Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica

Brian M. Kemp[a,b,1], Angélica González-Oliver[c], Ripan S. Malhi[d], Cara Monroe[b,e], Kari Britt Schroeder[f], John McDonough[f,2], Gillian Rhett[g], Andres Resendéz[h], Rosenda I. Peñaloza-Espinosa[i], Leonor Buentello-Malo[j], Clara Gorodesky[k], and David Glenn Smith[f]

[a]Department of Anthropology, Washington State University, Pullman, WA 99164-4910; [b]School of Biological Sciences, Washington State University, Pullman, WA 99164-4236; [c]Departmento de Biología, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; [d]Department of Anthropology, University of Illinois, Urbana, IL 61801; [e]Department of Anthropology, University of California, Santa Barbara, CA 93106-3210; [f]Department of Anthropology, University of California, Davis, CA 95616; [g]Vertebrate Ecology Lab, Moss Landing Marine Laboratories, Moss Landing, CA 95039; [h]Department of History, University of California, Davis, CA 95616; [i]Unidad de Investigación Médica en Genética Humana, Centro Médico Nacional, Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 04510, Mexico; [j]Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México, Mexico City 06703, Mexico; and [k]Departamento de Inmunología e Inmunogenética, Secretaria de Salud, Instituto Nacional de Diagnóstico y Referencia Epidemiológico, Mexico City 11340, Mexico

Edited by Francisco Mauro Salzano, Instituto de Biociencias, Porto Alegre, Rio Grande do Sul, Brazil, and approved February 25, 2010 (received for review June 23, 2009)

The Farming/Language Dispersal Hypothesis posits that prehistoric population expansions, precipitated by the innovation or early adoption of agriculture, played an important role in the uneven distribution of language families recorded across the world. In this case, the most widely spread language families today came to be distributed at the expense of those that have more restricted distributions. In the Americas, Uto-Aztecan is one such language family that may have been spread across Mesoamerica and the American Southwest by ancient farmers. We evaluated this hypothesis with a large-scale study of mitochondrial DNA (mtDNA) and Y-chromosomal DNA variation in indigenous populations from these regions. Partial correlation coefficients, determined with Mantel tests, show that Y-chromosome variation in indigenous populations from the American Southwest and Mesoamerica correlates significantly with linguistic distances ($r = 0.33–0.384$; $P < 0.02$), whereas mtDNA diversity correlates significantly with only geographic distance ($r = 0.619$; $P = 0.002$). The lack of correlation between mtDNA and Y-chromosome diversity is consistent with differing population histories of males and females in these regions. Although unlikely, if groups of Uto-Aztecan speakers were responsible for the northward spread of agriculture and their languages from Mesoamerica to the Southwest, this migration was possibly biased to males. However, a recent in situ population expansion within the American Southwest (2,105 years before present; 99.5% confidence interval = 1,273–3,773 YBP), one that probably followed the introduction and intensification of maize agriculture in the region, may have blurred ancient mtDNA patterns, which might otherwise have revealed a closer genetic relationship between females in the Southwest and Mesoamerica.

mtDNA | Uto-Aztecan | Y chromsome | maize | migration

Scholars have long noted the close connections, both past and present, between indigenous cultures of the U.S. Southwest and Mesoamerica (1), which are exemplified by the spread of maize from Mesoamerica to the Southwest and the distribution of speakers of Uto-Aztecan (UA) languages across the regions (2). It has been proposed that UA speakers were, in fact, responsible for the northward spread of maize cultivation (3–5). However, previous genetic and morphologic studies have failed to provide direct evidence for an ancient spread of UA speakers (6–8).

The American Southwest is both a cultural and geographic region that spans from the Mexican states of Baja California and Durango in the south to Southern Utah and Colorado in the north and west to east from approximately Las Vegas, NV to Las Vegas, NM (9). The southern boundary, however, may extend as far south as the Mexican states of Nayarit and Jalisco, encompassing an area called the Greater Southwest (10). Mesoamerica is neither a geographic region nor a socio-political unit (11), but rather, it is an area occupied by populations that shared a suite of cultural characteristics, which was first defined by Kirchhoff (12). The northern limit of Mesoamerica has fluctuated throughout prehistory, both expanding opportunistically and contracting under the threats of nomadic tribes to the north, and at its peak, it overlapped the southern frontier of the Southwest (11, 12).

The earliest widely accepted evidence of maize comes from San Marcos Cave in the Tehuacán region of Oaxaca, accelerator mass spectrometry dated to ~5,600 years before present (YBP) (dates are presented as calibrated calendar years unless otherwise noted) (13). However, recent phytolith data have pointed to an origin in the lowland tropics of Tabasco more than 7,000 YBP (14). Whereas the origin and timing of domestication continues to be debated in Mesoamerica, the introduction of maize into the Southwest before 4,000 YBP is no longer disputed (15, 16). The early appearance of this cultigen in the Southwest, soon after its widespread use in Mesoamerica, has been used to support the hypothesis that maize cultivation spread with humans from Mesoamerica and therefore, was not spread primarily through cultural diffusion (3–5).

