

# Gourds and squashes (*Cucurbita* spp.) adapted to megafaunal extinction and ecological anachronism through domestication

Logan Kistler<sup>a,b,c,1</sup>, Lee A. Newsom<sup>a,d</sup>, Timothy M. Ryan<sup>a,e</sup>, Andrew C. Clarke<sup>f</sup>, Bruce D. Smith<sup>g,h</sup>, and George H. Perry<sup>a,b,1</sup>

<sup>a</sup>Department of Anthropology, The Pennsylvania State University, University Park, PA 16802; <sup>b</sup>Department of Biology, The Pennsylvania State University, University Park, PA 16802; <sup>c</sup>School of Life Sciences, University of Warwick, Coventry, CV4 7AL United Kingdom; <sup>d</sup>Institutes of Energy and the Environment, The Pennsylvania State University, University Park, PA 16802; <sup>e</sup>Center for Quantitative Imaging, Earth and Mineral Sciences Energy Institute, The Pennsylvania State University, University Park, PA 16802; <sup>f</sup>McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, United Kingdom; <sup>g</sup>Program in Human Ecology and Archaeobiology, Department of Anthropology, Smithsonian Institution, National Museum of Natural History, Washington, DC 20560; and <sup>h</sup>Santa Fe Institute, Santa Fe, NM 87501

Edited by Paul Minnis, University of Oklahoma, Norman, OK, and accepted by the Editorial Board October 13, 2015 (received for review August 16, 2015)

The genus *Cucurbita* (squashes, pumpkins, gourds) contains numerous domesticated lineages with ancient New World origins. It was broadly distributed in the past but has declined to the point that several of the crops' progenitor species are scarce or unknown in the wild. We hypothesize that Holocene ecological shifts and megafaunal extinctions severely impacted wild *Cucurbita*, whereas their domestic counterparts adapted to changing conditions via symbiosis with human cultivators. First, we used high-throughput sequencing to analyze complete plastid genomes of 91 total *Cucurbita* samples, comprising ancient ( $n = 19$ ), modern wild ( $n = 30$ ), and modern domestic ( $n = 42$ ) taxa. This analysis demonstrates independent domestication in eastern North America, evidence of a previously unknown pathway to domestication in north-eastern Mexico, and broad archaeological distributions of taxa currently unknown in the wild. Further, sequence similarity between distant wild populations suggests recent fragmentation. Collectively, these results point to wild-type declines coinciding with widespread domestication. Second, we hypothesize that the disappearance of large herbivores struck a critical ecological blow against wild *Cucurbita*, and we take initial steps to consider this hypothesis through cross-mammal analyses of bitter taste receptor gene repertoires. Directly, megafauna consumed *Cucurbita* fruits and dispersed their seeds; wild *Cucurbita* were likely left without mutualistic dispersal partners in the Holocene because they are unpalatable to smaller surviving mammals with more bitter taste receptor genes. Indirectly, megafauna maintained mosaic-like landscapes ideal for *Cucurbita*, and vegetative changes following the megafaunal extinctions likely crowded out their disturbed-ground niche. Thus, anthropogenic landscapes provided favorable growth habitats and willing dispersal partners in the wake of ecological upheaval.

evolutionary ecology | sensory ecology | TAS2R genes | ancient DNA | archaeogenomics

The wild precursors of domestic squashes (*Cucurbita* spp.) were adapted for a landscape inhabited by large herbivores. Their robust pepo fruits (1) were dispersed by large mammals, as revealed by intact *Cucurbita* seeds in mastodon dung deposits (2). Furthermore, *Cucurbita* is a weedy genus (3) well suited to the mosaic-like landscapes maintained by megafauna, which offer an abundance of disturbed habitat in a niche-diverse ecosystem (4). However, all megafauna >1,000 kg disappeared from the Americas by the early Holocene—through ecological shifts, human predation, or some combination of both (5)—leaving *Cucurbita* in a turbulent ecosystem lacking its mutualistic partners. *Cucurbita* is among many such anachronistic New World taxa (6), including avocado (*Persea americana*), chocolate (*Theobroma cacao*), Osage orange (*Maclura pomifera*), honey locust (*Gleditsia triacanthos*), and tree calabash (*Crescentia cujete*). Distributions of ancient samples show that some species of *Cucurbita* were distributed much more broadly in the past (e.g., refs. 7 and 8), suggesting that

the Holocene has witnessed extirpation and refugiation of certain wild types. The decline of wild *Cucurbita* may have stemmed from both its loss of dispersal mutualists and the changing habitat. Other anachronistic plants have partnered with substitute dispersers (9), but most wild *Cucurbita* have not adapted in this way (10).

Although *Cucurbita* species declined in the wild, they thrived in domestication. The genus contains at least five domesticated species and has been used extensively throughout the Holocene (11), beginning around 10,000 B.P. in Mexico (12). *Cucurbita* domestication was a widespread phenomenon involving numerous wild precursor lineages throughout the Americas (11, 13–15), and these taxa now contain hundreds of cultivars and landraces grown worldwide (11, 15). In this study, we use a two-stage approach to study the simultaneous decline of wild *Cucurbita* and rise of its domestic counterparts. First, to assess aspects of *Cucurbita* domestication history and phylogeographic patterning, we recover and analyze complete plastid genomes from 91 total archaeological ( $n = 19$ ) and modern wild ( $n = 30$ ) and domestic ( $n = 42$ ) *Cucurbita* specimens. Second, wild *Cucurbita* may have adapted poorly to dispersal anachronism partly for physiological reasons.

## Significance

Squashes, pumpkins, and gourds belonging to the genus *Cucurbita* were domesticated on several occasions throughout the Americas, beginning around 10,000 years ago. The wild forms of these species are unpalatably bitter to humans and other extant mammals, but their seeds are present in mastodon dung deposits, demonstrating that they may have been dispersed by large-bodied herbivores undeterred by their bitterness. However, *Cucurbita* may have been poorly adapted to a landscape lacking these large dispersal partners. Our study proposes a link between the disappearance of megafaunal mammals from the landscape, the decline of wild *Cucurbita* populations, and, ultimately, the evolution of domesticated *Cucurbita* alongside human cultivators.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. P.M. is a guest editor invited by the Editorial Board.

