Archaeological Central American maize genomes suggest ancient gene flow from South America

Logan Kistler\textsuperscript{a,1}, Heather B. Thakar\textsuperscript{b}*, Amber M. VanDerwerker\textsuperscript{c}, Alejandro Domic\textsuperscript{d,\textasteriskcentered}, Anders Bergström\textsuperscript{e}, Richard J. George\textsuperscript{e}, Thomas K. Harper\textsuperscript{d}, Robin G. Allaby\textsuperscript{f}, Kenneth Hirth\textsuperscript{d}, and Douglas J. Kennett\textsuperscript{f,1}

\textsuperscript{a}Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560; \textsuperscript{b}Department of Anthropology, Texas A&M University, College Station, TX 77843; \textsuperscript{c}Department of Anthropology, University of California, Santa Barbara, CA 93106; \textsuperscript{d}Department of Anthropology, The Pennsylvania State University, University Park, PA 16802; \textsuperscript{e}Department of Geosciences, The Pennsylvania State University, University Park, PA 16802; \textsuperscript{f}Ancient Genomics Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom; and \textsuperscript{g}School of Life Sciences, University of Warwick, CV4 7AL Coventry, United Kingdom

Edited by David L. Lentz, University of Cincinnati, Cincinnati, OH, and accepted by Editorial Board Member Elsa M. Redmond November 3, 2020 (received for review July 24, 2020)

Maize (\textit{Zea mays ssp. mays}) domestication began in southwestern Mexico ∼9,000 calendar years before present (cal. BP) and humans dispersed this important grain to South America by at least 7,000 cal. BP as a partial domesticate. South America served as a secondary improvement center where the domestication syndrome became fixed and new lineages emerged in parallel with similar processes in Mesoamerica. Later, Indigenous cultivators carried a second major wave of maize southward from Mesoamerica, but it has been unclear until now whether the deeply divergent maize lineages underwent any subsequent gene flow between these regions. Here we report ancient maize genomes (2,300–1,900 cal. BP) from El Gigante rock shelter, Honduras, that are closely related to ancient and modern maize from South America. Our findings suggest that the second wave of maize brought into South America hybridized with long-established landraces from the first wave, and that some of the resulting newly admixed lineages were then reintroduced to Central America. Direct radiocarbon dates and cob morphological data from the rock shelter suggest that more productive maize varieties developed between 4,300 and 2,500 cal. BP. We hypothesize that the influx of maize from South America into Central America may have been an important source of genetic diversity as maize was becoming a staple grain in Central and Mesoamerica.

Maize is a global food staple with great economic and cultural importance. Archaeogenomic studies have revealed a process of protracted maize domestication and multiple waves of human-mediated dispersal in the Americas. Maize first arrived in South America as a partial domesticate, where the domestication syndrome became independently fixed and improved varieties developed away from the influence of wild gene flow. We demonstrate that hybrids of some of these improved varieties were likely reintroduced back to Central America. We hypothesize that this backflow of South American genetic material may have contributed to the development of a more productive staple, which was related to the growth and aggregation of human populations, and the formation of more complex social and political structures regionally.

Significance

Maize is a global food staple with great economic and cultural importance. Archaeogenomic studies have revealed a process of protracted maize domestication and multiple waves of human-mediated dispersal in the Americas. Maize first arrived in South America as a partial domesticate, where the domestication syndrome became independently fixed and improved varieties developed away from the influence of wild gene flow. We demonstrate that hybrids of some of these improved varieties were likely reintroduced back to Central America. We hypothesize that this backflow of South American genetic material may have contributed to the development of a more productive staple, which was related to the growth and aggregation of human populations, and the formation of more complex social and political structures regionally.

Author contributions: L.K. and D.J.K. designed research; L.K. performed research; L.K. and K.H. contributed new reagents/analytic tools; L.K., H.B.T., A.M.V., A.D., A.B., R.J.G., T.K.H., and R.G.A. analyzed data; L.K. and D.J.K. wrote the paper; and L.K., K.H., and D.J.K. supervised research.

The authors declare no competing interest.

This article is a PNAS Direct Submission. D.L.L. is a guest editor invited by the Editorial Board.

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

\textsuperscript{1}To whom correspondence may be addressed. Email: KistlerL@si.edu or kennett@anth.ucsb.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2015560117/-/DCSupplemental.

