PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

The expansion of agriculture has shaped the recent evolutionary history of a specialized squash pollinator

Nathaniel S. Pope[a,b,1], Avehi Singh[a], Anna K. Childers[c], Karen M. Kapheim[d], Jay D. Evans[c], and Margarita M. López-Uribe[a,1]

Edited by Gene Robinson, University of Illinois at Urbana-Champaign Institute for Genomic Biology, Urbana, IL; received May 10, 2022; accepted February 16, 2023

The expansion of agriculture is responsible for the mass conversion of biologically diverse natural environments into managed agroecosystems dominated by a handful of genetically homogeneous crop species. Agricultural ecosystems typically have very different abiotic and ecological conditions from those they replaced and create potential niches for those species that are able to exploit the abundant resources offered by crop plants. While there are well-studied examples of crop pests that have adapted into novel agricultural niches, the impact of agricultural intensification on the evolution of crop mutualists such as pollinators is poorly understood. We combined genealogical inference from genomic data with archaeological records to demonstrate that the Holocene demographic history of a wild specialist pollinator of *Cucurbita* (pumpkins, squashes, and gourds) has been profoundly impacted by the history of agricultural expansion in North America. Populations of the squash bee *Eucera pruinosa* experienced rapid growth in areas where agriculture intensified within the past 1,000 y, suggesting that the cultivation of *Cucurbita* in North America has increased the amount of floral resources available to these bees. In addition, we found that roughly 20% of this bee species' genome shows signatures of recent selective sweeps. These signatures are overwhelmingly concentrated in populations from eastern North America where squash bees were historically able to colonize novel environments due to human cultivation of *Cucurbita pepo* and now exclusively inhabit agricultural niches. These results suggest that the widespread cultivation of crops can prompt adaptation in wild pollinators through the distinct ecological conditions imposed by agricultural environments.

crop cultivation | agricultural adaptation | *Cucurbita* | bees

Significance

The conversion of natural to agricultural environments results in a dramatic modification of existing ecological conditions, and there are well-studied examples of crop pests that have rapidly evolved to fill novel agricultural niches. However, the degree to which agricultural intensification influences the evolution of wild insect pollinators is unknown, despite the importance of these mutualists to the global food supply and the persistence of plant populations. This study demonstrates that historical human agriculture in North America has had a profound impact on the recent evolutionary history of a wild, squash-specialized bee that is an essential pollinator of cucurbit crops. This provides a clear example of the role of agriculture as an evolutionary force acting on wild insect pollinators.

Large-scale agriculture is a leading cause of worldwide declines in biodiversity (1) but has also facilitated the population growth and spread of many nondomesticated ("wild") species (2–5). The agriculturalization of landscapes alters the abiotic and biotic conditions experienced by organisms, creating novel niches spanning large geographic areas. Agricultural environments characteristically possess managed populations of genetically homogeneous crop species, artificial supplementation of water, and extreme environmental conditions due to dramatic changes in microclimate, soil structure, and plant density relative to the original environment (3). These changes have ecological and evolutionary consequences for wild species that interact with crops. For example, as agriculture has expanded into new geographic regions, many insect herbivores have experienced range expansions and undergone speciation through host plant shifts or adaptation to agricultural environments (6, 7). However, the impacts of the expansion of agriculture on associated wild pollinators are poorly understood. In part, this is because wild pollinators are rarely exclusive to agricultural areas or agricultural plants, making it difficult to study the role of agriculturalization as an evolutionary force in these beneficial species (3).

Bees are the most important pollinators of flowering plants and are common floral visitors for 70% of the world's leading crop species (8, 9). As agricultural lands now cover approximately 40% of the Earth's surface, these pollinating insects increasingly inhabit novel agricultural environments and collect floral resources from crop plants. Managed pollinators such as honey bees have been purposely introduced into agricultural systems by humans (10), but the extent to which the conversion of natural to agricultural ecosystems has altered the recent evolutionary history of wild bee species is practically unknown (11). It is likely that the artificial selection and widespread cultivation of crops have facilitated the adaptation of insect pollinators to agricultural environments through selective pressures driven by changes in the phenology, availability and composition of floral resources (6). Despite the essential role that wild bees play in global food production, their ability to transition into agricultural niches—where most available floral resources

Author contributions: N.S.P. and M.M.L.-U. designed research; N.S.P., A.S., A.K.C., K.M.K., J.D.E., and M.M.L.-U. performed research; N.S.P. contributed new reagents/analytic tools; N.S.P. analyzed data; and N.S.P. and M.M.L.-U. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: mml64@psu.edu or natep@uoregon.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2208116120/-/DCSupplemental.

Published April 3, 2023.

belong to crop plants that have been bred for agronomically desirable traits—has not been previously investigated.

The squash bee *Eucera* (*Peponapis*) *pruinosa* offers a unique opportunity to investigate how bee pollinators evolve when restricted to agricultural habitats. Squash bees—members of the subgenera *Eucera* (*Peponapis*) and *Eucera* (*Xenoglossa*) (12)—are narrow specialists on the pollen of the plant genus *Cucurbita*, including economically important crops such as pumpkins, squash, zucchini, and gourds. Historically, *E. pruinosa* used the perennial wild buffalo gourd *C. foetidissima* in the deserts of Mexico and the southwestern United States as its primary source of pollen (13). Following the domestication and widespread cultivation of *Cucurbita* crops in North America (14), *E. pruinosa* began to collect pollen from domesticates in addition to wild plants, and its contemporary distribution extended beyond that of *C. foetidissima* (11). The disconnect between the distribution of the bee and its wild host is most striking in the eastern United States and Canada (15–17). As early as 7,000 y ago, *C. pepo* ssp. *ovifera* was cultivated in the Eastern Woodlands (Missouri, the United States, and vicinity), and by 1,000 y ago, *C. pepo* ssp. *pepo* (independently domesticated 10,000 y ago in central Mexico) had become an essential component of large-scale maize cropping systems in the region (18). Thus, the current abundance and geographic distribution of the bee *E. pruinosa* are consequences of the widespread cultivation of domesticated squash plants. However, the timeframe of the transition of this wild pollinator from natural to agricultural habitats is unknown, as are any genetic and phenotypic changes that may have occurred as a result.

In this study, we take advantage of the specialized relationship between squash bees and *Cucurbita* to investigate the evolution of a wild insect pollinator into a novel agricultural niche. To do so, we used a combination of whole genome and reduced representation sequencing to infer genetic structure and genealogical relationships across the range of the bee. We then developed an algorithm to infer historical demographic parameters from genealogical coalescence rates, that allows joint estimation of time-varying migration and effective population size across an arbitrary number of populations. In doing so, we were able to demonstrate that *E. pruinosa* is a complex composed of historically isolated lineages that originated long before the development of agriculture in North America. However, the widespread cultivation of domesticated *Cucurbita* that started approximately 1,000 y ago significantly shaped the recent demographic history of *E. pruinosa*, causing dramatic increases in effective population size across its range. That is to say, the spread of domesticated host plants cultivated in agricultural habitats at a continental scale led to extreme population expansions in the bee, *E. pruinosa*. We used a model of background selection as a null hypothesis against which to detect regional adaptation and found that signatures of positive selection in *E. pruinosa* are largely consistent with selective sweeps that initiated within the past 5,000 y. Further, these signatures of recent selection are overwhelmingly concentrated in a lineage in eastern North America that inhabits agricultural areas and depends entirely on cultivated *Cucurbita* for pollen. Within this lineage, genes associated with sensory function are overrepresented in genomic regions predicted to be under selection. To our knowledge, this is the clearest evidence to date of the effect of agricultural expansion on the evolution of a wild, crop mutualist. These findings provide evidence that the expansion of agriculture can shape the evolutionary trajectories of wild pollinators in profound ways and identify candidates for functional traits that may facilitate the successful adaptation of pollinators to novel agricultural environments.

Results

Phylogeographic Structure in *E. pruinosa* Predates Squash Cultivation. We used genotypes from 111,296 restriction-site associated SNPs and five microsatellite loci in 26 populations (1,079 individuals, *SI Appendix*, Table S1) to delimit genetic structure across the range of *E. pruinosa*. These bees separate into five major genetic clusters that are highly divergent from one another and have unambiguous geographic identities (Fig. 1 *A* and *B*). The phylogenetic relationships and geographic locations of clusters suggest that *E. pruinosa* originated in northern Mexico and the southwestern United States and then split into southern, western, and eastern lineages separated by mountains in the center of the continent (Fig. 1*A* cladogram). Populations in Arizona and west Texas—the probable northern range limit of *E. pruinosa* during the Last Glacial Period (LGP)—contain genetic material associated with all these lineages. To infer the timing of this diversification, we generated a 409-Mb chromosome-scale reference for *E. pruinosa* and sequenced the genomes of 44 haploid males from five populations (labels in Fig. 1*A*). These targeted populations were chosen to represent the three major phylogeographic divisions and the bee's ancestral range and to cover areas where the historical cultivation of *Cucurbita* could have facilitated range expansion. We inferred sample genealogies and recombination breakpoints (19) across these whole genome sequences using 3,674,130 biallelic SNPs; we then used fine-scale recombination maps, a model of background selection (*SI Appendix*, section 3), and a scan for selective sweeps to mask parts of the genome likely to be subject to purifying or positive selection. Next, we developed a method to jointly estimate time-varying effective population sizes and migration rates from the coalescence times of trios embedded within neutral genealogies (*SI Appendix*, section 4, Figs. S1–S5). The demographic parameter estimates generated by this method accurately recapitulated observed summary statistics (*SI Appendix*, Table S2) and indicate that the division between the southern and northern lineages occurred prior to the start of the LGP, while the separation of the northern lineage into eastern and western branches occurred between 90,000 and 50,000 y ago (Fig. 1*C* and *SI Appendix*, Fig. S6). These dates fall long prior to the Holocene, indicating that the origins of these major lineages predate human agriculture (20).

Population Expansions in *E. pruinosa* Are Concurrent with *Cucurbita* Agriculture. The contemporary range of *E. pruinosa* extends far beyond that of its primary wild host *C. foetidissima*, especially in eastern North America where the bee's only source of pollen is from cultivated *Cucurbita* (Fig. 2*A*). To characterize the geographic distribution of *E. pruinosa* prior to this range expansion, we used a species distribution model and historical climate projections to hindcast the range of *C. foetidissima*. This wild gourd was largely restricted to the southwestern United States and Mexico at the end of the Last Glacial Maximum (21,000 y ago) before spreading northward into the center of the continent during the Holocene (Fig. 2*A*). We next collated records of radiocarbon-dated *Cucurbita* remains from archaeological sites to determine when human cultivation of squash could have facilitated the dispersal of *E. pruinosa* into novel mesic habitats. These indicate that cultivated *C. pepo* were widespread in eastern North America as early as 7,000 y ago but were introduced into the southwestern United States only 2,000 to 3,000 y ago (Fig. 2*B*).

Finally, to relate the recent history of *E. pruinosa* to major transitions in human agricultural practices, we fit a second

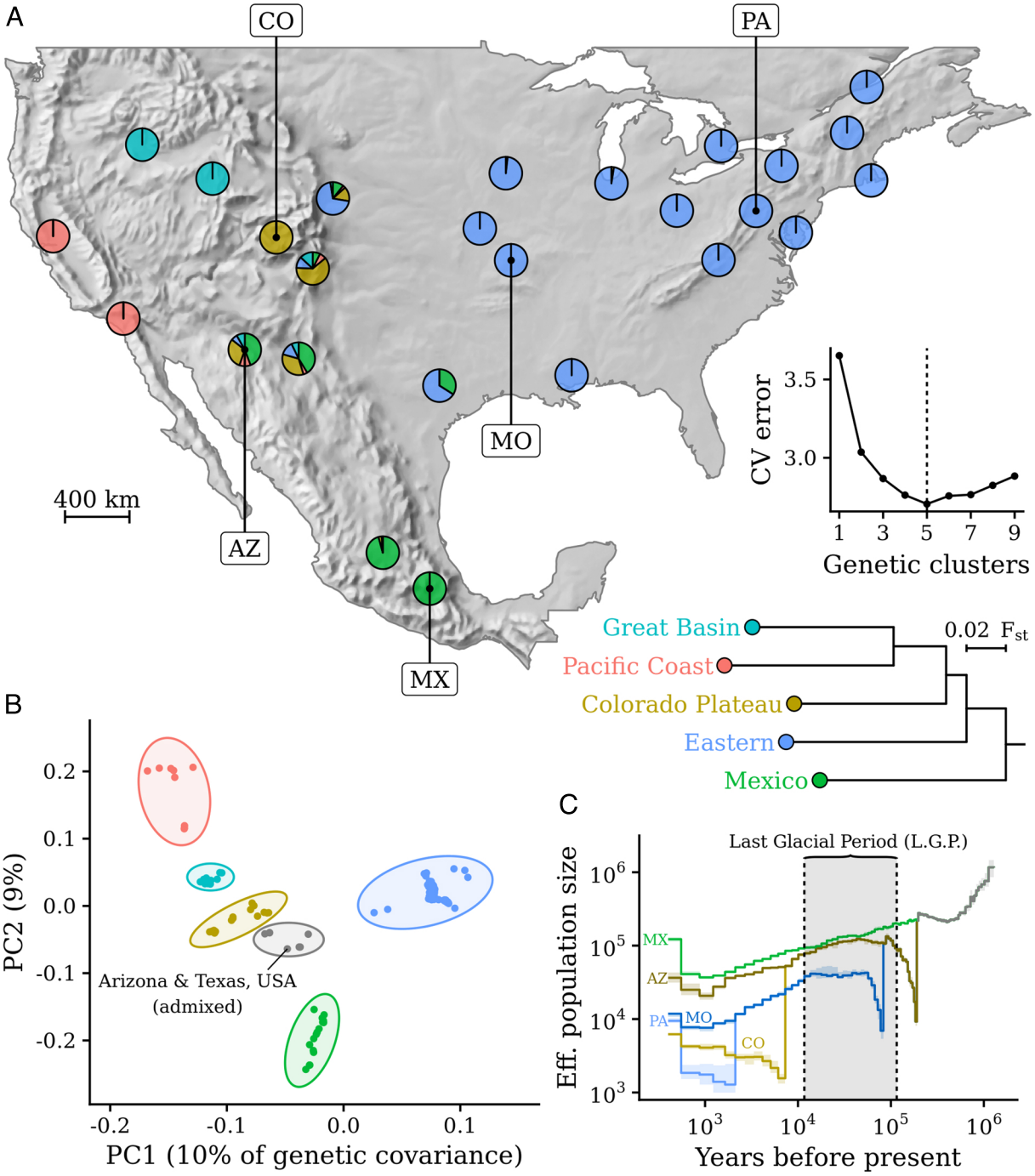

Fig. 1. The diversification of *E. pruinosa* predates squash cultivation. (*A*) Geographic genetic structure in *E. pruinosa*, shown as admixture coefficients per population (pie charts) for five genetic clusters (colors), inferred from 111,296 restriction-site associated SNPs and 5 microsatellite loci. The number of clusters was selected by cross-validation. The cladogram shows phylogenetic relationships among clusters inferred using maximum likelihood estimates of allele frequencies for each cluster and is scaled so that the cophenetic distance between tips equals the F_{st} between the associated clusters (*SI Appendix*, Table S2 for pairwise F_{st} between labeled populations). These indicate that *E. pruinosa* expanded northward out of Mexico, splitting into eastern and western lineages in the process. On the basis of this delimitation of genetic structure, the labeled populations were targeted for whole genome sequencing. These were chosen to cover the major phylogeographic divisions and to include geographically intermediate localities. (*B*) Genetic differentiation of *E. pruinosa* individuals sampled from across the species' range, shown by the positions of individuals along the first two principal components of normalized SNP genotypes. Assignment to clusters was done via the largest admixture coefficient of the associated population, excluding the highly admixed populations in Arizona/Texas. The orientation of clusters in principal component space closely resembles their geographic positions. (*C*) Haploid effective population sizes for the labeled populations across 100 epochs, jointly estimated using trio coalescence rates in 44 whole genome sequences (*SI Appendix*, section 4). These populations were chosen to represent the major phylogeographic divisions in the bee and are colored according to the genetic cluster. Shaded regions are 95% bootstrap confidence intervals, and population merger times were estimated via cross-validation. According to this demographic reconstruction, the major phylogeographic divisions of northern *E. pruinosa* originated during the Last Glacial Period (LGP). A geographic interpretation for the fitted model based on the predicted locations of ancestors is given in *SI Appendix*, Fig. S6.

demographic model to trio coalescence rates within 250-y epochs spanning the past 18,000 y (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S7). *Eucera pruinosa* populations outside of the range of wild *Cucurbita* (CO and PA in Fig. 2) originated during the Holocene, undergoing bottlenecks in the process (Fig. 2*C*). *Eucera pruinosa* in the southwestern United States (AZ in Fig. 2)

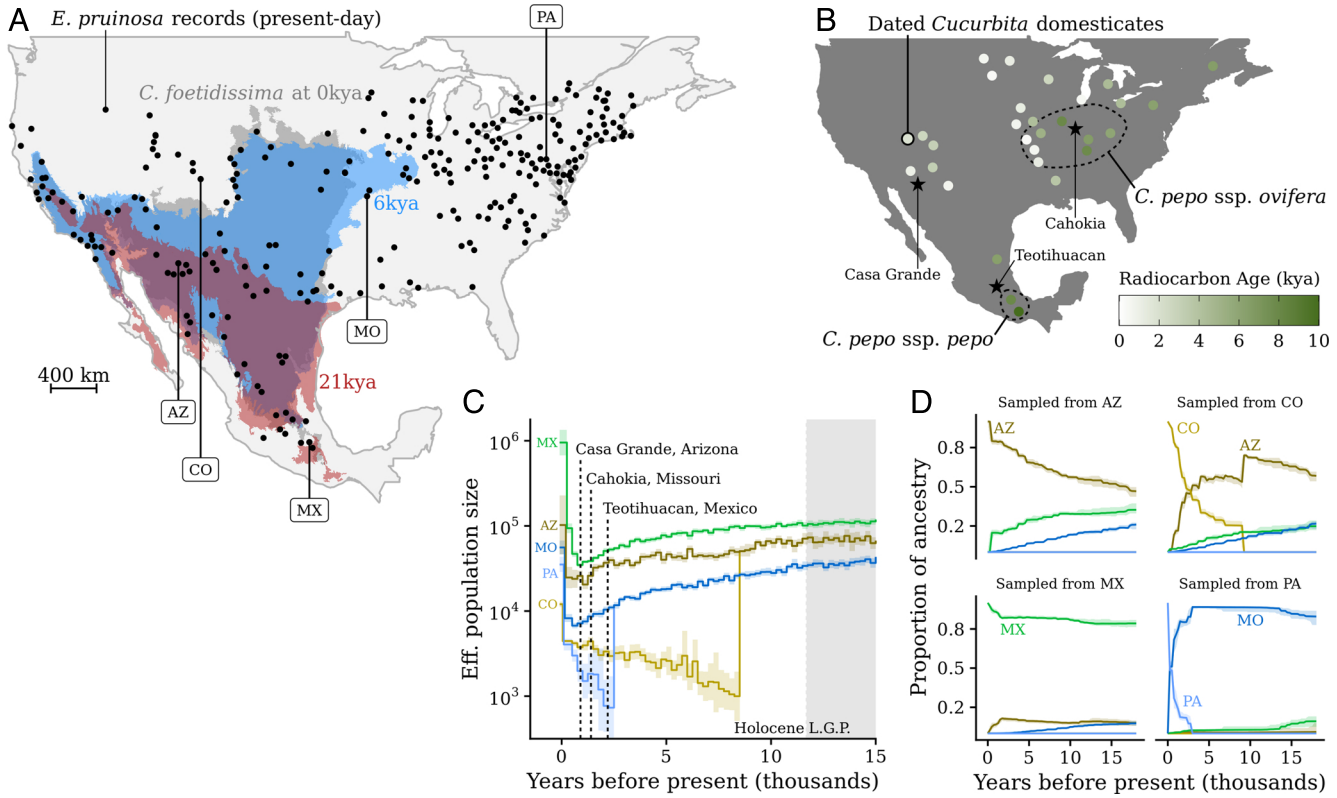

Fig. 2. Population expansions of *E. pruinosa* in North America follow transitions to large-scale agriculture. (*A*) The present-day distribution of *E. pruinosa* (points) overlaid on past and present distribution of its wild host *C. foetidissima* (red, blue, and gray polygons), the latter inferred via species distribution modeling applied to historical climatic projections. Labeled *E. pruinosa* occurrences are populations from which whole haploid genomes were used for demographic reconstruction. (*B*) Radiocarbon-dated *Cucurbita* tissue (rinds, stems, and seeds) from archaeological sites. Records are thinned to the oldest occurrences that are at least 150 km apart. The outlined areas are the putative independent origins for domesticated *C. pepo* ssp *pepo* and *C. pepo* ssp *ovifera* in Mexico and eastern North America, respectively (21). (*C*) Haploid effective population sizes for focal populations of *E. pruinosa* within 250-y intervals over the past 15,000 y, jointly estimated using trio coalescence rates across 44 whole genome sequences (*SI Appendix*, section 4). Shaded regions are 95% bootstrap confidence intervals around the estimates. These show rapid, range-wide population growth within the past 1,000 y, contemporary with agricultural intensification across pre-Columbian human civilizations in North America. Annotations show approximate dates for major human settlements made possible by large-scale agriculture. (*D*) Predicted ancestry for *E. pruinosa* populations given the fitted demographic model, moving backward in time from the present day. Each panel shows the proportion of ancestors of a sample that were located in a particular population at a particular time in the past (*SI Appendix*, section 4E). In samples from AZ, for example, the rapid increase in MX ancestry over the past 1000 y indicates a recent pulse of migration from the southern to the western lineages. These predictions of recent ancestry suggest that the eastern (PA and MO) and southern (MX) lineages were largely isolated during the Holocene but that the western lineage has experienced recent immigration from both eastern and southern lineages, consistent with the pattern of admixture in haplotypes from the southwestern United States shown in Fig. 1*A*. A geographic interpretation for these ancestry curves is shown in *SI Appendix*, Fig. S6.

diverged from the western *E. pruinosa* complex (CO in Fig. 2) at the beginning of the Holocene and gradually admixed with immigrants from the South and East (Fig. 2*D*). In contrast, the eastern bee lineage remained relatively isolated for much of the past 15,000 y, which may have resulted from its movement from northern Mexico into the eastern United States along with *C. foetidissima* (*SI Appendix*, Fig. S6). The northeastern lineage (PA in Fig. 2) originated 3,000 y ago, concurrent with the domestication of *C. pepo* ssp *ovifera* and the origins of agriculture in eastern North America (18). Population size increased rapidly and dramatically in all populations over the last 1,000 y (Fig. 2*C* and *SI Appendix*, Fig. S36), as did migration between the southwestern United States and central Mexico (Fig. 2*D*). The timing of this demographic shift coincided with a transition to large-scale agricultural systems across pre-Columbian North America that supported human settlements of unprecedented size, such as Teotihuacan in central Mexico, Cahioka in the eastern United States, and Casa Grande in the Sonoran desert (Fig. 2 *B* and *C*). This correspondence suggests that early agriculture had a profound impact on Holocene populations of this bee and provides an independent line of evidence for the timeline of *Cucurbita* agriculture in North America.

Signatures of Positive Selection Are Concentrated in *E. pruinosa* Exclusive to Cultivated *Cucurbita*. We used the SWEEPFINDER2 (22) composite likelihood ratio (CLR) statistic to infer genomic regions undergoing lineage-specific selective sweeps, given a null model of background selection (BGS) and variable recombination (Fig. 3). Observed nucleotide diversity (π) in the southern lineage closely matched predicted π under this null model. However, BGS was not sufficient to explain reductions in π observed within the two northern lineages, especially in *E. pruinosa* from eastern North America (Fig. 3*A*). Instead, many of the unusually low-diversity regions in this lineage were inferred to contain selective sweeps (on the basis of calibrated CLR scores, *SI Appendix*, Fig. S8). Reconstructed genealogies of these putative sweeps suggest that the haplotypes that swept to fixation may have originated within the last 5,000 y (Fig. 3*B*). On all major chromosomes, there are large genomic intervals (in some cases spanning Mb) that are essentially devoid of genetic polymorphism in the eastern lineage (Fig. 3 *C* and *D*, *SI Appendix*, Figs. S9–S28). The width of these "footprints" is driven by local recombination rates (*SI Appendix*, Fig. S29), especially at the highly repetitive edges of chromosomes where recombination is reduced. Thus, linkage to selective sweeps appears to have

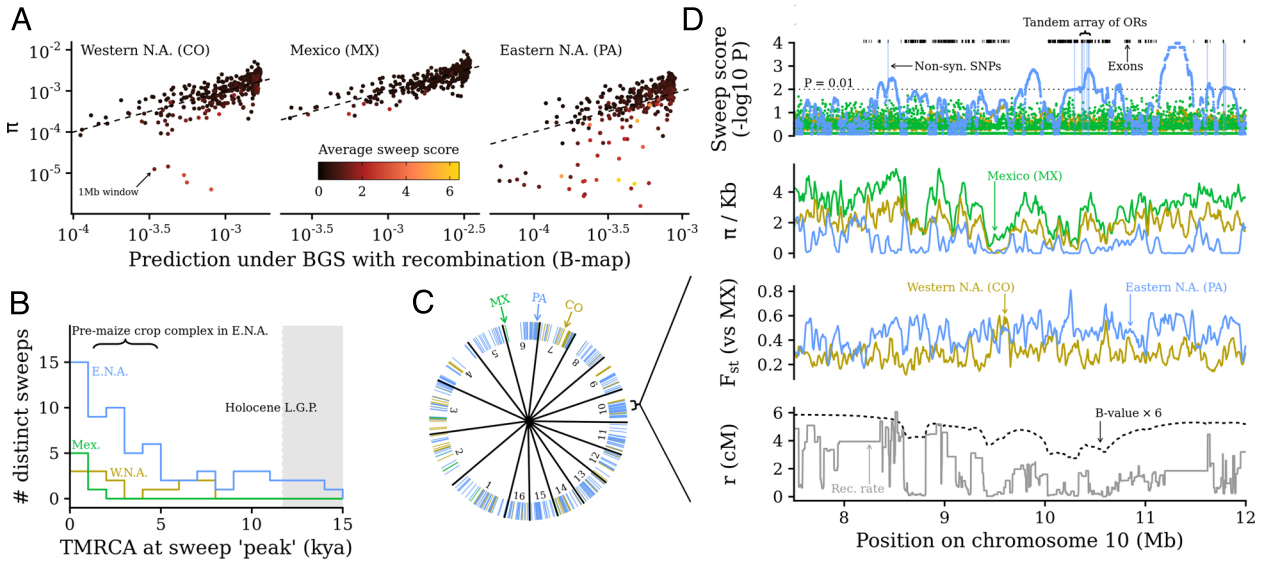

Fig. 3. Signatures of recent selection are concentrated in the *E. pruinosa* lineage in the eastern United States where bees subsist on *Cucurbita* agriculture without access to the wild host *C. foetidissima*. (*A*) Observed nucleotide diversity (π) against predicted diversity from a two-parameter model of background selection (BGS) and recombination (23) that incorporates annotations of conserved elements and coding regions along with inferred recombination rates (*SI Appendix*, section 3). The model provides an explanation for most low-diversity regions in MX and CO but cannot explain the extreme reductions in diversity observed in PA. These low-diversity "outliers" are typically enriched for putative selective sweeps, identified by comparing sweepfinder2 composite likelihood ratios (CLRs) to a simulated null distribution for each population. "Sweep score" refers to the negative $\log_{10}$ P-values calculated with reference to these null distributions. (*B*) The distribution of the genealogical time to the most recent common ancestor (TMRCA) of each population, at the highest-scoring position in each distinct sweep. Only top-scoring sweeps ($P_{CLR} < 0.01$) are included, with the additional requirement that samples from the swept population form a monophyletic clade. The branch lengths of genealogies were reestimated locally around each sweep by relate. The TMRCAs provide lower bounds for the origination of the swept haplotypes (assuming a single origin for the beneficial allele). The bracket delimits the existence of a premaize crop complex in eastern North America (present-day Missouri, Fig. 2*B*) that involved the domestication of *C. pepo* ssp *ovifera*. (*C*) The 16 scaffolds larger than 10 Mb in the reference genome have large "footprints" of sweeps in each population marked by colored lines (contiguous intervals ≥50 kb, where $P_{CLR} < 0.05$). (*D*) An example of the large sweeps mapped in (*C*), shown along a segment of the tenth-largest chromosome. Sweep scores ($-\log_{10} P_{CLR}$, *Top* panel) were calculated using inferred B-values and recombination maps (*Bottom* panel) to help distinguish signatures of positive and background selection. The bracketed exons are a tandem array of three olfactory receptors (ORs) with an unusual number of nonsynonymous substitutions in the eastern *E. pruinosa* lineage.

driven a large reduction in genetic polymorphism within this bee species. In the eastern lineage, 19.5% of the genome is inferred to be impacted by recent sweeps (i.e., with CLR scores exceeding the 95th quantile of a simulation-derived null distribution). Within these regions, the average nucleotide diversity is reduced by an order of magnitude ($\pi = 0.0002$ inside sweeps versus 0.00173 outside), substantially reducing levels of polymorphism for at least 1,600 protein-coding genes (*SI Appendix*, Fig. S30). Putative sweeps in the western lineage have reductions in diversity of a similar magnitude but cover less than 5% of the genome. Thus, signatures of directional selection are the most prominent in bees from eastern North America that subsist entirely on cultivated *C. pepo* and are concurrent with the development of a crop complex that included *C. pepo* ssp *ovifera* as a primary component (18).

Selection on Sensory Function Accompanies Agricultural Transition. To infer possible functional consequences of recent selection in *E. pruinosa*, we generated an annotation of protein-coding genes in the reference genome on the basis of RNA and protein evidence. To identify Gene Ontology (GO) terms associated with sweeps in each population across a range of critical values for the CLR, we used a test for gene set overrepresentation based on interval resampling (24) that accounts for the spatial clustering of genes with similar function (*SI Appendix*, Fig. S31). The only GO terms to show significant enrichment after this correction were associated with sweeps in the eastern lineage (PA in Fig. 2), and all were specific to sensory (especially olfactory) systems (Table 1 and *SI Appendix*, Table S3). For example, thirty-one (26%) of the 120 odorant receptor (OR) orthologs identified in *E. pruinosa*'s genome had CLR scores exceeding the 99th null quantile (as opposed to 8% expected under random interval sampling). Thirteen of these were dispersed across chromosomes in one- to three-gene arrays, and the remaining eighteen formed a large tandem array on chromosome 6 (inferred to be under selection in both eastern and western lineages, *SI Appendix*, Fig. S33). We identified 521 nonsynonymous, highly differentiated SNPs that were located within possible selective sweeps (*SI Appendix*, Fig. S32 and Data 3). These included substitutions in odorant receptors (e.g., Fig. 3*D*) as well as in UDP-glucuronosyltransferase, a detoxification gene associated with pesticide resistance (25) and inferred to be

Table 1. Significantly overrepresented gene ontology terms associated with recent selective sweeps in eastern *Eucera pruinosa*

Gene ontology term	# genes*	*P*-value (corrected)
Odorant binding (GO:005549)	30/132	1×10^{-6} (0.085)
Olfactory receptor activity (GO:004984)	31/120	5×10^{-8} (0.018)
Perception of chemical stimulus (GO:007606)	38/180	1×10^{-7} (0.008)
Perception of smell (GO:007608)	33/145	5×10^{-7} (0.018)
Sensory perception (GO:007600)	47/260	5×10^{-7} (0.022)

*In a sweep footprint (at $P_{CLR} \leq 0.01$) / in genome.

under positive selection in both eastern and western lineages (*SI Appendix*, Fig. S34).

Discussion

Here, we provide evidence that the expansion and intensification of agriculture has played a crucial role in shaping demographic and adaptive processes in a crop pollinator. Our results indicate that squash cultivation has increased habitat availability for *E. pruinosa* across its range but that recent adaptation has occurred primarily in eastern North America where the bees' habitat is exclusively agricultural. Specifically, we show that the effective population size of *E. pruinosa* began to rapidly increase starting approximately 1,000 y ago, across lineages that were historically isolated by topographic barriers. The timing of these increases roughly coincides with the origins of large-scale agriculture across North America (18) and may reflect the availability of cultivated *C. pepo* in these agroecosystems. In contrast to these synchronized changes in population size, we find that signatures of recent directional selection are overwhelmingly concentrated in a single lineage distributed across eastern North America. For most of the range of the eastern lineage of *E. pruinosa*, wild *Cucurbita* are absent, and squash bees exclusively occupy agricultural habitats. Signatures of adaptation in this lineage are consistent with selective sweeps initiating within the past 5,000 y, concurrent with anthropogenic alterations to the bees' environment: the development of a crop complex in eastern North America that included *C. pepo* ssp *ovifera* as a major component; the spread of agricultural habitats across the previously inaccessible eastern part of the continent; and the intensification of agriculture over the past millennium. Signatures of recent selection in *E. pruinosa* are far less evident in populations from the southwestern United States and central Mexico, where wild and crop *Cucurbita* co-occur (26). Thus, this study identifies an adaptive radiation of *E. pruinosa* facilitated by human agricultural practices and highlights the importance of agricultural niche creation as a key evolutionary force for insect pollinators.

These conclusions rely on the interpretation of the demographic and adaptive history of *E. pruinosa* within the context of the past distribution of its wild and domesticated hosts. Prior to the development of agriculture in the Holocene, *E. pruinosa*'s potential range would have matched the combined geographic distributions of wild *Cucurbita* spp in North America. On the basis of species distribution modeling and past climate projections, *Cucurbita* spp are predicted to have been restricted to lower elevations in Mexico and the southwestern United States during the Last Glacial Maximum (LGM, 21,000 y ago) (27). Thus, topographic barriers combined with cycles of warming and cooling during the Last Glacial Period may have driven the pre-Holocene diversification of *E. pruinosa* (Fig. 1). We hypothesize that the eastern *E. pruinosa* lineage originated in present-day Texas in a range overlap between *C. foetidissima* and the wild ancestor of *C. pepo*, where the bee would have been separated from southern and western lineages by the Sierra Madre Oriental and the southern Rocky Mountains (*SI Appendix*, Fig. S6). Increasing temperatures at the beginning of the Holocene enabled *C. foetidissima*—by far the most widespread wild cucurbit in North America—to expand northward into the arid western portion of the Great Plains (Fig. 2*A*). Presumably, the wild ancestor of domesticated *C. pepo* ssp *ovifera* was also able to colonize mesic environments from the Gulf Coast to the Great Lakes during this time (27). A warmer climate may have also enabled wild *Cucurbita* to grow at higher elevations, facilitating gene flow between the previously isolated eastern, western, and southern lineages of *E. pruinosa* and resulting in a zone of genetic admixture in northern Mexico and the southern United States. By 6,000 to 7,000 y ago, Middle Archaic hunter-gatherers utilized wild *C. pepo* ssp *ovifera* for containers and fish floats as far northeast as New England (Fig. 2*B*) and probably cultivated the squash to a limited degree outside of its native range (28). Between 3,800 and 5,000 y ago, a crop complex was developed in the Eastern Woodlands (outline in Fig. 2*B*) that involved the domestication of *C. pepo* ssp *ovifera* for edible seeds (18). Agriculture during the ensuing "Woodland Period" was small scale and subsistence based, revolving around seasonal settlements in river valleys where annual seed crops could be grown without irrigation (29). However, during this time, domesticated *C. pepo* ssp *ovifera* spread throughout the southeast (Fig. 2*B*), probably driven by increasing trade and agronomic exchange between loosely affiliated cultural groups (29). We hypothesize that the declining effective population size of and low rates of immigration into the eastern *E. pruinosa* lineage during the Holocene (Fig. 2 *C* and *D*) reflect the movement of the squash bees alongside cultivated *C. pepo* ssp *ovifera* into mesic habitats that were both ecologically distinct and geographically distant from those in its original range. Finally, the introduction of maize and domesticated *C. pepo* ssp *pepo* from Mexico led to the adoption of intensive squash-maize-beans agriculture 1,000 y ago, that ultimately supported high human population densities across eastern North America (30) and coincided with rapid population growth and elevated gene flow across the entire range of *E. pruinosa* (Fig. 2 *C* and *D* and *SI Appendix*, Fig. S6).

Our findings indicate that the transition from wild xeric to mesic agricultural habitats by eastern *E. pruinosa* was accompanied by selective sweeps that resulted in substantial reductions in genetic diversity across its genome. Within this lineage, nearly a fifth of the genome appears to be linked to recent sweeps. The impacted tracts are found across all major chromosome-scale scaffolds and contain roughly 15% of identified protein-coding sequences, wherein most derived alleles have hitchhiked to fixation. The breadth of the sweeps—in some cases spanning multiple megabases of a 400-Mb genome—makes it challenging to identify particular variants or loci that may have conferred fitness advantages in agricultural settings. However, we found an overrepresentation of protein-coding genes associated with chemosensation within genomic regions linked to sweeps, some of which contain amino acid substitutions private to the eastern lineage. The ecological differences between *E. pruinosa*'s agricultural and wild habitats provide hypotheses regarding adaptive pressures that could drive selection on sensory function in this lineage. Outside of eastern North America, the majority of *E. pruinosa*'s geographic distribution consists of desert and dry scrub with historically little cropland. In these environments, *C. foetidissima* forms perennial clumps in depressions and produces flowers following the monsoon rains (13). These patches of natural habitat for *E. pruinosa* are fairly consistent from year to year but are sparsely distributed, flower asynchronously, and have a low ratio of flowering to vegetative biomass relative to annual *Cucurbita* species (31). In contrast, the annual *C. pepo* in agroecosystems are phenologically uniform and grown at high densities (32). Thus, *C. pepo* crop fields provide an effectively unlimited food supply for squash bees but are spatially unpredictable on an annual basis because of crop rotation (33). Domesticated *C. pepo* exude a mixture of volatile compounds that is simpler than and distinct from those produced by *C. foetidissima*

(*SI Appendix*, Fig. S39 and Dataset 4). We hypothesize that selection on olfaction in the eastern squash bee lineage is associated with adaptation to a distinct sensory environment in agricultural habitats, driven by high densities and phenotypic uniformity of an alternate host species. Chemosensory adaptation has occurred in agricultural pests that have experienced similarly extreme population expansions following a host switch onto a crop plant (34). These results suggest that agricultural environments dominated by domesticated plants may favor distinct behavioral and sensory phenotypes in crop pollinators, similarly to insect agricultural pests.

The productivity of industrial agriculture is possible because of farming technologies that require large, regular plantings of a single crop species. The widespread adoption of these technologies during the mid-20th century drastically transformed existing agricultural landscapes. For example, the agricultural output of the United States quadrupled since the 1930s, while the amount of agricultural acreage remained constant (35). This timeframe implies that many univoltine insect species have experienced dramatic changes in ecological context over the course of the past 100 generations or less. One of the most profound consequences of this agricultural intensification for insect pollinators is the replacement of diverse, heterogeneous food resources with dense plantings of a few crop species (36). Because roughly 38% of worldwide crop production depends on pollination by wild insects (37, 38), the capacity of insect pollinators to adapt into agricultural niches has broad implications for the global food supply. *Eucera pruinosa* provides a system uniquely suited to investigate the long-term consequences of monocrop agriculture on insect pollinators because its coevolution with *Cucurbita* has facilitated an intimate dependence on a single crop that has lasted thousands of years. Further, the archaeological record of *Cucurbita* agriculture in North America provides a historical geographic and ecological context that is lacking for any other insect pollinator aside from domesticated honey bees. This work suggests that human manipulation of the density and composition of flowering vegetation has imposed strong selective pressures on insect pollinators throughout the history of agriculture and can result in profound changes to levels of polymorphism across much of their genomes.

Materials and Methods

Sequencing, Assembly, and Annotation of Reference Genome. We extracted high-molecular-weight (HMW) genomic DNA from the thorax of a haploid *E. pruinosa* male collected near State College, Pennsylvania (the United States) using a MagAttract HMW DNA Kit (Qiagen). Illumina HiSeq 4000 and PacBio Sequel platforms (University of Maryland Institute for Genome Sciences) were used to generate short- and long-read data, respectively. For Illumina sequencing, a randomly sheared paired-end library was generated using a KAPA HyperPrep Kit (Roche), yielding 88.3 Gb of high-quality data in 150-bp paired-end short reads at 220x coverage. For PacBio sequencing, a library was prepared from unsheared HMW DNA using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences), size-selected with a 10-kb cutoff on a BluePippin (Sage Sciences) and sequenced across six cells to generate 48.9 Gb of continuous long reads at 120x coverage. Contigs were assembled from long reads using CANU v1.8 (39) followed by two rounds of polishing with ARROW through SMRT software suite v7.0.0 (Pacific Biosciences). To construct a chromosome-scale assembly from contigs, an in vitro chromatin library (Dovetail Chicago Hi-C) was prepared with cross-linked HMW DNA extracted from the head of the same bee used for the Illumina and PacBio libraries and then sequenced on an Illumina Hiseq X instrument to generate 50 Gb of 150-bp read pairs at 240x coverage. The interaction frequency between Hi-C pairs was quantified and used to merge, break, and orient contigs into scaffolds with the HIRISE pipeline (Dovetail Chicago). Scaffolding was followed by five rounds of polishing with short reads using NEXTPOLISH (40), and contaminant sequences were removed on the basis of average sequencing depth and similarity to known bacterial, fungal, protist, and plant sequences. Notably, the 16-Mb genome (two chromosomes) of an undescribed *Apicystis* (Apicomplexa) parasite was recovered from the scaffolds. The final *E. pruinosa* assembly was 409.1 Mb across 696 contigs (N/L50 of 8/20.638 Mb), with 91% on 20 chromosome-sized scaffolds. Assembly completeness was 97.9% (5870 of 5991 BUSCOs), assessed via BUSCO v4.0.6 using the Hymenoptera reference set (41). Consensus quality (QV) was 41.9 (e.g. per-base error rate 6.5×10^{-5}), assessed via MERQURY v1.3 (42).

To annotate protein-coding features in this reference genome, we extracted RNA from four individuals: two adult males (whole heads), one adult female (head and abdomen), and one larva of unknown sex (head, middle section, and tail). Extractions were performed with a Qiagen RNeasy kit following the manufacturer's recommendations, including DNA degradation with DNase I and a final elution in 50 μl of nucleic acid-free water. Concentration of each sample was measured with a Qubit fluorometer, using the high-sensitivity RNA reagent. Quality was checked with an Agilent TapeStation. RNA from all eight extractions was pooled in equal concentrations and sent to Novogene (Davis, CA) for library preparation and sequencing. A single mRNA library was prepared using poly-A enrichment and a custom second-strand synthesis buffer (Illumina). Library insert size was checked on an Agilent 2100, and concentration was quantified with qPCR. Sequencing was performed on an Illumina NovaSeq 6000 machine, resulting in 6.3 Gb of raw data (42,196,074 paired-end 150-bp reads), with 92.92% of bases with a Phred quality score > 30. The BRAKER2 annotation pipeline (43) was run separately for mapped RNA and protein evidence, and predictions from both runs were combined using TSEBRA (44). We used the ORTHODB v10 arthropod collection (45) and the PROTHINT pipeline to generate protein evidence (43). We then assembled and quantified transcripts for the predicted gene models from mapped RNA reads using STRINGTIE (46). Finally, we used PASA (47) to correct exon/intron boundaries, add UTRs, and identify splice variants with the assembled transcripts. For subsequent analyses, we used only gene models that were either supported by RNA evidence or by homology to other proteins in the ORTHODB arthropod collection and the NCBI RefSeq protein database (48).

Genome Reduction and Inference of Spatial Genetic Structure. To infer genetic structure across the range of *E. pruinosa*, we augmented an existing five-locus microsatellite dataset of 938 bees (11) with 100-bp restriction site-associated sequences from 142 bees (*SI Appendix*, Table S1), covering a total of 26 populations. We extracted genomic DNA from ethanol-preserved samples using a standard phenol:chloroform extraction protocol, digested molecules with EcoRI and MspI, and then ligated barcoded adapters to the sheared ends. Fragments were size-selected using a BluePippin (Sage Science Inc, Beverly, MA) targeting an insert size of 350 bp, amplified in four separate high-fidelity PCR reactions, purified with AMPure XP beads, and then sequenced on a single lane of an Illumina Hiseq 2500 machine. After demultiplexing, reads were mapped to the reference genome using MINIMAP2 (49) and variants identified using SAMTOOLS and BCFTOOLS (50). To infer genetic clusters using both SNP and microsatellite markers, we used a clustering model similar to ref. 51 that employs genotype likelihoods instead of called genotypes but modified to allow for multiple marker types and haplodiploidy (*SI Appendix*, section 1). Genotype likelihoods were calculated for 111,296 SNPs via SAMTOOLS (50) and for microsatellites following the genotyping error model in ref. 52. Cross-validation was used to select the number of clusters, by setting a random subset of genotypes to missing (e.g., uniform genotype likelihoods) prior to fitting the model for each of ten folds and then averaging the holdout likelihood across folds. To infer phylogenetic relationships among the clusters, we use the maximum likelihood estimates of allele frequencies and the tree-building procedure of ref. 53. To visualize genetic structure across individuals, we used the first two principal components of the normalized genotypes (54).

Genome Resequencing Within Major Phylogeographic Divisions. After determining the geographic distribution of the major genetic clusters of *E. pruinosa*, we selected a set of 44 haploid males across 5 populations (*SI Appendix*, Table S1) for whole genome resequencing (WGS). We also sequenced the genomes of two congeneric squash bee species, *E.* (*Peponapis*)

utahensis and *E.* (*Xenoglossa*) *strenua*, as outgroups. Genomic DNA was extracted from specimens using a standard phenol/chloroform protocol, and a 150-bp randomly shared paired-end shotgun library was sequenced to 30x coverage on an Illumina Hiseq 4000 machine by Novogene (Davis, CA). Genomic reads were mapped to the reference genome using MINIMAP2 (49), after trimming adapters and low-quality bases. Base quality scores were recalibrated using consensus base calls using BBTOOLS (55). The GATK pipeline v4.3 (56) was used to remove PCR and optical duplicates, jointly call haplotypes across samples, and filter low-confidence SNP calls. We used NGSPARALOG (57) to detect and remove putative collapsed paralogs and mismapped regions, after modifying the model for use with haploid genomes (fork at https://github.com/nspope/ngsParalog). This resulted in 3,674,130 biallelic SNPs used in subsequent analyses of the whole genome data.

Estimation of Ancestral States and Recombination Rates. We estimated the average per-based mutation rate and the ancestral state at each SNP by fitting the general time-reversible mutation model (58) to the whole-genome sequences of *E. pruinosa* individuals collected from Mexico and the two outgroup taxa, using the divergence dates estimated by Dorchin et al. (12) (*SI Appendix*, section 2). We removed SNPs that were triallelic in *E. pruinosa* or for which the ancestral state could not be determined (where one or both of the outgroups was missing or the putative ancestral state was inconsistent across outgroups), and then polarized SNPs according to the inferred ancestral state. As all subsequent analyses required long contiguous sequences, we retained the 22 scaffolds that were 1 Mb or longer (a total of approximately 350 Mb after masking sites according to the criteria above). We estimated fine-scale recombination rates across these scaffolds using PYRHO (59) on the Mexican samples, given an effective population size trajectory inferred with SMC++ (60). Samples from Mexico were used for recombination rate estimation as these showed high levels of nucleotide diversity across the entire genome, did not appear to be admixed, and possessed very few signatures of recent selective sweeps.

Inference of Purifying and Positive Selection. We adopted the approach of McVicker et al. (23) to predict local reductions in nucleotide diversity due to background (purifying) selection across the genome. We constructed a multigenome alignment of *E. pruinosa*, the outgroup species, and 22 bee genomes available on GenBank (*SI Appendix*, Fig. S35) with CACTUS (61); then, we used PHASTCONS (62) to predict conserved elements in *E. pruinosa*'s genome under a 21-category discrete-gamma mutation rate model (where "conserved element" is defined as a genomic tract belonging to the lowest rate category). Together with the recombination maps and exon annotations, these conserved elements were used as inputs to the model of background selection described in McVicker et al. (23) and implemented in CALCBKGD, that produces per-base predictions of the strength of background selection ("B-values"). The two free parameters in this model were estimated by fitting the B-values to observed nucleotide diversity in the Mexican samples (*SI Appendix*, section 3 and Fig. S36).

We used SWEEPFINDER2 (22) to infer genomic intervals impacted by recent selective sweeps in three populations (Puebla, Mexico; Colorado, the United States, and Pennsylvania, the United States) that represent the major phylogeographic divisions in *E. pruinosa* (Mexico, Western United States, and Eastern United States). SWEEPFINDER2 was run with inferred recombination maps and B-values to help distinguish signatures of recent positive selection from those of background selection and recombination dead zones (Fig. 3*A* and *SI Appendix*, Fig. S29), producing a composite likelihood ratio (CLR) measuring the strength of evidence for a selective sweep at a particular genomic coordinate. An analysis without using B-values led to similar conclusions, except that large portions of scaffolds with very low recombination rates were inferred to contain sweeps.

Reconstruction of the Ancestral Recombination Graph and Demographic Inference. We used polarized SNPs from the whole genome sequences in combination with the recombination maps as input to RELATE v1.1.9 (19) to infer the Ancestral Recombination Graph (ARG; the set of all sample genealogies across the genome, including recombination events), after imputing missing variants with BEAGLE v5.1 (63) and masking SNPs in coding sequences. For demographic inference, we filtered genealogies belonging to 177 Mb ($\approx$ 50%) of the genome that was predicted to be either under strong purifying selection (i.e., a B-value < 0.3) or assigned a high score by SWEEPFINDER2 (highest 30% of CLR scores in each population). Less stringent thresholds produced similar results.

To jointly infer population size histories without reference to a prespecified demographic model (e.g., population splits/mergers/size changes), we developed a composite likelihood method that uses the rates of first coalescence events for subtrees (trios) of the inferred ARG and returns piecewise-constant estimates of migration rates and effective population sizes at a fixed temporal resolution (*SI Appendix*, section 3). The core idea behind this approach is to calculate expected rates of coalescence in time intervals (epochs) for trios with different population labelings, using a continuous-time Markov process where the rate matrix in each epoch is parameterized by effective population sizes and migration rates. These time-varying demographic parameters are then optimized to maximize the similarity between expected coalescence rates and those extracted from the ARG. This approach can be viewed as a generalization of the single-population effective population size estimators described in Speidel et al. (19) and is implemented in the R package coaldecoder (https://github.com/nspope/coaldecoder). We used TSKIT (64) to calculate genealogical statistics (such as coalescence times) and to query specific marginal genealogies.

We applied this method to jointly infer effective population sizes and migration rates for the five populations with whole genome sequences, at two temporal scales: across 100 variable-sized epochs over the entire genealogical history of the samples–with epochs chosen to contain roughly the same number of pairwise coalescence events–and within 250-y epochs over the past 18,000 y. The purpose of fitting models to two different time discretizations was to minimize computational costs, as complexity of the algorithm scales with the number of epochs. We used cross-validation to identify the times of population mergers by holding out single chromosomes greater than 10 Mb, refitting the demographic model for a given merger time and averaging log-likelihood across holdout sets. To calculate approximate confidence intervals around demographic parameters, we generated bootstrap replicates by resampling contiguous blocks of trees each spanning 1 Mb of the genome and then reestimated trajectories for each bootstrap sample. To provide a separate line of evidence for demographic shifts in the sequenced populations, we used MOMENTS.LD (65) to estimate changes in population size and divergence times from linkage disequilibrium statistics, which led to similar conclusions to the method based on coalescence rates (*SI Appendix*, section 5, Figs. S37–S38 and Table S4).

Scoring and Dating of Putative Selective Sweeps. We used MSPRIME (66) v1.0 to generate realistic chromosome-scale simulations given the demographic history, recombination maps, and site accessibility from the observed data. These simulations served two purposes: first, to ensure that the fitted demographic model produced summary statistics (Fst, nucleotide diversity, etc) that were comparable to those observed in the actual data (*SI Appendix*, Table S2); second, to calibrate the CLR scores produced by SWEEPFINDER2 in terms of a realistic null model. For the latter, we simulated mutations with local variation in mutation rates according to the inferred B-values, applied SWEEPFINDER2 across a 1-kb grid of test sites over each simulation, used the results to construct a null distribution for the CLR in each population, and finally calculated null quantiles (*P*-values) for the CLR scores from the observed data. These calibrated scores were used to delimit sweep "footprints" in each population, defined as contiguous intervals with a CLR exceeding the 95th quantile of the lineage-specific null distribution ($-\log_{10} P_{\text{CLR}} > 1.3$). To propagate uncertainty from the demographic inference into the CLR null distributions, we used a random bootstrap replicate (demographic trajectory) per scaffold for each of ten simulations. To estimate the timing of putative sweeps, we used RELATE to reestimate branch lengths locally around each sweep footprint and extracted the genealogy at the coordinate with the highest CLR (the "peak" of the footprint). If the samples from the swept population did not form a monophyletic clade, we discarded the putative sweep; otherwise, we extracted the time to the most recent common ancestor (TMRCA) of this clade, as a crude estimate of the sweeps' onset time.

Gene Set Enrichment Analysis and Nonsynonymous Substitutions. We used INTERPROSCAN v5 (67) to assign "molecular function" and "biological process" Gene Ontology annotations (GO v2021-07-02) to each predicted gene feature

in the *E. pruinosa* genome annotation. Where possible, we also assigned gene symbols and GO annotations of orthologs from *Drosophila melanogaster* (68) via reciprocal blast (subject and query cover $> 20\%$, E-value $< 10^{-6}$), excluding automatically curated annotations (evidence code "IAE"). The final annotated gene set consisted of 8093 genes with 8092 associated GO terms. We used a hypergeometric test per GO term to assess whether particular functional annotations were overrepresented in sweep footprints. Depending on the CLR threshold used to delimit sweeps, these may contain many genes that are proximal to but not themselves targets of positive selection. Thus, we repeated the overrepresentation analysis across a range of CLR thresholds (*SI Appendix*, Table S3). Because genes of similar function are often arranged in tandem arrays, we employed the permutation-based global test procedure described in ref. 24–implemented in GOFUNCR (69)–that resamples random genomic intervals of a similar size distribution to the observed sweep footprints and also controls the family-wise error rate.

To identify potential coding changes in genes occurring within sweep footprints that may be targets of recent selection, we used SPADES (70) to create separate short-read genome assemblies for the Colorado and Mexico populations of *E. pruinosa*, used CACTUS (61) to align these to the reference genome, and then used AUGUSTUS-CGP (71) to lift over the reference annotation to the new assemblies, thus ensuring that gene models for the nonreference populations formed complete coding sequences (e.g., accounting for indels and changes in splice sites or start/stop codons). We used HAL (72) to map SNPs onto these population-specific coding sequences, removing any paralogs. We did not find any large structural rearrangements between populations in these alignments, aside from a 100-kb transposition on the largest chromosome that did not contain coding genes. We used SNPEFF (73) to identify nonsynonymous substitutions for each population–retaining only those that were highly differentiated between lineages and that were located in putative sweeps (*SI Appendix*, Fig. S32)–and then used BLASTP to identify possible orthologs in *Drosophila melanogaster* and *Apis mellifera*.

Past and Present Geographic Distribution of *C. foetidissima* and Domesticates. To hindcast the distribution of *C. foetidissima*, we used 430 collection records from Castellanos-Morales et al. (27) in combination with WORLDCLIM climatic layers from the present day, 6 kya, and 21 kya (74, 75). To construct a prior that reflects variable intensity in collection efforts across space, we fit an areal Poisson Cox-process model using R-INLA (76) to all available records of herbaceous angiosperms on GBIF (77). These records were binned into a raster of North America, and a random-walk prior was used to model spatial autocorrelation between raster cells. The (present-day) climatic variables at twenty thousand background points sampled from this prior and the *C. foetidissima* records were orthogonalized via singular value decomposition to be used as covariates; the associated rotation matrix was then used to map the entirety of the present and historical climatic variables onto the same coordinate space. The orthogonalized covariates were then used to fit a species distribution model by maximum entropy (78) and predict log-intensity scores across the study area in all time periods. To create polygons representative of the species' present range, we thresholded these continuous surfaces using the 95th quantile of log-intensity scores across *C. foetidissima* accessions following Castellanos-Morales et al. (27). To characterize the geography and timing of human usage of *Cucurbita* domesticates, we collected records of radiocarbon-dated *Cucurbita* tissue from archaeological sites using the P3K14C database (79) and published phytoliths from Mexico (21, 80–82). We removed records for which the direct dating of tissue (as opposed to dating of a surrounding archaeological feature) could not be confirmed from a written reference.

Data, Materials, and Software Availability. All sequence data are publicly available through the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov). NCBI BioProject PRJNA637784 contains the reference genome assembly (GenBank accessions JAMYCR010000001–JAMYCR010000697) and associated raw long-read (SRA accessions SRR18337983–SRR18337988), short-read (SRR19855294–SRR19855296), HiC (SRR19886896), and RNA (SRR23612349) sequence data. NCBI BioProject PRJNA836616 contains the raw Illumina whole genome (SRR23618692–SRR23618742) and reduced representation (SRR23619470–SRR23619539) sequence data used for population genomic analyses. *SI Appendix*, Table S5 contains the accession numbers of the 22 chromosome-sized scaffolds used for population genomic analyses. The *E. pruinosa* microsatellite genotypes, collated archaeological *Cucurbita* radiocarbon dates, *E. pruinosa* gene annotations, lineage-specific nonsynonymous substitutions, and *Cucurbita* volatile profiles are included as *SI Appendix*. R and C++ libraries implementing the admixture and demographic inference algorithms are available through the first author's GITHUB repositories and are linked to in the Methods. Genome assembly, raw sequence data have been deposited in NCBI GenBank and SRA (BioProject PRJNA637784, BioProject PRJNA836616). Previously published data were used for this work, Microsatellite data (provided by lead author): (11). *Cucurbita foetidissima* records for species distribution modelling (provided by lead author): (27).

ACKNOWLEDGMENTS. This research was supported by a NSF CAREER Award (DEB-2046474); the USDA NIFA Appropriations under Projects PEN04716; the US Department of Agriculture, Agricultural Research Service (USDA-ARS); the Pennsylvania State University Lorenzo L. Langstroth Endowment, and a Dovetail Genomics matching funds grant awarded to MML-U. We thank Kelly G. McGowan for sending *E. pruinosa* collections from Missouri, Gabriella Castellanos-Morales and Luis Eguiarte for sharing data on *Cucurbita foetidissima* occurrences, Kristen Brochu for providing data on *Cucurbita* floral volatiles, and Rob Dunn and four anonymous reviewers for suggestions on clarifying the analyses and interpretation. This research used resources provided by the SCINet project of the USDA-ARS (project number 0500-00093-001-00-D). All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of USDA. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer.

Author affiliations: [a]Department of Entomology, The Pennsylvania State University, University Park, PA 16802; [b]Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403; [c]Bee Research Laboratory, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705; and [d]Department of Biology, Utah State University, Logan, UT 84322

1. T. Newbold *et al.*, Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science* **353**, 288–291 (2016).
2. D. P. Bebber, Range-expanding pests and pathogens in a warming world. *Annu. Rev. Phytopathol.* **53**, 335–356 (2015).
3. M. M. Turcotte, H. Araki, D. S. Karp, K. Poveda, S. R. Whitehead, The eco-evolutionary impacts of domestication and agricultural practices on wild species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372** (2017).
4. K. Kébé *et al.*, Global phylogeography of the insect pest *Callosobruchus maculatus* (Coleoptera: Bruchinae) relates to the history of its main host *Vigna unguiculata*. *J. Biogeogr.* **44**, 2515–2526 (2017).
5. Z. Du *et al.*, Global phylogeography and invasion history of the spotted lanternfly revealed by mitochondrial phylogenomics. *Evol. Appl.* **14**, 915–930 (2021).
6. K. E. Filchak, J. B. Roethele, J. L. Feder, Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* **407**, 739–742 (2000).
7. B. S. Coates, E. B. Dopman, K. W. Wanner, T. W. Sappington, Genomic mechanisms of sympatric ecological and sexual divergence in a model agricultural pest, the European corn borer. *Curr. Opin. Insect. Sci.* **26**, 50–56 (2018).
8. J. Ollerton, R. Winfree, S. Tarrant, How many flowering plants are pollinated by animals? *Oikos* **120**, 321–326 (2011).
9. A.-M. Klein *et al.*, Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B Biol. Sci.* **274**, 303–313 (2007).
10. M. H. Carpenter, B. A. Harpur, Genetic past, present, and future of the honey bee (*Apis mellifera*) in the United States of America. *Apidologie* **52**, 63–79 (2021).
11. M. M. López-Uribe, J. H. Cane, R. L. Minckley, B. N. Danforth, Crop domestication facilitated rapid geographical expansion of a specialist pollinator, the squash bee *Peponapis pruinosa*. *Proc. R. Soc. B Biol. Sci.* **283**, 20160443 (2016).
12. A. Dorchin, M. M. López-Uribe, C. J. Praz, T. Griswold, B. N. Danforth, Phylogeny, new generic-level classification, and historical biogeography of the *Eucera* complex (Hymenoptera: Apidae). *Mol. Phylogenet. Evol.* **119**, 81–92 (2018).

13. P. D. Hurd Jr, E. G. Linsley, T. W. Whitaker, Squash and gourd bees (*Peponapis, Xenoglossa*) and the origin of the cultivated *Cucurbita*. *Evolution* **25**, 218–234 (1971).
14. B. D. Smith, Eastern North America as an independent center of plant domestication. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12223–12228 (2006).
15. C. M. McGrady, R. Troyer, S. J. Fleischer, Wild bee visitation rates exceed pollination thresholds in commercial *Cucurbita* agroecosystems. *J. Econ. Entomol.* **113**, 562–574 (2020).
16. D. R. Artz, B. A. Nault, Performance of *Apis mellifera*, *Bombus impatiens*, and *Peponapis pruinosa* (hymenoptera: Apidae) as pollinators of pumpkin. *J. Econ. Entomol.* **104**, 1153–1161 (2011).
17. D. S. W. Chan, N. E. Raine, Hoary squash bees (*Eucera pruinosa*: Hymenoptera: Apidae) provide abundant and reliable pollination services to *cucurbita* crops in Ontario (Canada). *Environ. Entomol.* **50**, 968–981 (2021).
18. B. D. Smith, R. A. Yarnell, Initial formation of an indigenous crop complex in Eastern North America at 3800 BP. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6561–6566 (2009).
19. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
20. T. D. Price, Ancient farming in eastern North America. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6427–6428 (2009).
21. L. Kistler *et al*., Gourds and squashes (*Cucurbita* spp) adapted to megafaunal extinction and ecological anachronism through domestication. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15107–15112 (2015).
22. M. DeGiorgio, C. D. Huber, M. J. Hubisz, I. Hellmann, R. Nielsen, Sweepfinder 2: Increased sensitivity, robustness and flexibility. *Bioinformatics* **32**, 1895–1897 (2016).
23. G. McVicker, D. Gordon, C. Davis, P. Green, Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
24. K. Prüfer *et al*., FUNC: A package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinform.* **8**, 1–10 (2007).
25. X. Cui *et al*., Molecular mechanism of the UDP-glucuronosyltransferase 2b20-like gene (*AccUGT2B20-like*) in pesticide resistance of *Apis cerana cerana*. *Front. Genet.* **11**, 592595 (2020).
26. T. W. Whitaker, R. J. Knight, Collecting cultivated and wild cucurbits in Mexico. *Econ. Bot.* **34**, 312–319 (1980).
27. G. Castellanos-Morales *et al*., Historical biogeography and phylogeny of *Cucurbita*: Insights from ancestral area reconstruction and niche evolution. *Mol. Phylogenet. Evol.* **128**, 38–54 (2018).
28. J. B. Petersen, N. A. Sidell, Mid-holocene evidence of *Cucurbita* sp. from Central Maine. *Am. Antiq.* **61**, 685–698 (1996).
29. B. D. Smith, The cultural context of plant domestication in eastern North America. *Curr. Anthropol.* **52**, S471–S484 (2011).
30. B. D. Smith, N. Boivin, M. Petraglia, R. Crassard, "Tracing the initial diffusion of maize in North America" *Human Dispersal and Species Movement* (2017), pp. 332–348.
31. J. A. Winsor, S. Peretz, A. G. Stephenson, Pollen competition in a natural population of *Cucurbita foetidissima* (Cucurbitaceae). *Am. J. Bot.* **87**, 527–532 (2000).
32. Y. H. Chen, L. R. Shapiro, B. Benrey, A. Cibrián-Jaramillo, Back to the origin: In situ studies are needed to understand selection during crop diversification. *Front. Ecol. Evol.* **5** (2017).
33. T. H. Roulston, K. Goodell, The role of resources and risks in regulating wild bee populations. *Annu. Rev. Entomol.* **56**, 293–312 (2011).
34. Y. H. Chen, R. Gols, B. Benrey, Crop domestication and its impact on naturally selected trophic interactions. *Annu. Rev. Entomol.* **60**, 35–58 (2015).
35. C. Dimitri, A. Effland, N. C. Conklin, "The 20th Century transformation of US agriculture and farm policy" (Tech Rep. EIB No. 3, United States Department of Agriculture, 2005).
36. P. H. Raven, D. L. Wagner, Agricultural intensification and climate change are rapidly decreasing insect biodiversity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2002548117 (2021).
37. R. Winfree, N. M. Williams, J. Dushoff, C. Kremen, Native bees provide insurance against ongoing honey bee losses. *Ecol. Lett.* **10**, 1105–1113 (2007).
38. L. A. Garibaldi *et al*., Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* **339**, 1608–1611 (2013).
39. S. Koren *et al*., Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
40. J. Hu, J. Fan, Z. Sun, S. Liu, NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
41. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
42. A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
43. T. Brůna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
44. L. Gabriel, K. J. Hoff, T. Brůna, M. Borodovsky, M. Stanke, TSEBRA: Transcript selector for BRAKER. *BMC Bioinform.* **22**, 566 (2021).
45. E. V. Kriventseva *et al*., OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
46. M. Pertea *et al*., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
47. B. J. Haas *et al*., Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
48. N. A. O'Leary *et al*., Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
49. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
50. H. Li *et al*., The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
51. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
52. J. Wang, Sibship reconstruction from genetic data with typing errors. *Genetics* **166**, 1963–1979 (2004).
53. M. Lipson *et al*., Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* **30**, 1788–1802 (2013).
54. K. J. Galinsky *et al*., Fast Principal-Component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
55. B. Bushnell, J. Rood, E. Singer, BBMerge: Accurate paired shotgun read merging via overlap. *PloS One* **12**, e0185056 (2017).
56. A. McKenna *et al*., The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
57. T. Linderoth, *Identifying Population Histories, Adaptive Genes, and Genetic Duplication from Population-Scale Next Generation Sequencing* (University of California, Berkeley, 2018).
58. D. Catanzaro, R. Pesenti, M. C. Milinkovitch, A non-linear optimization procedure to estimate distances and instantaneous substitution rate matrices under the GTR model. *Bioinformatics* **22**, 708–715 (2006).
59. J. P. Spence, Y. S. Song, Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci. Adv.* **5**, eaaw9206 (2019).
60. J. Terhorst, J. A. Kamm, Y. S. Song, Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
61. J. Armstrong *et al*., Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
62. A. Siepel *et al*., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
63. B. L. Browning, Y. Zhou, S. R. Browning, A one-penny imputed genome from next-generation reference panels. *Am. J. Human Genet.* **103**, 338–348 (2018).
64. J. Kelleher *et al*., Inferring whole-genome histories in large population datasets. *Nat. Genet.* **51**, 1330–1338 (2019).
65. A. P. Ragsdale, S. Gravel, Models of archaic admixture and recent history from two-locus statistics. *PLoS Genet.* **15**, e1008204 (2019).
66. F. Baumdicker, et al., Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* **220**, iyab229 (2022).
67. P. Jones *et al*., InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
68. A. Larkin *et al*., FlyBase: Updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res.* **49**, D899–D907 (2021).
69. S. Grote, GOfuncR: Gene ontology enrichment using FUNC, 2021. R package version 1.13.2.
70. A. Bankevich *et al*., SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
71. S. Nachtweide, M. Stanke, "Multi-genome annotation with AUGUSTUS" in *Gene Prediction* (Springer, 2019), pp. 139–160.
72. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, Hal: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).
73. P. Cingolani *et al*., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPeff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
74. S. E. Fick, R. J. Hijmans, WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
75. P. Braconnot *et al*., Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum, part 1: Experiments and large-scale features. *Clim. Past* **3**, 261–277 (2007).
76. F. Lindgren, H. Rue, Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63**, 1–25 (2015).
77. T. Robertson *et al*., The GBIF integrated publishing toolkit: Facilitating the efficient publishing of biodiversity data on the internet. *PLoS One* **9**, e102623 (2014).
78. S. J. Phillips, R. P. Anderson, M. Dudík, R. E. Schapire, M. E. Blair, Opening the black box: An open-source release of Maxent. *Ecography* **40**, 887–893 (2017).
79. D. Bird *et al*., p3k14c: A synthetic global database of archaeological radiocarbon dates. *Sci. Data* **9**, 1–19 (2022).
80. B. D. Smith, The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* **276**, 932–934 (1997).
81. B. D. Smith, Reconsidering the Ocampo Caves and the era of incipient cultivation in Mesoamerica. *Lat. Am. Antiq.* **8**, 342–383 (1997).
82. B. D. Smith, Reassessing Coxcatlan Cave and the early history of domesticated plants in Mesoamerica. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9438–9445 (2005).