Because UA is the only language family to extend across Mesoamerica and the Southwest, members of this language family may have played a prominent role in the interactions that took place between the two regions. The structure of the language family and its diversity as well as reconstructed Proto-Uto-Aztecan (PUA) vocabulary suggest a northern origin for UA somewhere in the Southwest or Southern California (2, 17–20). Recently, Merrill et al. (16) used linguistic data to support a PUA homeland in the Great Basin, which also provides additional support for a northern origin. In contrast, Hill (3) argued that UA originated in the vicinity of where maize was domesticated, perhaps in Central Mexico. As part of this southern-origin hypothesis, Hill (3) argued that the innovation of maize agriculture precipitated a population expansion northward,

Author contributions: B.M.K., A.G.-O., R.S.M., and D.G.S. designed research; B.M.K., A.G.-O., R.S.M., C.M., K.B.S., J.M., G.R., A.R., R.I.P.-E., L.B.-M., and C.G. performed research; B.M.K., A.G.-O., and R.S.M. analyzed data; and B.M.K., A.G.-O., R.S.M., C.M., K.B.S., and D.G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The mitochondrial DNA sequences reported in this paper have been deposited in the Genbank database (accession nos. EU719668–EU720177, EU720188–EU720206, and EU20221–EU720379).

[1]To whom correspondence should be addressed. E-mail: bmkemp@wsu.edu.

[2]Deceased September 2, 2005.

This article contains supporting information online at www.pnas.org/cgi/content/full/0905753107/DCSupplemental.

bringing maize agriculture and UA languages to the Southwest. This is a prediction of the Farming/Language Dispersal Hypothesis (FLDH) that has been put forth to explain how the most widely spread language families came to be distributed at the expense of those that have more restricted distributions today (21–23).

Additional cultural similarities attest to similar world perspectives held by individuals residing in both areas (1, 10, 24–27). Moreover, turquoise mined in the Southwest was traded southward to Mesoamerica, and goods of Mesoamerican origin were traded northward to the Southwest along the trade route known as the Turquoise Road (1, 24, 26, 28, 29).

With the wealth of cultural and linguistic similarities between the regions and especially the hypothesis of whole-scale population movement from Mesoamerica to the Southwest, it is reasonable to hypothesize close genetic ties between populations in the two regions. Although studies of protein polymorphisms have identified similarities in gene frequencies between populations in the two regions (7, 30–33), previous mitochondrial DNA (mtDNA) data provided no evidence of a population expansion northward from Mesoamerica (6, 7). Southwest populations are characterized by high frequencies of mitochondrial haplogroup B and very low frequencies, or the complete absence, of haplogroup A. The exception to this pattern is the high frequency of haplogroup A found among Southwest Athapaskan populations (Navajo and Apache) that entered the region from the north within last ~500 years (6). As found in *Results and Discussion*, this population movement played little role in our understanding of the relationships between populations of the Southwest and Mesoamerica. In contrast to non-Athapaskan Southwest populations, Mesoamerican populations exhibit high frequencies of haplogroup A with moderate to low frequencies of haplogroup B (6, 7, 34–37). These haplogroup patterns crosscut linguistic and geographic boundaries within both regions, and predate European contact in the Southwest (38, 39) and Central Mexico (40, 41). In addition, phylogeographic analyses of mtDNA haplotypes have not supported a close relationship among UA-speaking populations in the Southwest and Central Mexico (6). Together with the mtDNA variation exhibited by populations in the Great Basin (42) and California (43), one finds that UA-speaking populations exhibit greater similarity within than across these regions.

Based on mtDNA patterns and the presence of the rare variant Albumin*Mexico in populations of Mesoamerica and the Southwest (7), it has been hypothesized that male movements have been the source for transmission of Albumin*Mexico, maize cultivation, and the UA language(s) across the regions but not mtDNA, which is maternally inherited (6, 7). Bellwood (23), a proponent of the FLDH, was not entirely convinced by this conclusion and stated that the "prospect of females staying close to home and males migrating makes a degree of sense, but one wonders how the claimed results reflect sampling and other factors" (23). Indeed, mtDNA of Mexican populations north of Mexico City had previously only been sampled from two cities that border the United States (44) and the Seri (6). Moreover, mitochondrial haplogroup affiliation of only 120 Native Americans from Mesoamerica had been identified before the comment made by Bellwood (23), and hypervariable region I (HVRI) haplotypes of only 15 of these individuals have been determined (6, 34, 35, 45–47). To date, no analyses of Y-chromosome variation have been conducted to directly investigate prehistoric movement of males between these regions.

If the conditions behind the FLDH explain how UA was spread across these regions, it is first predicted that UA speakers in the Southwest and Mesoamerica should be more closely related to each other than with non-UA speakers. In other words, genetic distances should correlate in a positive manner with linguistic but not geographic distances. Moreover, shared derived variation should be identified among UAs. Although these predictions might hold for a Southwest origin of UA as well, they are essential to the FLDH. It is these predictions of the hypothesis that are evaluated here with a large-scale examination of mtDNA and Y-chromosome variation.