Data deposition: Illumina short read sequence data reported in this paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive (study accession no. [SRP064244](https://www.ncbi.nlm.nih.gov/sra/SRP064244)). Annotated plastid genome sequences for taxonomic representatives have been deposited in GenBank database (accession nos. [KT898803](https://www.ncbi.nlm.nih.gov/nuclot/KT898803)–[KT898821](https://www.ncbi.nlm.nih.gov/nuclot/KT898821)).

<sup>1</sup>To whom correspondence may be addressed. Email: [L.J.Kistler@Warwick.ac.uk](mailto:L.J.Kistler@Warwick.ac.uk) or [gph3@psu.edu](mailto:gph3@psu.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516109112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516109112/-DCSupplemental).

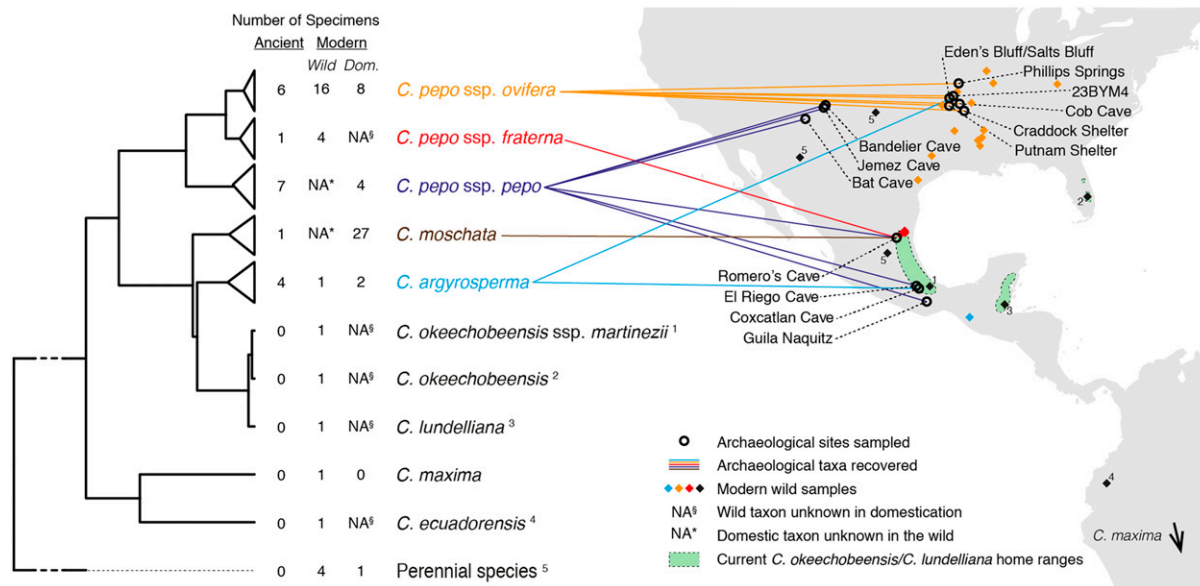
Specifically, it produces a cytotoxic suite of cucurbitacins—harshly bitter triterpenoid compounds (16)—that deter small mammals (17) but appear to be harmless to megafauna. As such, *Cucurbita* was palatable to some megafauna (2) but possibly not to the majority of smaller New World mammals that survived into the Holocene; thus, *Cucurbita* survival was threatened by the disruption of its propagation strategy. By screening 46 mammal genomes for bitter taste receptor-encoding *taste receptor type 2 (TAS2R)* genes, we evaluate the hypothesis that mammal body size may be related directly to the palatability of *Cucurbita* and other bitter species. If smaller mammals have adapted mechanisms for better detecting and avoiding bitter, potentially toxic compounds in lower doses, *Cucurbita* may have coadapted specifically for a landscape containing large mammals; such adaptation would increase the potential impacts of a major shift toward an ecosystem lacking them. In sum, we combine archaeogenetic and comparative genomic techniques to probe the natural history of a genus that is ecologically anachronistic in the wild but is prolific in cultivation.

## Results

**Plastid Genome Analysis.** Using Illumina shotgun sequencing, we recovered and assembled complete plastid genomes from 12 of the 14 accepted extant *Cucurbita* species, representing 18 total species, subspecies, and varietal taxa. Using our completed assemblies for *C. pepo* ssp. *pepo* and *C. moschata*, we then designed and synthesized biotinylated RNA probes (18) to enrich additional modern and ancient DNA isolates for targeted sequencing of complete plastid genomes. Using this enrichment strategy, we tested 32 archaeological *Cucurbita* samples, yielding 19 ancient plastid genomes sufficiently complete for analysis. We also sequenced an additional 53 modern plastid genomes using a combination of shotgun sequencing and targeted enrichment, bringing the total sample size to 91 ( $n = 19$  ancient;  $n = 72$  modern) (Tables S1–S3), and reconstructed a *Cucurbita* plastid genome phylogeny from this dataset (Fig. 1).

*C. pepo* comprises three subspecies: ssp. *pepo*, ssp. *ovifera*, and ssp. *fraterna*. Of these, *C. pepo* ssp. *pepo* (zucchini, pumpkins, and summer and winter squashes) appears to have been domesticated in the Oaxaca Valley region around 10 kya (12) and has no known extant wild type. *C. pepo* ssp. *ovifera* contains wild varieties found in patches throughout the southeastern United States, including *C. pepo* ssp. *ovifera* var. *ozarkana*, and *C. pepo* ssp. *ovifera* var. *texana*, as well as a wide variety of domestic types (scallop and acorn squashes, ornamental gourds). Archaeologically, it appears that ssp. *ovifera* domestication took place in eastern North America (ENA) alongside a suite of other crop species (e.g., 19). *C. pepo* ssp. *fraterna* is a wild form restricted to a narrow range in northeastern Mexico and unknown as a domesticate (20). However, it is genetically similar to ssp. *ovifera* and therefore has been hypothesized as an alternate source population of domesticated *C. pepo* in ENA (14).

