Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico

Kraig H. Kraft[a,1], Cecil H. Brown[b], Gary P. Nabhan[c], Eike Luedeling[d], José de Jesús Luna Ruiz[e], Geo Coppens d'Eeckenbrugge[f], Robert J. Hijmans[g], and Paul Gepts[a,2]

[a]Department of Plant Sciences, Section of Crop and Ecosystem Sciences, University of California, Davis, CA 95616-8780; [b]Department of Anthropology, Northern Illinois University, Pensacola, FL 32503-6634; [c]Southwest Center, University of Arizona, Tucson, AZ 85721-0185; [d]World Agroforestry Centre, Nairobi, Kenya; [e]Centro de Ciencias Agropecuarias, Universidad Autónoma de Aguascalientes, Aguascalientes, Mexico CP 20131; [f]Centre de coopération Internationale en Recherche Agronomique pour le Développement, Unité Mixte de Recherche 5175 Centre d'Ecologie Fonctionnelle et Evolutive, Campus Centre National de la Recherche Scientifique, Montpellier Cedex 5, France; and [g]Department of Environmental Science and Policy, University of California, Davis, CA 95616

The study of crop origins has traditionally involved identifying geographic areas of high morphological diversity, sampling populations of wild progenitor species, and the archaeological retrieval of macroremains. Recent investigations have added identification of plant microremains (phytoliths, pollen, and starch grains), biochemical and molecular genetic approaches, and dating through ^{14}C accelerator mass spectrometry. We investigate the origin of domesticated chili pepper, *Capsicum annuum*, by combining two approaches, species distribution modeling and paleobiolinguistics, with microsatellite genetic data and archaeobotanical data. The combination of these four lines of evidence yields consensus models indicating that domestication of *C. annuum* could have occurred in one or both of two areas of Mexico: northeastern Mexico and central-east Mexico. Genetic evidence shows more support for the more northern location, but jointly all four lines of evidence support central-east Mexico, where preceramic macroremains of chili pepper have been recovered in the Valley of Tehuacán. Located just to the east of this valley is the center of phylogenetic diversity of Proto-Otomanguean, a language spoken in mid-Holocene times and the oldest protolanguage for which a word for chili pepper reconstructs based on historical linguistics. For many crops, especially those that do not have a strong archaeobotanical record or phylogeographic pattern, it is difficult to precisely identify the time and place of their origin. Our results for chili pepper show that expressing all data in similar distance terms allows for combining contrasting lines of evidence and locating the region(s) where cultivation and domestication of a crop began.

The analysis of plant macroremains, morphological variation in crop varieties, and identification of wild progenitor species (as determined through their ability to hybridize with the crop) constitute traditional methods for studying crop origins (1, 2). Currently, analysis of microremains such as starch grains, accelerator mass spectrometry (AMS) ^{14}C radiocarbon dating, along with biochemical and molecular genetic analyses of wild and domesticated populations are also used to date and locate geographic areas of domestication (3, 4).

This set of approaches is extended here with two additional methods, species distribution modeling and paleobiolinguistics, integrating these in a comprehensive study of the origin of domesticated chili pepper, *Capsicum annuum* L., the world's most widely grown spice. *C. annuum* is one of five domesticated pepper species, which also include *Capsicum baccatum* L., *Capsicum chinense* Jacq., *Capsicum frutescens* L., and *Capsicum pubescens* Ruiz & Pav. The ~30 species of *Capsicum* are all native to the Americas (5). Comparing karyotypes of wild and domesticated *C. annuum* (var. *glabriusculum* and var. *annuum*, respectively), Pickersgill (6) identified Mexico as the general

region of domestication of this pepper. Loaiza-Figueroa et al. (7) used allozyme similarity to identify putative wild ancestral populations for chili pepper in a larger collection of wild and domesticated populations. They narrowed the likely domestication area to the eastern Mexican states of Tamaulipas, Nuevo León, San Luís Potosí, Veracruz, and Hidalgo. Since these investigations, others have sought to determine genetic relationships among wild and domesticated populations of chili pepper (8, 9).

The oldest macroremains unambiguously identified as *Capsicum* pepper were retrieved from preceramic strata of dry caves in two states of Mexico: Puebla (Tehuacán Valley; refs. 10, 11) and Tamaulipas (Ocampo caves; ref. 12) (Fig. 1A). These were found with macroremains of maize (*Zea mays*), squash (*Cucurbita* spp.), and other species used by humans, all of which, at both sites, were indirectly dated through associations in archaeological strata, suggesting a rough date for the chili pepper macroremains of around 9000–7000 B.P. (13). Subsequently, remains of maize from Tehuacán were dated directly by AMS and found to be more recent, 5600 y calibrated B.P. (14). AMS dating applied to bottle gourd and squash from Ocampo also yielded more recent ages, 6400–6000 y calibrated B.P. (15). Whereas no AMS dates