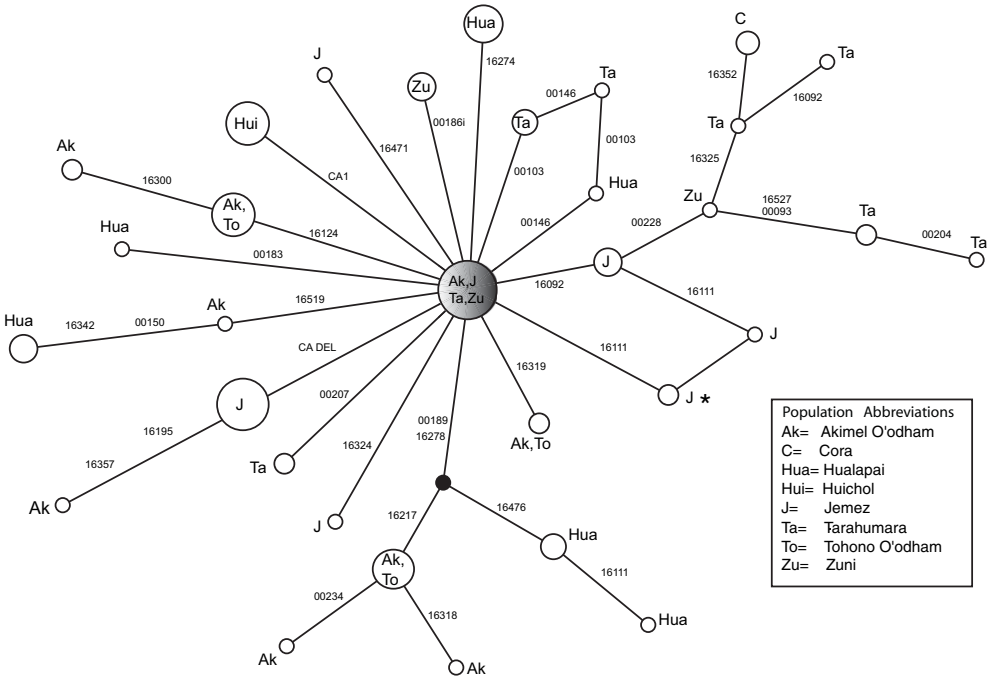

Fig. 1. Haplogroup B network containing all haplotypes that exhibit an adenine (A) at np 16483. The central node (shaded gray) exhibits the following mutations relative to the Cambridge reference sequence (75, 76): 16111, 16189, 16217, 16483, 16519, 00073, 00263, and 00499. Mutational positions from this haplotype are noted in small print. The haplotype marked with an asterisk links to the gray-shaded haplotype in Fig. S3 by the transition at np 16483. Black circles represent median vectors, haplotypes that existed at one time but are now extinct, or haplotypes that were simply not sampled in this study.

Table 1. Results of the Mantel tests

Mitochondrial DNA, geography, and language		Language Miller	*P* value	Language Hale/Hill	*P* value	Language Simple	*P* value
Correlation coefficient	mtDNA/geography	0.551*	0.000*	0.551*	0.000*	0.551*	0.000*
	mtDNA/language	0.273*	0.020*	0.259	0.057	0.245	0.081
	Geography/language	0.318*	0.004*	0.276*	0.009*	0.219	0.082
Partial correlation coefficient	mtDNA/geography	0.508*	0.000*	0.516*	0.001*	0.525*	0.000*
	mtDNA/language	0.124	0.204	0.134	0.223	0.153	0.196
	Geography/language	0.211*	0.033*	0.165	0.061	0.153	0.196

*Significant correlations and their associated *P* values.

Results and Discussion

In general, mitochondrial haplogroup frequencies differ substantially between Southwest and Mesoamerican populations; the former rarely exhibits mitochondrial haplogroup A, but it predominates in the latter (Table S1 and Fig. S1). In contrast, haplogroup B is very common in Southwest populations, but it is much less common in Mesoamerican populations. Two exceptions to this pattern are (*i*) the Nahua–Atocpan, a Mesoamerican population that exhibits slightly more haplogroup B than haplogroup A, and (*ii*) the Tarahumara, a Southwest population that exhibits a higher frequency of haplogroup A than haplogroup B (Table S1 and Fig. S1). The Cora and Huichol also exhibit over 20% greater frequencies of haplogroup B than haplogroup A. Because these two populations are grouped in Mesoamerica by some (11) and in the Southwest by others (10), the geographic distributions of haplogroup A and B across the Southwest and Mesoamerica are more clinal than previously described (6), but this does not provide any support for particularly close relations among UA speakers across the regions, a prediction of the FLDH.

The principal coordinates analysis (PCoA) plot of F_{ST} values based on mtDNA haplogroup frequencies (*SI Text*) and haplotypes (Fig. S2*B*) reflects the same geographic division of haplogroups A and B as just described, dividing the Southwest from Mesoamerica on the first principal coordinate. The second principal coordinate reflects substructure in the Southwest between Pueblo groups (Anasazi, Jemez, and Zuni) and Pimans/Yumans (Akimel O'odham, Hualapai, and Tohono O'odham) (Fig. S2 *A* and *B*).

The Y-chromosome haplogroup frequencies exhibited by the populations in this study are provided in Table S1. Haplogroup Q-M3 was detected in all populations and ranged in frequency from ~43–100%. Haplogroup Q-M242, ranging in frequency from ~5–43%, was in UA populations in both the Southwest and Mesoamerica as well as the Jemez. Non-Native American admixture (i.e., PM-45, M-173, and Y-chromosome Alu Polymorphism) ranging from ~5–45% was detected in five populations.

The vast majority of mtDNA and Y-chromosome haplotypes were unique to individuals and/or shared by individuals from the same population. Three hundred twenty-three mitochondrial haplotypes were identified among the 689 sequences belonging to haplogroups A, B, C, D, and X (*SI Text*). The majority of these haplotypes (60.4%), equally distributed among the haplogroups, are unique to individuals, whereas 30.3% of them are shared among individuals of the same population. Thus, 90.7% of the haplotypes are population-specific, and 9.3% are shared between or among populations. Ninety-eight haplotypes were identified among the 121 Y chromosomes studied for their short tandem repeat variation (*SI Text*). The majority of these haplotypes (83.7%) are unique to individuals; 10.2% were shared among individuals within the same population, and only 6.1% were shared between populations. The PCoA plot based on the Y-chromosome data shows that the closest Y-chromosome relationships between the Southwest and Mesoamerica predominantly involve UA speakers, consistent with predictions of the FLDH (Fig. S2*C*).

The mtDNA haplotype networks are found in Fig. 1 and *SI Text*. The vast majority of subclades within the networks were specific to regions but not language families. In particular, few UA-specific mtDNA clades were seen in the networks, contrary to expectations of the FLDH.

None of the haplogroup A haplotypes sampled here in the Southwest exhibited the characteristic Athapaskan markers (16233G and 16331G), which confirms that little mtDNA gene flow occurred from Navajo and Apache into other populations in the Southwest and supports previous studies (6, 7, 48). Of further note, within haplogroup A, one clade containing a Tohono O'odham and Zuni [derived by nucleotide positions (nps) 16257T and 16263A] is of interest, because this form of haplogroup A is only found elsewhere among the Chumash of Southern California (43). Combined with the absence of shared derived forms of haplogroup A across the regions, this cautions against the interpretation that haplogroup A in the Southwest was introduced by farmers from Mesoamerica (49), where haplogroup A is far more common.

Within haplogroup B, 130 of 250 individuals (52%) were derived at np 16483A, and most of these individuals were also derived at np 16111T. This subhaplogroup is termed B2a (50), and the network depicting the relationships between the B2a lineages is particularly interesting (Fig. 1). First, these derived lineages are found in every Southwest population sampled in addition to the Cora and Huichol. Although 126 of 246 individuals (51.2%) that belonged to haplogroup B in the Southwest belonged to this clade (or ~29% of all of the individuals of these populations), not a

Table 2. Results of the Mantel tests

Y chromosome, geography, and language*		Language Miller	*P* value	Language Hale/Hill	*P* value	Language Simple	*P* value
Correlation coefficient	Y chromosome/geography	−0.014	0.503	−0.014	0.503	−0.014	0.502
	Y chromosome/language	0.278†	0.045†	0.375†	0.037†	0.252	0.084
	Geography/language	0.305†	0.045†	0.251	0.101	0.196	0.141
Partial correlation coefficient	Y chromosome/geography	−0.108	0.712	−0.015	0.705	−0.067	0.625
	Y chromosome/language	0.296†	0.037†	0.342†	0.034†	0.360	0.082
	Geography/language	0.321†	0.039†	0.270	0.088	0.207	0.125

*These Mantel tests only included mtDNA data from the 10 populations for which Y-chromosome data were available.
†Significant correlations and their associated *P* values.

Table 3. Results of the Mantel tests

Y chromosome, mitochondrial DNA, and language*		Language Miller	*P* value	Language Hale/Hill	*P* value	Language Simple	*P* value
Correlation coefficient	Y chromosome/mtDNA	−0.200	0.838	−0.200	0.840	−0.264	0.928
	Y chromosome/language	0.278†	0.043†	0.328†	0.033†	0.252	0.086
	mtDNA/language	0.214	0.139	0.205	0.211	0.171	0.236
Partial correlation coefficient	Y chromosome/mtDNA	−0.276	0.928	−0.288	0.940	−0.322	0.970
	Y chromosome/language	0.33†	0.019†	0.384†	0.018†	0.313	0.051
	mtDNA/language	0.286	0.053	0.290	0.111	0.254	0.128

*These Mantel tests only included mtDNA data from the 10 populations for which Y-chromosome data were available.
†Significant correlations and their associated *P* values.

single Mesoamerican individual exhibited the 16483A transition. This form of haplogroup B has previously been detected in low frequency among Yavapai, Kumeyaay, Cochimi (6), the Washo (51), and the Turtle Mountain Chippewa (52). It might also be present in various Yuman and UA populations from Southern California (43) and the Apache and Navajo of the Southwest (53), because they exhibit the 16111T mutation; however, np 16483 was not sequenced for these populations.

Second, available D-loop sequences indicate that haplogroup B haplotypes derived at np 16483A are not found in Asian populations [based on the samples screened by Kemp et al. (54)], and whole-genome sequences reveal that the mutation is unique to Native Americas (55). This suggests that subhaplogroup B2a evolved in the Americas. Nucleotide diversity (π) within our samples of B2a was estimated in Mega 3.0 (56) with 10,000 bootstraps of the data to be 0.0020 [95% confidence interval (CI) $\pm$ 0.000]. The approximate age of this clade is 2,105 YBP (99.5% CI = 1,273–3,773 YBP), which was estimated by using Howell et al.'s (57) average rate of 47.5% per site/myr (99.5% CI = 26.5–78.5% per site/million years) for the evolution of the D-loop (nps 16024–00576). This pedigree-based rate is appropriate for estimating the age of events that have occurred within the past 15,000 years (58), a period that encompasses most, if not all, of the occupation of the Americas (59). The estimated date of this expansion closely coincides with the date of dramatic increase in population size in the Southwest estimated from archaeologic evidence (15).

Third, these data indicate that there has been substantial interlanguage family admixture in the Southwest, possibly since (and probably during) the time of the initial expansion. The expansion seems to have been one that was region-wide, encompassing all of the diverse populations within the greater Southwest. It is possible that risk management associated with dry farming of maize may have lead to substantial social and political reformation (60), including increased reliance on long-distance relationships that facilitated migration (61) and presumably gene flow as a consequence. In addition, increase in population size could have lead to more interpopulation contact and as a consequence, increased admixture.

Fourth, it is possible that because of the magnitude of this recent expansion, genetic patterns of greater antiquity have been obscured. If an in situ Southwest expansion did occur within the past 4,000 years, it may inform us little, if at all, about the proposed earlier movement of UA females from central Mexico into the Southwest (3). However, if UA originated in the Southwest, this population expansion may have spread the language family south as far as Nayarit and Jalisco where the Cora and Huichol reside. Currently, the only available ancient DNA (aDNA) evidence from the Southwest cannot address this issue, because the oldest samples analyzed are only ~1,600 years old (38, 39). These data confirm only that haplogroup B has been the most common haplogroup in the region for at least 1,600 years. The analysis of aDNA from early farmers and populations that predate the expansion will be required to characterize the gene pool of the Southwest at these times in relationship to Mesoamerica and provide a further test of this hypothesis.

It is unclear why this expansion is largely detectable in only mitochondrial haplogroup B and does not eliminate correlations between Y-chromosome variation and linguistic distances. It is possible that members of mitochondrial haplogroups A and D in the Southwest are so infrequent (Table S1) that they do not display a similar pattern. Although mitochondrial haplogroup C is relatively common, the haplotype network of this haplogroup (*SI Text*) may be confounded by containing members of at least four newly described founding lineages that cannot be differentiated by D-loop sequence alone (59, 62). As to the difference in pattern between the uniparentally inherited markers, Balaresque et al. (63) recently discovered that the advent of farming allowed for an expansion of Neolithic males in Europe, but a similar expansion was not detected in the mtDNA. Our case seems to be just the opposite.

The results of the Mantel tests show that mtDNA distances across the Southwest and Mesoamerica are positively and significantly correlated at the 0.05 level of probability with geographic distance but not with linguistic distances (Table 1). This result is in support of the recent report that linguistic and mtDNA diversity are not correlated among indigenous Mexican populations (36). The results of the Mantel tests support the overall trend of a major subdivision between populations from Mesoamerica and the Southwest on the direct maternal line.

In contrast, Y-chromosome diversity correlates positively and significantly with linguistic distances when the internal structure of UA was considered (i.e., when using the Miller and Hale/Hill models) but not when it was ignored (i.e., when using the Simple model) (Tables 2 and 3). Previous research has also indicated some agreement between Y-chromosome variation and membership in UA (64), supporting the notion that the internal structure is correlated with genetic distances between populations. However, it is clear from the present study that Y-chromosome diversity is not correlated with either geography or mtDNA distances (Tables 3 and 4). Overall, these tests elude to differing population histories of

Table 4. Results of the Mantel tests

Y chromosome, mitochondrial DNA, and geography*		Coefficient	*P* value
Correlation coefficient	Y chromosome/mtDNA	−0.200	0.841
	Y chromosome/geography	−0.014	0.503
	mtDNA/geography	0.619†	0.002†
Partial correlation coefficient	Y chromosome/mtDNA	−0.243	0.900
	Y chromosome/geography	0.142	0.191
	mtDNA/geography	0.629†	0.002†

*These Mantel tests only included mtDNA data from the 10 populations for which Y-chromosome data were available.
†Significant correlations and their associated *P* values.

males and females in these regions, especially with regards to UA prehistory. In other words, the Y-chromosome variation is predicted by the structure of UA and its relationship to non-UA language families, whereas mtDNA variation is not. Interestingly, because the correlation between linguistic distances and Y-chromosome variation disappears as the internal structure of UA is ignored (i.e., when using the Simple model), it seems that the degree of relation between UA males is as predicted by linguists. It does not, however, provide any sense of direction for the movement of UA. These results are also interesting with regards to Belle and Barbujani's (65) discovery that, at a global level, linguistic patterns have a nonnegligible, albeit small, correlation with autosomal genetic variation. Although their data cannot point to any sex bias, we detected one in our study area.

Because of the number of cultural and linguistic connections between populations of the Southwest and Mesoamerica, it was hypothesized that populations in these two regions would exhibit close genetic ties as well. In this study, we found evidence that males in the Southwest and Mesoamerica are as genetically related to one another as predicted by the proposed relationships of the languages that they speak. Yet, despite a substantial increase in sampling, mtDNA variation across these regions remains strongly correlated with geography and not with language family. However unlikely, these data suggest that if a migration of UAs was responsible for introducing maize agriculture to the Southwest as previously hypothesized by linguists and archaeologists (3, 4, 66), it was predominantly comprised of males and likely not as a result of a demographic expansion.

This scenario would explain why previous mtDNA studies failed to find a connection between the regions, whereas a rare nuclear DNA maker, Albumin*Mexico, clearly unites them (6, 7). However, it is also possible that the recent expansion of mtDNA haplogroup B within the Southwest blurred the preexisting mtDNA structure of Southwest populations (i.e., >4,000 years ago) that may have evinced a genetic relationship between the mtDNA of the two regions at one time. Thus, future studies of aDNA may alter our understanding of the prehistory of Southwest and Mesoamerican populations.

Materials and Methods

Samples. MtDNA variation was studied in 848 individuals from 13 populations from Mesoamerica and the American Southwest (*SI Text* provide sources of samples). These data were combined with those from seven previous studies (Table S1) totaling 960 individuals. Y-chromosome variation was studied in 178 males belonging to 11 of these populations (Table S1). Although the sampling was focused on populations that speak languages belonging to the UA language family, non-UA samples were included for comparative purposes (Table S1). A single sample from the Tepehuan, a UA population, was included only in the network analysis. Populations were assigned to the Southwest or Mesoamerica (Table S1). The Cora and Huichol, because of their intermediate geographic locations and their previous placement in either the Southwest (10) or Mesoamerica (11), are considered as special cases.

DNA Extraction, Haplogroup Determination, and Sequencing. DNA was extracted from buccal swabs and blood samples with the Qiagen Blood Amp Kit. All of the samples were screened for the polymorphisms that define Native American mitochondrial haplogroups A, B, C, D, and X. nps 16011–00684 (*SI Text* and Table S4) of the mitochondrial genome, representing the entire D-loop, was determined for 716 individuals. The 178 male samples were screened for six Y-chromsome binary polymorphisms: M3, M242, $RPS4Y_{711}$, M45, M173, and YAP (DYS287). In addition, eight Y-STR loci (DYS19, DYS390, DYS391, DYS392, DYS393, DYS389I, DYS389II, and DYS439) were typed in 121 males belonging to haplogroups Q-M3 and Q-M242 (*SI Text* has detailed laboratory methods).

Data Analysis. For mtDNA analysis, all individuals that do not belong to haplogroup A–D or X (Table S1) were excluded, because they most likely represent non-Native American admixture.

F_{ST} values were calculated for all pairs of populations in Arlequin (version 2.000) (67). A distance matrix was constructed of all pair-wise F_{ST} values and used to conduct PCoA performed in the program DistPCoA using the Caillez method to correct for negative eigenvalues (68). Mantel tests were performed in Arlequin (version 2.000) (67) with 100,000 permutations of the data to test for correlations between genetic, geographic, and linguistic distances between populations (*SI Text*) at the 0.05 level of probability. The genetic distances were estimated as pair-wise F_{ST} values. The geographic distances were calculated in GenAlEx (69) using latitude and longitude coordinates (*SI Text*) determined to be central to precontact population ranges or from the locations where the samples were collected. Precontact ranges of the Southwest populations were inferred from the *Handbook of North American Indians* (2). The precontact ranges of the Mixe, Mixtec, and Zapotec are found in Hollenbach et al. (70).

Linguistic distances were estimated in three ways. First, the Miller estimate was based on time estimates of UA language splits taken from Fig. S2*B* of Miller (2). A 5,500-year age was assigned to the Mixtec–Zapotec split (71), and language-family splits were assigned time depths of 8,000 years. Second, the Hale/Hill estimate was taken from Hale (72) with the Nahua de Mecayapan used to represent both Nahua–Atocpan and Nahua–Cuetzalan; although Hale (72) noted some problems with the Tarahumara dates, we have taken his numbers exactly. The Cora–Huichol and Mixtec–Zapotec splits were assigned time depths of 1,350 (72) and 5,500 (71) years, respectively, and language-family splits were assigned time depths of 6,000 years, reflecting the shorter time depth of UA proposed by Hale (73) compared with Miller (2). Lastly, the Simple estimate was made by assigning values of 0 (minimum) to intralanguage family distances and 1 (maximum) to interlanguage family distances. This estimate ignores any intralanguage family structure.

Median-joining haplotype networks were constructed in Network (version 4.1.1.2) (74) separately for haplogroups A, B, C, and D. No network was constructed for haplogroup X, because only one haplotype of this haplogroup was detected in this study (exhibited by eight Jemez). *SI Text* has details of network construction.

ACKNOWLEDGMENTS. We thank all the individuals who provided and authorized use samples in this study. We are grateful to Jane Hill for proving us expert advice on the Uto-Aztecan language family. Thanks to Tim Kohler for some important and compact advice on the possible link between dry farming and gene flow. This project was funded by The University of California Institute for Mexico and the United States and Wenner–Gren grants (B.M.K.). Additional support was provided by Grant BCS 0422144 from the National Science Foundation (to B.M.K., R.S.M., A.R., and D.G.S.) and Grant R24RR005090 from the National Center for Research Resources of the National Institutes of Health (to D.G.S.). A.G.-O. was supported by Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México.

1. Hedrick BC, Kelley JC, Riley CL (1974) The Mesoamerican Southwest:a Readings in Archaeology, Ethnohistory, and Ethnology (Southern Illinois University Press, Carbondale, IL), p 178.
2. Miller WR (1983) *Uto-Aztecan Languages. Handbook of North American Indians, Southwest*, ed Sturtevant WC (Smithsonian Institution, Washington, DC), Vol 10, pp 113–124.
3. Hill JH (2001) Proto-Uto-Aztecan: A community of cultivators in central Mexico? *Am Anthropol* 103:913–934.
4. Bellwood P (2000) The time depth of major language families: An archaeologist's perspective. *Time Depth in Historical Linguistics*, eds Renfrew C, McMahon AMS, Trask RL (McDonald Institute for Archaeological Research, Cambridge, UK), Vol 1, pp 109–140.
5. Matson RG (2002) The spread of maize agriculture in the U.S. Southwest. *Examining the Farming/Language Dispersal Hypothesis*, eds Bellwood P, Renfrew C (McDonald Institute for Archaeological Research, Cambridge, UK), pp 331–340.
6. Malhi RS, et al. (2003) Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* 120:108–124.
7. Smith DG, et al. (2000) Implications of the distribution of Albumin Naskapi and Albumin Mexico for new world prehistory. *Am J Phys Anthropol* 111:557–572.
8. González-José R, et al. (2007) Detection of a population replacement at the Classic-Postclassic transition in Mexico. *Proc R Soc Lond B Biol Sci* 274:681–688.
9. Cordell LS (1997) Archaeology of the Southwest (Academic, San Diego), 2nd Ed, p 522.
10. Beals RL (1974) Cultural relations between Northern Mexico and the Southwest United States: Ethnologically and archaeologically. The Mesoamerican Southwest: *R*eadings in *A*rchaeology, *E*thnohistory, and *E*thnology, eds Hedrick BC, Kelley JC, Riley CL (Southern Illinois University Press, Carbondale, IL), pp 52–57.
11. Suárez JA (1983) The Mesoamerican Indian Languages (Cambridge University Press, New York), p 206.
12. Kirchhoff P (1952) Mesoamerica: Its geographic limits, ethnic composition and cultural characteristics. Heritage of Conquest, ed Tax S (The Free Press Publishers, Glencoe, IL), pp 17–31.
13. Long A, Benz BF, Donahue DJ, Jull AJT, Toolin LJ (1989) First direct AMS dates on early maize from Tehuacán, Mexico. *Radiocarbon* 31:1035–1040.

14. Pohl ME, Piperno DR, Pope KO, Jones JG (2007) Microfossil evidence for pre-Columbian maize dispersals in the neotropics from San Andres, Tabasco, Mexico. *Proc Natl Acad Sci USA* 104:6870–6875.
15. Kohler TA, Glaude MP, Bocquet-Appel J-P, Kemp BM (2008) The Neotlithic Demographic Transition in the U.S. Southwest. *Am Antiq* 73:645–669.
16. Merrill WL, et al. (2009) The diffusion of maize to the southwestern United States and its impact. *Proc Natl Acad Sci USA* 106:21019–21026.
17. Fowler CS (1983) Some lexical clues to Uto-Aztecan prehistory. *Int J Am Linguistics* 49: 224–257.
18. Romney AK (1957) The genetic model and Uto-Aztecan time perspective. *Davidson J Anthropology* 3:35–41.
19. Lamb SM (1958) Linguistic prehistory in the Great Basin. *Int J Am Linguistics* 24: 95–100.
20. Miller WR (1984) The classification of the Uto-Aztecan languages based on lexical evidence. *Int J Am Linguistics* 50:1–24.
21. Bellwood P, Renfrew C, eds (2002) Examining the *F*arming/*L*anguage *D*ispersal *H*ypothesis (McDonald Institute for Archaeological Research, Cambridge, UK).
22. Diamond J, Bellwood P (2003) Farmers and their languages: The first expansions. *Science* 300:597–603.
23. Bellwood P (2005) *First Farmers: The Origins of Agricultural Societies* (Blackwell, Oxford).
24. Turner CG, Turner JA (1999) *Man Corn: Cannibalism and Violence in the Prehistoric American Southwest* (University of Utah Press, Salt Lake City), p 547.
25. Parsons EC (1974) Some Aztec and Pueblo parallels. The Mesoamerican Southwest: *R*eadings in *A*rchaeology, *E*thnohistory, and *E*thnology, eds Hedrick BC, Kelley JC, Riley CL (Southern Illinois University Press, Carbondale, IL), pp 131–146.
26. Kelley JC (1995) Trade goods, traders and status in Northwestern Greater Mesoamerica. *The Gran Chichimeca: Essays on the Archaeology and Ethnohistory of Northern Mesoamerica*, ed Reyman JE (Avebury, Aldershot, UK), pp 102–145.
27. Haury EW (1974) The problem of contacts between the Southwestern United States and Mexico. *The Mesoamerican Southwest: Readings in Archaeology, Ethnohistory, and Ethnology*, eds Hedrick BC, Kelley JC, Riley CL (Southern Illinois University Press, Carbondale, IL), pp 92–102.
28. Coe MD (1994) *Mexico: From the Olmecs to the Aztecs* (Thames and Hudson, New York), 4th Ed, p 215.
29. Di Peso CC (1979) *Prehistory: Southern Periphery. Handbook of North American Indians, Southwest*, ed Sturtevant WC (Smithsonian Institution, Washington), Vol 9, pp 152–161.
30. Brown KS, Hanna BL, Dahlberg AA, Strandskov HH (1958) The distribution of blood group alleles among Indians of Southwest North America. *Am J Hum Genet* 10: 175–195.
31. Spuhler JN (1979) Genetic distances, trees, and maps of North American Indians. *The First Americans: Origins, Affinities, and Adaptations*, eds Lauglin WS, Harper AB (Gustav Fischer, New York), pp 135–184.
32. Niswander JD, Brown KS, Iba BY, Leyshon WC, Workman PL (1970) Population studies on southwestern Indian tribes. I. History, culture, and genetics of the Papago. *Am J Hum Genet* 22:7–23.
33. Callegari-Jacques SM, Salzano FM, Constans J, Maurieres P (1993) Gm haplotype distribution in Amerindians: Relationship with geography and language. *Am J Phys Anthropol* 90:427–444.
34. Schurr TG, et al. (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613–623.
35. Torroni A, et al. (1994) MtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. *Am J Hum Genet* 54:303–318.
36. Sandoval K, et al. (2009) Linguistic and maternal genetic diversity are not correlated in Native Mexicans. *Hum Genet* 126:521–531.
37. Penaloza-Espinosa RI, et al. (2007) Characterization of mtDNA haplogroups in 14 Mexican indigenous populations. *Hum Biol* 79:313–320.
38. Carlyle SW, Parr RL, Hayes MG, O'Rourke DH (2000) Context of maternal lineages in the Greater Southwest. *Am J Phys Anthropol* 113:85–101.
39. Parr RL, Carlyle SW, O'Rourke DH (1996) Ancient DNA analysis of Fremont Amerindians of the Great Salt Lake Wetlands. *Am J Phys Anthropol* 99:507–518.
40. De la Cruz I, et al. (2008) Sex identification of infants sacrificed to the ancient Aztec raingods in Tlatelolco, Mexico. *Curr Anthropol* 49:519–526.
41. Kemp BM, Resendez A, Román Berrelleza JA, Malhi RS, Smith DG (2005) An analysis of ancient Aztec mtDNA from Tlatelolco: Pre-Columbian relations and the spread of Uto-Aztecan. *Biomolecular Archaeology: Genetic Approaches to the Past*, ed Reed DM (Southern Illinois University, Carbondale, IL), pp 22–46.
42. Kaestle FA, Smith DG (2001) Ancient mitochondrial DNA evidence for prehistoric population movement: The Numic Expansion. *Am J Phys Anthropol* 115:1–12.
43. Johnson JR, Lorenz JG (2006) Genetics, linguistics, and prehistoric migrations: An analysis of California Indian mitochondrial DNA lineages. *J Calif Gt Basin Anthropol* 26:31–62.
44. Green LD, Derr JN, Knight A (2000) mtDNA affinities of the peoples of north-central Mexico. *Am J Hum Genet* 66:989–998.
45. Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci USA* 91:1158–1162.
46. Torroni A, et al. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162.
47. Horai S, et al. (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 10:23–47.
48. Malhi RS, et al. (2008) Distribution of Y chromosomes among native North Americans: A study of Athapaskan population history. *Am J Phys Anthropol* 137:412–424.
49. LeBlanc SA, et al. (2007) Quids and aprons: Ancient DNA from artifacts from the American Southwest. *J Field Archaeol* 32:161–175.
50. Achilli A, et al. (2008) The phylogeny of the four pan-American MtDNA haplogroups: Implications for evolutionary and disease studies. *PLoS One* 3:e1764.
51. Kaestle FA (1998) Molecular evidence for prehistoric Native American population movement: The Numic Expansion. PhD dissertation (University of California, Davis, CA).
52. Malhi RS, Schultz BA, Smith DG (2001) Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. *Hum Biol* 73:17–55.
53. Budowle B, et al. (2002) HVI and HVII mitochondrial DNA data in Apaches and Navajos. *Int J Legal Med* 116:212–215.
54. Kemp BM, et al. (2007) Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am J Phys Anthropol* 132:605–621.
55. Ingman M, Gyllensten U (2006) mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res* 34: D749–D751.
56. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163.
57. Howell N, et al. (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* 72:659–670.
58. Henn BM, Gignoux CR, Feldman MW, Mountain JL (2009) Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol* 26: 217–230.
59. Tamm E, et al. (2007) Beringian standstill and spread of Native American founders. *PLoS One* 2:e829.
60. Kohler TA (1993) News from the Northern American Southwest: Prehistory on the edge of chaos. *J Archaeol Res* 1:267–321.
61. Cordell LS, Van West CR, Dean JS, Muenchrath DA (2007) Mesa Verde settlement history and relocation: Climate change, social networks, and ancestral Pueblo migration. *Kiva* 72:379–405.
62. Malhi RS, et al. (2010) Brief communication: Mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *Am J Phys Anthropol* 141:494–497.
63. Balaresque P, et al. (2010) A predominantly neolithic origin for European paternal lineages. *PLoS Biol* 8:e1000285.
64. Páez-Riberos LA, et al. (2006) Y-linked haplotypes in Amerindian chromosomes from Mexican populations: Genetic evidence to the dual origin of the Huichol tribe. *Leg Med* 8:220–225.
65. Belle EM, Barbujani G (2007) Worldwide analysis of multiple microsatellites: Language diversity has a detectable influence on DNA diversity. *Am J Phys Anthropol* 133:1137–1146.
66. Matson RG (1999) The spread of maize to the Colorado Plateau. *Archaeology Southwest* 13:10–11.
67. Schneider S, Roessli D, Excoffier L (2000) *Arlequin ver. 2.000: A Software for Population Genetics Data Analysis* (Genetics and Biometry Laboratory, University of Geneva, Geneva).
68. Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* 69:1–24.
69. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295.
70. Hollenbach JA, et al. (2001) HLA diversity, differentiation, and haplotype evolution in Mesoamerican Natives. *Hum Immunol* 62:378–390.
71. Flannery KV, Marcus J, eds (1983) *The Cloud People: Divergent Evolution of the Zapotec and Mixtec Civilizations* (Academic Press, New York).
72. Hale K (1958) Internal diversity in Uto-Aztecan: I. *Int J Am Linguistics* 24:101–107.
73. Kemp BM (2006) Mesoamerica and Southwest prehistory, and the entrance of humans into the Americas: Mitochondrial DNA evidence. PhD dissertation (University of California, Davis, CA).
74. Bandelt H-J, Forster P, Roehl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
75. Anderson S, et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.
76. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.