Environmental productivity predicts migration, demographic, and linguistic patterns in prehistoric California

Brian F. Codding1,2 and Terry L. Jones

1Department of Anthropology, University of Utah, Salt Lake City, UT 84112 and 2Department of Social Sciences, California Polytechnic State University San Luis Obispo, CA 93407

Global patterns of ethnolinguistic diversity vary tremendously. Some regions show very little variation even across vast expanses, whereas others exhibit dense mosaics of different languages spoken alongside one another. Compared with the rest of Native North America, prehistoric California exemplified the latter. Decades of linguistic, genetic, and archaeological research have produced detailed accounts of the migrations that aggregated to build California’s diverse ethnolinguistic mosaic, but there have been few attempts to explain the process underpinning these migrations and why such a mosaic did not develop elsewhere. Here we show that environmental productivity predicts both the order of migration events and the population density recorded at contact. The earliest colonizers occupied the most suitable habitats along the coast, whereas subsequent Mid–Late Holocene migrants settled in more marginal habitats. Other Late Holocene patterns diverge from this trend, reflecting altered dynamics linked to food storage and increased sedentism. Through repeated migration events, incoming populations replaced resident populations occurring at lower densities in lower-productivity habitats, thereby resulting in the fragmentation of earlier groups and the development of one of the most diverse ethnolinguistic patterns in the Americas. Such a process may account for the distribution of ethnolinguistic diversity worldwide.

Native California has long stood out as a region of exceptionally high ethnolinguistic diversity, a pattern generally recognized as the end product of repeated in-migrations by successive groups (1–6) (Fig. 1 and Fig. S1). Explaining why migrations led to the aggregation of so many ethnic groups in California, but not in neighboring regions, has been a longstanding challenge for North American prehistorians. Previous research worldwide has found correlations between environmental productivity, population density, and linguistic diversity (7–9), but these studies fail to explain the processes that fostered such patterning. Proposed explanations tend to focus on population replacement events, where incoming groups equipped with more intensive subsistence strategies out-compete in situ groups (10, 11). Although of great use in understanding cultural patterns in prehistory, these explanations provide less-than-adequate explanations for circumstances where in-migrations resulted not in full-scale replacement, but in the buildup of ethnic diversity.

Here we propose a simple explanation for this patterning based on predictions from an ideal free distribution model (IFD) from behavioral ecology (12, 13) (Fig. S2). Recent anthropological applications of the IFD have proven useful for explaining patterning in prehistoric colonization and settlement (14–18). The basic model assumes that environments vary in their suitability and that habitats decline in suitability as a function of population density. Assuming that individuals should attempt to maximize habitat suitability, incoming colonizers and migrants should occupy the highest-ranking habitats until a point where suitability declines to a level equal to the next highest ranking habitat. As populations increase through either migration or in situ growth, lower-ranking habitats should fill in rank order, with higher-ranking habitats always occupied by more individuals per area. From these dynamics, the IFD provides two main qualitative predictions: (i) the most suitable habitats should always be occupied first, and (ii) they should always have the highest population densities.

Observations that do not conform to these expectations may result from a number of factors, one of which is caused by a violation of the “free” assumption of the IFD. If for some reason individuals are no longer free to select the most suitable habitat, then IFD dynamics would give way to those of the ideal despotic (or dominance) distribution model (IDD) (12). Archaeological studies have used the IDD to better understand the emergence of hierarchies and intergroup resource competition (19, 20). IDD dynamics can emerge from any exclusionary tactics, including territoriality or even strongly sedentary adaptations that provide some advantage against potential competitors. This is more likely to occur where resources are concentrated and predictable, making resource-bearing habitats defensible (21).

Based on these model dynamics, we hypothesize that the first people to colonize California would have occupied the most suitable habitats. Individuals in these sweet spots were more likely to stay in place due to the greater demographic potential of these highly suitable environments. Subsequent migrants would have been best off settling in adjacent, although less productive, regions, resulting in the sequential occupation of increasingly marginal habitats. Because populations in more marginal habitats were likely to have lower population densities, they may have been susceptible to replacement by incoming migrants whose population densities were more likely to be at parity. In contrast, those occupying more suitable habitats would have been susceptible to replacement only in circumstances where incoming populations adopted exclusionary tactics ranging from greater sedentism to territorial aggression (10). In this way, the prehistory of regions with greater heterogeneity in habitat suitability should be characterized by a mix of migration outcomes that gradually produce the aggregation of ethnolinguistic diversity. Regions characterized by more homogeneous distributions of habitat suitability should experience zero-sum outcomes resulting either in stasis or full-scale replacement, leading to very limited ethnolinguistic diversity (11).

