mexican samples from across Mexico. Coastal lowland (Reventador, Chapalote) and highland (Palomero Toloqueño) landraces are highlighted on the map. Further details are available in Table 1 and Supplementary Tables 4 and 5. Allele frequency-based D-tests suggest an initial highland diffusion route from Mexico to the Southwest of the United States followed by extensive gene flow from the Pacific coast Chapalote race (Supplementary Table 6 and Supplementary Fig. 12); positive values of D indicate gene flow from the coastal varieties into the Southwest maize; thick and thin bars correspond to 2 and 3 standard errors, respectively. **b**, TreeMix maximum likelihood tree depicting the expected signal of gene flow from *Z. m. mexicana* into the highland landraces (also Supplementary Fig. 12) and gene flow from the coastal Chapalote into the SW2K. **c**, A subset of the population structure plot determined by NGSadmix with $K = 5$ (full plot in Supplementary Fig. 5); each individual is represented by a stacked column of the five proportions.

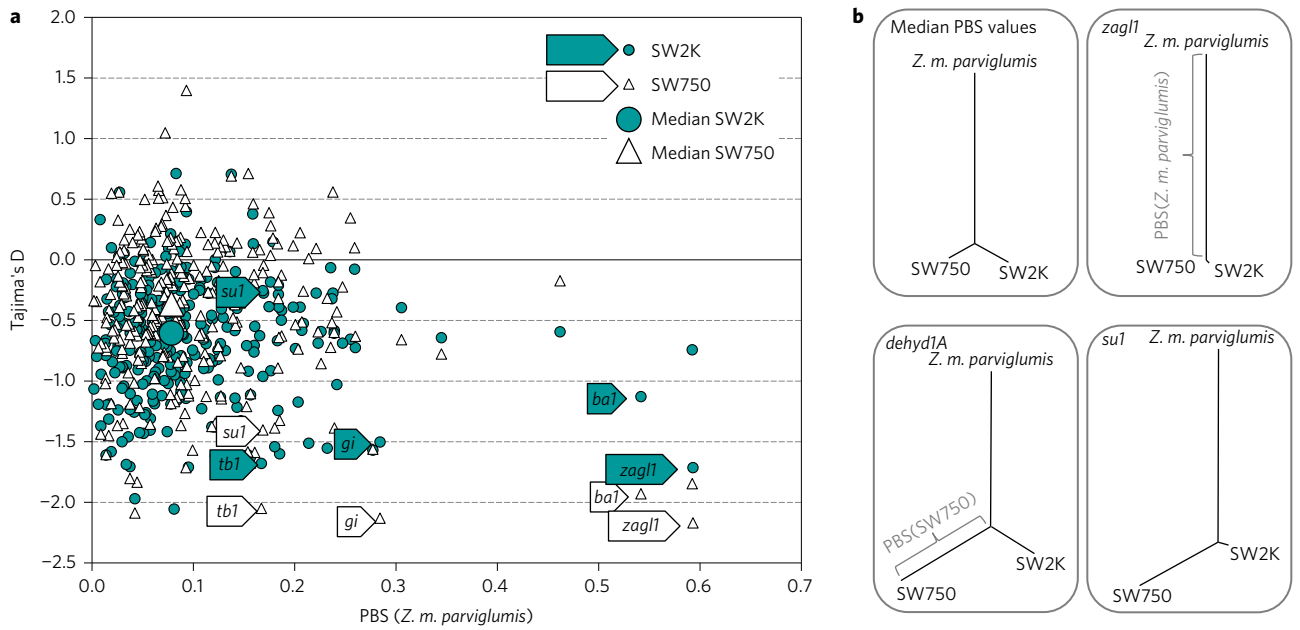

Figure 2 | Potential targets of selection during domestication. a, Tajima's D for the two Southwest populations dated to ~2,000 (coloured dots) and ~750 BP (white triangles) plotted against the PBS distance for *Z. m. parviglumis*. *zagl1* shows the highest dissimilarity between *Z. m. parviglumis* and the ancient Southwest landraces, that is, the largest PBS (*Z. m. parviglumis*). The gene with the lowest Tajima's D value for the SW750 population is also *zagl1*. Genes with major roles in domestication traits are depicted in trapezoids. **b**, Gene trees built using PBS distances. *dehyd1A* is the top outlier for PBS(SW750) (Supplementary Fig. 10) and *su1* displayed the highest decrease in nucleotide diversity between the SW2K and the SW750 populations.

before the arrival of Europeans and the high frequency of a *su1* mutation in Southwest maize could help explain the early appearance and maintenance of sweetcorn varieties by Native Americans.

The study of domestication and early crop evolution has largely been limited to the identification of key phenotypic, morphological and genetic changes between extant crops and their wild relatives. As demonstrated here, the application of new paleogenomic approaches to well-documented temporal sequences of archaeological assemblages opens a new chapter in the study of domestication: it is now possible to move beyond a simple distinction of 'wild' versus 'domesticated'[29,30] and track sequence changes in a wide range of genes over the course of thousands of years.

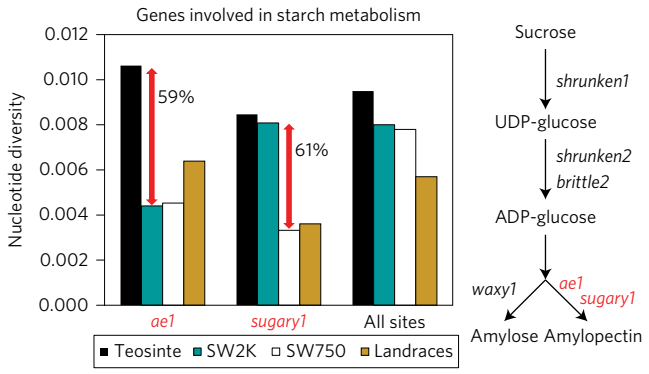

Figure 3 | Timing of selective pressures on genes involved in the starch metabolism. Nucleotide diversity variation for two key elements of the starch metabolism pathway *ae1* and *su1*. Comparison between the two Southwest populations dated to ~2,000 and ~750 BP (Table 1), and modern landraces and teosintes (Supplementary Table 4) plotted against the PBS distance for *Z. m. parviglumis*. There is a steep decrease in nucleotide diversity before 2,000 BP for *ae1*, whereas the reduction in nucleotide diversity (π) for *sugary1* to less than half occurred after 2,000 and before 750 BP.

Methods

Materials. Twenty-five archaeological maize cob samples from the Southwest United States dating from 4,300 to 740 years BP, three from Mexico dating from 5,910 to 1,410 BP, and four ancient Arica samples were obtained from the repositories and individuals listed in Supplementary Table 7 following established policies and procedures for destructive sampling. In addition, previously published sequence data[12] corresponding to an ancient sample from Mexico, was also used (Supplementary Table 7).

With the exception of the Turkey House Ruin sample, all of the archaeological cob samples from the Southwest United States and Mexico were recovered from dry cave contexts, and the Chilean (Arica) samples came from the dry desert coast of South America. All of the archaeological samples were desiccated, uncarbonized and in an excellent state of preservation. The cobs recovered from sites in the Southwest United States fall into two distinct morphological and temporal categories. These two temporally separated and morphologically distinct forms of maize correlate quite closely with the structural analysis groupings based on aDNA. The early southwestern maize, including samples from McEuen and Bat Caves, and from the early occupation at Tularosa Cave (1,850–1,750 BP), variously labelled as 'Chapalote' or 'small cob maize'[4] is a small cob, small kernel form having a thick midsection (1.9–2.5 cm diameter) and tapered ends (Pineapple shape) and 10–12 rows of kernels. The maize from the later occupation at Tularosa Cave (700–900 BP), as well as the Turkey House Ruin sample (670 BP), is a larger cob, larger kernel form, having parallel sides (cylinder shape), eight to ten rows of kernels, and a much smaller diameter than the earlier form (1.3–1.6 cm) (Table 1).

Data for modern samples (maize landraces, *Z. m. parviglumis* and tripsacum) were obtained from the HapMap2 set and downloaded from Panzea's website (www.panzea.org). Additionally, we generated shotgun data from an individual from the highlands of northern Mexico. Information about modern samples can be found in Supplementary Tables 4 and 5.

Reads mapping to the target regions were extracted from HapMap2 bam files and remapped and filtered in the same way as the ancient maize samples (Supplementary Table 4).

Target selection and bait design. A total of 348 genes were targeted: 318 genes were chosen because their similarity to sorghum was between 70% and 95% (a conservation level that is indicative of high functional relevance, and avoiding genes that are potentially invariable in maize), and they had some kind of functional annotation (Supplementary Table 9). The other 30 genes have been suggested to have an important role in traits selected during maize domestication[20,22,31,32] (Supplementary Table 8). Maize gene sequences were downloaded from ENSEMBL (annotation version ZmB73_5b). An extra 120 base pairs (bp) flanking region was added to each bait; 120 bp probes were designed with 20 bp tiling, resulting in a final number of 53,063 probes.

aDNA extraction. Archaeological maize remains were processed at a dedicated clean laboratory facility at the Centre for GeoGenetics, University of Copenhagen. All steps prior to library amplification were conducted in an isolated laboratory that utilizes nightly UV radiation and air filtration systems to avoid contamination, thereby conforming to the requirements of aDNA research[33].

To minimize modern DNA contamination, maize kernels were washed in 5% commercial bleach solution (NaClO) and rinsed in molecular grade water before extraction. Maize cobs could not be washed with bleach because they would absorb the solution, potentially leading to degradation of endogenous DNA. Instead, sterile scalpels were used to remove the external surface of cobs to expose material with presumably lower levels of contamination. Maize kernels were pulverized using a sterilized hammer and maize cob samples were sliced into fine slivers using a sterile scalpel. Either one kernel or ~0.1 g of cob shavings were used for an extraction.

DNA extractions were conducted according to an established protocol originally designed for extracting DNA from ancient hair samples[34], but which has also been applied to ancient grape pips and maize[12,35]. Recent testing has demonstrated the method generally outperforms other extraction techniques for a broad range of archaeobotanical remains, including maize cobs and kernels[36]. Pulverized samples were placed in 750 µl of extraction buffer (850 µl for cobs), as described previously[12], and incubated overnight at 55 °C. The following day, a phenol and chloroform extraction was conducted, followed by purification in Qiagen MinElute silica spin columns.

Library construction and amplification. DNA extracts were converted to Illumina-compatible DNA libraries using NEBNext library building kits for second-generation sequencing (New England Biolabs, Ipswich, MA; catalogue numbers: E6070L, E6090S). Libraries were prepared according to manufacturer's directions, except that no DNA size selection or fragmentation steps were undertaken.

Libraries were amplified with either Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA) or AmpliTaq Gold (Life Technologies, Carlsbad, CA). Libraries constructed in the later phases of the project were always first amplified using AmpliTaq Gold to incorporate molecules with damaged nucleotides. Apparent C to T transitions at the 5′ and 3′ ends of aDNA molecules resulting from the paring of adenine with deaminated cytosine (uracil) can thereby be used to investigate for characteristic aDNA damage patterns and help authenticate the presence of endogenous aDNA[37]. Nonetheless, libraries amplified during the earlier phases of the project were overall similar to those amplified with AmpliTaq Gold, and therefore should not lead to biases in analyses. Libraries were amplified 12–18 initial cycles, depending on the sample.

To reach DNA concentrations required for in-solution hybridization captures, libraries were amplified again, using a subset of the first amplification. These second amplifications were exclusively done with Phusion High-Fidelity PCR Master Mix because the polymerase replicates DNA with higher fidelity than AmpliTaq Gold, thereby reducing erroneous sequence polymorphisms. The second amplifications were conducted using 10–18 cycles. When necessary, libraries were size selected on a 2% agarose gel to remove adapter dimers. Libraries were characterized on a Qubit 2.0 fluorometer (Life Technologies) and Agilent 2100 Bioanalyzer (Santa Clara, CA).

Targeted capture. Enrichment of relevant genetic loci[38] was conducted using a custom-designed MYBait-3 target enrichment kit (MYcroarry, Ann Arbor, MI; 120 bp length RNA baits). The manufacturer of the kit recommends 100–500 ng of amplified library to be used for a capture, and all were performed at the higher end of this range, generally 300–500 ng of DNA. Libraries were hybridized for 24 hours at 65 °C in an Applied Biosystems Veriti thermal cycler (Life Technologies) using a heated lid to prevent condensation. Following hybridization with RNA probes, the samples were processed according to the manufacturer's protocol. Post-capture amplification was done with Phusion High-Fidelity PCR Master Mix, using 12–18 cycles. Samples were sequenced on an HiSeq 2000 in the single read 100 bp mode, three samples per lane.

This procedure resulted in a depth within the target regions of around 10×, a fivefold increase relative to other sites in the genome (Table 1).

See the Supplementary Information for more Methods.

Received 6 June 2014; accepted 8 October 2014; published 8 January 2015; corrected 14 January 2015

References

1. Gregory, D. *Excavations in the Santa Cruz River Floodplain* (Center for Desert Archaeology, 1999).
2. Huckell, L. in *Histories of Maize* (eds Staller, J., Tykot, R. & Benz, B. F.) 97–106 (Elsevier, 2006).
3. Cutler, H. in *Mogollon Cultural Continuity and Change: The Stratigraphic Analysis of Tularosa and Cordova Caves* (eds Martin, P., Rinaldo, J., Bluhm, E., Cutler, H. & Grange, R.) 461–479 (Chicago Natural History Museum, 1952).
4. González, J. in *Corn and Culture in the Prehistoric New World* (eds Johannessen, S. & Hastorf, C.) 135–157 (Westview, 1994).
5. Haury, E. in *Courses Toward Urban Life* (eds Braidwood, R. & Willey, G.) 106–131 (Aldine, 1962).
6. Ford, R. in *Prehistoric Food Production in North America. Museum of Anthropology Anthropological Papers* (ed. Ford, R.) 341–364 (University of Michigan, 1985).

7. Merrill, W. L. *et al.* The diffusion of maize to the southwestern United States and its impact. *Proc. Natl Acad. Sci. USA* **106,** 21019–21026 (2009).
8. Gross, B. L. & Zhao, Z. Archaeological and genetic insights into the origins of domesticated rice. *Proc. Natl Acad. Sci. USA* **111,** 6190–6197 (2014).
9. Van Heerwaarden, J. *et al.* Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl Acad. Sci. USA* **108,** 1088–1092 (2011).
10. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA* **106,** 5019–5024 (2009).
11. Matsuoka, Y. *et al.* A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl Acad. Sci. USA* **99,** 6080–6084 (2002).
12. Avila-Arcos, M. C. *et al.* Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* **1,** 74 (2011).
13. Chia, J., Song, C., Bradbury, P. & Costich, D. Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genet.* **44,** 803–807 (2012).
14. Patterson, N. J. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
15. Fang, Z. *et al.* Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191,** 883–894 (2012).
16. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8,** e1002967 (2012).
17. Hufford, M. B. *et al.* The genomic signature of crop-wild introgression in maize. *PLoS Genet.* **9,** e1003477 (2013).
18. Li, Y. *et al.* Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nature Genet.* **42,** 969–972 (2010).
19. Hufford, M. B. *et al.* Comparative population genomics of maize domestication and improvement. *Nature Genet.* **44,** 808–811 (2012).
20. Weber, A. L. *et al.* The genetic architecture of complex traits in teosinte (*Zea mays* ssp. parviglumis): new evidence from association mapping. *Genetics* **180,** 1221–1232 (2008).
21. Gallavotti, A. *et al.* The role of barren stalk1 in the architecture of maize. *Nature* **432,** 630–635 (2004).
22. Weber, A. *et al.* Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. parviglumis). *Genetics* **177,** 2349–2359 (2007).
23. Wang, H. *et al.* The origin of the naked grains of maize. *Nature* **436,** 714–719 (2005).
24. Liu, S. *et al.* Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. *PLoS Genet.* **9,** e1003790 (2013).
25. Wilson, L. M. *et al.* Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* **16,** 2719–2733 (2004).
26. Schultz, J. A. & Juvik, J. A. Current models for starch synthesis and the sugary enhancer1 (se1) mutation in *Zea mays*. *Plant Physiol. Biochem.* **42,** 457–464 (2004).
27. Brien, M. J. O. *et al.* Drought survival of tropical tree seedlings enhanced by non-structural carbohydrate levels. *Nature Clim. Change* **4,** 710–714 (2014).
28. Dinges, J. R., Colleoni, C., Myers, A. M. & James, M. G. Molecular structure of three mutations at the maize sugary1 locus and their allele-specific phenotypic effects. *PLANT Physiol.* **125,** 1406–1418 (2001).
29. Zeder, M. A., Emshwiller, E., Smith, B. D. & Bradley, D. G. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* **22,** 139–155 (2006).
30. Larson, G. & Burger, J. A population genetics view of animal domestication. *Trends Genet.* **29,** 197–205 (2013).
31. Yamasaki, M. *et al.* A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* **17,** 2859–2872 (2005).
32. Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S. & Buckler, E. S. Genetic diversity and selection in the maize starch pathway. *Proc. Natl Acad. Sci. USA* **99,** 12959–12962 (2002).
33. Cooper, A. & Poinar, H. N. Ancient DNA: do it right or not at all. *Science* **289,** 530–531 (2000).
34. Gilbert, M. T. P. *et al.* Ancient mitochondrial DNA from hair. *Curr. Biol.* **14,** R463–R464 (2004).
35. Cappellini, E. *et al.* A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* **97,** 205–217 (2010).
36. Wales, N., Andersen, K., Cappellini, E., Avila-Arcos, M. C. & Gilbert, M. T. P. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9,** e86827 (2014).
37. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).
38. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnol.* **27,** 182–189 (2009).