First published December 14, 2020.
modern maize for inflorescence and seed architecture (id1 – tassel dwarf1; th1 – teosinte branched1, ba1 – barren stalk1), circadian clock and flowering time (zmg1), and glycogen biosynthesis (bt2 – brittle endosperm2) (23, 24). However, some alleles controlling ear shattering and starch biosynthesis (zagl – MADS-box gene; sul – sugary 1, and wx1 – waxy 1) were more teosintelike at that point—four millennia after the onset of domestication. Introgression with mexicana favored adaptation to drier and cooler conditions in the Mexican highlands, and most modern landraces carry strong signals of postdomestication mexicana admixture (13, 25). Continued introgression within the natural range of parviglumus and mexicana may also explain the gradual changes in cob size evident in the Tehuacán Valley (19).

The initial first wave of maize dispersal (7,500–7,000 cal. BP) (5) through Central and South America likely occurred when maize was partially domesticated, and before the domestication syndrome—the suite of characters setting a domesticated species apart from its wild counterpart—was fixed (26). Multiple waves of dispersal brought maize out of southwestern Mexico into South America (26, 27), and may have episodically increased admixture and diversity. There is also evidence that deeply structured maize lineages in South America underwent independent fixation of the domestication syndrome and secondary improvement outside the range of parviglumus and mexicana (26). Reduced crop-wild gene flow outside the range of parviglumis and mexicana also enhanced selection for increased cob and seed size (6), and was instrumental in the development of more productive staple grain varieties and greater consumption in Central America after 4,700 cal. BP (28, 29). Furthermore, adaptations in the US Southwest—outside the domestication center—resulted in the shorter growing season varieties with earlier flowering times required for dispersal through more temperate parts of North America after 2,000 cal. BP (30). In total, it is clear from the available data that maize diversity and biogeography is complex, and resulted from multiple episodes of human-mediated selection and dispersal throughout the Americas.

Research has focused on the outward dispersal of maize from the original domestication center in southwestern Mexico. However, crop movements were complex, and the archaeological record shows clear evidence of two-way movements of plants and people lasting millennia between Central and South America. It is reasonable to suspect that maize, the most widespread crop species of the precolumial Americas, traveled back toward the domestication center in the hands of skilled farmers as part of this complex history. Here, we sequenced maize genomes from three archaeological samples from El Gigante rock shelter in western Honduras dating to between 2,300 and 1,900 cal. BP and compared these data to published modern landraces of maize and archaeological samples from North, Central, and South America. We use these genomes as a temporal anchor to test the hypotheses that humans moved maize from South America into Central America. In this scenario, the reintroduced germplasm may have been impactful for the development of highly productive varieties. Isotopic evidence from Central America demonstrates substantial maize consumption as a staple grain beginning between 4,700 and 4,000 cal. BP (29).

Finally, we use morphological comparisons within the El Gigante maize assemblage to help constrain the timing of this gene flow.

El Gigante and Maize Samples
El Gigante rock shelter is located in the highlands of western Honduras (88.06° W, 14.22° N; 1,300 masl; Fig. 1) outside the contemporary range of parviglumis and mexicana (28). The rock shelter is 42 m wide, 17 m deep, 12 m high and was used episodically during the last 11,000 y (31). Dry conditions inside the dripline promoted the preservation of desiccated plant material, including locally available wild plants, tree crops (e.g., avocado [Persea americana Mill.]; hog plum [Spondias sp.]; custard apple [Annona sp.]) and field cultigens (e.g., maize, beans [Phaseolus spp.], and squash [Cucurbita spp.]). Eight phases of occupation were identified based on a Bayesian chronological model of 89 radiocarbon dates through the sequence (28). Wild plant foods dominate the botanical assemblages in the earliest four phases (11,010-7,430 cal. BP; Early and Late Esperanza, Early and Middle Marcala Phases). Squash (Cucurbita spp.) first appears in the Middle Marcala Phase (7,600-7,430 cal. BP) just prior to a long hiatus in rock-shelter use (7,430-4,340 cal. BP). The earliest maize in the sequence dates just after this hiatus in the Late Marcala phase (4,340 and 4,020 cal. BP; Stratum IId-Ia) in a botanical assemblage dominated by wild plant foods. No maize has been directly dated in the subsequent, poorly defined Early Estanzuela Phase deposits (3,390-2,780 cal. BP), but returns as a
dominant feature in the Late Estanzuela Phase (2,350 and 1,820 cal. BP; Stratum Id-Ia) found in association with beans and squash.