Author contributions: B.F.C. and T.L.J. designed research; B.F.C. and T.L.J. performed research; B.F.C. analyzed data; and B.F.C. and T.L.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: brian.codding@anthro.utah.edu.

2This article contains supporting information online at www.pnas.orglookup/supp/doi/10.1073/pnas.1302008110/-/DCSupplemental.
Here we test these predictions with estimates of habitat suitability, settlement chronologies, and population densities for populations in California. We use terrestrial net primary productivity (NPP), an approximation of plant productivity, as a proxy for habitat suitability. To test the first prediction of the IFD that the most suitable habitats are occupied first, we use genetic (22–26) and linguistic (4–6, 27, 28) data to estimate the timing of colonization and migration events for each language group in California (Table 1). We then use estimates of contact population density for each ethnolinguistic group (2, 9, 29–31) (Table S1) to test the second prediction that the most suitable habitats always have the highest population densities.

**Results**

Fig. 1 shows the distribution of environmental productivity (measured by NPP) summarized for each ethnolinguistic territory (Table S1). As has been shown globally for hunter–gatherer populations (7), NPP negatively covaries with the size of ethnographic territories throughout California ([linear: \( r^2 = 0.19, p = 0.0002, \) log linear: \( r^2 = 0.287, p < 0.0001 \) (Fig. S3)]. Mean environmental productivity (NPP) varies significantly across linguistic groups (ANOVA, df = 8, \( F = 10.52, p < 0.0001 \)). These differences are not a function of spatial autocorrelation (Moran’s \( I = 0.30, z = 0.94, p = 0.35 \)).

**Prediction 1: Most Suitable Habitats Are Occupied First.** To test this prediction, we examine mean NPP for each ethnographic territory by linguistic grouping rank-ordered by the estimated colonization or migration date. As shown in Fig. 2, the variability is largely distributed through time following our chronological predictions. The earliest groups to enter California occupied some of the most productive habitats. It appears that these habitats remained occupied until the contact era. Mean productivity declines significantly with subsequent migrations through the Middle to Late Holocene (Table 2). The Late Holocene spread of Numic speaking peoples into the lowest-ranking habitats continues the declining trend. However, subsequent migrations of Algic and Athabaskan groups diverge from this patterning, suggesting that a different process was driving these last migration events.

**Prediction 2: Most Suitable Habitats Have the Highest Population Densities.** To test this prediction, we examine population density as a function of mean NPP for each ethnographic group. As shown in Fig. 3, there is a highly significant relationship between the two variables: as environmental productivity increases, population density increases at an exponential rate (linear: \( r^2 = 0.62, p < 0.0001 \); exponential: \( r^2 = 0.63, p < 0.0001 \)). In agreement with the IFD predictions, habitats with higher suitability support larger populations per unit area. Coupled with the chronological results, this suggests that the earliest migrants into a region are likely to grow to higher densities, thereby reducing the probability of replacement from subsequent migrations. However, again, Algic and Athabaskan groups are an anomaly: although they are the latest arrivals in the region, they still exhibit some of the highest population densities.

**Discussion**

**Implications for California Prehistory.** The first people to settle California, the ancestors of Chumash and Yukian speakers, occupied some of the most productive habitats, likely traveling to the region along the coast via boats (32). These findings provide ancillary support for a coastal colonization model for the Americas showing that the earliest migrants into the region settled along highly
productive shorelines with adjacent productive terrestrial resource patches. Geographic hot spots like estuaries and river mouths probably acted as draws for people where marine and terrestrial resources were concentrated (33). Although archaeological support for such patterning is greater in the southern portion of the state (32), the limited support for Early Holocene dates along the northwestern coast may be due to biased erasure of the record resulting from sea level rise at a steeper gradient along the interface between land and sea (34).

In the second wave of migrations, predecessors of the speakers of Hokan languages came to occupy less productive habitats, but in contiguous territory covering much of the region (2, 6). After their initial movement into these moderately productive areas, populations likely outgrew their resource base, causing individuals to spread into adjacent lands. Such in-filling of un- or under-occupied lands helps to explain the putative expansion of this second wave of migrants into moderately productive contiguous territory across the region.

This contiguous territory occupied by Hokan-speaking ancestors was later fragmented by migrations of Penutian (Yok-Utian, Wintuan, and Maiduan) and perhaps Takic-speaking ancestors. These groups occupied habitats that did not differ significantly from one another or from the lands inhabited by Hokan speakers until the time of contact. These replacement and fragmentation events were a potential consequence of environmental degradation in the Great Basin, which lowered landscape productivity and pushed populations westward (6, 35).

In the Late Holocene, Numic-speaking peoples spread throughout southeastern California and across the rest of the Great Basin. This expansion did not spill into adjacent highly productive areas, suggesting that these locations were already occupied at high densities by earlier migrants. Instead, these migrations filled in regions of low productivity, replacing previous populations that occupied the region at low densities. The previous occupants were likely out-competed by the intensive foraging practices of these migrants (11), a trend perhaps common to many of the Late Holocene migrations.

The subsequent Late Holocene migrations of Algic- and Athabaskan-speaking peoples into northwestern California departs from our predictions: these groups took over highly productive habitats that were initially settled by some of the earliest colonizers. This departure from IFD predictions likely signals a shift to IDD dynamics, which could have been driven by the intensive fishing practices previously adopted by these incoming populations who were focused on acquiring and storing anadromous fish (salmon) (36, 37). Such a subsistence focus coupled with more sedentary adaptations and notions of ownership over productive resource patches would have provided a competitive advantage for the incoming populations, who would have been able to remain in place throughout the year, thereby excluding the previous populations who were more seasonally mobile (2, 5, 37, 38). Furthermore, these migrations also likely brought the bow and arrow to the region (38), a technology that potentially provided greater foraging returns and possibly directly aided in population displacement (39, 40).

A General Framework. The dynamics embedded within IFD and IDD models provide a framework to explain why global patterns reveal a correlation between ecological and linguistic diversity. In ecologically homogenous regions where habitat suitability does not vary extensively across space, populations are likely to be more evenly distributed across the landscape. With roughly equitable population densities, colonization attempts by incoming migrants are likely either to fail outright and be completely repelled or to succeed and result in the complete replacement of the resident population (11).

In ecologically diverse regions with alternating habitats of varying suitability, migration events by mobile hunter–gatherers should be repelled by large populations that occupy highly suitable habitats, but should be more likely to succeed in replacing low-density populations that reside in low-suitability habitats. This will result in the fragmentation of a population whose occupations span habitats of varying suitability. As a result, regions characterized by greater environmental diversity will come to exhibit greater linguistic diversity through repeated migration events that end in partial replacements.

We expect these dynamics to hold until populations adopt exclusionary tactics that restrict individuals from occupying more suitable habitats. At that point, all of these rules would be swept aside. Whether these tactics are adopted by resident or migrant populations will result in very different outcomes. As such, this would represent a potential bifurcation point in historical trajectories. If resident populations are the first to adopt such practices, then they would be more likely to stay in place. However, if
incoming populations bring such practices with them, then they would have a competitive advantage and could replace in situ populations. This may result from factors as simple as increased sedentism around predictable, dense resources, thereby excluding more mobile foragers during their seasonal rounds or as complex as the in-migration of intensive agricultural groups (10).

**Conclusion**

This approach provides simple predictions about human population movements without relying on complex models or assumptions. Following the IFD and IDD, we suggest that foraging individuals will tend to distribute themselves across landscapes in ways that provide the greatest benefit at minimal cost. These simple predictions provide a framework to explain the emergence of a complex mosaic of ethnolinguistic groups not found elsewhere in North America. Intergroup dynamics that include practices of exogamy, networks of exchange, and episodes of violence complicate this picture. Stochastic environmental shocks, which may have helped initiate many of these movements, also restructure habitat suitability in significant ways (35, 41). However, overall, broad patterning in migration aggregations appears to meet the predictions from this simple model.

As an initial test of this hypothesis, this work outlines broad patterning in the prehistory of western North America, including an explanation of spatial patterning in the colonization of the continent. Given patterns in continental NPP, our findings highlight the potential of coastal habitats (42), which itself lends support for a coastal corridor as one of the first entry routes into the Americas (32, 43). Applied elsewhere, this approach may aid in the explanation of prehistoric hunter–gatherer migrations across the globe, including the initial spread of people out of Africa into Europe, Asia, and across to Sahul (Australia/New Guinea) (17, 44, 45). Although many of these linguistic records have been erased by the migrations of agricultural peoples (10), archaeological patterning coupled with estimates of environmental productivity could eventually provide a global test of our hypothesis and help elucidate why and how humans spread across the planet, creating a patchwork of linguistic and ethnic diversity.

**Materials and Methods**

**Environmental Productivity.** As a proxy for habitat suitability, we relied on terrestrial NPP. NPP is a measure of the initial step in the carbon cycle where energy is turned into mass; it is frequently used to approximate plant growth. Remote sensing data used to calculate NPP came from the Moderate Resolution Imaging Spectroradiometer (MODIS) collected from NASA’s Terra satellite. MODIS data processing followed the MOD17 Photynthesis and Net Primary Productivity algorithm were made available by the Numerical Terradynamics Simulation Group at the University of Montana (46). The raster image consists of average NPP calculated from 2000 to 2011 in 1-km resolution. Inland waters and urban areas were excluded and appear white on the map (Fig. 1). A map that estimates the distribution of California’s ethnographic groups at contact (2, 6) (Fig. S1) was used to estimate mean NPP values for each linguistic group with the Zonal Statistics tool in ArcGIS 10.47. The magnitude of NPP values was then reduced by four orders to aid interpretation.

**Chronology.** A rank-ordered chronology for the timing of group migrations into the region was developed by combining linguistics (19), genetic (22, 23) estimates. Chumashan and Yukian appear to represent the oldest linguistic stratum in the region (4–6). This broadly corresponds with genetic findings, with an estimated age of the clade around 7,353 (13, 233–3, 333) years ago (22). Hokan languages are estimated to be the second oldest group in California with linguistic diversity, suggesting an arrival date between 8,000 and 6,000 y ago. Hokan languages were likely fragmented by the first Penutian intrusion which brought Yok-Utian languages from California to the Great Basin between 5,000 and 3,000 y ago (6, 24). Linguistic evidence suggests that the Takic branch of Uto-Aztecan languages expanded sometime about 3,500 y ago (28); genetic evidence comparing ancient and modern mtDNA confirms this patterning (25). Linguistic analysis on the other Uto-Aztecan branch represented in the region places the divergence of Numic languages at about 2,000 (27) or 1,000 y ago (6). Based on a comparison of ancient mtDNA from burials recovered at Pyramid Lakes and Stillwater Marsh to modern mtDNA, this population replacement in the eastern Great Basin is thought to have occurred just short of 1,000 y ago (24). Tubatulabal were left out of this analysis due to their complicated and debated origin and migration estimates. Wintuan and Maiduan groups were probably pushed south based on the expansion of Algic and Athabaskan migrations in the Late Holocene; it is estimated that they settled into their historic territories by about 1,500 y ago (6). Their entrance into northern California was followed by the continued expansion of Algic and Athabaskan groups between 1,450 and 1,050 and between 1,250 and 750 y ago, respectively (6). A summary of these data along with rank order estimates is provided in Table 1.

**Population Estimates.** To approximate population densities at the time of contact, we drew on established estimates in the published literature (2, 3, 9, 10, 48). Where estimates for a group were not available, we used California’s population density (49).

Remote sensing data used to calculate NPP came from the Moderate Resolution Imaging Spectroradiometer (MODIS) collected from NASA’s Terra satellite. MODIS data processing followed the MOD17 Photosynthesis and Net Primary Productivity algorithm were made available by the Numerical Terradynamics Simulation Group at the University of Montana (46). The raster image consists of average NPP calculated from 2000 to 2011 in 1-km resolution. Inland waters and urban areas were excluded and appear white on the map (Fig. 1). A map that estimates the distribution of California’s ethnographic groups at contact (2, 6) (Fig. S1) was used to estimate mean NPP values for each linguistic group with the Zonal Statistics tool in ArcGIS 10. (47). The magnitude of NPP values was then reduced by four orders to aid interpretation.

**Chronology.** A rank-ordered chronology for the timing of group migrations into the region was developed by combining linguistics (19), genetic (22, 23) estimates. Chumashan and Yukian appear to represent the oldest linguistic stratum in the region (4–6). This broadly corresponds with genetic findings, with an estimated age of the clade around 7,353 (13, 233–3, 333) years ago (22). Hokan languages are estimated to be the second oldest group in California with linguistic diversity, suggesting an arrival date between 8,000 and 6,000 y ago. Hokan languages were likely fragmented by the first Penutian intrusion which brought Yok-Utian languages from California to the Great Basin between 5,000 and 3,000 y ago (6, 24). Linguistic evidence suggests that the Takic branch of Uto-Aztecan languages expanded sometime about 3,500 y ago (28); genetic evidence comparing ancient and modern mtDNA confirms this patterning (25). Linguistic analysis on the other Uto-Aztecan branch represented in the region places the divergence of Numic languages at about 2,000 (27) or 1,000 y ago (6). Based on a comparison of ancient mtDNA from burials recovered at Pyramid Lakes and Stillwater Marsh to modern mtDNA, this population replacement in the eastern Great Basin is thought to have occurred just short of 1,000 y ago (24). Tubatulabal were left out of this analysis due to their complicated and debated origin and migration estimates. Wintuan and Maiduan groups were probably pushed south based on the expansion of Algic and Athabaskan migrations in the Late Holocene; it is estimated that they settled into their historic territories by about 1,500 y ago (6). Their entrance into northern California was followed by the continued expansion of Algic and Athabaskan groups between 1,450 and 1,050 and between 1,250 and 750 y ago, respectively (6). A summary of these data along with rank order estimates is provided in Table 1.

**Population Estimates.** To approximate population densities at the time of contact, we drew on established estimates in the published literature (2, 3, 9, 10, 48). Where estimates for a group were not available, we used California’s population density (49).

**Analytical Methods.** To determine if patterning in mean NPP across ethnolinguistic territories was biased by nonrandom neighboring relationships, we used the Spatial Autocorrelation (Morans I) function in ArcMap 10 (47). To examine the relationship between rank-order migration and NPP values, we used the linear model function in R to run a series of paired ANOVAs to test for significant departures in mean NPP values between each ethnographic group; linear models were also used to examine the effect of productivity on population density (49).

**ACKNOWLEDGMENTS.** We thank the Running Laboratory at the University of Montana for generating and making available the NPP database, and V. Golla, the California Department of Water Resources, R. W. Goode, and J. D. Aldern for estimates of ethnolinguistic distributions in California. This paper benefited from conversations with R. L. Bettinger, J. Coltrain, B. Hildebrandt, D. J. Kennett, F. K. Lake, D. O’Rourke, S. Simms, N. E. Stevens, and S. Tushingham. J. F. O’Connell and two anonymous reviewers provided extensive comments on earlier versions of this manuscript, which greatly improved the final result. The complete dataset is available in Table S1.

### Table 2. Summary of paired models comparing mean NPP between each ethnolinguistic group

<table>
<thead>
<tr>
<th>Ethnolinguistic Group</th>
<th>Hokan</th>
<th>Numic</th>
<th>Algic</th>
<th>Athabaskan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hokan</td>
<td>3.06</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wintuan/Yukian Hokan</td>
<td>0.37</td>
<td>6.86</td>
<td>12.09</td>
<td>3.05</td>
</tr>
<tr>
<td>Wintuan/Wokan Takic</td>
<td>0.03</td>
<td>0.81</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Wintuan/Maiduan Numic</td>
<td>0.10</td>
<td>0.13</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Numic</td>
<td>97.05</td>
<td>6.77</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Algic</td>
<td>2.43</td>
<td>4.78</td>
<td>23.01</td>
<td>23.01</td>
</tr>
<tr>
<td>Athabaskan</td>
<td>20.53</td>
<td>33.88</td>
<td>185.69</td>
<td>194.74</td>
</tr>
</tbody>
</table>

F-statistic values resulting from paired ANOVAs are listed. Significance at <0.1, <0.05, <0.01, and <0.0001 are available in Table S2.
