npg

ORIGINAL ARTICLE

Complete mitogenome analysis of indigenous populations in Mexico: its relevance for the origin of Mesoamericans

Fuzuki Mizuno[1], Jun Gojobori[2], Li Wang[3], Keisuke Onishi[1], Saburo Sugiyama[4], Julio Granados[5], Celta Gomez-Trejo[6], Víctor Acuña-Alonzo[6] and Shintaroh Ueda[1]

Mesoamerica has an important role in the expansion of Paleoamericans as the route to South America. In this study, we determined complete mitogenome sequences of 113 unrelated individuals from two indigenous populations of Mesoamerica, Mazahua and Zapotec. All newly sequenced mitogenomes could be classified into haplogroups A2, B2, C1 and D1, but one sequence in Mazahua was D4h3a, a subclade of haplogroup D4. This haplogroup has been mostly found in South America along the Pacific coast. Haplogroup X2a was not found in either population. Genetic similarity obtained using phylogenetic tree construction and principal component analysis showed that these two populations are distantly related to each other. Actually, the Mazahua and the Zapotec shared no sequences (haplotypes) in common, while each also showed a number of unique subclades. Surprisingly, Zapotec formed a cluster with indigenous populations living in an area from central Mesoamerica to Central America. By contrast, the Mazahua formed a group with indigenous populations living in external areas, including southwestern North America and South America. This intriguing genetic relationship suggests the presence of two paleo-Mesoamerican groups, invoking a scenario in which one group had expanded into South America and the other resided in Mesoamerica.

INTRODUCTION

Mesoamerica, first defined by Kirchhoff,[1] is an area with wide linguistic and cultural variations, extending from central Mexico to Guatemala, Belize, El Salvador, and parts of Honduras, Nicaragua, and Costa Rica. It is popularly in agreement held that the ancestors of the present-day indigenous people of America entered the New World across the Bering Strait from Asia as founders. Although many studies have reported on the genetic structures of the indigenous American populations, genetic relationships among indigenous populations after dispersal into America remain unclear. Mitochondrial DNA (mtDNA) of the indigenous American people has been traced back to four major haplogroups and a fifth minor haplogroup. These haplogroups were initially named A, B, C, D and X[2–5] that are now termed as A2, B2, C1, D1 and X2a.[4,6] Complete mitochondrial genome (mitogenome) sequences have been published for indigenous American people,[7–11] allowing the identification of other different subhaplogroups as ancestral founding lineages.[12–16] At least fifteen haplogroups have been recognized as maternal-founding lineages of

the indigenous American populations.[17] Fine analyses at the level of entire mitogenomes, namely, analyses based on a greatly refined mtDNA phylogenetic tree, allow more comprehensive evaluation of the migration and timing of post-last glacial events.[12,17–21]

Previous studies of human mitogenome sequences have generally been carried out in a similar manner. mtDNAs are first screened for variants to define haplogroups, using either a limited number of samples from each haplogroup or many samples from a particular haplogroup of interest are selected, and then complete mitogenome sequencing is carried out for these selected samples. However, limitations of this approach, especially problems in demographic inferences about populations, such as population size changes, population divergence times and migration/admixture events, have been pointed out.[22–25] Thus, to examine the peopling of Mesoamerica and to facilitate the reconstruction of ancient dispersal of indigenous people through Mesoamerica from North to South America, we showed here complete mitogenome sequences at the highest level of molecular resolution for two indigenous Mesoamerican populations,

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan; [2]School of Advanced Sciences, The Graduate University for Advanced Studies, Hayama, Kanagawa, Japan; [3]School of Medicine, Hanzhou Normal University, Zhejiang, China; [4]Graduate School of International Cultural Studies, Aichi Prefectural University, Aichi, Japan; [5]Department of Transplants, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico and [6]Molecular Genetics Laboratory, National Institute of Anthropology and History, Mexico City, Mexico
Correspondence: Professor L Wang, School of Medicine, Hanzhou Normal University, Zhejiang 311121, China.
E-mail: liwang@hznu.edu.cn
or Professor S Ueda, Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.
E-mail:sueda@biol.s.u-tokyo.ac.jp
Received 3 September 2013; revised 10 April 2014; accepted 14 April 2014; published online 8 May 2014

Table 1 Haplogroup frequencies of 26 indigenous populations in Mesoamerica and its surrounding areas

Tribe	n	A	B	C	D	Others	References
Cora	72	0.306	0.514	0.139	0.042	0.000	Kemp et al.[42]
Huichol	62	0.306	0.532	0.161	0.000	0.000	Kemp et al.[42]
Mazahua	25	0.160	0.520	0.080	0.240	0.000	Present study
Mixe	52	0.308	0.288	0.288	0.115	0.000	Kemp et al.[42]
Mixtec	67	0.672	0.209	0.075	0.045	0.000	Kemp et al.[42]
Nahua Atocpan	59	0.475	0.356	0.119	0.051	0.000	Peñaloza-Espinosa et al.[44]
Nahua Chilacachapa	41	0.463	0.341	0.073	0.122	0.000	Peñaloza-Espinosa et al.[44]
Nahua Ixhuatlancillo	47	0.553	0.277	0.000	0.170	0.000	Peñaloza-Espinosa et al.[44]
Nahua Necoxtla	37	0.514	0.405	0.081	0.000	0.000	Peñaloza-Espinosa et al.[44]
Nahua Xochimilco	43	0.721	0.186	0.093	0.000	0.000	Peñaloza-Espinosa et al.[44]
Nahua Zitlala	46	0.652	0.304	0.022	0.022	0.000	Peñaloza-Espinosa et al.[44]
Otomi	35	0.600	0.200	0.114	0.086	0.000	Peñaloza-Espinosa et al.[44]
Purepecha	37	0.568	0.216	0.162	0.054	0.000	Peñaloza-Espinosa et al.[44]
Tzeltal	35	0.600	0.143	0.143	0.114	0.000	Peñaloza-Espinosa et al.[44]
Zapotec	88	0.670	0.182	0.114	0.034	0.000	Present study
Pima	30	0.067	0.500	0.433	0.000	0.000	Torroni et al.[54]
Tarahumara	73	0.342	0.288	0.315	0.055	0.000	Kemp et al.[42]
Embera	44	0.227	0.523	0.250	0.000	0.000	Kolman and Bermingham[39]
Kuna	63	0.714	0.286	0.000	0.000	0.000	Batista et al.[55]
Ngobe	46	0.674	0.326	0.000	0.000	0.000	Kolman et al.[56]
Wounan	31	0.290	0.194	0.484	0.032	0.000	Kolman and Bermingham[39]
Ingano	43	0.116	0.465	0.419	0.000	0.000	Torres et al.[57]
Piapoco	39	0.179	0.026	0.744	0.051	0.000	Torres et al.[57]
Ticuna	74	0.149	0.230	0.378	0.243	0.000	Torres et al.[57]
Wayuu	40	0.250	0.350	0.400	0.000	0.000	Torres et al.[57]
Zenu	36	0.222	0.417	0.306	0.056	0.000	Torres et al.[57]

Mazahua and Zapotec. To our knowledge, these are the first population-based data composed of complete mitogenome sequences for indigenous people in America, including Mesoamerica.

MATERIALS AND METHODS

DNA samples
DNAs were extracted from saliva or blood of unrelated adults. A total of 113 individuals were examined from two indigenous populations in Mexico: 25 Mazahua and 88 Zapotec. All the examined indigenous individuals and their ancestors (two generations) had been born in the same community and spoke their own native language. All the individuals were informed about the objectives of research and signed institutional review board-mandated consent forms. Approval for the present study was provided by National Institute of Anthropology and History, Mexico and the Ethics Committee, The University of Tokyo, Graduate School of Science.

Nucleotide sequence determination of entire mitogenome
We amplified the entire mitogenome as four fragments (each about 4 kb long) and determined their complete mtDNA sequences by PCR direct sequencing. The primers used are described in Supplementary Tables S1 and S2. PCR amplification was carried out in a 50-μl reaction mixture containing 1 mM Mg^{2+}, 200 μM each dNTP, 0.26 μM each primer and 1.25 units of PrimeSTAR GXL DNA polymerase (TAKARA). The conditions for the first-round PCR were as follows: an initial denaturation step for 1 min at 98 °C, followed by 30 cycles of denaturation for 10 s at 98 °C, annealing for 15 s at 60 °C and extension for 5 min at 68 °C, with a final extension for 10 min at 68 °C. The conditions for the second-round PCR were as follows: an initial denaturation step for 1 min at 98 °C, followed by 30 cycles of denaturation for 10 s at 98 °C, annealing for 15 s at 63 °C and extension for 5 min at 68 °C, with a final extension for 5 min at 68 °C. The amplified fragments were analyzed by electrophoresis on a 1% agarose gel and visualized by staining with ethidium bromide. PCR products were purified using a High Pure PCR Purification Kit

(Roche, Basel, Schweiz). For sequencing, we referred to the primers used by Kivisild et al.[9] for coding regions, while we additionally used other primers for control regions, as shown in Supplementary Table S2. Sequencing reactions were carried out with BigDye Terminator v3.1 (Applied Biosystems, Carlsbad, CA, USA). After excess dye terminators were removed by ethanol precipitation, DNA was dried and re-suspended in formamide. The dissolved DNA was heated for 2 min at 95 °C for denaturation, then immediately cooled on ice. Sequences were analyzed using a 3100 Genetic Analyzer capillary sequencer (Applied Biosystems) under the default settings. For singletons, we repeated PCR amplification and direct sequencing, and confirmed their nucleotide variants.

Phylogenetic and statistical analyses
Each mitogenome sequence was assigned to haplogroup, and its specific haplotype was defined based on mutations observed in both the control and coding regions. Because of hypervariability, C homopolymeric tract polymorphisms in regions 303–315 and 522–523 and variation at 16 519 were disregarded for tree reconstruction. C homopolymeric tract polymorphism in region 16 180–16 193 was also disregarded except for variation at 16 189 that was one of the defining mutations of haplogroup B. The haplogroup nomenclature is based on Phylotree Build 15 (http://www.phylotree.org/).[26]

In order to compare haplotype variation within each haplogroup belonging to indigenous American ancestry, we used published indigenous American haplogroups (A, B, C and D) for analysis.[11–13,17] We investigated the genetic relationships of Mazahua and Zapotec with other indigenous populations in Mesoamerica and its neighboring areas using mtDNA haplogroup frequency. There have been many studies of mtDNA variation in indigenous populations of Mexico and Central America, but we excluded populations that contain European and African maternal origin, such as haplogroups H and L, and populations of which data contain only control region sequences. We further selected only populations with sample size of ⩾25 individuals. We also added five indigenous populations from Colombia, located at the Northern end of South America. As a result, 26 indigenous populations consisting of 1265 individuals were used for phylogenetic analysis (Table 1 and Figure 1).
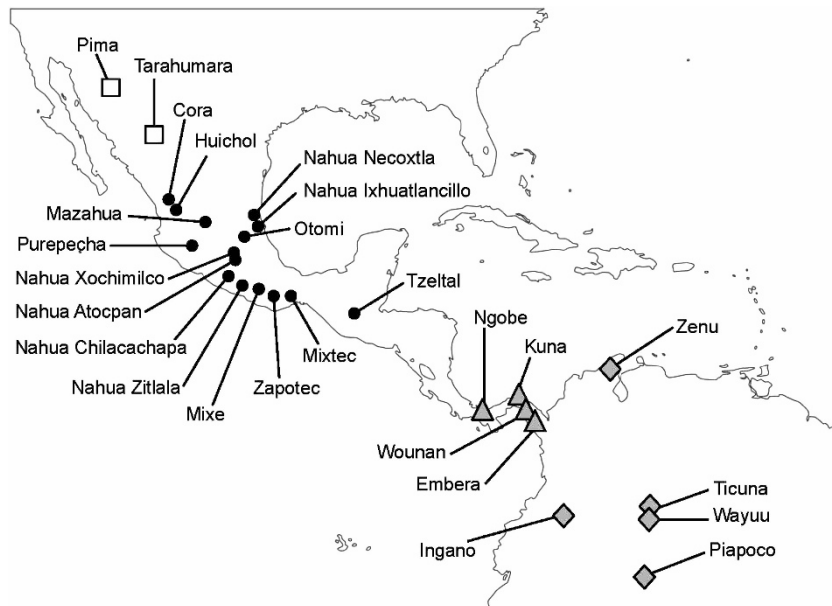
Figure 1 Geographical locations of the 26 indigenous American populations used for the present study. Dots: Mesoamerican populations, Open squares: southwestern North American populations, Shaded triangles: Central American populations and Shaded rhombuses: South American populations. The classification is based on the geographical locations of their present-day residences. According to Torroni et al.,[54] the Pima are descendants of the Hohokam who migrated north from northwestern Mexico around 300 B.C., although they are now living in southwestern Arizona.

We constructed neighbor-joining (NJ) tree using F_{ST} distances estimated from mtDNA haplogroup frequencies, using Arlequin.[27] We obtained a scattergram of the principal component analysis (PCA). Construction of NJ phylogenetic tree and PCA scattergram were carried out using MEGA4[28] and R Ver 2.11.1,[29] respectively. We also constructed NJ tree based on population pairwise F_{ST} distances of entire control region sequences (nucleotides from nps 16024 to 576 of the revised Cambridge Reference Sequence, rCRS[30]). Sequence data are available from the following indigenous populations in Mesoamerica and its surrounding areas: Huichol, Maya, Mayo, Nahua, Otomis, Pima and Tepehua, besides Mazahua and Zapotec (Supplementary Figure S1).

RESULTS

Complete mitogenome sequences belonging to haplogroups A, B, C and D

It is shown that more than 70% of haplogroup polymorphic diversity is located outside the control region, namely in the coding region.[31] Control region sequence data alone cannot provide sufficient evidence for founder status owing to a number of highly mutable sites.[32] It has also been pointed out that founder analysis should use not just coding region sequences but complete mitogenome sequences.[6] Sequence analysis using not only coding regions but also control regions of mitogenome makes it possible to do precise definitions as indigenous American-specific haplogroups. Namely, haplogroups A2j1, A2r, B2b, B2c, B2c1, B2c1a, C1b, C1b1, C1b9, C1b9a, C1c1 and C1c1b have been dissected into branches and sub-branches, most of which were distinguished by singletons.

Using the complete mitogenome sequences, all the Mazahua and Zapotec sequences were classified into the four major indigenous American haplogroups, A2, B2, C1 and D1, with the exception of one individual (Table 1). The variation of all mitogenome sequences is displayed in the phylogeny of each haplogroup (Figure 2). Sixty-three individuals belonged to haplogroup A2, among whom 45 different haplotypes were identified. Subhaplogroups A2d, A2g, A2p, A2v, A2j, A2m and A2r were found in Zapotec, while subhaplogroup A2o was found in Mazahua. The Mazahua and the Zapotec shared haplogroup

A2t, but their sub-branches were different from each other. Twenty-nine individuals belonged to haplogroup B2, among whom 22 different haplotypes were identified. About one-third of individuals belong to known B2 subhaplogroups, B2b, B2c1 and B2g1. Although six Mazahua and three Zapotec individuals belong to subclade of B2 classified by variations/transitions at np 16 278, most of the B2 subhaplogroups in our samples were so far undefined. Eight Zapotec individuals belong to subhaplogroup B2b, which is defined by variations/transitions at nt 6755. One individual each from Mazahua and Zapotec belongs to the B2c1 subhaplogroup, and the Mazahua was further classified into B2c1a. We found that one Mazahua individual was classified into subclade B2g1, which is defined by variation at nps 114, 3766, 6164, 1002 and 16 298. As for seven sequences of 12 individuals with haplogroup C, all the sequences were classified into C1 and furthermore into one of its subhaplogroups C1b and C1c. These two subhaplogroups are known to be widely distributed among indigenous people from North to South America.[12] The Mazahua and the Zapotec shared the C1b subhaplogroup, but their subclades within C1b were different from each other. Mazahua individuals were found in a subclade C1b9a with variations at nps 6297 and 8047, whereas Zapotec individuals were classified into subhaplogroup C1b1 defined by variation at np 11 147. Among four Zapotec individuals classified into subhaplogroup C1c, three individuals belong to subhaplogroup C1c1b, which is defined by np 215 and 5773. As for eight sequences of nine individuals with haplogroup D, all of them were classified into D1, except for one sequence of a single individual. The one exception belonged to D4h3a. Four Mazahua sequences were classified into subhaplogroup D1c that shared additional variation at nps 8674 and 15 805.

Haplogroup D4h3a in American people

Interestingly, one sequence from Mazahua was classified into D4h3a. D4h3a and D1 have a common ancestor, haplogroup D4. Phylogenetic tree of 44 complete mitochondrial genome haplotypes belonging to haplogroup D4h3a among Americans was shown in Figure 3.
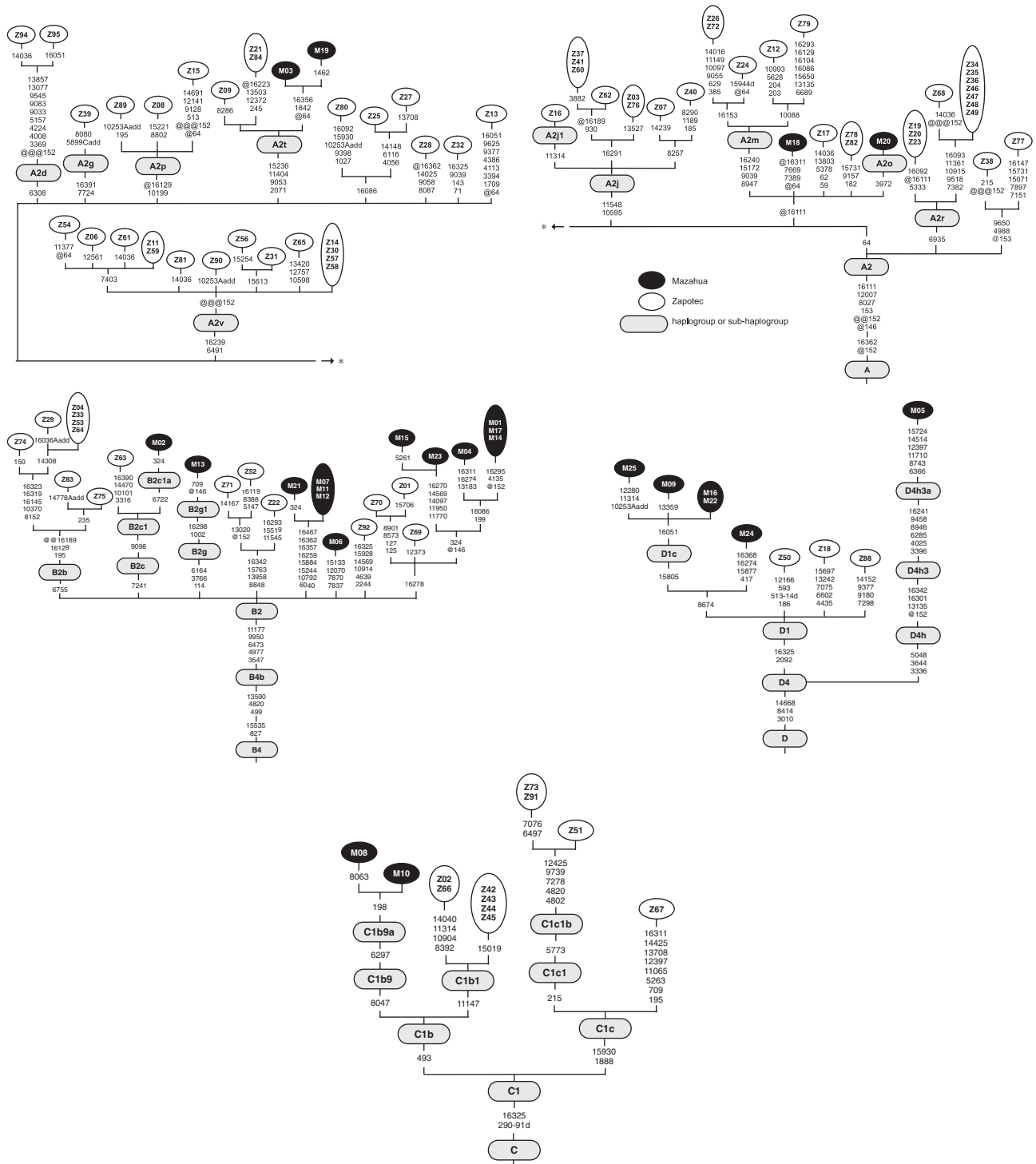
Figure 2 Phylogenetic tree of complete mitogenome sequences belonging to four major indigenous Americans haplogroup A, B, C and D among of the Mazahua and Zapotec. Ellipse of gray shows haplogroup or sub-haplogroup name. Black and white circle represents Mazahua and Zapotec, respectively. Alphabet and number inside are sample names. The nucleotide substitutions are listed relative to the revised Cambridge reference sequence, rCRS.[30] The suffixes (d) and (A, G, C and T add) indicate deletions or insertions, respectively. The prefix '@' indicates back mutation. C homopolymeric tract polymorphisms in regions 303–315 and 522–523 and variation at 16 519 were disregarded for tree reconstruction. C homopolymeric tract polymorphism in region 16 180–16 193 was also disregarded except for variation at 16 189 that is one of the defining mutations of haplogroup B. We followed van Oven's mtDNA tree Build 15.[26]

Two sequences that could be allocated to D4h3 have been found in Tarahumara and Nahua Ixhualtlancillo, indigenous people in Mexico,[33] and one D4h3a individual was also found in Mestizo of Ecuadol,[34] but their haplogroups were characterized as D4h3 or D4h3a by only control region sequences. Therefore, we could not compare our D4h3a with these sequences in detail. D4h3a3a is
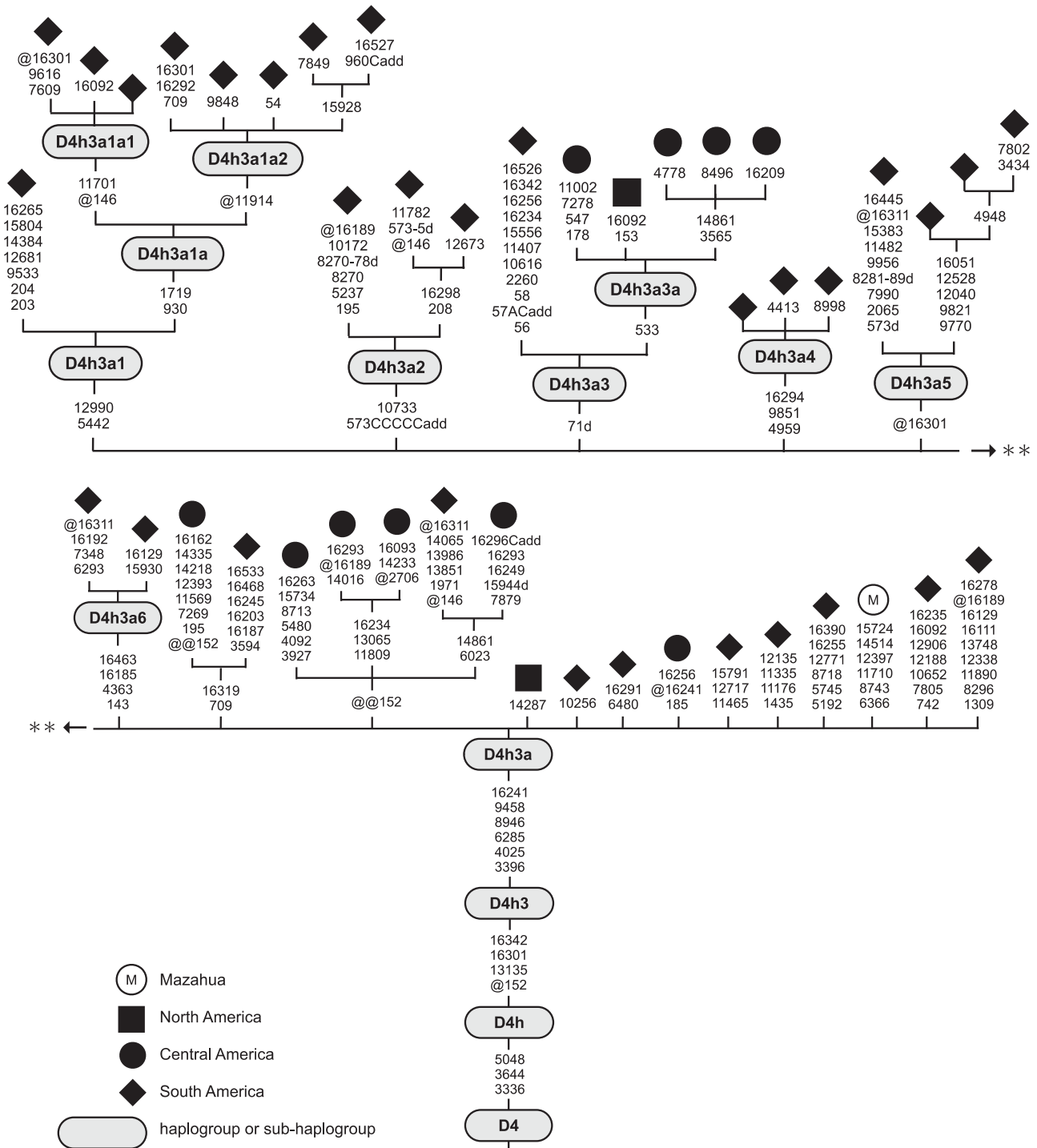
Figure 3 Phylogenetic tree of haplogroup D4h3a among Americans. The suffixes (d) and (A, G, C and T add) indicate deletions or insertions, respectively. The prefix '@' indicates back mutation. C homopolymeric tract polymorphisms in regions 303–315 and 522–523 and variation at 16 519 were disregarded for tree reconstruction. C homopolymeric tract polymorphism in region 16 180–16 193 was also disregarded except for variation at 16 189. M: Mazahua, Squares: North Americans, Circles: Central Americans, Rhombuses: South Americans. Mazahua D4h3a was not identical to any D4h3a so far reported.[11,12,14,21,35]

defined by a mutation at np 533 that is shared between California and Mexican populations.[21] Our D4h3a (M) was not identical to any subclade of D4h3a so far reported.[11,12,14,21,35] D4h3a found along the Pacific coast of the continent supports the Pacific coastal migration scenario previously proposed.[16,21] As a first population-based study of

complete mitogenome of indigenous Americans, we found an absence of shared sequences between Mazahua and Zapotec, showing a number of subclades unique to each indigenous population. This likely reflects that no or extremely low genetic exchange has occurred between Mazahua and Zapotec.
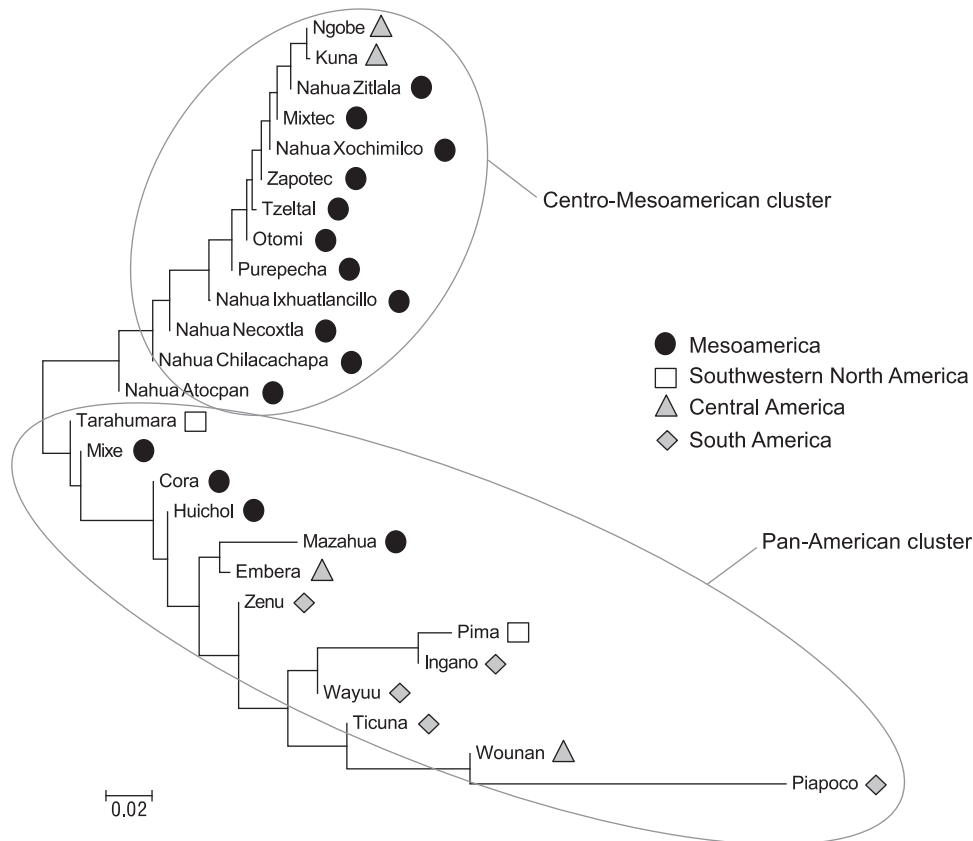
Figure 4 NJ tree of 26 indigenous populations. Designation is the same as that in Figure 1.

Relationships of Mazahua and Zapotec with other indigenous American populations

Next, we investigated the genetic relationships of Mazahua and Zapotec with other indigenous populations in Mesoamerica and its neighboring areas using mtDNA haplogroup frequency. Twenty-six indigenous populations consisting of 1265 individuals were used for analysis (Table 1 and Figure 1). Using mtDNA haplogroup frequencies, we obtained a NJ tree and a PCA scattergram. In the NJ tree, we found a large cluster including most of the indigenous Mesoamerican and Central American populations (Figure 4). The overall configuration of the 26 populations of the PCA scattergram was consistent with that of the NJ tree. Two-dimensional PCA scattergram explains 74.55% of the variation, with the first (X-axis) and the second coordinates (Y-axis) accounting for 45.37% and 29.18%, respectively (Figure 5). The cluster obtained is principally composed of populations that are located within a particular geographical area, namely, from central Mesoamerica to Central America, and we here named this cluster the Centro-Mesoamerican cluster. The out-of-Centro-Mesoamerican cluster populations, that is, Mazahua, Cora, Huichol, Embera and Wounan, appear to also form a cluster together with Pima and Tarahumara living in the northwestern external area of Mesoamerica (in other words, Southwestern North America) and Zenu, Ingano, Ticuna, Wayuu and Piapoco living in South America; we here tentatively name this cluster the Pan-American cluster.

DISCUSSION

Both the NJ tree and PCA scattergram show a large cluster including most of the indigenous Mesoamerican and Central American

populations (Centro-Mesoamerican cluster). The exceptions are the Mazahua, Cora, Huichol and Mixe of Mesoamerica, and the Embera and Wounan of Central America. According to Kirchhoff,[1] Mesoamerica is defined as neither a geographic region nor a socio-political unit, but rather, it is an area occupied by populations that share cultural characteristics. It is also considered that the northern limit overlapped with the southern frontier of southwestern North America at its peak.[36,37] The present-day Cora and Huichol geographically reside in Mesoamerica, but their attribution remains controversial. Suárez[37] linguistically classified them into the Mesoamerican group, whereas Beals[38] ethnologically classified them into the southwestern North American group. Kolman and Bermingham[39] mentioned in their original article that Embera and Wounan individuals examined that there are not admixture with other indigenous or non-indigenous polulations, and many of them are born in Colombia, reflecting the recent migration into Central America from South America.

It is now widely considered that (1) the Paleoamericans from Asia retreated into refuge areas during the Last Glacial Maximum, (2) recomposed their population genetic structure by genetic drift and gene flow from other newly arrived people, and (3) expanded into North America.[12,40,41] This resulted in the presence of certain mtDNA haplogroups. For example, haplogroup X2a shows a distribution restricted to northern and southwestern North America, whereas haplogroup D4h3a is detected at very low frequencies in indigenous populations from both North and South America, but is observed at locally high frequencies in restricted populations along the Pacific coast and on the western side of the Andes.[21] On the basis of the locally restricted distribution of these rare haplogroups, the presence
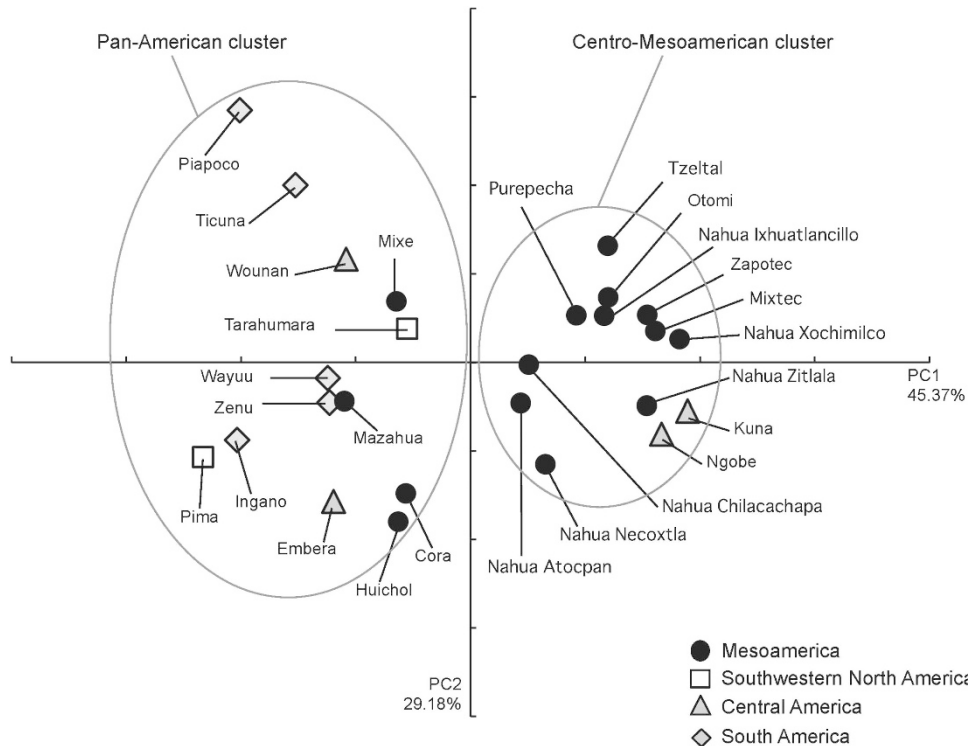
Figure 5 PCA scattergram of 26 indigenous populations. The two-dimensional scattergram explains 74.55% of the variation, with the first (*X*-axis) and the second coordinates (*Y*-axis) accounting for 45.37% and 29.18%, respectively. Designation is the same as that in Figure 1.

of two paths of migration from Beringia to North America is indicated. One is a path through the ice-free corridor between the Laurentide and Cordilleran ice sheets, and the other is a path along the Pacific coast. The overlapping sequence divergences among these two haplogroups also suggest almost concomitant spread into North America.[21] Recently, from the distribution patterns of subhaplogroups D1g and D1j, it has been proposed that the Pacific coast was the major entry and diffusion route for the early paleo-indigenous South Americans.[16] In this study, we found one sequence of D4h3a from Mazahua. D4h3a is reported in Tarahumara belonging to the Pan-American cluster, and is also found in indigenous populations belonging to the Centro-Mesoamerican cluster.[21,33] The fifth haplogroup X2a was not found, supporting previous observations that haplogroup X is present at a low frequency in North American indigenous people and absent in other American indigenous populations.[21,42,43] Haplogroup X, identified by np 1719 (loss of *Dde*I site at np 1715) but not control region sequence, was reported at frequencies of 20.0 and 1.9% for the indigenous populations of the Huichol and Tarahumara, respectively,[44] both of which belong to the Pan-American cluster. The variant at np 1719 has also been observed in other haplogroups besides haplogroup X. Recently, the same authors have shown contradictory results, namely, the absence of haplogroup X in both populations, by using np 14 470 (gain of *Acc*I site at np 14 465).[42] Accepting that, the latest results reflect the true haplogroup frequencies, haplogroup X is present neither in populations belonging to the Centro-Mesoamerican cluster nor those belonging to the Pan-American cluster in Mexico, Central and South America.[21,42,43] Therefore, both the Pan-American and the Centro-Mesoamerican clusters can be considered to be descendants of the people who had spread into America along the Pacific coast, but not through the ice-free corridor. Indigenous people in Southwestern North America, Pima and Tarahumara, bear greater genetic

resemblance to those in South America than to those in Mesoamerica and Central America, despite their geographical location. These findings indicate that the single-migration scenario cannot fully explain the current status of mtDNA haplogroup distribution in Mesoamerica, and also shows that only the Pan-American but not the Centro-Mesoamerican cluster populations succeeded in expanding into South America.

According to 364 470 SNPs data from 52 indigenous American populations, indigenous Americans descend from at least three streams of gene flow and effective population sizes in Mesoamerica have been relatively large since settlement of the region.[45] Their report also shows that the initial peopling followed a southward expansion was facilitated along the coast, in agreement with the Pacific coastal migration scenario by mtDNA haplogroup D4h3a. According to Y-chromosome diversity, Sandoval *et al.*[46] showed that Pima and Tarahumara indicated a clear differentiation with other indigenous Mesoamerican populations such as Mixtec, Otomi, Nahuas and Purepecha.[46] These paternal pattern is consistent with our present result.

Besides haplogroup frequencies, we also carried out pairwise analysis using nucleotide sequences of control region.[47] We constructed NJ tree based on pairwise F_{ST} distances of mtDNA control region (Supplementary Figure S2). Mazahua, Mayo and Pima were distinguished from a cluster including Zapotec, Maya, Tepehua, Nahua, Huichol, Otomis in Mesoamerica. It is consistent with the result obtained by using haplogroup frequencies that Mesoamerican populations are diverse.

The dual structure of indigenous Mesoamerican populations suggests two areas of paleo-indigenous people into Mesoamerica from North America. At present, the Centro-Mesoamerican people are dominant in Mexico, but the Pan-American people are supposed to have been dominant in the past. Unfortunately, it is not feasible to

deduce which of the Centro-Mesoamerican or Pan-American people are early immigrants from the data currently available. In either case, the Centro-Mesoamerican people should have experienced population expansion later, and maize domestication is likely to be a key factor behind this. The origin and timing of maize domestication continues to be debated, but recent archeological and genetic studies show that maize was domesticated in southwestern Mexico, an area of residence of the Centro-Mesoamerican people, during the early Holocene period and that maize cultivation was spread widely in Mesoamerica, probably in combination with the spread of humans.[48–52]

mtDNA remains the most widely studied and the best described locus in population genetics, despite its interpretation is limited to the maternal genetic history.[53] However, there have been no population data of complete mitogenome sequences either for Mesoamerica or for America. Here, we have applied complete mitogenome sequencing to two indigenous populations in Mesoamerica, Mazahua and Zapotec, and obtained the result that the indigenous people of Mesoamerica are divided into two clusters. The availability of population-based complete mitogenome sequences from other indigenous people should greatly refine the evolutionary scenario concerning the population history of Mesoamerica hypothesized here. In addition, population-based phylogeographic study will provide new insights into demographic inferences in addition to providing deep knowledge on the age of haplogroups and their geographic distribution at the subclade level. Furthermore, population-based ancient genome information enables us to reveal the details of the dual structure of Mesoamerican populations, including whether the Centro-Mesoamerican or the Pan-American people were the early immigrants into Mesoamerica.

CONFLICT OF INTEREST

The authors decalre no conflict of interest.

1 Kirchhoff, P. *Mesoamérica: sus límites geográficos composición étnica y caracteres culturales, Acta Americana* Vol 1, 92–107 (Inter-American Society of Anthropology and Geography, 1943).
2 Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M. et al. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* **53,** 563–590 (1993a).
3 Torroni, A., Sukernik, R. I., Schurr, T. G., Starikorskaya, Y. B., Cabell, M. F., Crawford, M. H. et al. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* **53,** 591–608 (1993b).
4 Forster, P., Harding, R., Torroni, A. & Bandelt, H. J. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.* **59,** 935–945 (1996).
5 Brown, M. D., Hosseini, S. H., Torroni, A., Bandelt, H. J., Allen, J. C., Schurr, T. G. et al. mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am. J. Hum. Genet.* **63,** 1852–1861 (1998).
6 Bandelt, H. J., Herrnstadt, C., Yao, Y. G., Kong, Q. P., Kivisild, T., Rengo, C. et al. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann. Hum. Genet.* **67,** 512–524 (2003).
7 Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
8 Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S. et al. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl Acad. Sci. USA* **100,** 171–176 (2003).
9 Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K. et al. The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172,** 373–387 (2006).
10 Fagundes, N. J., Kanitz, R., Eckert, R., Valls, A. C., Bogo, M. R., Salzano, F. M. et al. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* **82,** 583–592 (2008a).
11 Kumar, S., Bellis, C., Zlojutro, M., Melton, P. E., Blangero, J. & Curran, J. E. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol. Biol.* **11,** 293 (2012).

12 Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J. et al. Beringian standstill and spread of Native American founders. *PLoS ONE* **2,** e829 (2007).
13 Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q. P., Woodward, S. R. et al. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* **3,** e1764 (2008).
14 Hartmann, A., Thieme, M., Nanduri, L. K., Stempfl, T., Moehle, C., Kivisild, T. et al. Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* **30,** 115–122 (2008).
15 Malhi, R. S., Cybulski, J. S., Tito, R. Y., Johnson, J., Harry, H. & Dan, C. Brief communication: mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *Am. J. Phys. Anthropol.* **141,** 494–497 (2010).
16 Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A. W., Zimmermann, B. et al. Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res.* **22,** 811–820 (2012).
17 Perego, U. A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Hooshiar Kashani, B. et al. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res.* **20,** 1174–1179 (2010).
18 Fagundes, N. J., Kanitz, R. & Bonatto, S. L. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS ONE* **3,** e3157 (2008b).
19 Kitchen, A., Miyamoto, M. M. & Mulligan, C. J. A three-stage colonization model for the peopling of the Americas. *PLoS ONE* **3,** e1596 (2008).
20 Mulligan, C. J., Kitchen, A. & Miyamoto, M. M. Updated three-stage model for the peopling of the Americas. *PLoS ONE* **3,** e3199 (2008).
21 Perego, U. A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A. et al. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* **19,** 1–8 (2009).
22 Cox, M. P. Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum. Biol.* **80,** 335–357 (2009).
23 Nielsen, R. & Beaumont, M. A. Statistical inferences in phylogeography. *Mol. Ecol.* **18,** 1034–1047 (2009).
24 Gunnarsdóttir, E. D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res.* **21,** 1–11 (2011).
25 Schönberg, A., Theunert, C., Li, M., Stoneking, M. & Nasidze, I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur. J. Hum. Genet.* **19,** 988–994 (2011).
26 van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30,** E386–E394 (2008).
27 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1,** 47–50 (2005).
28 Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24,** 1569–1596 (2007).
29 R Development Core Team. R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org (2010)
30 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23,** 147 (1999).
31 Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C. et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* **4,** 1764 (2013).
32 Malhi, R. S., Eshleman, J. A., Greenberg, J. A., Weiss, D. A., Schultz Shook, B. A., Kaestle, F. A. et al. The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. *Am. J. Hum. Genet.* **70,** 905–919 (2002).
33 Sandoval, K., Buentello-Malo, L., Peñaloza-Espinosa, R., Avelino, H., Salas, A., Calafell, F. et al. Linguistic and maternal genetic diversity are not correlated in Native Mexicans. *Hum. Genet.* **126,** 521–531 (2009).
34 Baeta, M., Núñez, C., Sosa, C., Bolea, M., Casalod, Y., González-Andrade, F. et al. Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador. *Int. J. Legal. Med.* **126,** 299–302 (2012).
35 de Saint Pierre, M., Bravi, C. M., Motti, J. M., Fuku, N., Tanaka, M., Llop, E. et al. An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS ONE* **7,** e43486 (2012).
36 Kirchhoff, P. *Mesoamerica: Its Geographic Limits Ethnic Composition and Cultural Characteristics* (Heritage of Conquest ed. Tax, S.) 17–31 (The Free Press Publishers: Glencoe, IL, USA, 1952).
37 Suárez, J. A. *The Mesoamerican Indian Languages* 11–29 (Cambridge University Press: New York, NY, USA, 1983).
38 Beals, R. L. *Cultural relations between Northern Mexico and the Southwest United States: Ethnologically and archaeologically, The Mesoamerican Southwest: Readings in Archaeology Ethnohistory and Ethnology* (eds Hedrick, B. C., Kelley, J. C. & Riley, C. L.) 52–57 (Southern Illinois University Press: Carbondale, IL, USA, 1974).
39 Kolman, C. J. & Bermingham, E. Mitochondrial and nuclear DNA diversity in the Chocó and Chibcha Amerinds of Panamá. *Genetics* **147,** 1289–1302 (1997).
40 Gilbert, M. T., Kivisild, T., Grønnow, B., Andersen, P. K., Metspalu, E., Reidla, M. et al. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320,** 1787–1789 (2008).

41 Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757–762 (2010).

42 Kemp, B. M., González-Oliver, A., Malhi, R. S., Monroe, C., Schroeder, K. B., McDonough, J. *et al.* Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc. Natl Acad. Sci. USA* **107,** 6759–6764 (2010).

43 Dornelles, C. L., Bonatto, S. L., De Freitas, L. B. & Salzano, F. M. Is haplogroup X present in extant South American Indians? *Am. J. Phys. Anthropol.* **127,** 439–448 (2005).

44 Peñaloza-Espinosa, R. I., Arenas-Aranda, D., Cerda-Flores, R. M., Buentello-Malo, L., González-Valencia, G., Torres, J. *et al.* Characterization of mtDNA haplogroups in 14 Mexican indigenous populations. *Hum. Biol.* **79,** 313–320 (2007).

45 Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N. *et al.* A. Reconstructing Native American population history. *Nature* **488,** 370–374 (2012).

46 Sandoval, K., Moreno-Estrada, A., Mendizabal, I., Underhill, P. A., Lopez-Valenzuela, M., Peñaloza-Espinosa, R. *et al.* Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **148,** 395–405 (2012).

47 Gorostiza, A., Acunha-Alonzo, V., Regalado-Liu, L., Tirado, S., Granados, J., Sámano, D. *et al.* Reconstructing the history of Mesoamerican populations through the study of the mitochondrial DNA control region. *PLoS ONE* **7,** e44666 (2012).

48 Doebley, J. The genetics of maize evolution. *Annu. Rev. Genet.* **38,** 37–59 (2004).

49 Pohl, M. E., Piperno, D. R., Pope, K. O. & Jones, J. G. Microfossil evidence for pre-Columbian maize dispersals in the neotropics from San Andres Tabasco Mexico. *Proc. Natl Acad. Sci. USA* **104,** 6870–6875 (2007).

50 Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA* **106,** 5019–5024 (2009).

51 Ranere, A. J., Piperno, D. R., Holst, I., Dickau, R. & Iriarte, J. The cultural and chronological context of early Holocene maize and squash domestication in the Central Balsas River Valley Mexico. *Proc. Natl Acad. Sci. USA* **106,** 5014–5018 (2009).

52 van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., de Jesus Sanchez Gonzalez, J. *et al.* Genetic signals of origin spread and introgression in a large sample of maize landraces. *Proc. Natl Acad. Sci. USA* **108,** 1088–1092 (2011).

53 Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E. L., Silva, N. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90,** 675–684 (2012).

54 Torroni, A., Schurr, T. G., