Biological Continuity in the Central Valley: Evidence from Ancient and Modern Mitochondrial DNA

KARI BRITT SCHROEDER

Centre for Behaviour and Evolution, Institute of Neuroscience, Newcastle University, Framlington Place,
Newcastle Upon Tyne, NE2 4HH, United Kingdom; and Department of Anthropology, University of California, Davis,
330 Young Hall, One Shields Ave., Davis, CA, 95616

GLEN VILLA JR.

Ione Band of Miwok Cultural Heritage Committee, 9252 Bush Street, Suite 3, Plymouth, CA 95669

RIPAN SINGH MALHI

Department of Anthropology, University of Illinois, Urbana-Champaign,
185 Davenport Hall, 607 Matthews Ave., Urbana, IL, 61801

ALYSON R. RODE

Department of Anthropology, University of Illinois, Urbana-Champaign,
195 Davenport Hall, 607 Matthews Ave., Urbana, IL, 61801

DAVID GLENN SMITH

Department of Anthropology, University of California, Davis, 330 Young Hall, One Shields Ave., Davis, CA, 95616

Previous comparisons of modern maternal genetic lineages with those obtained from Early and Middle Horizon burial sites suggested population discontinuity in the Central Valley. This result was interpreted as support for a later Penutian expansion. This study re-addresses the question of biological continuity in the Central Valley with more modern samples and attention to genetic variants that offer higher resolution. Ample evidence of biological continuity in the Central Valley is found. This includes evidence of a Yok-Utian population expansion. The estimated timing of this population expansion is consistent with the estimated time depth of the Yok-Utian language family and the appearance of the Early Horizon in the archaeological record. The inclusion of ancient and modern genetic samples from the Columbia Plateau and Great Basin, motivated by evidence of cultural ties between these regions and the Central Valley, supports the hypothesis that the northwest Great Basin may have played a pivotal role in the spread of both genes and culture in the West.

THERE HAS BEEN CONSIDERABLE INTEREST IN the prehistoric processes that resulted in the pattern of linguistic and cultural diversity observed within California at the time of European contact. Linguistic and archaeological data have been used to formulate a number of hypotheses about *in situ* developments, migrations, and population replacements in western North America (see Breschini 1983).

The recent incorporation of modern and ancient genetic data, primarily mitochondrial DNA[1] (mtDNA; see endnote for more definitions used in the article),

in the anthropologist's toolkit has enabled the testing of specific migration and population replacement hypotheses. For example, comparisons between modern and ancient mtDNA haplogroup frequencies have suggested population replacements in the Central Valley of California (Eshleman 2002; Eshleman and Smith 2007) and in the Great Basin (Kaestle 1998; Kaestle and Smith 2001).

However, there are reasons to view these inferences with caution. Most ancient DNA data to date come from the mitochondrial genome. Unique properties of

the mitochondrial genome (i.e., lineal inheritance, high copy number per cell, and a fast mutation rate) make it an effective tool for anthropologists. However, exclusive consideration of mtDNA data results in inferences made on the basis of a very small window into a population's past. Aside from the strictly matrilineal inheritance of mtDNA, which precludes insights into prehistoric population events perpetuated by males, stochastic events lead to different genealogies at different places throughout the genome (Siepel 2009). Therefore, an accurate portrayal of population prehistory is best attempted by averaging over genealogies from many different places throughout the genome. MtDNA offers insight into only one of these genealogies.

Microevolutionary forces other than population replacement, such as genetic drift and gene flow, can also result in temporal changes in haplogroup frequencies. Indeed, Cabana et al. (2008) used computer simulations to demonstrate that the hypothesis of population continuity in the Great Basin of western Nevada need not be rejected (Kaestle 1998; Kaestle and Smith 2001) if female effective population sizes and rates of gene flow among demes were relatively low. Conversely, although a statistical association between haplogroup frequency distributions is not direct evidence of genetic relatedness, it is often interpreted as such.

Finally, the strength of any inferences about population continuity and population replacement rely heavily on the adequacy of sampling. Considerations include not only the comprehensiveness of temporal and spatial sampling but also the potentially problematic assignment of individuals to biological populations on the basis of linguistic, cultural, or geographic ties/relatedness. The latter implicitly incorporates hypotheses about how language, culture, and geography influence population structure.

POPULATION REPLACEMENT IN THE CENTRAL VALLEY

Penutian is a hypothetical grouping of languages and language families in western North America that has undergone multiple manifestations since Dixon and Kroeber's (1913) proposal that the Californian languages of Miwok, Yokuts, Costanoan, Wintun, and Maidu have a common origin. Historically, the geographic distribution of Penutian languages in California has been interpreted as the result of a migration of Penutian speakers into Central California, resulting in the geographic marginalization of indigenous people speaking languages in the hypothetical grouping of Hokan to the periphery of California. While solid evidence for the validity of Penutian is lacking, extensive research has provided support for some of the proposed relationships *within* Penutian (see DeLancey and Golla [1997] for a review). Recent work supports the validity of Yok-Utian, a proposed language family within Penutian that unites Yokutsan and Utian (Costanoan and Miwok) (Callaghan 1997, 2001). Yok-Utian speakers ethnographically inhabited the Bay, Delta, Valley, and foothills of Central California. The Windmiller and Berkeley cultural patterns have been interpreted as representing the appearance of Utians in the Lower Sacramento Valley, followed by a subsequent westward population expansion, respectively (Moratto 1984). Corroborating this interpretation of the archaeological record, the appearance of the Windmiller pattern ca. 4,400 years ago is coterminus with the suggested time depth of Utian, ca. 4,500 years ago (Callaghan 1997, 2001), and the reconstructed lexicon of flora and fauna for Proto-Utian is concordant with an inland homeland (Golla 2007). The shorter time depth and relative lack of ethnographic and linguistic diversity for Yokutsan could be consistent with either a later migration into Central California (Golla 2007) or an occupation of Central California contemporaneous with Utians but followed by a population (and language) bottleneck (Moratto 1984).

To evaluate the hypothesis that the Windmiller pattern represented a replacement of an older Hokan population by a Penutian population, Eshleman and Smith (2007) determined mtDNA haplogroups for a total of 16 individuals from one Windmiller burial site (CA-SJO-112, also referred to as the "Cecil" or "Bear Creek" site) and 29 individuals from two Middle Horizon burial sites (CA-SOL-270 and CA-AMA-56, also referred to as the "Cook" and "Applegate" sites, respectively) in the Central Valley. MtDNA haplotypes, based on Hypervariable Region I (HVI) sequences, were determined for a subset of these individuals. Population continuity between the people who used these sites was inferred from similar haplogroup frequencies and some haplotype sharing (Eshleman 2002; Eshleman

and Smith 2007). The combined haplogroup frequencies from the three burial sites were then compared to those from a sample of modern Yok-Utians (Yokuts, Miwok, and Costanoan), as well as to those obtained from other ancient and modern groups in western North America. There was no statistical association between the haplogroup frequencies of the ancient inhabitants of the Central Valley and those of modern Penutian speakers. In contrast, the haplogroup frequencies of the ancient Central Valley inhabitants were statistically indistinguishable from modern Takic (Uto-Aztecan) speakers.

These results were viewed as consistent with the hypothesis of a Uto-Aztecan presence in the Central Valley (Nichols 1981) and inconsistent with the hypothesis that Windmiller represents the appearance of Penutian-speaking people in the archaeological record (Eshleman 2002; Eshleman and Smith 2007).

In this study, we revisit the question of population replacement in the Central Valley but attempt to do so in a manner that incorporates a few precautionary observations. Because of the potential mismatch between biological and linguistic groups, a close biological relationship between an ancient group and a modern group should not be interpreted as evidence that the ancient group spoke the same language as the modern group. Furthermore, because mtDNA provides only a narrow slice of population prehistory, and microevolutionary forces can affect haplogroup and haplotype frequencies over time, the lack of evidence of biological continuity between an ancient group and a modern group is not necessarily evidence of a population replacement.

Hence, we reformulate the question of a Penutian population replacement: Is there evidence of biological continuity between ancient and modern inhabitants of the Central Valley? We evaluate this with new data and a different methodological approach.

GENETIC AFFINITIES AMONG MODERN AND ANCIENT INHABITANTS OF CALIFORNIA, THE COLUMBIA PLATEAU, AND THE GREAT BASIN

If the presence of Penutian-speaking groups in the Central Valley is the result of a migration from elsewhere, then linguistic and archaeological data suggest an origin on the Columbia Plateau or in the Great Basin. The geographic density of Penutian languages in western North America is consistent with a homeland in the region of the southern Columbia Plateau and northwestern Great Basin (Golla 2007). Similar ecological adaptations, technologies, and ornamentation are shared between Windmiller and contemporaneous cultures on the Plateau (Ragir 1972) and in the Great Basin (Hattori 1982). One or more migrations from the Plateau and/or Great Basin into the Central Valley have been hypothesized (Whistler 1977).

Genetic relationships among modern and ancient populations in these regions have been evaluated, primarily with mtDNA haplogroup frequency data. Kaestle and Smith (2001) note the similarity in mtDNA haplogroup frequencies between a grouping of modern Penutians in California and ancient populations in the Great Basin and suggest that this is consistent with a Penutian migration from northwest Nevada. Eshleman et al. (2004) further propose, based on the linguistic affiliations and geographic distributions of groups with a high frequency of mtDNA haplogroup D, that the ancient inhabitants of the Great Basin were ancestral to modern Penutians on the Columbia Plateau (Yakima and Wishram) and in the Central Valley (Yok-Utians), perhaps as part of an ancient widespread Penutian population inhabiting the Great Basin, California, and the Columbia Plateau. (However, Eshleman et al. [2004] note that Early and Middle Horizon inhabitants of the Central Valley had a markedly different haplogroup frequency distribution.) The similarly high frequency of haplogroup D in the Uto-Aztecan Northern Paiute of the Great Basin would then be the result of admixture between the indigenous inhabitants of the Great Basin and the ancestors of the Northern Paiute during the Numic spread (Eshleman et al. 2004). Consistent with these observations is an instance of haplotype-sharing (characterized by a mutation at nucleotide pair [np] 16294) among mtDNA haplogroup D lineages ascertained from protohistoric and ancient samples from the Plateau and the Great Basin, respectively, and from modern samples on the Northwest Coast (Malhi et al. 2004).

In this study, we further investigate the supposition that modern Penutian groups in Central California share more recent ancestry with Penutian groups on

the Columbia Plateau and with the ancient inhabitants of the Great Basin, than with non-Penutian groups in other regions of the western coast of the USA. A preponderance of haplotype-sharing and/or closely related matrilines among members of these groups would be consistent with this supposition, and if Windmiller represents a Penutian presence in Central California, this pattern of closely related matrilines might extend back in time to that period.