We found that members of ssp. *ovifera* form a well-supported clade that includes modern cultivars and wild gourds from ENA as well as all archaeological ENA *C. pepo* samples. There was no signal of ssp. *ovifera* among ancient samples from Mexico, the alternative possible source region (14), nor in modern free-living ssp. *fraterna* in that region. This absence strongly supports local evolution of domesticated *C. pepo* ssp. *ovifera* in ENA, in agreement with archaeological evidence of an increase in seed size over time and molecular data clustering modern ssp. *ovifera* cultivars with ENA wild populations (21–23). However, although ssp. *fraterna* is known only in the wild at present and is absent from the ENA archaeological record, we observed a ssp. *fraterna* plastid sequence in an archaeological rind fragment from Romero's Cave in northeastern Mexico. The specimen shows a thick (>2 mm) rind and a furrow typical only of domesticated *Cucurbita*. We obtained a direct date from the rind of 4,614–4,836 calibrated calendar years BP (2 $\sigma$  range). This result might represent cross-pollination of native wild ssp. *fraterna* by domesticated ssp. *pepo* leading to the expression of some domestic-type characters or an independent



**Fig. 1.** Sample map and plastid genome phylogeny of *Cucurbita*. All nodes shown are supported with Bayesian posterior probability of 1 and maximum likelihood support of 100/100 bootstrap replicates, with two exceptions: (i) *C. okeechobeensis*/*C. okeechobeensis* ssp. *martinezii* node: 93 bootstrap replicates; (ii) *C. argyrosperma*/*C. okeechobeensis*/*C. lundelliana* node: 95 bootstrap replicates. Tree tip labels in color represent taxa recovered among our archaeological samples, with color-matching lines indicating sampling sites and diamonds showing sampling locations of modern wild counterparts, if applicable (present for *C. pepo* ssp. *ovifera*, *C. pepo* ssp. *fraterna*, and *C. argyrosperma*). Labeled black diamonds show the sampling locations of other wild accessions analyzed as single taxonomic representatives. Among the four wild perennial species analyzed (Tables S1–S3), *C. cordata* is not shown, because sampling records for this accession are not specific regarding collection locality. *C. maxima*, a South American species with wild and domestic members, was sampled from a wild accession of *C. maxima* ssp. *andrea* from central Argentina and so is not shown.

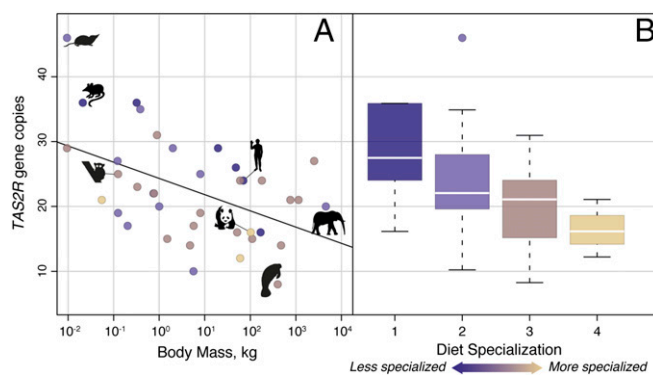
domestication trajectory that did not contribute to modern plastid diversity among cultivars. In either circumstance, this finding demonstrates a previously unknown role of *ssp. fraterna* in the crop fields of ancient Mexico. Additionally, we observed archaeological *ssp. pepo* in cave sites ranging from Oaxaca to the southwestern United States, suggesting either prolific human-mediated dispersal following domestication or a geographically widespread evolution of domestic forms.

Finally, we found that three endemic populations spanning the Gulf of Mexico, currently defined as distinct taxa (24), are essentially indistinguishable based on the plastid genome. *C. okeechobeensis* is restricted to the Florida peninsula, on the shores of Lake Okeechobee and the St. John's River basin, *C. okeechobeensis ssp. martinezii* is found on the north coastal plain of Veracruz, and *C. lundelliana* grows on the limestone plains of the Yucatan Peninsula (24), but the three taxa differ at only six single-nucleotide variants across the 156.6-kb plastid genome (average pairwise nucleotide diversity:  $\pi = 0.0043\%$ ) despite their geographic separation. Along with the undifferentiated plastid genome structure among patchy *C. pepo ssp. ovifera* in ENA (Fig. 1), this result may reflect recent fragmentation of a previously contiguous population.

In summary, we observed a pattern that supports hypotheses for widespread Holocene domestication of *Cucurbita* and suggests that this process coincided with dramatic habitat and range fractionation and wild plant extinction. *C. pepo ssp. pepo* is thought to be extinct in the wild, but we find seven archaeological examples from this lineage spanning more than 2,000 km and dating back nearly 10,000 y. We observe at least one other independent domestication episode of *C. pepo*, that of *ssp. ovifera* in ENA alongside a suite of other native crops, and a possible third, arrested domestication trajectory involving *ssp. fraterna* in northeastern Mexico. We also observe genetic similarity between disparate endemic taxa (*C. okeechobeensis*, *C. okeechobeensis ssp. martinezii*, and *C. lundelliana*), suggesting recent separation of these populations.

**Analysis of Ability to Detect Bitter-Tasting Compounds.** We screened 46 therian mammal genomes to assess general ability to detect bitter-tasting compounds by analyzing functional *TAS2R* gene number variation. We excluded the poorest quality assemblies [length-weighted midpoint scaffold size of the genomic assembly (scaffold N50) <250,000 nt;  $n = 8$ ] because of unreliable gene detectability (*Materials and Methods*). Among the 38 remaining genomes, we detected a combined total of 851 intact *TAS2R* genes, ranging from a low of 8 in the genome of the West Indian manatee (*Trichechus manatus latirostris*) to a high of 46 in the genome of the common shrew (*Sorex araneus*). We tested for relationships among *TAS2R* count, body mass, and dietary breadth (*Materials and Methods*) while controlling for phylogenetic nonindependence.

When all species are treated as independent, *TAS2R* count is significantly negatively correlated with both body mass (Pearson's product-moment correlation;  $r^2 = 0.236$ ;  $P = 0.002$ ) (Fig. 2A) and greater dietary specialization ( $r^2 = 0.194$ ;  $P = 0.0056$ ) (Fig. 2B). When all variables are analyzed simultaneously with a generalized linear model, body mass and dietary breadth are each significant predictors of *TAS2R* count when controlling for the other variable ( $P = 0.0043$  and  $P = 0.0122$ , respectively), but the significance of this relationship is strongest when both variables are considered together, and explains a substantial proportion of the variance in *TAS2R* count ( $P = 0.0004$ ; multiple  $r^2 = 0.363$ ). When taxonomic nonindependence among our sample is controlled for by using a phylogenetic generalized least squares test (*Materials and Methods*), the associations of body mass and dietary breadth with *TAS2R* count are still significantly stronger than expected from chance (*TAS2R*-size:  $P = 0.044$ ; *TAS2R*-diet:  $P = 0.015$ ; overall model:  $P = 0.007$ , multiple  $r^2 = 0.246$ ). The *TAS2R*-body size association is most pronounced among smaller-bodied taxa; when only the very



**Fig. 2.** *TAS2R* gene count and phenotypic traits in 38 high-quality mammal genomes. (A)  $\text{Log}_{10}$ -scaled body mass vs. *TAS2R* count. The regression line assumes taxonomic independence, and point shading corresponds to diet specialization as in B (blue: least specialized; tan: most specialized). Representative species shown from left to right are common shrew, mouse, tarser, human, giant panda, West Indian manatee, and African elephant. (B) Boxplots of *TAS2R* count by diet specialization. Category 1 consists of broad dietary generalists (e.g., human, mouse), and category 4 consists of highly specialized feeders (e.g., giant panda).

largest mammal species are excluded from the analysis, the inverse relationships between the number of intact *TAS2R* genes and both body size and diet are highly significant even when controlling for phylogeny (among  $n = 35$  mammals <1,000 kg, *TAS2R*-size:  $P = 0.0013$ ; *TAS2R*-diet:  $P = 0.001$ ; overall model:  $P = 4.08 \times 10^{-5}$ ; multiple  $r^2 = 0.468$ ). To summarize, smaller species, and especially those with more diverse diets, tend to have a larger repertoire of functional bitter taste receptor genes, and thus they likely are better able to detect and avoid potentially toxic bitter compounds while foraging.

## Discussion

**Holocene Anachronism.** The Holocene decline of wild *Cucurbita* was likely driven, at least in part, by the nearly complete disappearance of herbivores  $\geq 1,000$  kg, with whom they appear to have had an important symbiotic relationship. Wild *Cucurbita* are “too bitter for humans and well-fed livestock” (10), but larger mammals are much more tolerant of bitter plant compounds than smaller ones. Larger species are likely able to metabolize or pass moderate toxins harmlessly as part of a high daily biomass intake. For example, even if we ignore any additional physiological or behavioral detoxification strategies and the large digestive throughput of hindgut fermenters, a fully-grown African elephant [averaging 4,540 kg (22)] would need to ingest an estimated 7.5–23 kg of wild *Cucurbita* gourds—approximately 75–230 whole fruits—over a short time-span (25, 26) to approach a lethal dose of cucurbitacins (27). Indeed, several bitter species of Cucurbitaceae are eaten and dispersed by African elephants in the present (10). As another illustration of these concepts, the one-horned rhinoceros in lowland Nepal [1,500–2,000 kg (22)] preferentially feeds on the bitter fruits of trewia (*Trewia nudiflora*, Euphorbiaceae), daily consuming up to hundreds of fruits that are unpalatable to smaller mammals who are concomitantly incapable of dispersing the  $\sim 1$ -cm-diameter seeds (10, 28). Humans have reported severe symptoms of cucurbitacin toxicity after consuming only “one or two bites” of cucurbitacin-tainted squash (29). The presence of *Cucurbita* seeds in mastodon dung is clear evidence that, like their extant counterparts, North American megafauna were effective Cucurbitaceae dispersers, physiologically undeterred by cucurbitacin bitterness.

Larger numbers of intact *TAS2R* genes are associated with greater ability to detect and avoid toxins via their bitterness, and interspecific variation in copy number is hypothesized to reflect ecological and evolutionary diversity (30–32). Previous studies

have reported that herbivores have slightly higher *TAS2R* copies than carnivores, presumably to detect plant-based toxins more acutely (30), and that the common ancestor of modern whales completely lost functional *TAS2R* copies, probably as a result of feeding behaviors and the switch to a marine environment (33). Similarly, in our analysis, dietary breadth is significantly correlated with *TAS2R* count, presumably because a wider range of possible food species necessitates sensitivity to a more diverse range of toxic compounds, and because narrow dietary specialists have little need for toxin detection. Because very large-bodied species are more physiologically resilient to moderately toxic compounds, they too require fewer *TAS2R* gene copies. That is, bitter compounds signal toxicity to foraging animals, but in the absence of serious toxic threat, large mammals do not benefit by acutely recognizing this danger signal. Therefore, selection to maintain a broad *TAS2R* repertoire in larger species may be relaxed, leading to gene loss. Furthermore, although we cannot test this hypothesis in the present study, *TAS2R* pseudogenization could even be adaptive for megafauna if the consumption of bitter fruit was a benefit to survival, i.e., if acuity to bitter tastes would be an unnecessary barrier to herbivory.

The now-extinct megafaunal herbivores of the Americas were sufficiently large that they were likely to have been protected against moderate toxins, whereas the smaller surviving mammals are not. Accordingly, smaller mammals, and especially those with broad dietary habits, tend to have larger suites of *TAS2R* genes to detect and avoid toxic compounds. These smaller species generally would be unable to consume *Cucurbita* fruits whole and pass the intact seeds in the manner of mastodons; therefore, as seed predators instead of dispersal mutualists, the smaller species pose a threat to *Cucurbita*. Thus the toxicity of wild *Cucurbita* and related species may be, in part, an adaptive strategy to facilitate beneficial seed dispersal by very large mammals while warding off smaller seed predators. The resulting unpalatability of *Cucurbita* to small-bodied mammals, combined with the megafaunal extinctions, would have left the plants lacking plausible dispersal partners and the associated nutrient-rich dung deposits important in seed germination and seedling development (28); thus intergenerational seed propagation would be curtailed and offspring would be more vulnerable to localized stressors.

Moreover, *Cucurbita* prefers field edges, floodplains, and other disturbed habitats and would have thrived in the niche-diverse, mosaic-like landscapes maintained by large herbivores (4). In the wake of these herbivores' extinction, large-scale zonal vegetative patterns replaced the fine-grained variation maintained by the megafauna (34), and weedy taxa would have been crowded out as their disturbed-ground niche disappeared. Taken together, these transitions could have been instrumental in driving wild *Cucurbita* into relict, refugial patches, causing fragmentation of previously contiguous distributions, and ultimately, in some cases, leading to the extinction of wild forms.

**Domestication as an Adaptive Strategy.** Against the backdrop of these ecological changes, the anthropogenic landscape provided a disturbed habitat readily colonized by weedy species, and humans fulfilled the dispersal role. Our phylogenetic results support previous hypotheses that the domestication process was widespread, occurring independently many times throughout the Americas, including in ENA (*C. pepo* ssp. *ovifera*), Mesoamerica (*C. pepo* ssp. *pepo* and *C. argyrosperma*, possibly *C. pepo* ssp. *fraterna*, *C. moschata*, and *C. ficifolia*), and South America (*C. maxima* and possibly *C. ecuadorensis* and *C. moschata*) (35). ENA is a long-debated putative source region for *Cucurbita pepo* ssp. *ovifera* var. *ovifera*, a hypothesis that is strongly supported by the clustering of *ovifera* cultivars with modern and ancient ENA gourds in our analysis. *C. pepo* ssp. *pepo*—unknown in the wild—originated in the Valley of Oaxaca around 10,000 y B.P. (12) but also was prominent among our samples from northern Mexico and the southwestern

United States, several hundred miles to the north. This result suggests either (i) a widespread ancient distribution of this subspecies and the emergence of domestic forms across a broad geographic range, illustrating a range-wide shift to human landscapes as an adaptive strategy; or (ii) prolific human-mediated dispersal of this taxon throughout Mesoamerica, illustrating the massive reproductive and dispersal potential for plants under cultivation. Either possibility underscores the adaptive potential of domestication (36). One ancient *C. pepo* ssp. *fraterna* sample from northeastern Mexico also indicates the possible presence of this taxon in ancient crop fields. Presently, *C. pepo* ssp. *fraterna* exists only in the wild, suggesting that our ancient *fraterna* sample may represent a lost domestic lineage or an independent, subsequently arrested, pathway to domestication.

**Reproductive Isolation.** The fixation of domesticated phenotypes requires a level of reproductive isolation between cultivated and wild forms to prevent the constant back-crossing of unfavorable traits into populations under human selection (37, 38). For example, emmer wheat required removal from its source region to express the phenotypic effects of domestication robustly (39), and it has been suggested that maize flourished into its dramatic, familiar, hyperdomesticated forms only after being transported out of southern Mesoamerica (40). This requirement may vary by species, with the level of isolation being linked to the genetic configuration and cultural importance of domestication traits. In *Cucurbita*, the requirement for reproductive isolation is especially pronounced, with cross-fertilization from wild forms producing unpalatable fruits (10), likely through the up-regulation of cucurbitacin production. That is, without isolation from wild stands, it is unlikely that *Cucurbita* could have evolved into a form suitable for human consumption. During the Holocene, the collapse of wild populations following the onset of a relationship with humans ultimately may have prevented gene flow from wild types constantly recharging crop fields with the allelic predisposition for a tough rind and bitter cucurbitacins. *Cucurbita* species have thrived in the Holocene largely by partnering with humans through domestication as wild populations dwindled and have been bred to lower their natural defenses for human palatability. At the same time, the fixation of important domestication alleles conferring palatability might have been impossible if wild plants had not largely receded and disappeared.

## Materials and Methods

### Plastid Genome Analysis.

**Sample materials.** Details of all modern and ancient *Cucurbita* samples analyzed are reported in Tables S1–S3. Modern samples were provided by L.A.N. and the US Department of Agriculture Agricultural Research Service National Plant Germplasm System. Access to ancient samples was granted by B.D.S., the Illinois State Museum, the National Park Service, and the University Museum at the University of Arkansas.

**DNA isolation, sequencing, and assembly.** In modern samples, we used sterile razor blades to excise embryonic tissue from seeds and extracted whole genomic DNA from the embryos using Qiagen DNeasy Plant Mini Kits, following the manufacturer's protocol after grinding tissue in the kit lysis buffer with sterile pellet pestles. In modern-sample representatives from each separate taxon, we first PCR-amplified long, overlapping fragments of the plastid genome with primers in conserved regions (see Table S4 for primer sequences) designed using an alignment of cucumber (*Cucumis sativus*, Cucurbitaceae) and karaka (*Corynocarpus laevigatus*, Corynocarpaceae) plastid genomes, both order Cucurbitales. We were unable to amplify complete plastid genomes in any taxon using long-range PCR (LR-PCR), but we recovered several long fragments to assist with plastid genome assembly. We separately sheared LR-PCR product pools and whole genomic DNA to ~350 bp using a Covaris model S2, size-selected the sheared DNA at ~350 bp on a 1.6% low-melt agarose gel, and prepared barcoded Illumina libraries (41). We pooled all barcoded libraries equally among samples, with 20% of estimated pool molarity comprising LR-PCR products and the remainder whole genomic DNA. We sequenced the pool on one lane of an Illumina HiSeq 2000 flow cell with a 101-nt paired-end configuration. To acquire additional *C. pepo* and *C. moschata* plastid genome sequences to compare with the initial assemblies, we shotgun-sequenced eight additional plastid

genomes on one lane of an Illumina HiSeq 2500 flow cell with a 151-nt paired-end configuration in Rapid Run mode.

We divided LR-PCR-derived sequence data into groups of 100,000 read pairs and shotgun sequence data into 1,000,000 pair subsets to normalize coverage for de novo assembly. We used Velvet 1.2.08 to iterate over LR-PCR subsets to produce long, high-quality contigs from the high-coverage enriched plastid regions. We combined all LR-PCR contigs per sample and then used Velvet (42) to iterate over shotgun data subsets using stringent parameters (long k-mer value and high coverage cutoff) and included LR-PCR contigs as long reads to anchor the assembly. We combined all resulting contigs and reiterated over shotgun subsets using new contigs as long reads until no substantial increase in contig length and quality was observed (two to five iterations). We used Burrows–Wheeler Aligner (BWA) (43) to align resulting contigs from one taxon, *Cucurbita pepo* ssp. *fraterna*, to the plastid genome of cantaloupe (*Cucumis melo*, Cucurbitaceae), a relative with an estimated 30-Mya divergence date from *Cucurbita* (44). We then used the resulting alignment to identify contigs covering the majority of the plastid genome and used pairwise alignment to assemble them to the cantaloupe reference sequence. We used bash utilities to close the remaining short gaps by reiteratively querying contig terminal sequences from the short-read data and realigning matching short reads to the working assembly until reaching the next adjacent contig. We used YASRA (45) with LASTZ (46) to map shotgun reads from other taxa to the circularized *C. pepo* ssp. *fraterna* plastid reference assembly reiteratively and closed gaps using de novo contigs and manual gap-filling as described above. Two short stretches (estimated <200 nt based on comparison with the cantaloupe reference sequence) in the long single-copy region containing tandem repeats were unresolvable using short-read data, one in the *rps4-ndhJ* intergenic spacer region and the other apparently within the coding region of *accD*. These regions were hard-masked for analysis. After the initial assembly of plastid genomes across all taxa, we used BWA (43) and SAMtools (47) to align short reads from subsequent samples to the appropriate reference and call consensus sequences.

Following plastid genome assembly for all available taxa, we used the *C. pepo* ssp. *pepo* and *C. moschata* plastid consensus sequences to design target capture probes for plastid DNA enrichment by RNA hybridization (18). We extracted DNA and prepared sequencing libraries from 22 additional modern specimens of *C. pepo* and 23 additional modern *C. moschata*. In addition, we extracted DNA from 33 archaeological *Cucurbita* rind, seed, and peduncle specimens using an *N*-phenacylthiazolium bromide (PTB)-based extraction protocol (48) and prepared sequencing libraries (41). Because of the fragmentary, low-copy, and contamination-prone nature of ancient DNA, stringent protocols were observed during ancient DNA extraction and library preparation. All pre-PCR handling of ancient tissue and DNA was carried out in a dedicated sterile laboratory physically isolated from any molecular biology building, with HEPA filtration, positive pressure, frequent decontamination using a strong bleach solution, and workflow protocols to prevent modern contamination of the clean laboratory. We enriched modern and ancient DNA sequencing libraries for plastid DNA using the MYbaits kit (Mycroarray), according to the manufacturer's protocol but with two extra wash repetitions and an 18-cycle PCR reamplification using primers IS5 and IS6 (41) after hybridization. For the modern samples, we combined the barcoded libraries into pools of 10 samples before the capture step and carried each pool through a single capture reaction. The enriched libraries were sequenced in parallel on an Illumina HiSeq 2500.

For modern libraries, we used BWA (43) to map reads to the appropriate reference plastid assembly and then used the SAMtools (47) mpileup function and a custom Perl script to call majority-rule haplotype consensus sequences. For ancient samples, we first merged overlapping paired-end reads after Kircher (49) and then BWA-mapped (43) merged reads to the *C. pepo* ssp. *fraterna* plastid reference sequence. We estimated the level of cytosine deamination at fragment ends—characteristic ancient DNA damage most prevalent in single-stranded overhangs (50)—by analyzing base mismatches to the reference sequence relative to base position on the fragment, revealing an average 8-nt interval of elevated 5' C->T bias and 3' G->A bias (Fig. S1). This observation helped authenticate our ancient DNA data and indicated that the region within 8 nt of read ends, on average, might be prone to damage-derived sequence error. To mitigate any effects of these biases in our analysis, we hard-masked all potentially damaged positions (5' thymines and 3' adenines) within appropriate distance of fragment ends (Fig. S1 and Tables S1–S3). We then called haplotype consensus sequences as above, enforcing a minimum 2× nonredundant coverage and 80% site identity for a consensus base call. Under these requirements, we recovered plastid assemblies between 10.9% and 99.8% complete for all ancient samples (Tables S1–S3). We discarded samples with <65% coverage of the single-copy regions and one of the two inverted repeat regions, leaving a total of 19 ancient samples for analysis, 15 of which had ≥97% coverage.

We aligned all plastid genome sequences ( $n = 91$ ) using MAFFT (51), manually checked the alignment, and manually masked any short regions of poor mapping quality caused by long homopolymer repeats prone to polymerase and sequencing error (aligned and curated sequences provided as Dataset S1). We used a log-likelihood ratio test to select the Hasegawa, Kishino, and Yano (HKY) substitution model with four gamma-distributed rate categories, and used PhyML (52) with 100 bootstrap replicates to reconstruct a phylogenetic tree (Fig. 1). We also confirmed the tree topology by running a Bayesian phylogenetic analysis to convergence using BEAST 2 (53), enforcing a strict molecular clock and implementing the same evolutionary model described for the maximum likelihood analysis.

### Mammalian *TAS2R* Analysis.

***TAS2R* gene identification.** We retrieved 46 mammal genome assemblies from public repositories (Tables S5 and S6), representing all scaffold- or chromosome-level therian mammal genome assemblies available at the time of data acquisition (April, 2014), excluding the bottlenose dolphin. Extant whales possess only pseudogenized *TAS2R* homologs because of the complete loss of functional *TAS2R* genes in a common ancestor living 36–53 Mya (33). Therefore, dolphins have never possessed a suite of functional bitter taste receptors (33), rendering analysis of subsequent evolution of functional *TAS2R* suites alongside phenotypic and behavioral variables moot. To identify intact, putatively functional *TAS2R* gene copies in the mammalian genomes, we first adapted the strategy of Li and Zhang (30) by generating BLAST databases (54) from each genome assembly and queried them for the annotated human ( $n = 25$ ) and mouse ( $n = 35$ ) *TAS2R* protein sequences (30) using TBLastN, enforcing an e-value threshold of  $\leq 10^{-10}$ . This approach previously has proven robust for localizing *TAS2R* variants across diverse species, with conserved transmembrane domains in *TAS2R* genes providing reliable amino acid targets for the TBLastN queries (30). From the TBLastN output, we generated a bed file for each species containing genomic coordinates of candidate *TAS2R* regions, merged overlapping regions, added 100 nucleotides to each flank, and used BEDtools (55) to generate nucleotide FASTA files for the corresponding regions from the genome assemblies. We then used a Perl script to translate these nucleotide data into amino acid sequences in all six reading frames and return the longest putative intact gene from each region, defined previously (30) as having at least 270 amino acid residues, starting and ending with start and stop codons, and being uninterrupted by premature stop codons. We used MAFFT (51) to align all resulting protein sequences and visually inspected the result to ensure plausible multiple alignment among all copies, validating these genes as *TAS2R* homologs. This strategy resulted in the recovery of 933 intact *TAS2R* copies among the 46 species analyzed. To minimize any remaining taxonomic or dietary acquisition biases associated with using human and mouse query sequences, we repeated the entire screening process a second time, using all *TAS2R* genes recovered in the first pass as input queries. In the second pass, we recovered 20 additional genes across 14 different species, bringing the total *TAS2R* count to 953 to carry into our analyses (provided as Datasets S2 and S3).

**Phenotypic character assignments.** Body mass was assigned based on typical values for adult males of each species from the University of Michigan Animal Diversity Web (ADW) (56) or from adults in general when sexual dimorphism data were not available. For dietary breadth, we assigned each mammal species a value from 1 to 4, where 1 signifies a broad-spectrum dietary generalist (omnivorous species with a wide range of foraging behaviors, e.g., mouse, pig, human) and 4 indicates a narrow dietary specialist (species with extremely specialized dietary behaviors, especially species for which the diet may consist almost entirely of a single species or a few similar species, e.g., bamboo for the giant panda or highland grasses for the alpaca), using ADW descriptions of diet as guidelines (Tables S5 and S6). Diet breadth assignments were made without prior knowledge of *TAS2R* count to avoid undue influence on category choice.

**Analysis of *TAS2R* copy number variation.** To test for the effects of genome assembly quality on the sensitivity of our technique to detect *TAS2R* copies in poorer assemblies, we first tested for a relationship between  $\log_{10}$  scaffold N50 and *TAS2R* count using a linear model alongside  $\log_{10}$  body mass and diet breadth, using the R statistical package (57). We identified a highly significant correlation between scaffold N50 and *TAS2R* count ( $P = 0.003$ ). This effect apparently is caused by difficulty in identifying intact *TAS2R* variants in more fragmentary, lower-quality genome assemblies. When assemblies with N50 <50 kb ( $n = 5$ ) were excluded, significance decreased dramatically ( $P = 0.103$ ). Conservatively, we chose to exclude assemblies with N50 <250 kb ( $n = 8$ ) from analysis, eliminating the significant relationship between assembly quality and intact *TAS2R* count ( $P = 0.863$ ). To analyze associations between phenotypic characters and *TAS2R* gene count while

correcting for phylogenetic influence, we implemented a phylogenetic generalized least squares (pGLS) approach within the “caper” R package (58). We constructed an ultrametric phylogenetic mammal supertree manually from the literature (Fig. S2 and Dataset S4) (59–61). We used maximum likelihood optimizations of the delta, kappa, and lambda branch length transformations implemented in the pGLS (detailed in ref. 58).

**ACKNOWLEDGMENTS.** We thank Craig Praul and Candace Price from the Pennsylvania State University Huck Institutes DNA Core Laboratory for

assistance with sequence data collection and the US Department of Agriculture Agricultural Research Service National Plant Germplasm System, the Illinois State Museum, the National Park Service, and the University Museum at the University of Arkansas for access to modern and ancient sample materials. Research was supported by The Pennsylvania State University Huck Institutes of the Life Sciences and College of the Liberal Arts (G.H.P.), Wenner-Gren post-PhD Research Grant 8770 and Natural Environment Research Council Independent Research Fellowship NE/L012030/1 (to L.K.), and the Smithsonian Institution (B.D.S.). Instrumentation was funded by the National Science Foundation through Grant OCI-0821527.

- Simpson MG (2006) *Plant Systematics* (Elsevier/Academic, Boston).
- Newsom LA, Muhlbacher MC (2006) Mastodons (*Mammot americanum*) diet foraging patterns based on analysis of dung deposits. *First Floridians and Last Mastodons: The Page-Ladson Site in the Aucilla River*, ed Webb SD (Springer, Dordrecht, The Netherlands), pp 263–331.
- Cowan CW, Smith BD (1993) New perspectives on a wild gourd in Eastern North America. *J Ethnobiol* 13(1):17–54.
- Johnson CN (2009) Ecological consequences of Late Quaternary extinctions of megafauna. *Proc Biol Sci* 276(1667):2509–2519.
- Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of late Pleistocene extinctions on the continents. *Science* 306(5693):70–75.
- Janzen DH, Martin PS (1982) Neotropical anachronisms: The fruits the gomphotheres ate. *Science* 215(4528):19–27.
- Fritz GJ (1999) Gender and the early cultivation of gourds in Eastern North America. *Am Antiq* 64(3):417–429.
- Crook MR (2008) Bilbo (9Ch4) and Delta (38Ja23): Late Archaic and Early Woodland shell mounds at the mouth of the Savannah River. *Occasional Papers in Cultural Resource Management no. 17*.
- Guimarães PR, Galetti M, Jordano P (2008) Seed dispersal anachronisms: Rethinking the fruits extinct megafauna ate. *PLoS One* 3(3):e1745.
- Barlow C (2000) *The Ghosts of Evolution* (Basic Books, New York).
- Nee M (1990) The domestication of *Cucurbita* (Cucurbitaceae). *Econ Bot* 44(3):56–68.
- Smith BD (1997) The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* 276(5314):932–934.
- Decker-Walters DS (1990) Evidence for multiple domestications of *Cucurbita pepo*. *Biology and Utilization of the Cucurbitaceae*, eds Bates DM, Robinson RW, Jeffrey C (Cornell Univ Press, Ithaca, NY), pp 96–101.
- Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L (2002) Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proc Natl Acad Sci USA* 99(1):535–540.
- Ferriol M, Picó B, Belen P (2008) Pumpkin and winter squash. *Handbook of Plant Breeding. Vol 1. Vegetables I.*, eds Prohens J, Nuez F (Springer, Heidelberg), pp. 317–349.
- Tallamy DW, Krischik VA (1989) Variation and function of cucurbitacins in *Cucurbita*: An examination of current hypotheses. *Am Nat* 133(6):766–786.
- David A, Vallance DK (1955) Bitter principles of Cucurbitaceae. *J Pharm Pharmacol* 7(1):295–296.
- Gnirke A, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27(2):182–189.
- Smith BD (2006) Eastern North America as an independent center of plant domestication. *Proc Natl Acad Sci USA* 103(33):12223–12228.
- Andres TC (1987) *Cucurbita fraterna*, the closest wild relative and progenitor of *C. pepo*. *Cucurbit Genet Coop Rep* (10):69–71.
- Decker-Walters DS, Staub JE, Chung SM, Nakata E, Quemada HD (2002) Diversity in free-living populations of *Cucurbita pepo* (Cucurbitaceae) as assessed by random amplified polymorphic DNA. *Syst Bot* 27(1):19–28.
- Decker-Walters DS, Walters T, Cowan CW, Smith BD (1993) Isozymic characterization of wild populations of *Cucurbita pepo*. *J Ethnobiol* 13(1):55–72.
- Smith BD (2006) Seed size increase as a marker of domestication in squash (*Cucurbita pepo*). *Documenting Domestication: New Genetic and Archaeological Paradigms*, eds Zeder MA, Bradley D, Emshwiller E, Smith BD (Univ of California Press, Berkeley, CA), pp 25–31.
- Walters TW, Decker-Walters DS (1993) Systematics of the endangered okeechobee gourd (*Cucurbita okeechobeensis*: Cucurbitaceae). *Syst Bot* 18(2):175–187.
- Ferguson JE, Metcalf RL (1985) Cucurbitacins: Plant-derived defense compounds for diabroticites (Coleoptera: Chrysomelidae). *J Chem Ecol* 11(3):311–318.
- Bemis WP, Curtis LD, Weber CW, Berry J (1978) The feral buffalo gourd, *Cucurbita foetidissima*. *Econ Bot* 32(1):87–95.
- Gry J, Søborg I, Andersson H (2006) *Cucurbitacins in Plant Food* (TemaNord, Copenhagen).
- Dinerstein E, Wemmer C (1988) Fruits rhinoceros eat: Dispersal of *Trewia nudiflora* (Euphorbiaceae) in lowland Nepal. *Ecology* 69(6):1768–1774.
- Kusin S, Angert T, Von Derau K, Horowitz BZ, Giffin S (2012) Toxic squash syndrome: A case series of diarrheal illness following ingestion of bitter squash. Poster presentation at the annual meeting of the North American Congress of Clinical Toxicology.
- Li D, Zhang J (2014) Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Mol Biol Evol* 31(2):303–309.
- Dong D, Jones G, Zhang S (2009) Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evol Biol* 9:12–20.
- Wang X, Thomas SD, Zhang J (2004) Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum Mol Genet* 13(21):2671–2678.
- Feng P, Zheng J, Rossiter SJ, Wang D, Zhao H (2014) Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol Evol* 6(6):1254–1265.
- Williams J, Shuman BN, Webb T, Bartlein PJ, Leduc PL (2004) Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecol Monogr* 74(2):309–334.
- Piperno DR, Stothert KE (2003) Phytolith evidence for early Holocene domestication in southwest Ecuador. *Science* 299(5609):1054–1057.
- Allaby RG, et al. (2015) Archaeogenomic insights into the adaptation of plants to the human environment: Pushing plant–hominin co-evolution back to the Pliocene. *J Hum Evol* 79:150–157.
- Jones M, Brown T (2009) Selection, cultivation, and reproductive isolation: A re-consideration of morphological and molecular signals of domestication. *Rethinking Agriculture: Archaeological and Ethnoarchaeological Perspectives*, eds Denham T, Iriarte J, Vrydaghs L (Left Coast, Walnut Creek, California), pp 36–49.
- Dempewolf H, Hodgins KA, Rummell SE, Ellstrand NC, Rieseberg LH (2012) Reproductive isolation during domestication. *Plant Cell* 24(7):2710–2717.
- Civán P, Ivaničková Z, Brown TA (2013) Reticulated origin of domesticated emmer wheat supports a dynamic model for the emergence of agriculture in the fertile crescent. *PLoS One* 8(11):e81955.
- Webster DL (2011) Backward Bottlenecks. *Curr Anthropol* 52(1):77–104.
- Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):t5448.
- Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Schaefer H, Heibl C, Renner SS (2009) Gourds afloat: A dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proc R Soc B* 276(1658):843–851.
- Ratan A (2009) *Assembly Algorithms for Next-Generation Sequence Data*. PhD dissertation (The Pennsylvania State University, State College, PA).
- Harris RS (2007) *Improved Pairwise Alignment of Genomic DNA*. PhD dissertation (The Pennsylvania State University, State College, PA).
- Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Kistler L (2012) Ancient DNA extraction from plants. *Ancient DNA: Methods and Protocols, Methods in Molecular Biology*, eds Shapiro B, Hofreiter M (Humana, New York), Vol 840, pp 71–79.
- Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Ancient DNA: Methods and Protocols, Methods in Molecular Biology*, eds Shapiro B, Hofreiter M (Humana, New York), Vol 840, pp 197–228.
- Gilbert MTP, et al. (2007) Recharacterization of ancient DNA miscoding lesions: Insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35(1):1–10.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059–3066.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704.
- Bouckaert R, et al. (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10(4):e1003537.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
- Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
- Myers P, et al. (2015) The Animal Diversity Web. Available at animaldiversity.org. Accessed January 1, 2015.
- R Development Core Team (2013) R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna) Available at www.R-project.org/. Accessed April 1, 2015.
- Orme D, et al. (2013) caper: Comparative Analyses of Phylogenetics and Evolution in R. R Package version 052. Available at CRAN.R-project.org/package=caper. Accessed April 1, 2015.
- Pozzi L, et al. (2014) Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Mol Phylogenet Evol* 75(1):165–183.
- Huchon D, et al. (2007) Multiple molecular evidences for a living mammalian fossil. *Proc Natl Acad Sci USA* 104(18):7495–7499.
- Meredith RW, et al. (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–524.