Significance

The novelty of the information of this manuscript resides in the addition of species distribution modeling and paleobiolinguistics data, combined with genetic and existing archaeobotanical data, to trace back the geographic origin of a crop, namely domesticated pepper, *Capsicum annuum*. Furthermore, the utilization of a geographic framework of reference for the four types of data has allowed us to combine these independent data types into a single hypothesis about the origin of this crop. Our results suggest that food crops in Mexico had a multiregional origin with chili pepper originating in central-east Mexico, maize in the Balsas River Basin and common bean in the Lerma–Santiago River Basin, resembling similar finds for the Fertile Crescent and China.
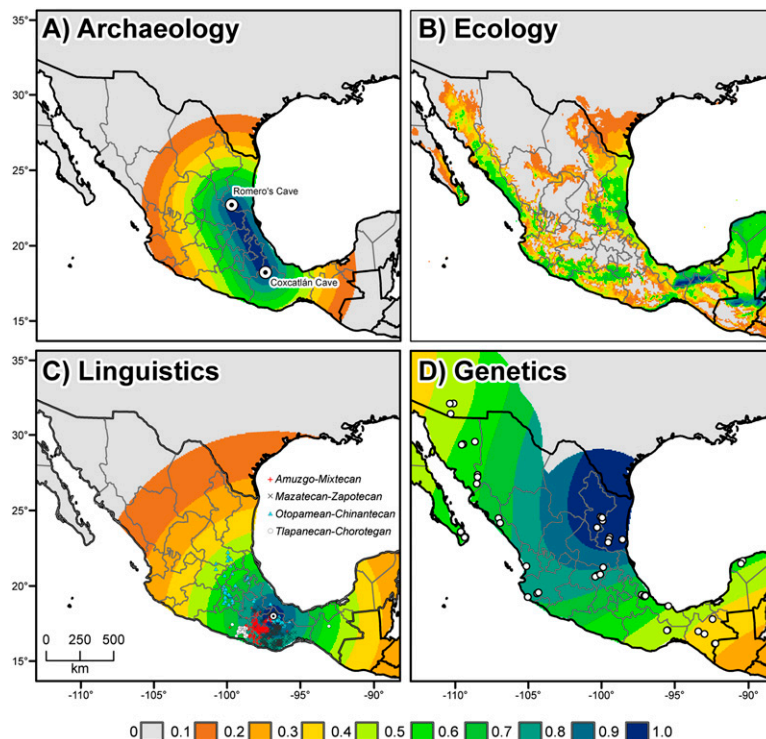
Fig. 1. Possible area of Mexico for *Capsicum annuum* domestication based on (*A*) archaeological, (*B*) paleoclimatic, mid-Holocene, (*C*) linguistic, and (*D*) genetic data. In addition to the strength of evidence (between 0 and 1), the maps show: (*A*) Location of the oldest archaeological remains of chili: Romero Cave, Ocampo, Tamaulipas; Coxcatlán Cave, Tehuacán Valley, Puebla. (*C*) Location of the homeland of Proto-Otomanguean (dotted circle) and of the four subgroups of current Otomanguean languages (see Fig. S5 legend and Table 1; for a more detailed view of the subgroup distribution, see Fig. S5). Open circles represent approximate locations of protolanguages with a reconstructed word for chili. (*D*) Open circles indicate the location of the wild chili samples used in the genetic distance analysis (29). For explanation of values, see *Materials and Methods*.

have been recorded for the Tehuacán and Ocampo remains of chili pepper, remains from Guilá Naquitz and Silvia's Caves in the arid eastern valley of Oaxaca state were dated indirectly by AMS to 1400–500 B.P. (16). Rock shelters in the seasonally dry tropical forest of the Central Balsas watershed (state of Guerrero) have produced phytoliths and starch grain residue for domesticated maize and squash (*Cucurbita* sp.) dated by association to around 9000 B.P. (17). However, no remains of *Capsicum* pepper have been found at that site.

Species distribution modeling (SDM) can be used to predict areas that are environmentally suitable for a species from the sites where it is known to occur (18). In SDM, locations of the known current distribution of a species are compiled; values for climatic predictor variables at these locations and a large set of random (background) locations are extracted from spatial databases; and the climatic values are used to fit a model that estimates the similarity of the climate in any location to climatic conditions at known occurrence locations, using a machine-learning algorithm such as MaxEnt (19). The model is then used to predict the climatic suitability for a species across an area of interest. This prediction can be made using current climate data, but the model can also be "transferred" in time, by using past or future climate data simulated by global climate models (GCMs). This approach has been used for many purposes, including to predict the effect of climate change on the geographic distribution of crop wild relatives (20) and to successfully locate unknown *Capsicum* populations (21).