Acknowledgements

The authors acknowledge the following grants: Marie Curie Actions IEF 272927 and COFUND DFF-1325-00136, Danish National Research Foundation DNRF94, Danish Council for Independent Research 10-081390 and 1325-00136, Lundbeck Foundation grant R52-A5062, Vand Fondecyt Grant 1130261, a grant from the UC Davis Genome Center for the highland maize sequence and NSF IOS-1238014. R.F. is supported by a Young Investigator grant (VKR023446) from Villum Fonden. P.S. was funded by the Wenner-Gren foundation. The authors thank Ângela Ribeiro, Shohei Takuno and Philip Johnson for comments and discussion and staff at the Danish National High-Throughput DNA Sequencing for technical support.

Author contributions

M.T.P.G., B.D.S. and R.R.F. conceived and headed the project. M.T.P.G., N.W. and E.C. designed the experimental research project setup. R.R.F. designed the bioinformatics and population genetics setup with input from M.T.P.G., A.A. and J.R.I. Both B.D.S. and B.A. provided ancient samples and associated context information. M.B.H. and J.R.I. provided sequence data for the highland Palomero de Jalisco landrace. B.D.S. provided the archaeological background and performed the radiocarbon dating. N.W., E.C. and C.C. performed the ancient DNA extractions, library construction and capture with input from M.T.P.G. Both M.C.A. and J.A.S. provided bioinformatics support for the optimization of the capture-related laboratory work. J.A.S. annotated the silent and non-synonymous sites. TSK designed the tool to filter transitions in bam files. R.R.F. chose the capture targets, performed the quality filtering and mapping of the ancient datasets, and prepared the maize HapMap2 data and the modern genome data for all downstream analyses. R.R.F. performed the error determination, neutrality tests, NGSadmix, TreeMix, phylogenetic and demographic inference analyses with input from A.A. and J.R.I. D-statistics analysis was performed by P.S. with input from M.J. Both R.R.F. and M.F. performed the PBS-based selection analyses with input from R.N. Both D.E.H. and M.B.H. performed the STRUCTURE analysis. F.G.V. performed the inbreeding analysis. R.R.F., B.D.S., M.B.H., J.R.I. and M.T.P.G. wrote the manuscript with critical input from all authors.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.R.F. and M.T.P.G.

Competing interests

The authors declare no competing financial interests.