Over 10,000 carbonized and uncarbonized maize specimens (cobs, kernels, stalks, leaves) occur in the assemblage. The earliest directly dated maize cobs have 12–14 rows and are relatively slender (conical to lanceolate in shape). There is substantial morphological overlap in cob width, kernel rows, and rachis segment length between the earliest and latest cobs in the assemblage dating to between 2,350 and 1,180 cal. BP. Seed row number is equivalent in the earliest and latest cobs and permutation tests show no statistical differences in cob diameter ($P = 0.0693$), rachis diameter ($P = 0.1006$), and cupule length ($P = 0.96104$) [SI Appendix, Fig. S1; data from ref. (28)]. Temporal differences in cupule width ($P = 0.00978$) and cupule wing width ($P = 0.04228$) differ significantly from expectations of random chance. This shift suggests these attributes increased on average between the earlier and later assemblage, suggesting selection for increasing seed size. Our permutation tests are consistent with a robustly domesticated variety of maize developing or arriving in southeastern Mesoamerica by at least 4,300 cal. BP (28) at a time when it was becoming a dietary staple (29).

We attempted to extract DNA from all of the directly dated Late Marcala maize cobs in the assemblage (4,340 and 4,020 cal. BP; $n = 20$) but preservation was unsuitable for genomic sequencing. We also attempted to extract DNA from 10 Late Estanzuela Phase (2,350–1,900 cal. BP) maize cobs and three specimens produced endogenous DNA sufficient for genomewide sequencing (Table 1).

### Genomics

We compared the Late Estanzuela Phase maize genomes with 120 previously published modern ($n = 109$) and ancient ($n = 11$) maize genomes [SI Appendix]. Model-based clustering revealed that the ~2,000 cal. BP El Gigante maize is most similar to the “Pan-American” cluster described previously (26) (Fig. 2B). This group is widespread in Mexico, Central America, and South America, with some presence in North America north of Mexico. These genomes carry a surplus of teosinte ancestry compared with other major maize lineages in North and South America, suggesting that this lineage was dispersed away from the domestication center later in time after additional ancient crop wild gene flow (26). Although model-based clustering primarily places El Gigante maize in the Pan-American cluster, it also identified a small but consistent component of South American ancestry in the El Gigante genomes (Fig. 2B and SI Appendix). We used outgroup-$f^3$ statistics to further interrogate the nearest genetic neighbors to El Gigante maize. We found that its genome most closely resembled modern landraces in South America with predominantly Pan-American ancestry, as well as archaeological genomes from South America (Fig. 2D). One Cuban genome from the maize Hapmap2 project (32) also matched closely. However, this variety (Cuban Flint) was introduced from Argentina in the early 20th century (33). The outgroup-$f^3$ results therefore reinforce a consistent link to South America.

Because maize was well-established in several regions of the Americas by this time, the El Gigante maize at ~2,000 cal. BP was likely fully domesticated. Therefore, we predicted domestication syndrome genes to be more maizelike than either 1) teosintelike or 2) a combination of maize- and teosintelike signifying partial domestication. Adapting an approach based on ancestry informative markers [AIMs; ref. (23) and SI Appendix], we quantified the proportion of teosintelike character states near domestication genes. The genewise proportion of teosintelike variants in the El Gigante domestication genes overlapped modern maize (Wilcoxon rank-sum test $P = 0.75$; Welch’s $t$ test $P = 0.63$), and deviated strongly from modern teosinte (both tests $P < 2 \times 10^{-16}$).

We also used a likelihood-based strategy to assess whether individual genes in El Gigante maize were more likely drawn from a population resembling modern maize or modern teosinte (Fig. 2C). Under this analysis, key domestication genes analyzed in previous aDNA studies—bt2, tga1, b1, sl1, zag1, and orf13—are all robustly maizelike in the El Gigante genomes (Bayes factors 3.84–78.93, signifying support ranging from “substantial” to “very strong”) (34) [SI Appendix]. In total, we found strong support that Late Estanzuela El Gigante maize was robustly domesticated.

We used admixture graph fitting to refine the relationships between maize populations (Fig. 3A). This method found support ($\text{maxZ}^2 = 2.7$) for a model where El Gigante maize falls just outside the range of variation of modern Pan-American maize but shares the historic parviglumis gene flow that characterizes the Pan-American cluster.

On the basis of $f^4$ statistics, all individual modern Pan-American maize genomes and the El Gigante maize carry this signal for excess teosinte ancestry compared with South American lineages (Fig. 3B and SI Appendix). This reinforces previous findings that Pan-American maize underwent additional crop wild gene flow in the domestication center after the first wave of maize was carried into South America (26). Although our analysis above demonstrates that El Gigante maize carries a full suite of maizelike alleles at domestication loci, this excess teosinte ancestry is consistently present at low levels on a genomewide basis in El Gigante maize and modern Pan-American genomes.

We also found evidence for introgression from first-wave South American maize into southern members of the Pan-American group, which occurred following the second wave of maize dispersal into South America. This finding suggests hybridization in South American maize fields between established and introduced lineages. This is not surprising in light of millennia of sympatric cultivation. Indigenous farmers apparently maintained the linkage identity of first-wave and second-wave maize to a significant degree in South America, while also integrating admixed varieties into their cropping systems.

We further explored the link between El Gigante maize and South America by identifying a set of AIMs that reliably distinguish between genomes with primarily South American ancestry and all other maize, and quantifying the proportion of SA-like ancestry in each sample (SI Appendix). On this basis, El Gigante maize has the higher proportion of SA-like alleles than any maize originating from north of South America (Fig. 3C; EG maize is 2.53 SD from the mean of the normally distributed southern set; $P = 0.0057$).

Finally, genome size is extremely variable in maize. We estimated relative genome size among modern maize using a previously established proxy in the heterochromatic knob fraction of the genome (32, 35) (Materials and Methods and Dataset S1). As shown previously (36), we observed that genome size in maize is significantly correlated with elevation in Central and South American highlands (linear model $r^2 = 0.11, P = 0.0012$, and

| Table 1. Maize cobs with endogenous DNA sufficient for genomewide sequencing |
|-----------------------------|----------------|------------------|---------------|-----------------|-----------------|
| ID | Provenience | Date (cal. BP) | Rows | Diameter (mm) | Shape |
| EG84 | Unit 2, Level 6, Stratum Ibb/c | 1870–1740 | 16 | 17.55 | Lanceolate |
| EG85 | Unit 2, Level 6, Stratum Ibb/c | 2300–2070 | 14 | 15.10 | Conical |
| EG90 | Unit 18, Level 14, Stratum Ibb | 2300–2120 | 10 | 20.50 | Lanceolate |
that maize genomes are significantly smaller than teosinte genomes on average (32) (Welch’s t test $P = 2.28 \times 10^{-4}$, Wilcoxon rank-sum test $P = 2.24 \times 10^{-4}$)—albeit with substantially overlapping ranges. Genome size also decreases significantly with distance from the domestication center (linear model $r^2 = 0.122$, $P = 6.06 \times 10^{-4}$), and a multivariate linear model including elevation and distance substantially increases predictive power ($r^2 = 0.281$, $P = 3.59 \times 10^{-7}$). Distance to the domestication center correlates strongly with both a decrease in genomic heterozygosity (25) and an increase in the genomic mutation load (26). Thus, we find that in addition to losing diversity and accumulation of deleterious variants during dispersal, the genome tends to physically contract down the dispersal gradient (Fig. 3D).

The members of the Pan-American group defy this broader pattern. Among low-elevation (<1,000 m) samples with ≥90% Pan-American ancestry on the basis of model-based clustering, genomes near the domestication center are significantly smaller than those in South America (linear model $r^2 = 0.501$, $P = 0.014$; Fig. 3D). In fact, the largest maize genomes we observe are South American samples with high proportions of Pan-American ancestry. By adding the proportion of Pan-American ancestry as an independent variable alongside distance to the domestication center, the model substantially increases predictive power ($r^2 = 0.645$, $P = 0.014$).

Fig. 2. Genetic affinities and domestication status of archaeological El Gigante maize. (A) Outgroup-$f_2$ statistics in the form $f_2(\text{Tripsacum}; X, \text{El Gigante})$ with all other maize samples in position $X$, showing that maize samples sharing the most drift with El Gigante maize are modern and ancient genomes from South America. The sample in Cuba with a high $f_2$ value is from a HapMap2 landrace with known origins in Argentina. (B) Ancestry proportions of modern and ancient maize estimated via model-based clustering. (C) Estimation of domestication status in El Gigante and other maize via AIMs located near and within domestication syndrome genes. The y axis displays normalized log-likelihood of a gene being drawn from a maizelike (1) vs. teosintelike (−1) reference panel, with significance thresholds marked where Bayes factors (BF) (and 1/BF) ≥ 3 and ≥10. Each column of dots shows a single genome with up to 278 individual gene lnL ratios, with ancestry proportions corresponding to $B$ above each column. In El Gigante maize (E.G.), domestication genes overlap the modern maize reference panel and deviate strongly from the teosinte panel, and all six specifically analyzed domestication genes are maizelike with at least BF ≥ 3. In contrast, Middle Holocene Tehuacán Valley maize (Teh.) carries a mixture of maizelike and teosintelike variants as previously reported (23). Mid-Holocene San Marcos maize (S.M.) was also previously shown to be a partial domesticate (24), although more maizelike than the Tehuacán specimen (26), a finding reinforced here.
through the Isthmus of Panama prior to Indigenous farmers carrying or dispersing maize northward. The South Americans signal in El Gigante maize provides evidence from earlier-dispersed South American lineages, revealing hybridization in South America. Instead, we hypothesize that American lineages, nor the outgroup-AIM-based links between El Gigante and earlier-dispersed South Pan-American genomes. However, this pattern does not explain the lineage where El Gigante maize acts as an outgroup to southern genomes and modern and ancient maize from South America. On the basis of links to southern Pan-American landraces upon arrival. First, by around 7,000 cal. BP, an initial wave of semidomesticated maize was carried southward, where it likely underwent secondary improvement in the southwestern Amazon before diversifying across the continent (26, 37). Later, a second population of maize was transported southward, carrying the signal for excess gene flow through the Isthmus of Panama prior to ∼2,000 y ago. Admixture graphs and f4 statistics reinforce multiple waves of dispersal southward out of the center of domestication. First, by around 7,000 cal. BP, an initial wave of semidomesticated maize was carried southward, where it likely underwent secondary improvement in the southwestern Amazon before diversifying across the continent (26, 37). Later, a second population of maize was transported southward, carrying the signal for excess gene flow through the Isthmus of Panama prior to ∼2,000 y ago. 

Alzheimer’s disease is a progressive neurological disorder that affects memory and cognitive function. It is characterized by the presence of amyloid plaques and neurofibrillary tangles in the brain. These deposits are thought to cause nerve cell damage and death, leading to cognitive decline and memory loss. The disease is associated with aging, and its prevalence increases with age. The exact cause of Alzheimer’s disease is unknown, but it is believed to be influenced by a combination of genetic and environmental factors. Early detection and treatment are crucial in managing the disease and improving quality of life. There are currently no cure or effective treatments to stop or reverse the progression of the disease. Prevention strategies, such as maintaining a healthy lifestyle and participating in cognitive exercises, are recommended to reduce the risk of developing Alzheimer’s disease. The information provided is a general overview, and individual cases may vary.
northward, as well as with spreading Indigenous farmers throughout South America.

Future work will be required to determine when the backflow of improved South American maize varieties occurred. Ancient DNA is poorly preserved in the early cobs in the El Gigante sequence (4,340–4,020 cal. BP), and other cobs dating to the Middle Holocene are rare in this region. However, the morphological evidence from El Gigante is comparable in size and row number to the later cobs (~2,300–1,900 cal. BP cob; ref. 28). The appearance of robustly domesticated cobs also corresponds well with dietary stable isotope data indicating that maize was becoming a staple grain after ~4,700 cal. BP (29), and with pollen records revealing maize cultivation in Honduras around the same time (38). The backflow of improved varieties from South America may have occurred by this time and contributed to the development of more productive staple grain varieties. Ultimately, the influx of improved varieties and resulting diversification of maize might have contributed to the growth and aggregation of human populations, helping pave the way for the formation of more complex social and political structures regionally.

Materials and Methods

Complete details of genomic methods and radiocarbon dating used in this study are provided in SI Appendix. Briefly, maize samples were prepared in the dedicated ancient DNA clean laboratory facilities at The Pennsylvania State University Anthropology Department, and the Smithsonian Institution’s Museum’s Support Center. Standard protocols to prevent and detect contamination were utilized (39), including strict workflow procedures, frequent cleaning with bleach and ethanol, use of complete personal protective equipment, and the preparation and sequencing of negative control reactions. We extracted DNA, prepared sequencing libraries, and screened samples following established protocols for highly degraded ancient DNA (40–42), identified single nucleotide polymorphism (SNP) calls, and small DNA databases are available on Dryad (DOI: 10.5061/dryad.xs39d).

Data Availability

Raw sequencing data for El Gigante maize specimens have been deposited in the NCBI Sequence Read Archive under BioProject PRJNA636370. Custom scripts, single nucleotide polymorphism (SNP) calls, and small DNA databases are available on Dryad (DOI: 10.5061/dryad.xs39d).