Genetic relationships of prehistoric Baja California populations

Cara Monroe and Brian M. Kemp
Washington State University, Pullman

Archaeological, ethnographic and linguistic evidence from Baja California suggest a complex culture history defined by dynamic regional interaction and migration as well as a considerable degree of peninsular isolation. In particular, the extent of the biological similarities between Yuman, Cochimí, Guaycura, and Pericú speakers is still unresolved. This study addresses the biological relationships of prehistoric Baja populations through the study of ancient DNA. Mitochondrial DNA (mtDNA) was extracted from 12 individuals from six archaeological sites that span the Guaycura and Pericú culture areas. Here we present preliminary results of this study and compare the results to modern mtDNA from modern Yuman speakers as well as other linguistic groups from California, the American Southwest, and Sonora, Mexico.

An outstanding issue in Baja prehistory is the degree of cultural and biological interaction that has occurred from the initial entrance of humans into the area to the time of European contact, and beyond. Theories about native Baja populations range from extreme isolation among indigenous groups, especially in Baja Sur, to a certain amount of cultural and biological diffusion. Reviewing the linguistic, biological and cultural evidence will help to set up predictive and testable hypotheses of whether indigenous groups in this region interacted, and to what degree, throughout prehistory.

Linguistically, the Baja area is divided into four major language groups: Yuman, Cochimí, Guaycura, and Pericú. While a clear linguistic connection exists between Cochimí-speaking groups of central Baja with their Yuman neighbors to the north, the nature of the relationship of these languages to Guaycura and Pericú (languages in southern Baja) is less clear. Moreover, while the relationship between Guaycurá and Pericú is equally unclear, with most ethnohistoric data suggesting that the languages originated from different language families (Mixco 2006), archaeological evidence suggests that these languages have long been in the region, dating to as early as the Late Paleolithic/Early Holocene (Laylander 2006). The split of the Cochimí/Yuman dialect, along with a subsequent language expansion (Mixco 2006), could have placed a population pressure on the Pericú and Guaycura groups, precipitating a migration or a concentration of populations into the southern portion of Baja. Indeed, this suggestion has also been used to explain the presence of the Seri along the coast of Sonora, as a result of crossing the Sea of Cortez (Hyland 1997).

While evidence indicates a great antiquity to the linguistic distance between Cochimí/Yuman and Pericú and Guaycura, this does not preclude substantial social and/or biological interaction. Language as a barrier to gene flow and social borrowing has been a predominate paradigm for years. However, recent genetic work in the American Southwest has shown that this is not always the case. Yuman, Athapaskan and Uto-Aztecan speakers as well as language isolates such as the Zuni share mitochondrial DNA types while still retaining their linguistic identity (Kemp 2006; Monroe et al. 2006a). Based on this information, one could predict...
that the Pericú, Guaycura, and Cochimí-Yuman look genetically similar to each other despite their linguistic divisions, and that such relatedness was created and reinforced by the geographic isolation of the peninsula in addition to the high mobility of indigenous groups. Alternatively, the documented territoriality and warfare among Pericú and Guaycura populations as well as their more geographically isolated position (Mathes 2006) may have resulted in genetic drift and isolation. A third, albeit less likely, scenario is a genetic relationship between Cochimí, Pericú or Guaycura and the Seri. Seri is a linguistic isolate, but little work has been done to show a possible linguistic relationship with the Guaycura or Pericú. DNA research on the Seri reveals extreme genetic isolation in both mtDNA and Y-chromosome markers and shows no genetic ties to Uto-Aztecan or Yuman groups in Sonora, the American Southwest or California. It may be possible, if the Seri did indeed migrate from Baja, that they share a genetic relationship with groups from the peninsula.

Research on skeletal remains from the cape region’s Las Palmas culture identified prevalence for dolichocephalic (or unusually long and narrow) cranial vaults. Investigation from the past decade has also suggested that these remains most resemble American Paleo-Indian remains, particularly from South America and that their presence in Baja is the result of an early Paleolithic occupation followed by genetic isolation (González-José et al. 2003). If the peoples of the Las Palmas culture represent members of an earlier and/or a separate migration into the Americas (e.g., as proposed by Perego et al. 2009) than that which gave rise to the vast majority of Native Americas (Tamm et al. 2007), the Pericú may show close genetic (DNA) level to other populations displaying such morphological cranial characteristics. Interestingly, such burials have been discovered from the La Jolla culture complex (Tyson 1980) as well as from the southern Channel Islands in California (Chumash and Uto-Aztecan territory) (Potter 2005; Walker 2006).

A DNA investigation of living Chumash descendants has also revealed a genetic relationship to a rare maternal lineage, now called D4h3a (Perego et al. 2009), dating back at least to the early Holocene in Alaska. The distribution of this haplotype is coastal and represents a founder lineage from the initial peopling of the Americas (Johnson and Lorenz 2006; Kemp et al. 2007). Based on the craniometric and genetic evidence, a hypothesis could then be made that the Pericú share genetic lineages (or mitochondrial haplotypes) with indigenous groups such as the Chumash.

Based on the discussion above, four distinct hypotheses could be made for the genetic relationship of indigenous groups in Baja California:

1. The Guaycura, Pericú and Cochimí are genetically similar to each other and resemble Yuman speaking groups.
2. The Guaycura, Pericú and Cochimí (particularly the first two groups) are genetically similar to the Seri in Sonora, Mexico.
3. The Pericú, or skeletal remains displaying dolichocephalic characteristics, share a genetic relationship to populations in southern California such as the Chumash, where burials with long and narrow skulls have been found as well as genetic types that are extremely rare and represent early mtDNA occupation in the Americas.
4. Genetic drift has been the dominating evolutionary force on the gene pools of peninsular groups, having led to great genetic distances between them.

To test these predictions, mtDNA was extracted from 12 individuals from six archaeological sites that span the Guaycura and Pericú culture areas (Table 1). The samples were obtained by Dr. Silvia Gonzales with permission from the Instituto Nacional de Antropología e Historia (INAH) in Baja California Sur.
While many may already be familiar with mtDNA research and its usefulness in addressing population history, a quick overview is provided, to describe how researchers use mtDNA and to discuss some terminology and applications from North America as well as to review previous mtDNA studies of Yuman and southern Californian populations.

Mitochondrial DNA is a small (16,569 bp) (bp = base pairs) extra-nuclear genome found in the mitochondria. Several distinct properties make this type of DNA unique for addressing population history. First, it is maternally inherited and does not undergo recombination, which allows researchers to easily track direct female lineages back in time (Cann et al. 1987). Second, there are approximately 700-1,000 mitochondria per cell versus one nucleus. This makes mtDNA especially useful in ancient DNA studies, as mtDNA -- due to its high copy number -- will preferentially stand the test of time and degradation (Kaestle and Horsburgh 2002). The non-coding hypervariable region is the most often targeted portion of the mtDNA genome. This is primarily due to the fact that this particular region does not code for any genes, and mutations occur in this region randomly rather than via natural selection. As such, this portion of the mtDNA tends to contain variation that is informative at the regional and sometimes tribal level. Mutations that are found outside this region are more conserved and when found are used to designate larger groups (haplogroups or macrogroups) rather than specific maternal lineages (Horai et al. 1993).

Haplogroup and haplotype are two terms commonly misunderstood by archaeologists, and as such the terms are discussed in detail.: Haplogroups are groups of haplotypes which exhibit mutations in common, denoting a common origin at sometime in the past. The nomenclature is often confusing, but haplogroups are usually labeled as a letter. For example, haplogroups A, B, C, D and X are Native American haplogroups. Haplotypes are specific lineages within a haplogroup, identified by sequence variation beyond that defining the haplogroup. Mostly, haplotypes in Native American populations have been identified by hypervariable region sequences. Since there is a very high rate of mutation rate in this region of the genome, it is useful in studying more recent population prehistory since there has been sufficient time for mutations to accumulate, and therefore it is possible to distinguish between populations (Horai and Hayasaka 1990). Haplogroup data for populations are often informative in recognizing regional patterns, while haplotype data gives a finer resolution to the interaction of particular maternal lineages.

As an example of how haplogroup data are informative, the haplogroup frequencies of the major regions of North America are shown in Figure 1. Most geneticists agree that there was a
single migration into the Americas, and the haplogroup frequencies in Figure 1 show the results of regional population history data, compiled from published and presented data, through subsequent millennia of genetic drift and gene flow. In many ways it is akin to patterns of linguistic or cultural variation that we see across North America. There are some quite obvious patterns of regional haplogroup diversity. For example, we see that populations in the Arctic and Subarctic are almost entirely represented by haplogroups A and D. However, if we further subdivide this area, we would find that Northern Athapaskan groups are almost entirely made up of haplogroup A, while Aleut populations have high frequencies of haplogroup A and haplogroup D. Similarly, we see that haplogroup X is almost entirely found among Algonquian speakers and very rarely outside the northeast or Great Lakes region. In the greater Southwest we see that populations have a large amount of haplogroups B and C, with a moderate amount of haplogroup A (which through haplotypic data is determine to be admixture from Athapaskans). Finally, in Mesoamerica, we see high frequencies of haplogroup A. In these examples, we can see how haplogroup information can be informative at a regional level.

When the haplogroup data are subdivided according to populations, there is an increase in specificity which gives insight into local separation and continuity when looking at the southwest coast and southwestern cultural area of North America (Figure 2). Some general trends are
Figure 2. Haplogroup distribution among specific groups in Western North America. Data were collected from both published sources and investigations by the authors. Yuman data from the Havasupai, Yavapai, Kumeyaay/Cochimi, Kumeyaay (Desert), Paipai, Kiliwa, Cocopa and Quechan were collected by Monroe in 2006. The Hualapai data were collected by Kemp et al. in 2006. The Ipai/Tipai, Luiseño, Gabrieleno, Serrano, Tubatulabal, Chumash, Sierra Miwok and Salinan were collected by Johnson and Lorenz (2006). The ancient California Valley samples were collected by Eshleman (2002). The ancient Channel Island data were collected by Potter (2005). The Northwest Coast data were by Mahli et al. (2004), Torroni et al. (1993), Shields et al. (1993) and Ward et al. (1993). The Great Basin data were collected by Kaestle and Smith (2001) and Mahli et al (2004).
apparent. Yuman speakers predominately consist of haplogroups B and C. However, there is an increase in haplogroup B in southern Yuman populations such as the Kiliwa, Paipai, and Kumeyaay/Cochimi, and the desert Kumeyaay. Conversely in the coastal California populations there is a higher amount of haplogroups A and D. In the California valley and the Great Basin there is a higher amount of haplogroups B and D. Haplotypic data (or specific sequence lineages) of Haplogroup D in these Californian groups are clearly different. Great Basin and Californian groups to the east share maternal lineages with each other, while the Coastal Californian groups share a relationship with other coastal groups ranging from Alaska to Tierra del Fuego in South America (Kemp et al. 2007).

Taking into consideration the previous mtDNA studies discussed above, we could predict that the haplogroup and haplotype data from the Guaycura and Pericú could match either the Chumash (seen by a high amount of haplogroups A and D, with specific lineages or haplotypes that match the haplogroup D individual from Alaska) or Yuman-speaking groups in southern California and northern Baja who have a high amount of haplogroup B. Again, a corresponding match to haplotypic (or lineage) data could be expected with a relationship to Yuman groups.

Results of Study

Out of the 12 samples in the study (see Table 1), only one sample, Sample 8 (La Matancita Entiero Multi I-J) was confirmed to contain amplifiable mtDNA. Sporadic amplification was obtained from Sample 4 (Isla San Jose La Paz BCS Entiero 1), Sample 6 (La Matancita BCS Entiero Multi I-G), and Sample 10 (El Conchalito La Paz Entiero 39), but none of these samples could be determined to belong to one of the common Native American haplogroups. It is most likely that these are false positives having originated from contaminating DNA. The poor success rate (8.3%) may be due to having had access to only bone samples. Future studies of DNA preserved in teeth may be met with higher success, as teeth generally preserve DNA better than bone. The enamel provides a protective coating that hinders DNA degradation and DNA contamination (Nunnally 2005), and in the case of the Baja samples has the benefit of not being treated with red ochre which leaches into the bone and prevents DNA amplification.

Preliminary results show that Sample 8 from La Matancita belongs to haplogroup B, but more specific haplotypic data have not been collected. While sample size here severely limits the interpretative value, it is notable that this individual belongs to a haplogroup found in higher frequency among Yuman speakers and does not belong to haplogroup A or D found along the California coast. However, it should be noted that ancient DNA haplogroup data collected from the southern Channel Islands does have a notable amount of haplogroup B among Uto-Aztecan speakers.

---

1 The Kumeyaay/Cochimi data were from blood samples collected in the late 1970 and early 1980s from La Huerta and San Antonio Necua. Individuals from this region self-reported themselves with the term Cochimi

2 The majorities of haplogroup B in the desert Kumeyaay from sequence (or haplotypic data) are related to American Southwest populations and are part of a Southwest population expansion related to the intensification of agriculture and possible cross-cultural interaction around Lake Cahuilla. The haplotype data from the other Kumeyaay-Cochimi, Kiliwa, and Paipai are not related to the Southwest and appear to be Yuman-specific.

3 To date, this particular individual has not been radiocarbon dated; however, isotope analysis is currently underway by Dr. Silvia Gonzales and colleagues from Liverpool John Moores University in England.
Conclusion

No concrete conclusions can be drawn, as the hypotheses proposed above still need to be tested and cannot be answered with the one datum point collected here. Perhaps the individual from La Matancita is more closely related to Yuman-speaking groups to the north, or conversely this individual may be more closely related to the ancient samples from the Channel Islands.

Future work will elucidate these questions and will include additional samples from the Massey collection, which are being collected from the Smithsonian. It is also our future goal to receive permits to sample teeth from skeletal material currently being held in INAH repositories in Baja California and Baja California Sur. Finally, haplotype sequence data will be collected from the previously analyzed skeletal material from the Channel Islands as well as burials from the Chumash region that have never been studied.

Acknowledgments

We would like to wish great thanks to Julia Bendímez Patterson and Elinora Topete for organizing the volume, David Glenn Smith from UC Davis for providing modern DNA samples and useful advice, and most especially Dr. Silvia Gonzales and Nicholas Beer from Liverpool John Moores University and the Centro INAH Baja California Sur for providing the ancient skeletal samples. And most importantly, Elinora Topete for organizing this event and her patience with my tardiness.

References cited

Bolnick, Deborah A. and David Glenn Smith  

Bruce Budowle, Marc W. Allard, Constance L. Fisher, Alice R. Isenberg, Keith L. Monson, John E. Stewart, Mark R. Wilson and Kevin W. Miller  

Cann Rebecca L., Mark Stoneking and Allan C. Wilson  

Carlyle, Shawn W., Ryan L. Parr, M. Geoffrey Hayes and Dennis H. O’Rourke  

De La Cruz, Isabel, Angélica González-Oliver, Brian M. Kemp, Juan A. Román, David Glenn Smith and Alfonso Torre-Blanco  

Eshleman, Jason Aaron  
2002 *Mitochondrial DNA and prehistoric population movements in western North America*, dissertation, University of California, Davis.
González-José, Rolando, Antonio González-Martín, Miquel Hernández, Héctor M. Pucciarelli, Marina Sardi, Alfonso Rosales and Silvina van der Molen

González-Oliver, Angélica, Lourdes Márquez-Morfin, José C. Jiménez and Alfonso Torre-Blanco

Green L.D., James N. Derr and Alec Knight

Horai, Satoshi and Kenji Hayasaka

Horai, Satoshi, Rumi Kondo, Yuko Nakagawa-Hattori, Seiji Hayashi, Shunro Sonoda and Kazuo Tajima

Huoponen, Kirsi, Antonio Torroni, P. R. Wickman, Daniele Sellitto, D. S. Gurley, Rosaria Scozzari and Douglas C. Wallace

Hyland Justin R.

Johnson, John R. and Joseph G. Lorenz

Kaestle, Frederika A. and K. Ann Horsburgh

Kaestle, Frederika A. and David Glenn Smith

Kemp, Brian M.

Kemp, Brian M., Ripan S. Malhi, John McDonough, Deborah A. Bolnick, Jason A. Eshleman, Olga Rickards, Cristina Martinez-Labarga, John R. Johnson, Joseph G. Lorenz, E. James Dixon, Terence E. Fifield, Timothy H. Heaton, Rosita Worl and David Glenn Smith

Kemp, Brian M., Andrés Reséndez, Juan Alberto Román Berrelleza, Ripan S. Malhi and David Glenn Smith
2005 “An analysis of ancient Aztec mtDNA from Tlatelolco: pre-Columbian relations and
the spread of Uto-Aztecan”, in Biomolecular archaeology: genetic approaches to the past, David M. Reed, ed., pp. 22-46, Southern Illinois University, Carbondale.

Laylander, Don

Lorenz, J. and David Glenn Smith

Malhi, Ripan S.
2001 Investigating prehistoric population movements in North America with ancient and modern DNA, dissertation, University of California, Davis.

Malhi, Ripan S., Katherine E. Breece, Beth A. Schultz Shook, Frederika A. Kaestle, James C. Chatters, Steven Hackenberger and David Glenn Smith

Malhi, Ripan S., Holly M. Mortensen, Jason A. Eshleman, Brian M. Kemp, Joseph G. Lorenz, Frederika A. Kaestle, John R. Johnson, Clara Gorodezky and David Glenn Smith

Mathes, W. Michael

Merriwether, D. Andrew and Robert E. Ferrell

Merriwether, D. Andrew, Francisco Rothhammer and Robert E. Ferrell

Mixco, Mauricio J.

Monroe, Cara, Brian M. Kemp and David Glenn Smith
2006a “Mitochondrial DNA evidence supports historic documentation of Yavapai-Apache interaction”, San Juan, Puerto Rico.

2006b “Mitochondrial DNA variation of Yuman speaking populations”, University of California, Santa Barbara.

Nunnally Brian K.
2005 Analytical techniques in DNA sequencing, Taylor & Francis, Boca Raton, Florida.

Parr, Ryan L., Shawn W. Carlyle and Dennis H. O'Rourke
Perego, Ugo A., Alessandro Achilli, Norman Angerhofer, Matteo Accetturo, Maria Pala, Anna Olivieri, Baharak Hooshiar Kashani, Kathleen H. Ritchie, Rosaria Scozzari, Qing-Peng Kong, Natalie M. Myres, Antonio Salas, Ornella Semino, Hans-Jürgen Bandelt, Scott R. Woodward and Antonio Torroni

Potter, Amiee B.

Rubicz, Rohina, Theodore G. Schurr, Paul L. Babb and Michael H. Crawford

1990 “Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages”, American Journal of Human Genetics 46(3):613-623.

Scozzari, Rosaria, Fulvio Cruciani, Piero Santolamazza, Daniele Sellitto, David E. C. Cole, Laurence A. Rubin, Damian Labuda, Elisabetta Marini, Valeria Succa, Giuseppe Vona and Antonio Torroni

Shields, Gerald F., Andrea M. Schmeichen, Barbara L. Frazier, Alan Redd, Mikhail I. Voevoda, Judy K. Reed and R. H. Ward

Smith, David Glenn, Ripan S. Malhi, Jason A. Eshleman, J. Lorenz and Frederika A. Kaestle

Stone, Anne C. and Mark Stoneking


Torroni, Antonio, Yu-Sheng Chen, Ornella Semino, Augusta Silvana Santachiara-Beneceretti, C. Ronald Scott, Marie T. Lott, Marcus Winter and Douglas C. Wallace
Torroni, Antonio, Theodore G. Schurr, Margaret F. Cabell, Michael D. Brown, James V. Neel, Merethe Larsen, David G. Smith, Carlos M. Vullo and Douglas C. Wallace

Torroni, Antonio, Theodore G. Schurr, Chi-Chuan Yang, Emőke J. E. Szathmary, Robert C. Williams, Moses S. Schanfield, Gary A. Troup, William C. Knowler, Dale N. Lawrence, Kenneth M. Weiss and Douglas C. Wallace
1992 “Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations”, Genetics 130(1):153-162.

Tyson, Rose A.

Walker, Philip

Ward, R. H., Barbara L. Frazier, Kerry Dew-Jager and Svante Pääbo

Ward, R. H., Alan Redd, Diana Valencia, Barbara L. Frazier and Svante Pääbo
1993 “Genetic and linguistic differentiation in the Americas”, Proceedings of the National Academy of Science USA 90(22):10663-10667.