Crop origins can also be studied using paleobiolinguistics (PBL), which employs the comparative method of historical linguistics to reconstruct the biodiversity known to human groups of the remote, unrecorded past (22–24). By comparing words for a species in modern languages, terms for plants and animals in ancestral languages can be retrieved. The presence of words for a species in an ancestral language is an indication of the species' significance to speakers of that language (25, 26), if not their status as domesticated plants. PBL uses Automated Similarity Judgment Program (ASJP) chronology for estimating the latest date at which a protolanguage was spoken based on lexical similarity (27). Lexical similarity found among related languages is calibrated with historical, epigraphic, and archaeological divergence dates for 52 language groups. In addition, the general area in which an ancestral language was spoken, i.e., the protolanguage homeland, can be approximately determined by locating the area where its modern descendant languages are found to be most diverse (28).

In this paper, we complement existing archaeobotanical data with ecological, paleobiolinguistic, and molecular diversity data to identify the region of initial intensification of human interest in chili pepper that led to crop domestication. The novelty of our approach resides in the addition of SDM and PBL to this type of analysis and the expression of all lines of evidence in comparable spatially explicit units (distance to the area of origin) that allows for their integration into a single prediction.

Results

Archaeological Evidence. The remains from Tehuacán and Ocampo constitute at present the oldest macrobotanical evidence for preceramic chili pepper in the New World. Although these chili specimens cannot be identified as cultivated or domesticated, their archaeological association with domesticated remains of important crops, such as maize and squash, is strongly suggestive of ancient intensive human interaction with chili in these areas. Based on this evidence, we assumed that the nearer a place may be to either of these sites, the more likely the location was part of

the region where the crop was first grown and domesticated (Fig. 1*A*).

Ecological Evidence. Wild chili pepper (*C. annuum* var. *glabriusculum*), the ancestor of domesticated *C. annuum* (6), is a perennial shrub that produces dozens of erect, globular, pea-sized fruits. The fruits are consumed and dispersed by frugivorous birds, which pass the seed through their digestive system. Generally found in the northern half of Mexico, the wild chili pepper is associated with a nurse plant—often a hackberry (*Celtis pallida* Torey), a mesquite (*Prosopis* sp.), or columnar cacti. As one moves further southwards, wild chili pepper is found more frequently in human-disturbed landscapes—fence rows, home gardens, and roadsides (29). Based on our own collecting localities and those of herbarium specimens and gene bank accessions (29), we estimate that wild chili peppers grow currently in environments with a median annual average temperature of 24 °C and between 20 °C and 26 °C for 90% of the locations. The coldest locations with known wild pepper populations are mostly in the central Mexican highlands, the warmest locations in the southern coastal regions of Mexico and Guatemala. The median annual rainfall of these locations is 907 mm, and between 495 and 2,253 mm for 90% of the locations, with the driest locations in the northwestern part of the distribution (e.g., Baja California and the Sonoran Desert) and the wettest locations in southeastern Mexico.

The MaxEnt species distribution model had an internal (training) fit area under the curve (AUC) of the receiver operating characteristic (ROC) curve of 0.89. The average cAUC (bias corrected) obtained with fivefold cross-validation was 0.80, which suggests that the model has very good predictive power (30). The two most important predictor variables (based on permutation

importance) were mean temperature of the coldest quarter (53%), followed by annual precipitation (14%).

Under the climate conditions of the mid-Holocene (about 6000 B.P.), the regions predicted to be most suitable for wild chili pepper include areas along the western and eastern coasts of Mexico, southeast Mexico and northern Guatemala (Fig. 1*B*). The central highlands were clearly unsuitable for this species during this period. The correlation coefficient between the predicted suitability for the current climate (Fig. S1) and the mid-Holocene climate was 0.92. Despite this overall similarity, there were important differences between these predictions, with areas in the southeast of Mexico more suitable and areas in the northeast less suitable during the mid-Holocene (Fig. S2).

Paleobiolinguistic Evidence. Brown (22) surveyed the reconstructed vocabularies of 30 protolanguages of Mesoamerica (southern half of Mexico and northern Central America) and abutting areas for terms for 41 different crops, including chili pepper. His survey presented for each protolanguage the estimated date it was spoken at the latest, making it possible to stratify reconstructed words for crops chronologically (Table 1) (27, 28).

Proto-Otomanguean is the oldest (∼6500 B.P.) protolanguage of the New World for which a word for chili pepper reconstructs (31). All daughter languages of Proto-Otomanguean, as defined by Kaufman (32), show reconstructed terms for chili pepper (Table 1). Given that estimated dates are to be understood as the latest dates at which ancestral languages were spoken, it is plausible that speakers of Proto-Otomanguean actually had a word for chili pepper hundreds, if not thousands, of years before ∼6500 B.P. The oldest protolanguage of Table 1 not belonging to the Otomanguean family is Proto-Totozoquean (∼4300 B.P.), for which a term for chili pepper does not reconstruct. Non-Otomanguean languages for

Table 1. Reconstruction of terms for *Capsicum* in selected protolanguages of Mesoamerica and abutting areas

Years before present	Protolanguage[†]	Reconstructed word for chili[‡]	Location of modern descendant languages	Genetic affiliation
6591	Otomanguean	*$?ki^3$	Mexico	Otomanguean
5976	Eastern Otomanguean	*$(h)saH^3$, *ki	Mexico	Otomanguean
5498	Popolocan-Zapotecan	*ki	Mexico	Otomanguean
5357	Amuzgo-Mixtecan	*$n?sah^3$	Mexico	Otomanguean
4542	Mixtecan	*(H)ya?, Hyah, $H\theta a?^2$	Mexico	Otomanguean
4274	Totozoquean	NR	Mexico	Totozoquean
4018	Uto-Aztecan	NR	US Southwest, Mexico, Central America	Uto-Aztecan
3654	Otopamean	*(m)?i	Mexico	Otomanguean
3472	Southern Uto-Aztecan	NR	Mexico, Central America	Uto-Aztecan
3434	Kiowa-Tanoan	NR	US Southwest	Kiowa-Tanoan
3149	Zapotecan	*ki:?na?	Mexico	Otomanguean
3036	Popolocan	*hña	Mexico	Otomanguean
3000	Lencan	NR	Central America	Lencan
2774	Misumalpan	kuma	Central America	Misumalpan
2576	Northern Uto-Aztecan	NR	US Southwest	Uto-Aztecan
2445	Chiapanec-Mangue	*ni:-^{n}gi?	Mexico	Otomanguean
2400	Sonoran	*ko?okoLi	Mexico	Uto-Aztecan
2220	Mayan	*i:hk	Mexico, Central America	Mayan
1935	Chinantecan	*$?u^{HL}$	Mexico	Otomanguean
1865	Yuman	NR	US Southwest	Yuman
1737	Numic	NR	US Southwest	Uto-Aztecan
1587	Takic	NR	US Southwest	Uto-Aztecan
1509	General Aztec	*či:l	Mexico	Uto-Aztecan
1435	Totonacan	*pi'n	Mexico	Totozoquean
1407	Mixe-Zoquean	*ni:wi	Mexico	Totozoquean

NR, not reconstructable.
[†]Sources for each language are listed in *SI Materials and Methods* under *Paleobiolinguistics*.
[‡]Explanations for phonetic representation of pepper words are listed in *SI Materials and Methods* under *Paleobiolinguistics*.

which a term for chili pepper reconstructs are Proto-Misumalpan (∼2800 B.P.), Proto-Sonoran (∼2400 B.P.), and Proto-Mayan (∼2200 B.P.). Thus, the earliest non-Otomanguean dates for *Capsicum* in Mesoamerica and abutting regions are over 3,700 y more recent than the oldest date, suggesting that speakers of a prehistoric Otomanguean language or languages may have been among the first cultivators or domesticators of chili pepper. Note that the current word—chili—is derived from the General Aztec language, Nahuatl, which reconstructs to a much more recent date (∼1500 B.P.; Table 1).

The area of maximum diversity of a language family has been viewed traditionally by linguists as suggestive of the location of a family's ancestral language (e.g., ref. 28).We use this phylogenetic diversity information in locating the Otomanguean homeland by identifying where languages of the four subgroups of the family—Mazatecan-Zapotecan, Amuzgo-Mixtecan, Tlapanecan-Chorotegan, and Otopamean-Chinantecan (32)—are currently spoken in closest proximity (Fig. 1*C*).

Genetic Evidence. During the fall of 2006 and 2007, expeditions were conducted in the southern United States and throughout Mexico to sample populations of wild *C. annuum* (29). This provided the most complete set of wild *C. annuum* from Mexico available to date. Based largely on this set, 139 wild types distributed over the entire exploration area were chosen as were 49 domesticated types that are endemic landraces (*ancho*, *puya*, and *guajillo*) (33). This collection was screened with 17 simple sequence repeat (SSR) DNA markers (34, 35). These markers were chosen for this study because of their consistency of amplification and polymorphism within our sample. For each wild plant, a distance was calculated to the domesticated group based on the average proportion of shared SSR alleles. These distances were then spatially interpolated to produce in each grid cell an estimated genetic similarity between wild pepper populations (if any occurred in the cell) and the group of domesticated chili peppers (regardless of where they occurred). This molecular-marker–based analysis of genetic similarity between wild and domesticated types revealed a broad area of high similarity in the northeastern quadrant of Mexico (Fig. 1*D*), including the states of Tamaulipas, Nuevo León, San Luís Potosí, and Veracruz. In contrast, genetic similarity between wild and cultivated types was generally low in southern and northwest Mexico, confirming earlier results (7).

Consensus Model. The four lines of evidence—archaeological, ecological, paleobiolinguistic, and genetic—were all expressed as a spatial model and they can therefore be combined into a single consensus model represented geographically through mapping. Each type of evidence has its particular strengths and weaknesses, discussed below, which need to be taken into consideration when producing a consensus model. Because these merits and demerits are difficult to quantify (some are simply unknown), assigning differential weights to each line of evidence is problematic. Our solution is to present a number of different consensus maps based on several different weighting combinations (Fig. 2 and Fig. S3).

The first map, Fig. 2*A*, was established using equal weighting for each type of evidence (each weighted as making a 1/4 contribution). According to this model, areas in central-east Mexico and northeastern Mexico are the most likely area of origin of chili pepper. The second model assigned a high weight to genetic evidence (weighted 1/2) and equal but lower weights to the other three lines of evidence (each weighted 1/6). This assumes that genetic data might be superior to one or more of the other lines of evidence used because, for example, it might suffer less from sampling bias. This results in primary support for northeastern Mexico and only secondary support for central-east Mexico (Fig. 2*B*). The third approach assigned a low weight to archaeology (1/10) and equal higher weights to the other three lines of evidence (each weighted 1/3). This weighting was motivated by the observation that the current archaeological data are assembled from macroremains of only two sites in Mexico. This weighting produces a consensus model resembling the equal weighting of Fig. 2*A* because both central-east Mexico and northeastern Mexico result as equally plausible geographic candidates for chili pepper domestication (Fig. 2*C*). Additional information, from other sites and microremains, yet to be discovered, would justify a stronger weighting for archaeobotanical data.

Another weighting strategy produces different models based on randomly assigning combinations of weights for the four types of evidence. This approach allows us to explore the universe of possible weight combinations given different interpretations of the individual lines of evidence. Fig. S3 shows the percentile distribution obtained for this approach. The resulting maps suggest again that either central-east Mexico or northeastern Mexico or, conceivably, both areas were locations of the domestication of *C. annuum*.

Discussion

We have embraced the template of multidisciplinary approaches to study crop origins proposed first by de Candolle (36) and later
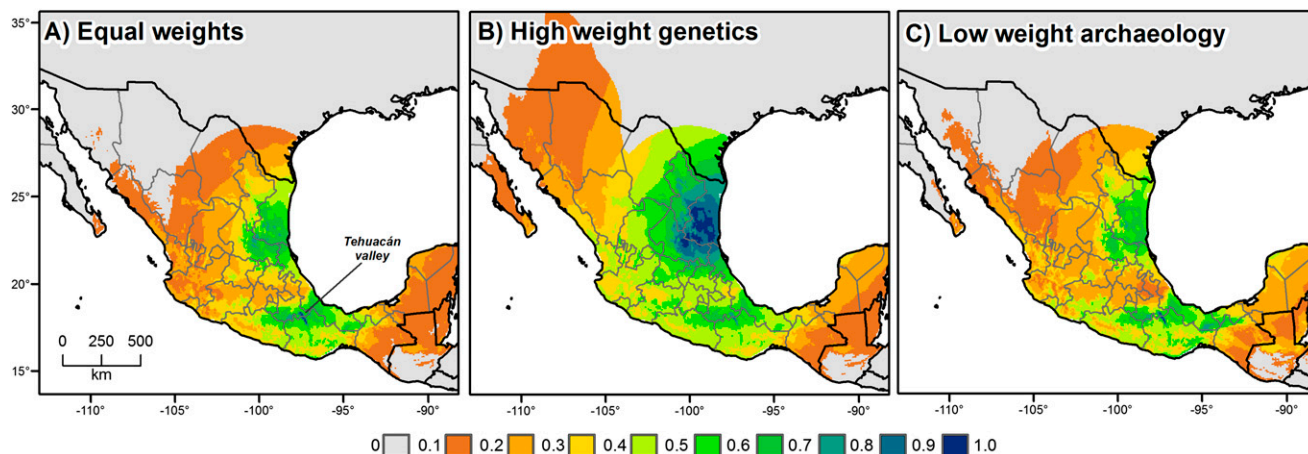

Fig. 2. Consensus models of the likelihood that cultivated chili pepper originated in an area. The models were obtained by combining the four lines of evidence for the origin of domesticated chili pepper (Fig. 1). (*A*) equal weights; (*B*) genetics 1/2, all others 1/6 weight; and (*C*) archaeology 1/10, all others 1/3 weight. After combining, the values were scaled between 0 and 1 and then squared to give more weight to the higher values.

by Harlan and de Wet (37). Confidence in a crop-origin hypothesis is increased when supported by multiple, independent lines of evidence, and improved understanding comes from new evidence in each field and concomitant predictions in other fields (38). Our multidisciplinary approach depends on the independence and strength of evidence from the different fields, each of which has its strengths and weaknesses.

Current archaeobotanical data for chili pepper is mainly based on macroremains from only two sites. In addition to the identification of ancient chili remains at additional sites, our understanding could benefit from the investigation of microfossil data such as starch grains (39) in Mesoamerican sites. Availability of microfossils may provide information on the more ancient distribution and importance of chili peppers and potentially also help distinguish domesticated from nondomesticated remains (as in the case of maize) (40).

The quality of species distribution models depends on having a representative sample of the current distribution of the wild species, the quality of climate data, particularly the modeled past climate data, and the algorithm used. Our sample size was large and the species is widespread, suggesting that the SDM approach should work well (41), as confirmed by a high cAUC score (30). Backcasted climate data for the mid-Holocene is, of course, uncertain; furthermore, we did not consider climate variation during that period. Nevertheless, because we use an ensemble of climate models (Fig. S4), our predictions should be relatively robust (42).

Utilization of linguistic data assumes an understanding of language development, including information relating to language origin, dispersal, and diffusion of traits across languages that is still emerging as new computer approaches are increasingly applied in linguistic analysis (43, 44). PBL provides an assessment of when species acquired substantial salience for prehistoric groups, whether they were merely harvested, cultivated, or eventually domesticated. If a word for a biological species reconstructs for a protolanguage, this is evidence that the species was known to and probably of considerable importance to speakers of the language as shown by Berlin et al. (25) for two closely related Mayan languages, Tzeltal and Tzotzil (Tzeltalan) and by Balée and Moore (26) in a study of plant names in five Eastern Amazonian Tupi-Guaraní languages.

Genetic data are generally based on the analysis of contemporary populations of the wild ancestor of the crop. The wild populations included in this study constitute the largest and most widespread sample used in genetic analyses for this species (29). However, we do not know to what extent the distribution and genetic structure of these populations have changed over the past 6,000 y. Hence, we modeled the past distribution of wild chili peppers based on the assumption that their climatic requirements are the same as today's wild chili pepper population. Correlation between the suitability scores for ancient and current distributions is high (0.92), suggesting that, whereas climate change over the past 6,000 y has likely shifted the species distribution, for the most part, the historical and current ranges of this species overlap. Another potentially confounding factor is gene flow between domesticated and wild chili peppers, which may cause similarities that are not due to ancestor–descendant relationships (45). However, this would not seem very important for chili peppers because they are mostly a self-fertilizing species with minimal outcrossing, which is confirmed by the high levels of homozygosity observed for wild chili pepper populations analyzed here Dataset S1, *Microsat info*).

The concept of origin of *C. annuum* used in this study encompasses wild plant protection, management, cultivation, and domestication. Within this continuum of increasingly close interaction between humans and plants, distinguishing among these four stages for most crops is difficult. However, with respect to chili pepper, the fact that a Proto-Otomanguean word for the crop was retained in daughter languages attests to its high salience for speakers of the protolanguage. Furthermore, Proto-Otomanguean speakers may have been actively engaged in cultivation, as suggested by the

reconstruction of words for a range of plants, including staple crops such as maize and squash, but also other crops such as avocado and nopal (22). Plausibly, then, the saliency of chili pepper among Proto-Otomanguean speakers reflects cultivation and perhaps incipient domestication and not merely use of a wild plant species.

When analyzed separately, our four lines of evidence do not all suggest the same geographic area as being the most likely place of chili pepper origin. Nevertheless, we identify central-east Mexico as a likely region of initial cultivation or incipient domestication because that interpretation most parsimoniously reconciles all evidence (Fig. 2*A*). This area extends from southern Puebla and northern Oaxaca to southern Veracruz and encompasses the valley of Tehuacán (Fig. 2*A*). The Coxcatlán Cave from which preceramic macroremains of chili pepper have been recovered (13) is situated in this valley. Species distribution modeling shows that many parts of the identified area were suited for the wild progenitor of *C. annuum* around the time of first cultivation or domestication in the mid-Holocene and there are currently populations of wild chili pepper that are genetically similar to the domesticated species (Fig. 1*D*). Near to the valley is the likely center of the Otomanguean homeland. Proto-Otomanguean, spoken in mid-Holocene times some 6,500 y ago, is the oldest ancestral language of the New World for which a term for chili pepper reconstructs. Speakers of contemporary Otomanguean languages live in or close to the region. Otomanguean people, then, may have been the first in the New World to transform wild chilies into the domesticated spice and condiment so widely enjoyed today.

By expressing all data as a distance, whether geographical (archaeological and linguistic data), climatic, or genetic, we have developed a method to bring together different lines of evidence about crop origins into a single framework of analysis. This approach has led to the discovery that the origin of domesticated chili peppers may have been located further south than previously thought (7) and in different regions of Mexico than proposed for common bean (46) or maize (47). Thus, our data do not suggest a single, nuclear area for crop domestication in Mesoamerica, but rather a multiregional model as suggested also for the Southwest Asian (48) and Chinese (49) centers of agricultural origins.

Materials and Methods

Archaeology. We used two locations for which there is evidence of the earliest use of chili: Romero's Cave (near Ocampo, Tamaulipas) and Coxcatlán Cave (Tehuacán Valley, Puebla) (Fig. 1*A*). We connected these locations by their shortest path, and then computed the distance *d* (in kilometers) to this path for cells on a raster with 1-km² spatial resolution. We truncated the distances at 1,000 km and used an inverse squared distance decay function, scaled between 0 and 1, (1 − (*d*/1,000)) as a measure of the likelihood that chili was domesticated in a location (grid cell).

Species Distribution Modeling. We used SDM to assess spatial variation in suitability for wild *C. annuum* var. *glabriusculum*, the ancestor of domesticated *C. annuum* (6), during climatic conditions of the mid-Holocene (about 6,000 y ago) (Fig. 1*B*). Locations where wild *Capsicum* populations currently occur were from collections made in the fall of 2006 and 2007 (Dataset S2, *Coordinates_w_SSR info*) (29) and from additional records obtained from the Global Biodiversity Information Facility (GBIF) (Dataset S3, *Coordinates GBIF*). We used the SDM algorithm MaxEnt (19) to predict suitability during the mid-Holocene according to nine global climate models (*SI Materials and Materials*).

Paleobiolinguistics. The different languages considered are listed in Table 1 and the respective information sources are compiled in *SI Materials and Methods*. Protolanguage dates (Table 1) were calculated through use of ASJP chronology (27). The center of phylogenetic diversity of Otomanguean languages was located from the distribution of places where Otomanguean languages are currently spoken (50), by determining the area where languages spoken in close geographic proximity to one another are found to be affiliated with the largest number of major divisions of the family.

Genetic Distance Analysis. Genetic distance between 139 wild and 49 domesticated pepper accessions (Datasets S2 and S3, *Coordinates GBIF*) were assessed

with data from 17 microsatellite markers (SSRs) developed before this study as described in Dataset S1, *Microsat info*).

Consensus Model. All four data sources were used to create a spatial model on a common raster. All models had values between 0 and 1, with higher scores indicating that a location is more likely to be the area where domestication occurred. We combined these four sources of data into a single consensus model by assigning weights to each indicator. To get a more pronounced differentiation between sites, we squared the values, after first rescaling them between 0 and 1.

1. Smith B (1995) *The Emergence of Agriculture* (Scientific American Library, New York).
2. Gepts P, et al. (2012) *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability* (Cambridge Univ Press, Cambridge, UK), pp 630.
3. Gepts P (2004) Domestication as a long-term selection experiment. *Plant Breed Rev* 24(Part 2):1–44.
4. Burke JM, Burger JC, Chapman MA (2007) Crop evolution: From genetics to genomics. *Curr Opin Genet Dev* 17(6):525–532.
5. Bosland PW (1994) Chiles: History, cultivation and uses. *Spices, Herbs and Edible Fungi*, ed Charambous G (Elsevier, New York).
6. Pickersgill B (1971) Relationships between weedy and cultivated forms in some species of chili peppers (genus *Capsicum*). *Evolution* 25:683–691.
7. Loaiza-Figueroa F, Ritland K, Cancino JAL, Tanskley SD (1989) Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Syst Evol* 165:159–188.
8. Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC (2009) Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *Am J Bot* 96(6):1190–1202.
9. Votava EJ, Nabhan GP, Bosland PW (2002) Genetic diversity and similarity revealed via molecular analysis among and within an in situ population and ex situ accessions of chiltepin (*Capsicum annuum* var. *glabriusculum*). *Conserv Genet* 3(2):123–129.
10. Smith CE (1967) Plant remains. *The Prehistory of the Tehuacan Valley*, ed Byers DS (Univ of Texas Press, Austin, TX), pp 220–255.
11. Smith CE, Jr. (1987) Current archaeological evidence for the beginning of American agriculture. *Studies in the Neolithic and Urban Revolutions, The V. Gordon Childe Colloquium*, ed Manzanilla L (British Archaeological Reports, Oxford, UK), pp 81–101.
12. Mangelsdorf PC, McNeish RS, Willey GR (1965) Origins of Middle American agriculture. *Natural Environment and Early Cultures*, ed West RC (Univ. of Texas Press, Austin, Texas), pp 427–445.
13. McClung de Tapia A (1992) The origins of agriculture in Mesoamerica and Central America. *The Origins of Agriculture: An International Perspective*, eds Cowan CW, Watson PJ (Smithsonian Institution Press, Washington, DC), pp 143–171.
14. Long A, Benz B, Donahue D, Jull A, Toolin L (1989) First direct AMS dates on early maize from Tehuacán, Mexico. *Radiocarbon* 31:1035–1040.
15. Smith BD (1997) Reconsidering the Ocampo Caves and the era of incipient cultivation in Mesoamerica. *Latin American Antiquity* 8(4):342–383.
16. Perry L, Flannery KV (2007) Precolumbian use of chili peppers in the Valley of Oaxaca, Mexico. *Proc Natl Acad Sci USA* 104(29):11905–11909.
17. Ranere AJ, Piperno DR, Holst I, Dickau R, Iriarte J (2009) The cultural and chronological context of early Holocene maize and squash domestication in the Central Balsas River Valley, Mexico. *Proc Natl Acad Sci USA* 106(13):5014–5018.
18. Elith J, Leathwick JR (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst* 40(1):677–697.
19. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190(3–4):231–259.
20. Jarvis A, Lane A, Hijmans RJ (2008) The effect of climate change on crop wild relatives. *Agric Ecosyst Environ* 126(1–2):13–23.
21. Jarvis A, et al. (2005) Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genet Resour Crop Evol* 52(6):671–682.
22. Brown CH (2010) Development of agriculture in prehistoric Mesoamerica: The linguistic evidence. *Pre-Columbia Foodways: Interdisciplinary Approaches to Food, Culture and Markets in Ancient Mesoamerica*, eds Staller JE, Carrasco MD (Springer, Berlin), pp 71–107.
23. Fowler CS (1972) Some ecological clues to Proto-Numic homelands. *Desert Research Institute Publications in the Social Sciences* 8:105–117.
24. Hill JH (2001) Proto-Uto-Aztecan: A community of cultivators in central Mexico? *Am Anthropol* 103:913–934.
25. Berlin B, Breedlove DE, Laughlin RM, Raven PH (1973) Cultural significance and lexical retention in Tzeltal-Tzotzil ethnobotany. *Meaning in Mayan Languages*, ed Edmonson MA (Mouton, The Hague), pp 143–164.
26. Balée WL, Moore D (1991) Similarity and variation in plant names in five Tupi-Guarani languages (Eastern Amazonia). *Bulletin of the Florida Museum of Natural History* 35(4):209–262.
27. Holman EW, et al. (2011) Automated dating of the world's language families based on lexical similarity. *Curr Anthropol* 52(6):841–875.
28. Wichmann S, Muller A, Velupillai V (2010) Homelands of the world's language families: A quantitative approach. *Diachronica* 27(2):247–276.
29. Kraft KH, Luna-Ruiz JD, Gepts P (2013) A new collection of wild populations of Capsicum in Mexico and the southern United States. *Genet Resour Crop Evol* 60(1):225–232.
30. Hijmans RJ (2012) Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology* 93(3):679–688.
31. Brown CH, Clement CR, Epps P, Luedeling E, Wichmann S (2013) The paleobiolinguistics of domesticated chili pepper (*Capsicum* spp.). *Ethnobiology Letters* 4:1–11.
32. Kaufman TS (1990) Early Otomanguean homeland and cultures: Some premature hypotheses. (University of Pittsburgh Working Papers in Linguistics 1), pp 91–136.
33. Kraft KH, de Jesús Luna-Ruíz J, Gepts P (2010) Different seed selection and conservation practices for fresh market and dried chile farmers in Aguascalientes, Mexico. *Econ Bot* 64(4):318–328.
34. Lee JM, Nahm SH, Kim YM, Kim BD (2004) Characterization and molecular genetic mapping of microsatellite loci in pepper. *Theor Appl Genet* 108(4):619–627.
35. Minamiyama Y, Tsuro M, Hirai M (2006) An SSR-based linkage map of *Capsicum annuum*. *Mol Breed* 18(2):157–169.
36. de Candolle A (1882) *L'Origine des Plantes Cultivées [The Origin of Cultivated Plants]* (Appleton, New York), pp 468.
37. Harlan JR, de Wet JMJ (1973) On the quality of evidence for origin and dispersal of cultivated plants. *Curr Anthropol* 14:51–62.
38. Ammerman AJ, Cavalli-Sforza LL (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton, NJ), pp 176.
39. Perry L, et al. (2007) Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. *Science* 315(5814):986–988.
40. Holst I, Moreno JE, Piperno DR (2007) Identification of teosinte, maize, and Tripsacum in Mesoamerica by using pollen, starch grains, and phytoliths. *Proc Natl Acad Sci USA* 104(45):17608–17613.
41. Wisz MS, et al. (2008) Effects of sample size on the performance of species distribution models. *Divers Distrib* 14(5):763–773.
42. Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22(1):42–47.
43. Perreault C, Mathew S (2012) Dating the origin of language using phonemic diversity. *PLoS ONE* 7(4):e35289.
44. Bouckaert R, et al. (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097):957–960.
45. Papa R, Gepts P (2004) Asymmetric gene flow and introgression between wild and domesticated populations. *Introgression from Genetically Modified Plants into Wild Relatives and Its Consequences*, eds Den Nijs D, Bartsch D, Sweet J (CABI, Oxon, UK), pp 125–138.
46. Kwak M, Kami JA, Gepts P (2009) The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago basin of Mexico. *Crop Sci* 49(2):554–563.
47. Matsuoka Y, et al. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci USA* 99(9):6080–6084.
48. Fuller DQ, Willcox G, Allaby RG (2011) Cultivation and domestication had multiple origins: Arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeol* 43(4):628–652.
49. Cohen DJ (2011) The beginnings of agriculture in China: A multiregional view. *Curr Anthropol* 52(S4):S273–S293.
50. Avila-Blomberg A de, Moreno-Díaz NG (2008) Distribución de las lenguas indígenas de México [Distribution of the indigenous languages of Mexico]. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico, DF.