DATA AND ANALYSES

Collection of Newly Reported MtDNA Sequences

Ione Band of Miwok Indians. In 2003, K. B. Schroeder contacted Dwight Dutschke of the Cultural Heritage Committee (CHC) of the Ione Band of Miwok and proposed collecting biological samples for mtDNA analysis in order to establish a broader, modern comparative base for ancient samples from the Central Valley that had previously been analyzed (Eshleman 2002). The CHC agreed to administer a voluntary written survey of members' opinions on human genetic research, which was made available during a tribal meeting (this survey was also part of a larger effort to survey Native Americans on their opinions about human genetic research, the results of which are published in Schroeder et al. [2006]). Following a review of the results of the survey and multiple meetings between members of the Smith Molecular Anthropology Laboratory (MAL) at the University of California, Davis (UCD) and the CHC, including a tour of the Smith MAL, the CHC agreed to participate in the study. Consent forms and project protocols were discussed with and approved by the CHC. Dwight Dutschke and both Glen Villa Jr. and Sr. of the CHC were instrumental throughout this process and in recruiting individuals to participate in the study.

Thirty-four buccal swabs from individuals of indigenous maternal ancestry from the IBMI were collected for this study (Fig. 1; Table 1). Glen Villa Jr., who has researched extensive genealogies for members of regional indigenous communities, was able to determine the specific origin for the maternal lineage of sampled individuals. The 34 individuals sampled originated from eight genealogically-determined maternal groups. Affiliations for these eight groups include Plains Miwok, Northern Miwok, Southern Nisenan, Waupumne

Nisenan, Costanoan, and Luiseño. Two of the maternal groups share ties with both Northern Miwok and Nisenan.

The extraction, amplification, and sequencing of HVI of the mtDNA genome was performed at the Smith MAL and the DNA Sequencing Facility in the Division of Biological Sciences at UCD. The results were shared with individual participants and presented at a tribal meeting.

Northern Paiute and Washo. The Northern Paiute and Washo samples (Fig. 1; Table 1) used for this study are those described in Smith et al. (2000). DNA was extracted from a total of 13 Northern Paiute and 20 Washo serum samples at the Smith MAL at UCD. HVI was amplified and sequenced at either UCD or the Malhi MAL at University of Illinois Urbana-Champaign; samples sequenced at the latter location were whole genome-amplified prior to sequencing to increase the quantity of DNA.

Previously Published MtDNA Sequences

Sources of Ancient MtDNA Sequences. Previously published mtDNA sequences from ancient individuals from the Columbia Plateau, Great Basin, Central Valley, and California coast were included in the analyses (Fig. 1; Table 2). The two protohistoric tooth samples from the Middle Columbia River are from regions traditionally occupied by Sahaptin (Penutian) speakers (Malhi et al. 2004). All dates and cultural affiliations for the ancient samples are exactly as referenced in the sources cited for the mtDNA data.

Sources of Modern MtDNA Sequences. Previously published modern mtDNA sequences from the Columbia Plateau, the Great Basin, and California (Figure 1; Table 1) were used in the analyses. Sequences from Chinookan speakers on the Columbia Plateau (Memaloose Island Cemetery) were taken from Malhi et al. (2004). Although these sequences were obtained from protohistoric tooth samples, we refer to these as modern because, unlike sequences obtained from the protohistoric samples from the Columbia Plateau, these sequences are in the literature (Eshleman et al. 2004; Malhi et al. 2004) as associated with a particular ethnolinguistic group (Wishram). The other modern sequences sampled from the Columbia Plateau are from the Yakima, as reported in Shields et al. (1993).
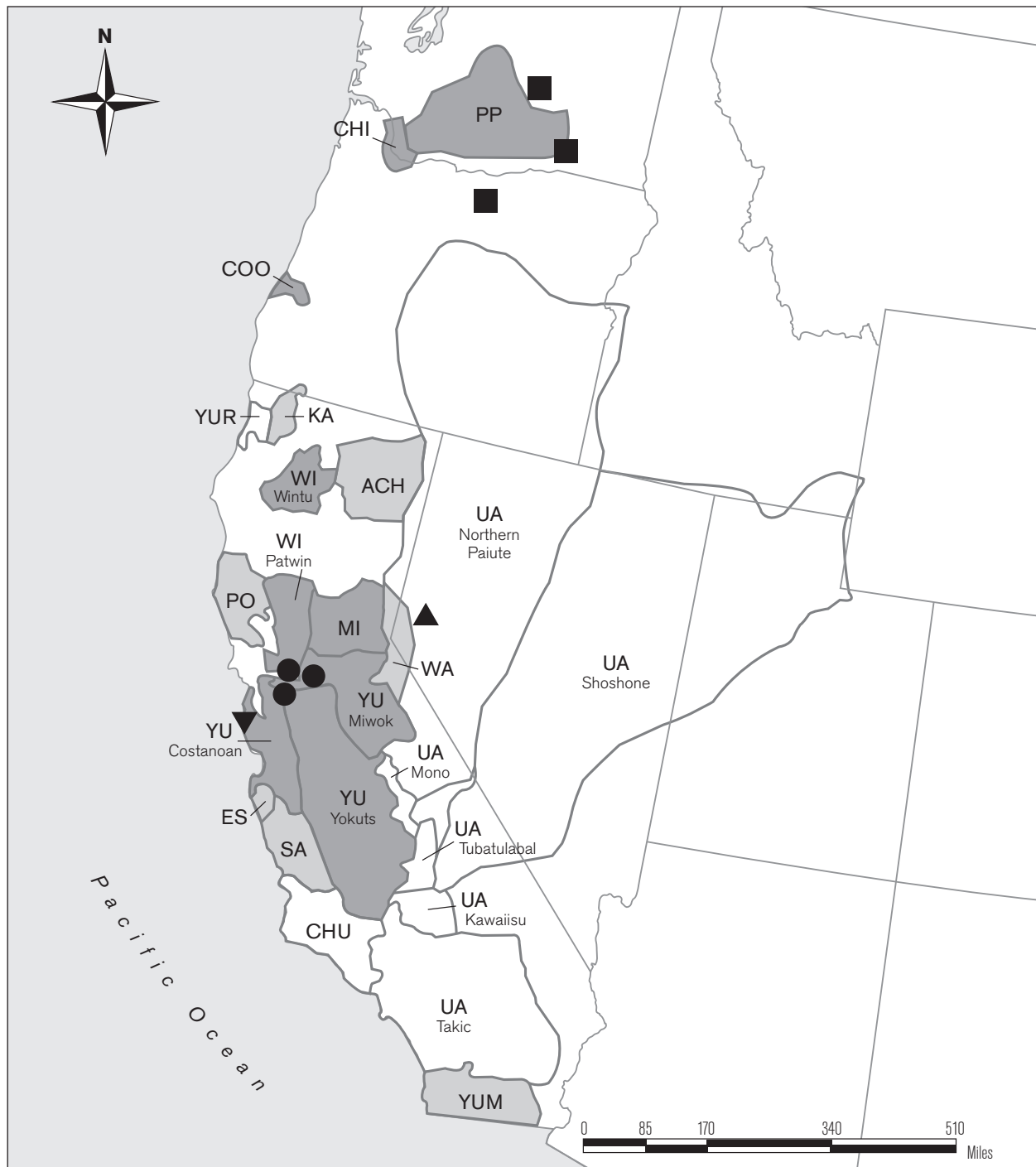
Figure 1. Map of sampling locations for ancient and modern samples included in this study. Approximate geographic locations of ancient samples on the Columbia Plateau, in the Great Basin, in the Central Valley, and on the California Coast (Table 2) are illustrated with black squares, triangles, circles, and an inverted triangle, respectively. Approximate geographic distributions at contact of linguistic groups from which modern individuals were sampled (Table 1) are drawn after Mithun (1999). Linguistic groups are labeled to the highest resolution given in Table 1. The following abbreviations are used to designate family-level linguistic affiliations: ACH = Achumawi/Atsugewi, ES = Esselen, KA = Karok, PO = Pomoan, SA = Salinan, WA = Washo, YUM = Yuman, CHI = Chinookan, COO = Coosan, PP = Plateau Penutian, WI = Wintuan, MI = Maiduan, YU = Yok-Utian, UA = Uto-Aztecan, CHU = Chumashan, YUR = Yurok. Those included in the proposed Hokan macrofamily are shaded light gray, while those in the proposed Penutian macrofamily are shaded dark gray.

<div style="text-align:center">

Table 1

**LINGUISTIC AFFILIATION AND SOURCE
OF MODERN SAMPLES USED IN ANALYSES**

</div>

Language family or (proposed) macrofamily	Number of Samples	Source
Hokan		
Achumawi/Atsugewi	1	Johnson and Lorenz 2006
Esselen	1	Johnson and Lorenz 2006
Karok	1	Kaestle 1998
Pomoan	3	Kaestle 1998
Salinan	6	Johnson and Lorenz 2006
Washo	20	This study
Yuman	10	Johnson and Lorenz 2006
Penutian		
Chinookan		
Wishram	16	Malhi et al. 2004
Coosan		
Coos	1	Johnson and Lorenz 2006
Plateau Penutian		
Yakima	38	Shields et al. 1993
Maiduan		
Nisenan[a]	9	This study
Wintuan		
Patwin	1	Johnson and Lorenz 2006
Wintu	2	Johnson and Lorenz 2006
Yok-Utian		
Costanoan	5	Johnson and Lorenz 2006; This study
Miwok[b]	25	Johnson and Lorenz 2006; This study
Yokuts	15	Johnson and Lorenz 2006
Uto-Aztecan		
Numic		
Kawaiisu	5	Johnson and Lorenz 2006
Mono	4	Johnson and Lorenz 2006
Northern Paiute	13	This study
Shoshone	2	Kaestle 1998
Takic	35	Johnson and Lorenz 2006; This study
Tulatulabal	4	Johnson and Lorenz 2006
Other		
Chumashan	19	Johnson and Lorenz 2006
Yurok	1	Johnson and Lorenz 2006

a. Six of these, in the current study, are Waupumne Nisenan with Northern Miwok ancestry as well.

b. Ten of these, in the current study, are Northern Miwok with Southern Nisenan ancestry as well. See Table 4 for further detail.

<div style="text-align:center">

Table 2

**CULTURAL OR GEOGRAPHIC AFFILIATION AND
SOURCE OF ANCIENT SAMPLES USED IN ANALYSES**

</div>

Geographic Region	Number of Samples	Age or Cultural Affiliation	Source
Columbia Plateau			
–	1	Protohistoric	Malhi et al. 2004
Middle Columbia River	5	1,500 – 500 B.P.[a]	Malhi et al. 2004
Middle Columbia River	2	Protohistoric	Malhi et al. 2004
Great Basin			
Pyramid Lake	17	9,260 –785 B.P.[b]	Kaestle 1998
California Coast			
Monterey	3	1,000 – 500 years old[c]	Eshleman 2002
Central Valley			
Amador County	4	2,090 –1,735 B.P.[d]	Eshleman 2002
San Joaquin County	10	3,350 –2,710 B.P.[d]	Eshleman 2002
Solano County	15	Middle Horizon	Eshleman 2002

a. Reportedly based on unpublished radiocarbon dates; no standard deviation was given.

b. Each individual at the site was radiocarbon dated; age range includes oldest and youngest individuals, with ± one standard deviation incorporated into range.

c. Age determination unspecified.

d. Based on radiocarbon dates from two individuals from the site, with ± one standard deviation incorporated into age range.

among those used to estimate the age of the haplogroup B subclade characterized by a transversion at np 16184. Chumash individual JJ429 was also excluded because of excessive missing data. Uncertainty was associated with some "Ethnolinguistic Group" assignments (as designated by a question mark by Johnson and Lorenz [2006]). For the current study, we did not incorporate any uncertainty about Ethnolinguistic Group assignment and used the information provided by Johnson and Lorenz (2006) as a "best guess." Four additional sequences from modern indigenous Californians (one Karok and three Pomo) and two additional sequences from modern individuals from the Great Basin (Shoshone) were taken from Kaestle (1998).

Construction of Networks

Unlike a phylogenetic tree, a phylogenetic network may include reticulation; that is, it allows the representation of multiple possible evolutionary pathways, or ancestor/descendant relationships, between nodes. It offers a more appropriate means of visualizing population-level evolutionary relationships than a tree, particularly for

The majority of published mtDNA sequences from modern indigenous Californians are provided in Johnson and Lorenz (2006). All samples labeled as "Unidentified" by Johnson and Lorenz (2006) were excluded, except for a single Unidentified sample which we included

the rapidly evolving mtDNA HVI, because it enables the incorporation of uncertainty due to homoplasy. Median-joining networks were constructed for each of the four major mtDNA haplogroups observed in the Americas (A, B, C, and D) with the software Network, version 4.5 (Bandelt et al. 1999; software available at http://www.fluxus-engineering.com).

Sites were weighted in Network to reflect lesser or greater importance given to mutations at sites with evolutionary rates relatively higher or lower than average, respectively. The default weight for each site was set at 10 to correspond to an average (over all sites) evolutionary rate in HVI of 0.6 (Meyer et al. 1999). All sites that were estimated by Stoneking (2000) and/or Meyer et al. (1999) to be evolving at a faster or slower rate were down-weighted or up-weighted in accordance with the estimated rate. Primacy was given to the estimates in Stoneking (2000) when both sources listed the same site as differing from the average evolutionary rate. For example, Stoneking (2000) estimates the evolutionary rate of np 16362 as 4.72, which is approximately 7.87 times faster than the average HVI evolutionary rate of 0.6. Hence, with a weight of 10 given to those sites with an average HVI evolutionary rate, np 16362 was down-weighted to 1.27, which was rounded down to a weight of 1 (Network only accepts integers from 1 to 99 as weight values). Meyer et al. (1999) gives estimates of evolutionary rates only relative to the average for HVI so, for example, np 16343 was down-weighted to 5 based on their estimate that this site is evolving twice as fast as the average for HVI. Table 3 shows the rates for specific, rapidly-evolving sites in HVI that are mentioned in the text and/or pointed out in the figures. Based on a transition/transversion ratio of 15.55 for HVI (Meyer et al. 1999), which means that transitions are estimated to be 15 times more common than transversions, transversions (at np 16184 and np 16213) were given the maximum possible weight of 99.

Sequences used in the construction of networks span the following sites in HVI for haplogroups A through D, respectively: 16086–16356, 16086–16360, 16086–16362, and 16069–16362. After preliminary network analyses, the following changes were made either to the weighting scheme or to the variable sites used in network construction. The "A" insertion at np 16183 was down-weighted from 10 to 1 for construction of the haplogroup A network to minimize reticulation

Table 3

EVOLUTIONARY RATE OF HYPERVARIABLE SITES MENTIONED IN TEXT OR NOTED IN FIGURES

Site	Source	Evolutionary Rate
16093	Stoneking 2000	5.0
16111	Stoneking 2000	4.3
16124	Stoneking 2000	1.4
16126	Meyer et al. 1999	4.0
16129	Stoneking 2000	8.0
16148	Meyer et al. 1999	2.0
16183	Stoneking 2000	5.0
16189	Stoneking 2000	8.0
16192	Stoneking 2000	6.0
16223	Meyer et al. 1999	4.0
16234	Stoneking 2000	1.4
16261	Meyer et al. 1999	2.0
16274	Meyer et al. 1999	2.0
16294	Meyer et al. 1999	4.0
16295	Stoneking 2000	1.2
16311	Stoneking 2000	8.0
16319	Meyer et al. 1999	2.0
16362	Stoneking 2000	7.9

Evolutionary rate is relative to an average rate of one for HVI; for example, np 16093 is estimated to evolve at five times the average rate for HVI.

and conform to the more parsimonious scenario of two independent mutations, one each in the ancestors of modern Yokuts and Chumash groups, rather than a shared mutation followed by two additional back-mutations in Yokuts. Note also that this insertion is in the rapidly mutating poly C tract of HVI, and that transitions at np 16183 had already been down-weighted from 10 to 2 in accordance with Stoneking (2000). Nucleotide pair 16311 appeared more times than any other site in an excessive reticulation, involving 12 distinct nodes that made the network for haplogroup C inscrutable. Hence, np 16311 was removed and the data for haplogroup C were re-analyzed. Nucleotide pair 16261 was down-weighted from 5 (following the estimation by Meyer et al. (1999) that this site is evolving roughly twice as fast as the average for HVI) to 1 for the haplogroup B network because this site caused reticulation between a subclade with three different haplotypes, all observed in Uto-Aztecans, and the haplogroup B "Yok-Utian" subclade demarcated by the transversion at 16184. All of the lineages in the subclade demarcated by a mutation at np 16261 also differed at one or more sites from the basal lineage in the haplgroup B Yok-Utian subclade. It was deemed more likely that recurrent mutation has occurred

at np 16261, a site known to evolve faster than average for HVI, rather than a recurrent transversion at 16184 in a similar geographic area, given that this transversion has not been observed elsewhere in the Americas.

Estimate of Age of Haplogroup B "Yok-Utian" Subclade. The transversion at np 16184, sampled primarily in Yok-Utian individuals, characterizes a star-shaped subclade, which is often caused by a population expansion. The minimum age of this mutation was estimated with the software Network as the rate of phylogenetic dispersion, ρ, or the average number of mutations separating the ancestral node and descendant nodes. This age also provides a maximum estimate for the age of the Yok-Utian subclade. Twenty-four sequences with the transversion at np 16184, representing nine distinct haplotypes, were used for this analysis, including a sequence from a single individual of unidentified etholinguistic origin (JJ214, reported in Johnson and Lorenz [2006]) that is identical to one observed in a Southern Valley Yokuts individual.

Two different calibrations of ρ for HVI were used to convert the age in mutational units to an age in years: (1) one transition per 20,180 years (Forster et al. 1996), and (2) one mutation per 9,213 years (Kemp et al. 2007). Kemp et al. (2007) also provide a slightly more rapid calibration (one mutation per 8,429 years) that is based on a greater number of sequences but a shorter section of HVI. However, while the point estimates are similar, the standard deviation for the more rapid estimate is almost twice as high as for the slower estimate, and the section of HVI used to arrive at the more rapid estimate does not cover the sites we used to estimate ρ in the Yok-Utian subclade, so we only use the slightly slower estimate in our analyses.

RESULTS

Haplogroup Diversity

Ione Band of Miwok Indians. Seven different haplotypes were observed among the 34 sequences, all of which belong to either haplogroup B (four haplotypes, one of which—Maternal Group 1 in Table 4—is basal B) or haplogroup D (three haplotypes). The seven observed haplotypes corresponded to individuals' membership in eight maternal groups, as determined genealogically by Villa Jr. (Table 4). Two of the maternal groups—one originating with the Waupumne Nisenan and the other with Costanoan from Alameda county—share the same haplogroup D haplotype. The prevalence of halpogroups B and D in this sample (which is primarily from individuals with Yok-Utian affiliation) is consistent with previously published reports of the predominance of these haplogroups among Yok-Utians (Eshleman 2002; Eshleman et al. 2004; Johnson and Lorenz 2006; Eshleman and Smith 2007). We abstain from using traditional estimates of diversity or frequency-based

Table 4

ORIGIN AND HAPLOTYPES OF MATERNAL GROUPS FROM IBMI INCLUDED IN STUDY

Maternal Group	Maternal Linguistic Group	Locality	Number Individuals	Haplogroup	111	183	184	189	217	223	234	256	260	311	319	325	362	391	519
CRS[a]					C	A	C	T	T	C	C	C	C	T	G	T	T	G	T
1	Northern Miwok	Locolumne	1	B	.	C	.	C	C	.	.	.	.	.	.	.	.	.	.
2	Northern Miwok/Southern Nisenan	Upper Cosumnes	10	B	.	C	.	C	C	.	.	T	.	.	.	.	.	.	.
3[b]	Northern Miwok	Fiddletown	7	B	.	C	A	C	C	.	.	.	.	.	.	.	.	.	.
4[b]	Plains Miwok	Lockeford/Galt	3	B	.	C	A	C	C	.	.	.	.	T	.	.	.	.	.
5[c]	Waupumne Nisenan	—	3	D	.	.	.	.	.	T	.	.	.	C	.	C	C	.	.
6[c]	Costanoan	Alameda County	1	D	.	.	.	.	.	T	.	.	.	C	.	C	C	.	.
7	Luiseño	Mission	3	D	T	.	.	.	.	T	.	.	.	.	A	A	A	.	C
8	Waupumne Nisenan/Northern Miwok	—	6	D	.	.	.	.	.	T	T	.	.	.	.	C	C	A	.

a. CRS = Cambridge Reference Sequence. Only nucleotide pairs that differ from the CRS appear in the table; nucleotide pairs are listed by the last three digits (i.e., "16111" is written as "111").

b. The haplotypes of these maternal groups belong in the haplogroup B "Yok-Utian" subclade.

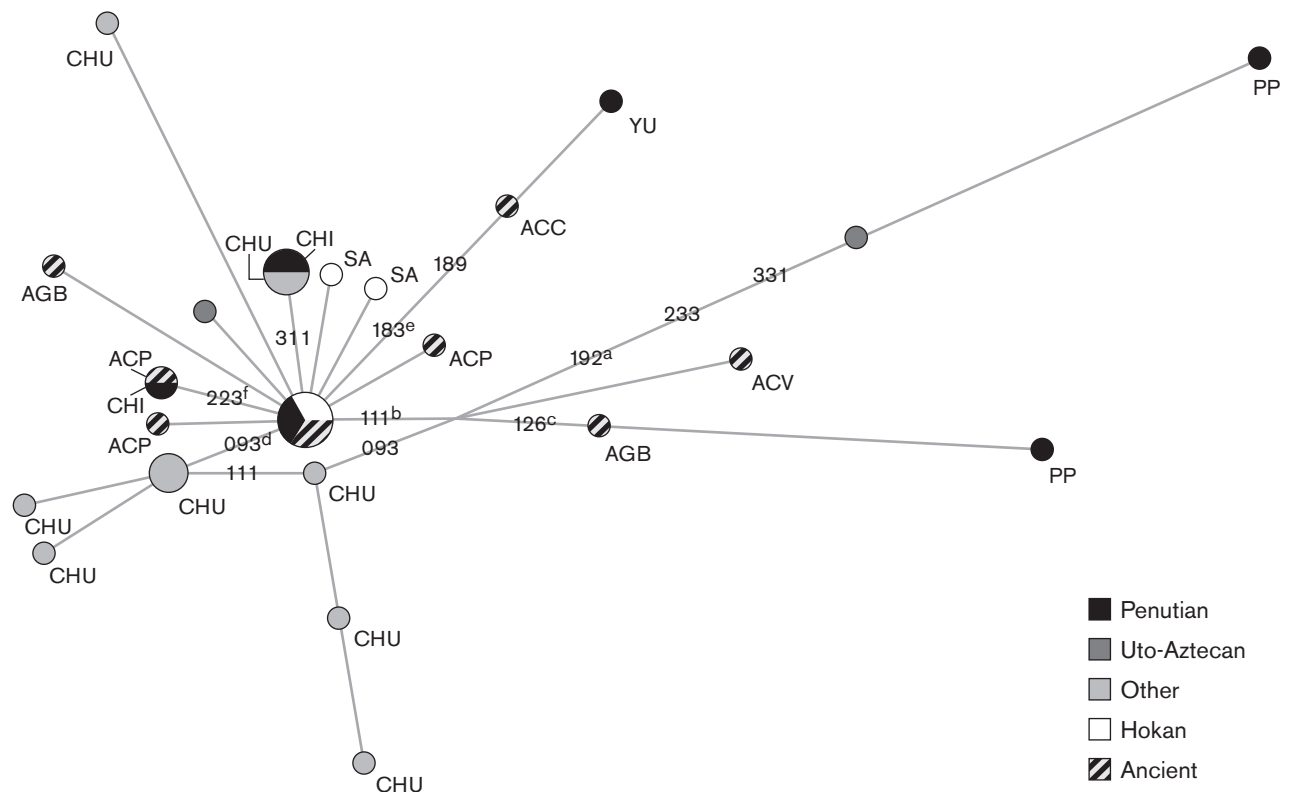c. These two maternal groups share the same haplotype.

Figure 2. Haplogroup A network. The following abbreviations are used to designate family-level linguistic affiliations (modern samples) or regional affiliations (ancient samples): SA = Salinan, CHI = Chinookan, PP = Plateau Penutian, YU = Yok-Utian, CHU = Chumashan, ACC = Ancient Central Coast, ACP = Ancient Columbia Plateau, ACV = Ancient Central Valley, AGB = Ancient Great Basin. The basal A node includes one Esselen, one Salinan, two Plateau Penutian, and two Ancient Central Coast samples. For Figures 2–6, node size corresponds to the number of occurrences of the haplotype in the dataset. The last three digits of the np for mutations referenced in the text are given; all mutations fall within nps 16000–16569. Nucleotide pairs at which observed mutations characterize potential subclades discussed in the text are marked with alphabetic superscript. Where mutations at multiple nps characterize the same potential subclade, only one np is marked.

analyses because sample collection was non-random. Multiple individuals from families were targeted because they were known to belong to matrilines with a long-term presence in the area. Hence, haplogroup and haplotype frequencies may not be representative of the population. We also abstain from using haplogroup and haplotype frequencies estimated from the data collected by Johnson and Lorenz (2006) because specific matrilines were targeted in sample collection, *in concordance with the particular purposes of that study*.

Northern Paiute and Washo. The Northern Paiute and Washo samples primarily represent haplogroups B and D as well. However, one of the Northern Paiute samples belongs to haplogroup A, and 15.4% of the Northern Paiute samples and 20.0% of the Washo samples belong to haplogroup C. Individuals wishing to

use the data for further analyses are requested to email the corresponding author.

Phlyogenetic Networks

Haplogroup A. Haplogroup A, which primarily occurs among coastal populations in western North America (Eshleman et al. 2004), was not observed among the newly collected IBMI or Washo sequences. One Northern Paiute sequence belongs to haplogroup A. This highly-derived haplotype shares a number of mutations (at nps 16192, 16233, and 16331) with a sequence observed in a Yakima individual that is further derived by three additional mutations (Fig. 2, subclade a). Both haplotypes belong to a subclade (Fig. 2, subclade b), characterized by the loss of a mutation at np 16111, that includes individuals from the Great Basin (both ancient,

$1,520 \pm 90$ B.P., and modern), the Columbia Plateau (modern Yakima), and the Central Valley (CA-SOL-270, associated with the Middle Horizon). While np 16111 is hypermutable, reducing confidence in this subclade, the ancient Great Basin and modern Yakima (Plateau Penutian) are further connected by a shared mutation at np 16126 (Fig. 2, subclade c).

Indeed, reticulation in this network suggests that a back-mutation at np 16111 may have occurred at three separate times among the ancestors of the individuals represented in the network. While nps 16111 and 16093, both of which are hypermutable (Stoneking 2000), cause reticulation in the network, homoplasy at np 16111, rather than at np 16093, would retain all the Chumash lineages with 16093 in a single subclade (Fig. 2, subclade d), while

homoplasy at np 16093 would result in the clustering of lineages sampled in geographically and linguistically distant groups. The former homoplasic event would be congruent with the observation that closely related lineages are often found within a single ethnolinguistic group, presumably the result of limited dispersal.

Another haplogroup A subclade that might demonstrate regional continuity includes an ancient individual from the California Coast (1,000–500 year old, Monterey) and an individual identified as Yokuts who share two mutations (Fig. 2, subclade e). However, both mutations are in the hypermutable poly C tract of HVI and often co-occur. In addition, a protohistoric individual from the Columbia Plateau and a Wishram individual (protohistoric, Memaloose Cemetery) share a back-
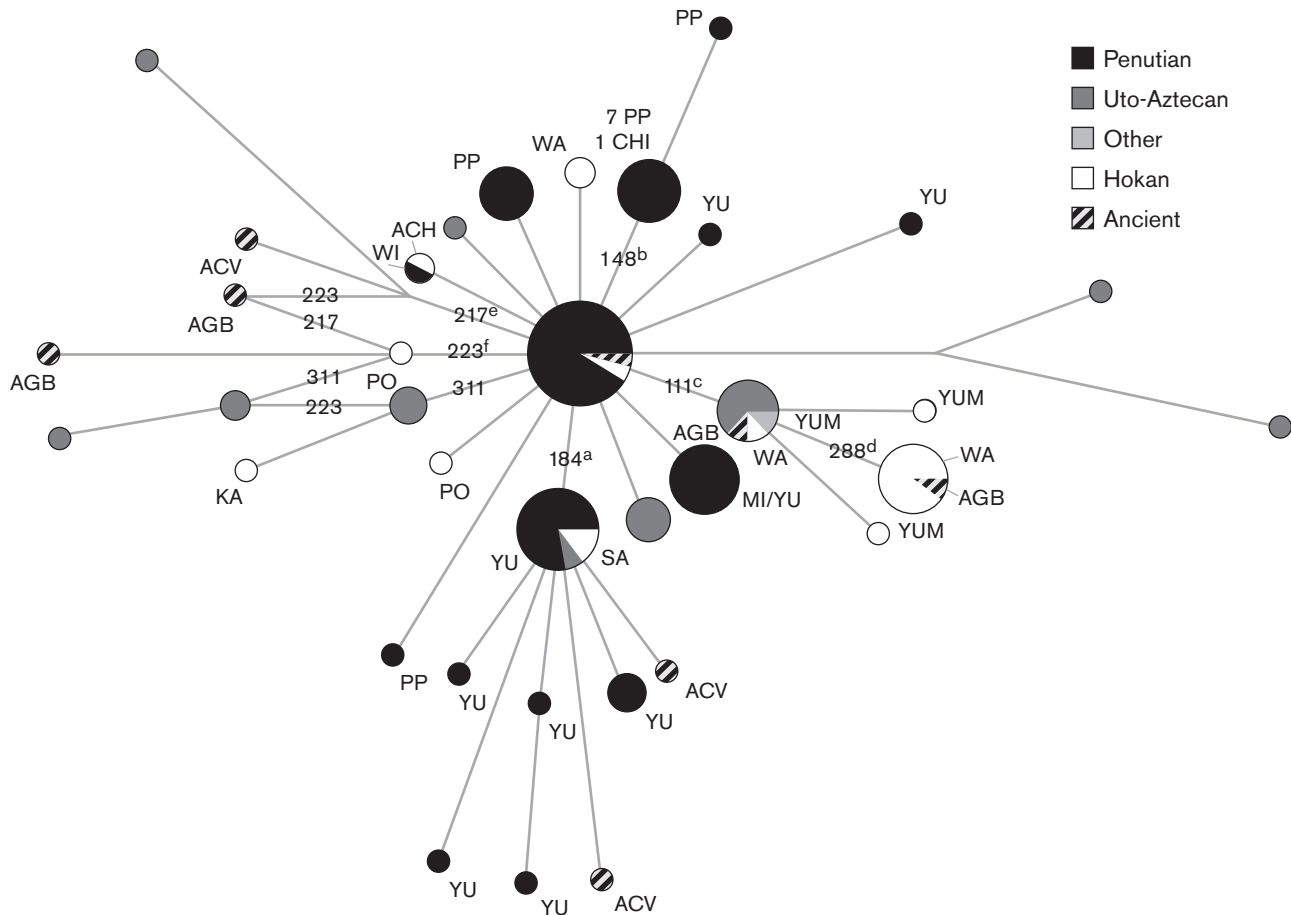

Figure 3. Haplogroup B network. ACH = Achumawi/Atsugewi, KA = Karok, PO = Pomoan, SA = Salinan, WA = Washo, YUM = Yuman, CHI = Chinookan, PP = Plateau Penutian, WI = Wintuan, MI = Maiduan, YU = Yok-Utian, CHU = Chumashan, ACV = Ancient Central Valley, AGB = Ancient Great Basin. The basal B node includes four Chinookan, eleven Plateau Penutian, two Wintuan, two Yok-Utian, one Ancient Columbia Plateau, and one Ancient Great Basin sample.

mutation at np 16223 (Fig. 2, subclade f). This lineage was reported by Malhi et al. (2004), who also observed the same mutation in the neighboring Nuu-Chah-Nulth.

Haplogroup B. Haplogroup B is common among modern individuals from the Central Valley, Great Basin, and Columbia Plateau as well as among Yuman speakers. Two star-shaped subclades, possibly indicative of population expansions, are evident in the network (Fig. 3). The subclade characterized by a transversion at np 16184 (Fig. 3, subclade a; Fig. 4) was noted by Johnson and Lorenz (2006). Despite the position of np 16184 in the poly C tract of HVI, the relative rarity of transversions, as well as the lack of observation of this mutation elsewhere in the Americas, lends credence to the supposition that this is a monophyletic subclade. Figure 4 shows that

this mutation is geographically widespread within both Yokuts and Utian and that most of the individuals belonging to this subclade are either Yok-Utian or from the ancient Central Valley. Three other individuals (two Salinan and one Mono) also belong to this subclade.

Further lineage-sharing among members of different language families in the hypothetical grouping of Penutian is observed. A mutation at np 16148 that is primarily observed among the Yakima (Plateau Penutian) was also reported among the Chinookan Wishram (Fig. 3, subclade b) (Malhi et al. 2004).

The other star-shaped subclade in the haplogroup B network, demarcated by a mutation at hypermutable site 16111 (Fig. 3, subclade c), unites Takic and Yuman individuals with Washo and individuals from the ancient
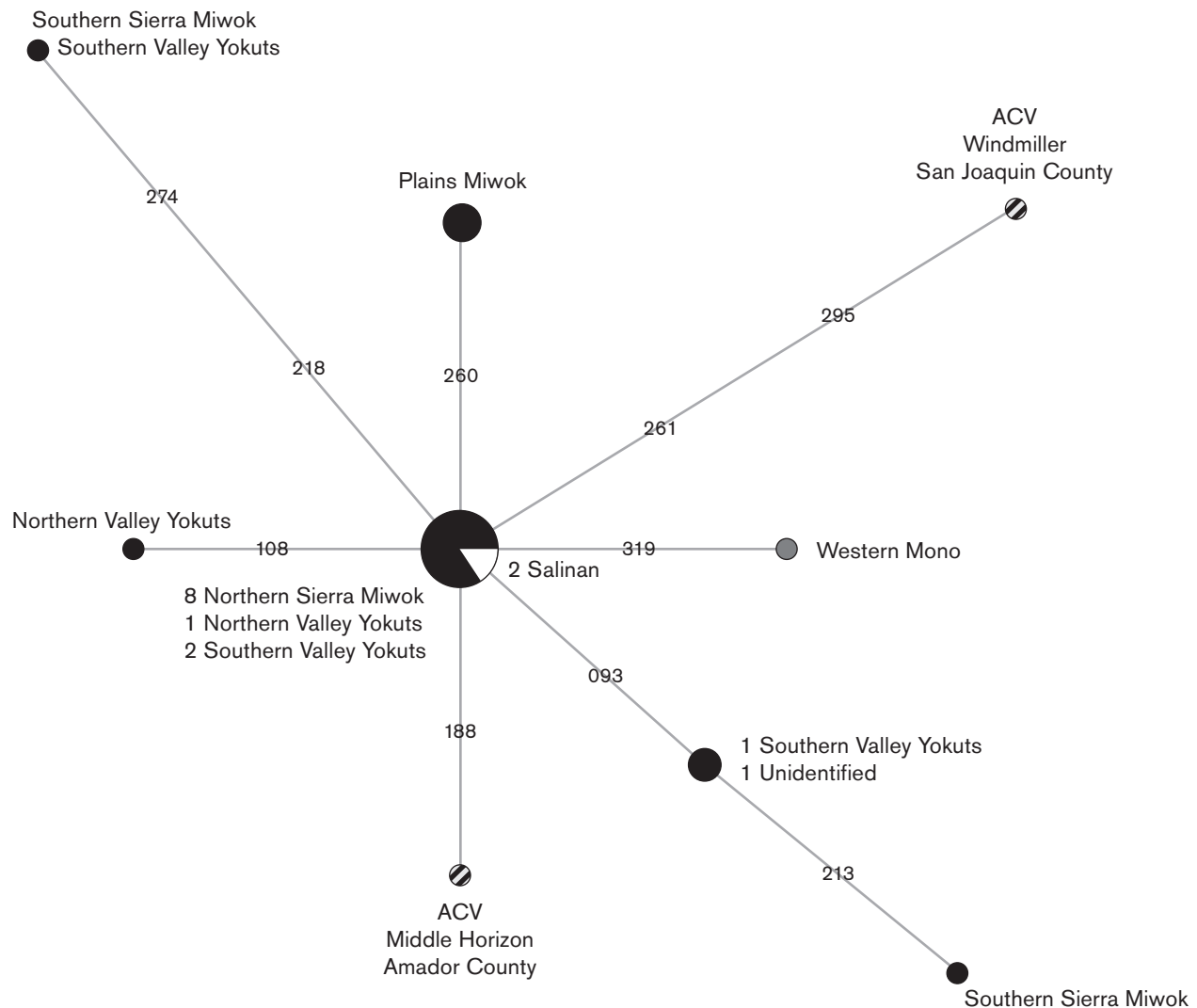

Figure 4. Yok-Utian subclade.

Great Basin, but only the nodal lineage was observed among Takic. The remains of the individuals in this subclade who are from the Great Basin are dated to 1,950±100 B.P. and 1,360±80 B.P. A derived haplotype within the subclade, characterized by a mutation at np 16288 (Fig. 3, subclade d), was observed in several Washo and one of the individuals from the Great Basin (the more recent individual), and provides direct evidence for biological continuity in the Great Basin.

While reticulation in the haplogroup B network is caused by nps 16217, 16223, and 16311, a few lineages of interest are still evident. Individuals from the ancient (3,540±95 B.P.) and modern (Shoshone) Great Basin and ancient (Middle Horizon) Central Valley share a back-mutation at np 16217 (Fig. 3, subclade e), and individuals from the ancient (860±75 B.P. and 3,540±95

B.P.) and modern Great Basin (Mono) share a mutation at np 16223 (which is, albeit, hypermutable) (Fig. 3, subclade f). This mutation at np 16223 is shared with a Pomo individual as well.

Haplogroup C. Haplogroup C reaches a high frequency among ancient inhabitants of the Central Valley, the Washo, and speakers of Yuman and Takic languages. The high frequency of haplogroup C observed among ancient inhabitants of the Central Valley and modern Takic speakers is consistent with the hypothesis that Proto-Uto-Aztecans were present in the Central Valley prior to the arrival of Yok-Utians (Eshleman and Smith 2007). However, the haplogroup C network (Fig. 5) does not provide strong support for this hypothesis. Evidence of haplogroup C lineages shared by Takic speakers and ancient inhabitants of the Central

Figure 5. Haplogroup C network. WA = Washo, YUM = Yuman, YUR = Yurok, COO = Coosan, PP = Plateau Penutian, YU = Yok-Utian, CHU = Chumashan, ACP = Ancient Columbia Plateau, ACV = Ancient Central Valley, AGB = Ancient Great Basin. The basal C node includes two Uto-Aztecan (one Numic and one Takic), one Plateau Penutian, six Ancient Central Valley, one Ancient Columbia Plateau, and one Ancient Great Basin sample. One lineage observed solely in the Washo that is two mutations away from basal C was removed from the network because of lack of space.

Valley rests primarily on np 16311 (Eshleman and Smith 2007), which was removed from consideration for this network because it caused excessive reticulation (see above). The shared presence of the mutation in Takic speakers and ancient inhabitants of the Central Valley may or may not be due to homoplasy.

There are, however, additional instances of closely-related lineages observed among Takic speakers and ancient inhabitants of the Central Valley. An older subclade, characterized by a mutation at np 16234 (Fig. 5, subclade a), unites an ancient (Middle Horizon, CA-SOL-270) individual from the Central Valley, Takic speakers (two Luiseño and one Cahuilla), and a Yuman speaker. Yet another Takic speaker is in the same subclade, defined by a mutation at np 16362 (Fig. 5, subclade b), as ancient (Windmiller and Middle Horizon) individuals from the Central Valley (CA-SJO-112 and CA-SOL-270, respectively). However, this subclade is also interesting in that it primarily contains modern individuals from language families designated as Penutian (Yokuts, Costanoan, and Yakima). While intriguing, it should be noted that np 16362 is highly mutable.

Another potential subclade also connects ancient individuals from the Central Valley (from sites associated with both the Windmiller and Middle Horizon cultures) to a modern Yok-Utian (Costanoan). While nps 16223 and 16124 cause reticulation, making this subclade irresolvable, the mutation at np 16124 (Fig. 5, subclade c) is shared by the Costanoan individual and by an individual from the Middle Horizon in the Central Valley (CA-AMA-56). Unlike np 16124, np 16223 is highly hypervariable. Hence, it is uncertain whether the lineage observed in the Costanoan individual descends from the lineages observed in both the Windmiller and Middle Horizon burials, with a back mutation at np 16223, or whether the Middle Horizon lineage with the mutation at np 16124 acquired the mutation at np 16223 separately from the other ancient Central Valley lineages with this mutation.

One of the haplogroup C lineages observed in the ancient (Windmiller) Central Valley, characterized by a mutation at 16356 (Fig. 5, subclade d), was also found in a modern Coos individual. Coosan is part of the hypothetical grouping of Oregon Penutian (Sapir 1921).

The haplogroup C network also provides evidence of distant relationships between ancient and modern individuals outside the Central Valley. A potential

subclade, characterized by a mutation at np 16086 (Fig. 5, subclade e), includes an ancient (500-1,500 B.P.) individual from the Columbia Plateau, a Numic speaker (Northern Paiute), and a Takic speaker (Kitanemuk). Although mutations at nps 16189 and 16086 cause reticulation in the network, the mutation at np 16086 should be given greater credence because np 16189 is in the hypermutable poly C tract.

Haplogroup D. Haplogroup D is common among modern and ancient individuals from the Great Basin, Central Valley, and Columbia Plateau. The haplogroup D network (Fig. 6) provides some evidence of lineage sharing spanning geography, (i.e., between two or more of these regions), time (i.e., between modern and ancient individuals within a region), and both geography and time (i.e., between modern individuals in one region and ancient individuals in another region).

A subclade defined by a back-mutation at np 16325 (Fig. 6, subclade a) appears to be old because all of the descendant nodes are both highly derived and represented primarily by ancient individuals (the exception is a highly derived lineage in four modern Chumash individuals, as described by Johnson and Lorenz [2006]). It may be evidence of a biological connection between ancient individuals in the Great Basin (dated to $3,630 \pm 60$ B.P. and $3,165 \pm 370$ B.P., with the more derived lineage observed in the more ancient individual) and those in the Central Valley (both from the Middle Horizon, CA-SOL-270). However, it has been demonstrated that this site experiences elevated post-mortem damage as well (Gilbert et al. 2003), so it is possible that what appears to be a mutation inherited from a common ancient ancestor is in fact due to post-mortem DNA damage in the ancient individuals and a separate mutation in the Chumash lineage.

Sharing of lineages in different geographic regions is also evident among modern individuals. Sequences sampled from Miwok and Yakima individuals share a mutation at np 16129 (Fig. 6, subclade b). However, this site is highly mutable (Stoneking 2000).

Biological continuity within the Central Valley is demonstrated by a haplotype, due to a mutation at np 16234 (Figure 6, subclade c), that is shared by an ancient individual (from the Middle Horizon, CA-SOL-270) and modern Yok-Utian individuals (descendants of intermarriage between Waupumne
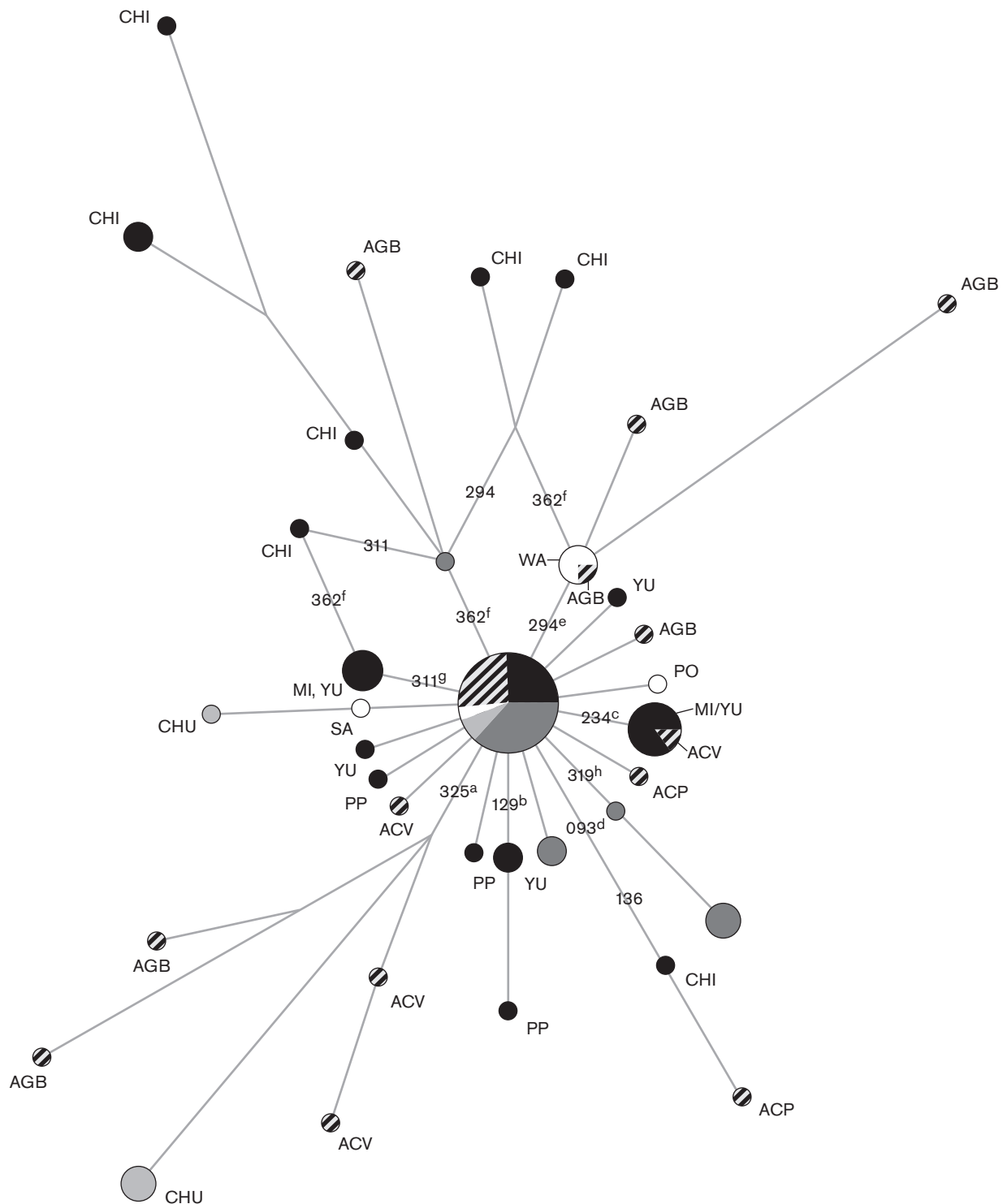
Figure 6 . Haplogroup D network. PO = Pomoan, SA = Salinan, WA = Washo, CHI = Chinookan, PP = Plateau Penutian, MI = Maiduan, YU = Yok-Utian, CHU = Chumashan, ACP = Ancient Columbia Plateau, ACV = Ancient Central Valley, AGB = Ancient Great Basin. The basal D node includes four Uto-Aztecan (two Numic, one Takic, and one Tubatulabal), five Yok-Utian, two Chumashan, five Ancient Central Valley, and two Ancient Great Basin samples.

Nisenan and Northern Miwok, Maternal Group 8 in Table 4). Stoneking (2000) demonstrates a slightly elevated evolutionary rate at np 16234 relative to the average for HVI. The modern Yok-Utian individuals also exhibit a transition at np 16391, but this site was not sequenced in the ancient individual from Solano county and was outside of the stretch of HVI used in haplotype analysis. Biological continuity on the Columbia Plateau is demonstrated by a subclade, defined by mutations at nps 16093 and 16136 (Fig. 6, subclade d), that includes an ancient individual (500–1,500 B.P.) and a modern individual (Wishram).

Potential subclades, with reticulation caused by nps 16311, 16362, and 16294, unite ancient inhabitants of the Great Basin with modern inhabitants of the Great Basin, Columbia Plateau, and Central Valley/San Francisco Bay Area. While all of these sites are hypervariable (Stoneking 2000), np 16294 exhibits the lowest rate of mutation of the three. The subclade defined by a mutation at np 16294 (Fig. 6, subclade e) links the ancient Great Basin (burials dated to $1,620 \pm 90$ B.P. , $2,434 \pm 85$ B.P., and $2,735 \pm 100$ B.P.) with modern Columbia Plateau (Wishram) as well as modern Washo. A back mutation at np 16362 (Fig. 6, subclade f—three potential evolutionary routes to this subclade are illustrated) potentially unites the ancient Great Basin ($4,745 \pm 115$ B.P.) with the modern Columbia Plateau (Wishram). This back mutation is also present in a modern Vanyumé (Uto-Aztecan) individual. Another individual from the ancient Great Basin ($3,630 \pm 60$ B.P.) has the back-mutation as well, but this individual is in subclade "a" by virtue of sharing both the back-mutation at np 16325 and a mutation at np 16069 with another individual from the ancient Great Basin. A mutation at np 16311 (Fig. 6, subclade g) potentially unites the modern Central Valley/San Francisco Bay Area (Waupumne Nisenan and Costanoan from Alameda County) with the modern Columbia Plateau (Wishram). However, as stated, these sites are hypervariable.

Subclade "h" (Fig. 6) unites the maternal group from IBMI with Luiseño heritage (Maternal Group 7, Table 4) with a Cahuilla individual. The haplotype associated with Maternal Group 7 is further derived by two mutations.

Age of Haplogroup B "Yok-Utian" Subclade

Phylogenetic dispersion for the Yok-Utian subclade (Fig. 4), defined by a transversion at np 16184, was calculated as $\rho = 0.58333$ ($\sigma = 0.21246$). Two calibrations for ρ, one transition/20,180 years (Forster et al. 1996) and one mutation/9,213 years (Kemp et al. 2007), provide the following estimated ages of the ancestral node in years: $11,772 \pm 4,287$ and $5,374 \pm 1,957$. Kemp et al. (2007) also provide a standard deviation for the estimate of ρ they used to calibrate the molecular clock. Incorporating one standard deviation into their (Kemp et al. 2007) estimate of ρ results in a range of 2,642 years ($\rho = 0.37087$ for the Yok-Utian subclade and one mutation every 7,123 years) to 10,375 years for the ancestral node of the subclade ($\rho = 0.75979$ for the Yok-Utian subclade and one mutation every 13,038 years).

It has been noted that dates for population events in the Americas based on molecular data are often substantially older than dates based on archaeological data. At the same time, pedigree-based estimates for the mtDNA molecular clock are substantially higher than phylogeny-based estimates. One reason for this may be that the observed rate of molecular evolution slows down with time (Ho and Larson 2006). The faster rate of HVI evolution estimated by Kemp et al. (2007) with the use of ancient and modern DNA and radiocarbon dating, and determined in accordance with the recommendations of Ho and Larson (2006), is intermediate between pedigree-based estimates of the mtDNA mutation rate and phylogeny-based estimates of the mtDNA substitution rate. It is more appropriate for dating events in recent human history, such as those within the Americas (Kemp et al. 2007). Similarly high rates for mtDNA HVI evolution, based on ancient DNA and radiocarbon dating rather than phylogenies, have also been observed in other animals besides humans (Lambert et al. 2002; Shapiro et al. 2004). Therefore, we have greater confidence in the age of the ancestral node of the Yok-Utian subclade based on the calibration provided by Kemp et al. (2007); i.e., between ca. 2,600 and 10,400 years, with a point estimate of ca. 5,400 years. This estimate is also a minimum age for the transversion at 16184 that defines the subclade.

A minimum age of the ancestral node is also provided by the radiocarbon dates from the burial sites of the two ancient individuals who belong to this subclade. Although these particular burials were not dated directly, other remains at these two sites provide dates of $1,765 \pm 30$ B.P. to $2,055 \pm 35$ B.P. (CA-AMA-56, Middle

Horizon, dates from two burials) and $2,860 \pm 150$ B.P. to $3,200 \pm 150$ B.P (CA-SJO-112, Windmiller, dates from two burials). Hence, an archaeologically-based provisional minimum age of the ancestral node is $2,860 \pm 150$ B.P. to $3,200 \pm 150$ B.P., which is slightly older than the lower range of the molecularly-based estimated age of the ancestral node. Assuming that the individual at CA-SJO-112 who belongs in the Yok-Utian subclade expired over ca. 3,000 years ago, the ancestral node of the subclade is likely substantially older than 3,000 years because the haplotype observed in this ancient individual is differentiated from the ancestral node by two mutations. However, it is also possible that the individual in this subclade from CA-SJO-112 was not a contemporary of the individuals from the same site whose remains were radiocarbon dated. Additionally, because the frequency of different haplotypes associated with the descendant nodes affects the estimate of phylogenetic dispersion, bias in the estimate of the age of the ancestral node of this subclade may have been introduced by the non-random sampling of matrilines in this study and in that of Johnson and Lorenz (2006).

IMPLICATIONS FOR PREHISTORY IN CALIFORNIA, THE GREAT BASIN, AND THE COLUMBIA PLATEAU

Biological Continuity in Central California

The inclusion in our study of additional modern samples from the Central Valley and of network analyses provides strong evidence for biological continuity between individuals buried in Windmiller and Middle Horizon sites and modern individuals indigenous to the Central Valley. The results from the current study are consistent with those from previous reports that demonstrate a high frequency of mtDNA haplogroups B and D among modern Yok-Utians (Eshleman 2002; Eshleman and Smith 2007; Johnson and Lorenz 2006), in contrast to the high frequency of haplogroups C and D observed among ancient inhabitants of the Central Valley (Eshleman 2002; Eshleman et al. 2007). This discrepancy, coupled with the high frequency of haplogroup C among Takic speakers in southern California (Eshleman 2002; Eshleman et al. 2004; Johnson and Lorenz 2006), was previously interpreted as tentative support for a Proto-Uto-Aztecan presence

in the Central Valley, suggesting that a Penutian expansion occurred more recently in time. However, processes other than population replacement can affect haplogroup frequencies, and analysis of mtDNA haplotypes allows the assessment of whether specific maternal lineages have persisted in a region over time.

Inspection of haplotypes reveals that several identical or closely related matrilines in haplogroups B, C, and D are shared by both ancient and modern inhabitants of Central California. There is far less evidence of closely related matrilines shared between the ancient inhabitants of the Central Valley and modern Takic speakers. The similarity in haplogroup C lineages observed in the prehistoric Central Valley and among Takic speakers is primarily due to a mutation at np 16311 (Eshleman and Smith 2007). Although this may represent shared ancestry, this site is hypervariable and may represent homoplasty; in the dataset analyzed here, as might be expected with homoplasy, it caused such extensive reticulation that it was excluded from the final haplogroup C network. Other than two distantly related lineages in a subclade of haplogroup C, we find no evidence that Takic lineages are closely related to those found in the prehistoric Central Valley. While the present data cannot be used to rule out a Proto-Uto-Aztecan presence in the Central Valley, haplotypic analyses reveal that, contrary to previous observation (Eshleman and Smith 2007), the mtDNA data do not provide compelling support for this hypothesis.

Some of the mutations that define subclades shared by both ancient and modern inhabitants of Central California are also at sites with elevated evolutionary rates. There is also the possibility that purported mutations in ancient individuals included in this study are actually due to post-mortem damage; human mtDNA sites that are more prone to post-mortem damage also tend to be those that are highly mutable (Gilbert et al. 2003). However, over a relatively short period of time, most shared mutations that result from descent from a common ancestor are expected to be at highly variable sites. Therefore, some of the shared mutations at hypervariable sites are probably the result of descent from a common ancestor, rather than homoplasy or post-mortem damage. Additionally, some of the shared mutations we observed are at less variable sites and are unlikely to be due to homoplasy or post-mortem damage.

One shared mutation that is unlikely to reflect homoplasy, due to the rarity of transversions relative to transitions, is the transversion at np 16184. This transversion defines the haplogroup B star-shaped subclade that primarily includes ancient inhabitants of the Central Valley and modern Yok-Utians. It is also possible that what appears to be a transversion is actually the result of the loss of a C residue in the poly C tract of HVI and the insertion of an A residue prior to the poly C tract. If this is the case, it is still unlikely that the presence of an A residue at np 16184 that was observed in a number of individuals in this study is due to homoplasy.

Johnson and Lorenz (2006) hypothesized that this subclade results from a Yokutsan population expansion. The inclusion in the present study of sequences from more Utian individuals (Johnson and Lorenz [2006] analyzed twice as many Yokutsan as Utian lineages) and of sequences from ancient inhabitants of the Central Valley leads us to a different conclusion. The transversion at np 16184 was found not only in diverse subfamilies of Yokuts but also in diverse subgroups of Utian. However, while this mutation has been segregating for thousands of years in populations within the Central Valley, it appears only rarely in neighboring modern groups (to date, Salinan and Mono) that are not of either Yokuts or Utian affiliation.

These observations suggest that the subclade is likely associated with a Yok-Utian, rather than Yokutsan, population expansion. Local gene flow may be responsible for the appearance of the transversion in neighboring groups (Johnson and Lorenz 2006). However, we cannot rule out the possibility that disproportionate sampling of Yok-Utians has led us to erroneously infer that the expansion primarily involved the ancestors of Yok-Utians.

This subclade is not directly informative as to the linguistic validity of the hypothesized Yok-Utian family. It does, however, support a relatively deep shared history between Yokuts and Utian groups, and the estimated age of the population expansion associated with this genetic subclade is intriguing in light of linguistic and archaeological data. The estimated age of the population expansion associated with this subclade can be bracketed by the point estimate of the ancestral node and the radiocarbon dates for the archaeological site

from which the older of the two ancient individuals carrying the derived lineages was recovered. This rough approximation of the age of the commencement of the expansion, between about 5,400 years ago and 3,000 years ago, is consistent with a probable split of pre-proto-Yokuts and Utian around 4,500 years ago (Golla 2007) and with a population expansion of Utians represented in the archaeological record by the Windmiller or Middle Horizon cultures (Moratto 1984).

While the results of this study are consistent with the hypothesis that the inhabitants of the Cecil, Applegate, and Cook sites were speakers of Proto-Utian or Proto-Yok-Utian languages, the genetic data presented cannot directly inform us as to the cultural or linguistic affiliations of these ancient groups. Nevertheless, the results suggest that if a biological population replacement has occurred in the Central Valley in the past few thousand years, it was an incomplete replacement that included extensive admixture with local females. Hence, any anthropological models for culture or language change in the Central Valley in the last few thousand years should accommodate some degree of biological continuity.

Matrilineal Ties Among Inhabitants of Central California, the Great Basin, and the Columbia Plateau

All four networks, for haplogroups A–D, demonstrate that closely related mtDNA lineages are shared among ancient and/or modern inhabitants of Central California, the Great Basin, and the Columbia Plateau. A number of the observed matrilineal relationships rest on mutations at highly variable sites and, correspondingly, are in reticulated subclades. However, as previously noted, most shared mutations at a shallow time depth are expected to occur at predominantly hypervariable sites, and the preponderance of matrilineal ties within and between groups in these regions suggests that at least some of the shared hypervariable mutations are the result of descent from a common ancestor.

Intriguingly, the ancient inhabitants of the Great Basin appear to provide a link between the Plateau and the Central Valley. This echoes the inference of Eshleman et al. (2004), which was based on the high frequency of haplogroup D among all three regions in modern individuals and in the Great Basin in ancient individuals. Closely related matrilines are shared primarily between

the ancient inhabitants of the Great Basin and either the ancient inhabitants of the Central Valley or the modern (Penutian) inhabitants of the Columbia Plateau. However, biological continuity is also evident within all three regions, and the apparent discrepancy in the temporal provenience of haplotypes in the Central Valley and on the Plateau that are closely related to haplotypes sampled in the Great Basin could be due to sampling biases; for example, few individuals from northern California or southern Oregon were sampled.

The observation that ancient maternal lineages in the Great Basin provide a link between lineages in the Plateau and in the Central Valley is consistent with hypotheses based on both archaeological and linguistic data that give the Great Basin a central role in the movement of culture and/or people in the West. That is, if it is assumed that the ancient inhabitants of the Central Valley for which we have mtDNA data spoke a Penutian language, then this observation would be consistent with the linguistic hypothesis of a Penutian homeland in the northwest Great Basin. However, this proposition must remain extremely tenuous. It defies (1) our admitted ignorance of the languages spoken by those long departed, (2) our acknowledgment that, given different means of inheritance, biological relationships cannot be used as sole evidence of linguistic relationships (and vice versa), and (3) the unestablished validity of the Penutian language group.

A central role for the ancient Great Basin, linking inhabitants of the Central Valley and Columbia Plateau, is also consistent with the hypothesis based on archaeological data that, spurred by the arid environment in the Great Basin during the Altithermal, groups moved from the Great Basin into the Central Valley and Columbia Plateau (Ragir 1972). However, the data offer only a limited slice of population prehistory in these regions and are uninformative as to the nature of these genetic relationships. That is, this shared biological prehistory could be the result of gene flow accompanying trade, migrations from one region to another, or an ancient cohesive population spread over multiple regions.

Evidence of biological continuity in the Great Basin, with instances of both the Northern Paiute and Washo clustering with ancient inhabitants of the Great Basin, adds yet another perspective on the hypothesized Numic

expansion from southeast California into the Great Basin about 1,000 years ago (Lamb 1958). Cabana et al. (2008) recently used computer simulations to demonstrate that, under certain demographic parameters, a population replacement is not necessary to explain observed changes in mtDNA haplogroup frequencies between ancient and modern inhabitants of the Great Basin (Kaestle 1998; Kaestle and Smith 2001). While this approach is very informative, it cannot be used to reject population replacement. Analysis of mtDNA haplotypes, however, reveals that a complete population replacement is unlikely. Whether a hypothesized Numic expansion was associated with a partial population replacement cannot be determined from our analysis or from that of Cabana et al. (2008). However, female biological continuity in the Great Basin would be consistent with a Numic spread that, due to the importance of a high ratio of females to males in a culture that relies heavily on processing of food (Bettinger and Baumhoff 1982), incorporated pre-Numic females.

NOTE

[1]Some of the specialized terms employed in this article are defined below.

Mitochondrial DNA (mtDNA): Small circular genome in the mitochondria of cells. Unlike the nuclear genome, it is haploid and inherited uniparentally (through the mother).

Hypervariable Region I (HVI): Also "HVRI" and "HVSI" (Hypervariable Segment I). One of two stretches in the mtDNA control region that are commonly sequenced for population studies because of their rapid rate of evolution.

Hypervariable site: Also "hypermutable." A site with an evolutionary rate that is higher than average for HVI.

Phylogenetic network: A graph of ancestor/descendant relationships. Unlike a phylogenetic tree, a network allows reticulation. Different algorithms can be used for constructing a network.

Phylogenetic dispersion: The average number of mutations separating ancestral and descendant nodes. Given a rate of evolution in HVI, phylogenetic dispersion can be used to estimate the time separating ancestral and descendant nodes

Reticulation: Existence of multiple possible evolutionary pathways, or ancestor/descendant relationships, between nodes in a phylogeny. That is, it is the presence of more than one possible ancestral haplotype for a given haplotype.

Basal: A (relatively) more ancestral node or haplotype.

Derived: A (relatively) more recently evolved node or haplotype, as indicated by the accumulation of additional mutations.

Haplogroup: A grouping of mtDNA lineages, or haplotypes, that descend from a common ancestor a relatively long time ago, as determined by a large number of shared mutations.

Haplotype: Can often be used interchangably with "lineage" and "node." Each haplotype belongs to a single haplogroup by virtue of sharing specific mutations in common with all other haplotypes in that haplogroup. A haplotye may be basal (that is, it may have no additional mutations beyond those that characterize the haplogroup) or it may be derived (that is, it may have additional mutations). Barring mutation, all individuals in a matriline should have the same haplotype.

Back-mutation: Also "reversal." A mutation that results in a reversal of a character to the state prior to mutation. For example, a "T" at nucleotide pair 16111 is typical for haplotypes in haplogroup A, but a back-mutation could result in a "C," the ancestral state, at np 16111 in a haplotype that still has a number of other markers indicative of haplogroup A.

Homoplasy: Shared character state (i.e, specific base such as "A") that is the result of separate mutations in separate lineages rather than inheritance from a common ancestor.

Transition: Mutations that involve a purine exchanged with a purine or a pyrimidine exchanged with a pyrimidine; i.e., mutations that exchange bases of similar shape.

Transversion: Mutations that involve a purine exchanged with a pyrimidine or vice versa; i.e., mutations that exchange bases of dissimilar shape.

Ancient: Sample taken from the bone or tooth of a deceased individual of undetermined ethnolinguistic affiliation. This incorporates individuals who lived from about 10,000 years ago to about 200 years ago.

Modern: Sample taken from a living individual. Note, however, that sequences from protohistoric samples from Memaloose Island Cemetery are treated here as "modern" because they were given a specific ethnolinguistic affiliation (Chinookan) in the published source from which they were taken.

REFERENCES

Bandelt, H. J., P. Forster, and A. Rohl
1999   Median-Joining Networks for Inferring Intra-specific Phylogenies. *Molecular Biology and Evolution* 16(1):37–48.

Bettinger, R. L., and M. A. Baumhoff
1982   The Numic Spread: Great Basin Cultures in Competition. *American Antiquity* 47(3):485–503.

Breschini, G. S.
1983   *Models of Population Movements in Central California Prehistory*. Salinas: Coyote Press.

Cabana, G. S., K. L. Hunley, and F. A. Kaestle
2008   Population Continuity or Replacement? A Novel Computer Simulation Approach and its Application to the Numic Expansion (Western Great Basin, USA). *American Journal of Physical Anthropology* 135(4):438–447.

Callaghan, C. A.
1997   Evidence for Yok-Utian. *International Journal of American Linguistics* 63(1):18–64.
2001   More Evidence for Yok-Utian: A Reanalysis of the Dixon and Kroeber Sets. *International Journal of American Linguistics* 67(3):313–345.

DeLancey, S., and V. Golla
1997   The Penutian Hypothesis: Retrospect and Prospect. *International Journal of American Linguistics* 63(1):171–202.

Dixon, R. B., and A. L. Kroeber
1913   Relationship of the Indian Languages of California. *Science* 37(945):225.

Eshleman, J. A.
2002   *Mitochondrial DNA and Prehistoric Population Movements in Western North America*. Ph.D. dissertation, University of California, Davis.

Eshleman, J. A., R. S. Malhi, J. R. Johnson, F. A. Kaestle, J. G. Lorenz, and D. G. Smith
2004   Mitochondrial DNA and Prehistoric Settlements: Native Migrations on the Western Edge of North America. *Human Biology* 76(1):55–75.

Eshleman, J. A., and D. G. Smith.
2007   Prehistoric Mitochondrial DNA and Population Movements. In *California Prehistory: Colonization, Culture, and Complexity*, T. L. Jones and K. Klar, ed., pp. 291–298. New York: AltaMira Press.

Forster, P., R. Harding, A. Torroni, and H. J. Bandelt
1996   Origin and Evolution of Native American mtDNA Variation: A Reappraisal. *American Journal of Human Genetics* 59(4):935–945.

Gilbert, M. T. P., E. Willerslev, A. J. Hansen, I. Barnes, L. Rudbeck, N. Lynnerup, and A. Cooper
2003   Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. *American Journal of Human Genetics* 72(1):32–47.

Golla, V.
2007 Linguistic Prehistory. In *California Prehistory: Colonization, Culture, and Complexity*, T. L. Jones and K. Klar, ed., pp. 71–82. New York: AltaMira Press.

Hattori, E. M.
1982 The Archaeology of Falcon Hill, Winnemucca Lake, Washoe County, Nevada. *Nevada State Museum Anthropological Papers* 18. Carson City, Nevada.

Ho, S. Y. W., and G. Larson
2006 Molecular Clocks: When Times Are A-changin'. *TRENDS in Genetics* 22(2):79–83.

Johnson, J. R., and J. G. Lorenz
2006 Genetics, Linguistics, and Prehistoric Migrations: An Analysis of California Indian Mitochondrial DNA Lineages. *Journal of California and Great Basin Anthropology* 26(1):33–64.

Kaestle, F. A., and D. G. Smith
2001 Ancient Mitochondrial DNA Evidence for Prehistoric Population Movement: The Numic Expansion. *American Journal of Physical Anthropology* 115(1):1–12.

Kaestle, F. A.
1998 *Molecular Evidence for Prehistoric Native American Population Movement: The Numic Expansion*. Ph.D. dissertation, University of California, Davis.

Kemp, B. M., R. S. Malhi, J. McDonough, D. A. Bolnick, J. A. Eshleman, O. Rickards, C. Martinez-Labarga, J. R. Johnson, J. G. Lorenz, E. J. Dixon, T. E. Fifield, T. H. Heaton, R. Worl, and D. G. Smith
2007 Genetic Analysis of Early Holocene Skeletal Remains from Alaska and Its Implications for the Settlement of the Americas. *American Journal of Physical Anthropology* 132(4):605–621.

Lamb, S.
1958 Linguistic Prehistory in the Great Basin. *International Journal of American Linguistics* 24(2):95–100.

Lambert, D. M., P. A. Ritchie, C. D. Millar, B. Holland, A. J. Drummond, and C. Baroni
2002 Rates of DNA Evolution in Ancient DNA from Adélie Penguins. *Science* 295(5563):2270–2273.

Malhi, R. S., K. E. Breece, B. A. S. Shook, F. A. Kaestle, J. C. Chatters, S. Hackenberger, and D. G. Smith
2004 Patterns of mtDNA Diversity in Northwestern North America. *Human Biology* 76(1):33–54.

Meyer, S., G. Weiss, and A. von Haeseler
1999 Pattern of Nucleotide Substitution and Rate Heterogeneity in the Hypervariable Regions I and II of Human mtDNA. *Genetics* 152(3):1103–1110.

Mithun, M.
1999 *The Languages of Native North America*. Cambridge, United Kingdom: Cambridge University Press.

Moratto, M. J.
1984 *California Archaeology*. Salinas: Coyote Press.

Nichols, M. J. P.
1981 Old California Uto-Aztecan. *Journal of the Steward Anthropological Society* 15: 23–46.

Ragir, S.
1972 The Early Horizon in Central California Prehistory. *Contributions of the University of California Archaeological Research Facility* 15. Berkeley, California.

Sapir, E.
1921 A Characteristic Penutian Form of Stem. *International Journal of American Linguistics* 2(1/2):58–67.

Schroeder, K. B., R. S. Malhi, and D. G. Smith
2006 Opinion: Demystifying Native American Opposition to Genetic Research. *Evolutionary Anthropology* 15(3):88–92.

Shapiro, B., A. J. Drummond, A. Rambaut, M. C. Wilson, P. E. Matheus, A. V. Sher, O. G. Pybus, M. T. P. Gilbert, I. Barnes, J. Binladen, E. Willerslev, A. J. Hansen, G. F. Baryshnikov, J. A. Burns, S. Davydov, J. C. Driver, D. G. Froese, C. R. Harington, G. Keddie, P. Kosintsev, M. L. Kunz, L. D. Martin, R. O. Stephenson, J. Storer, R. Tedford, S. Zimov, and A. Cooper
2004 The Rise and Fall of the Beringian Steppe Bison. *Science* 306(26):1561–1565.

Shields, G. F., A. M. Schmiechen, B. L. Frazier, A. Redd, M. I. Voevoda, J. K. Reed, and R. H. Ward
1993 MtDNA Sequences Suggest a Recent Evolutionary Divergence for Beringian and Northern North American Populations. *American Journal of Human Genetics* 53(3):549–562.

Siepel, A.
2009 Phylogenomics of primates and their ancestral populations. *Genome Research* 19:1929–1941.

Smith, D. G., J. G. Lorenz, B. K. Rolfs, R. L. Bettinger, B. Green, J. A. Eshleman, B. A. Schultz, and R. S. Malhi
2000 Implications of the Distribution of Albumin Naskapi and Albumin Mexico for New World Pre-history. *American Journal of Physical Anthropology* 111(4): 557–572.

Stoneking, M.
2000 Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots. *American Journal of Human Genetics* 67(4):1029–1032.

Whistler, K.
1977 Wintun Prehistory: An Interpretation Based on Linguistic Reconstruction of Plant and Animal Nomenclature. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 3:157–174.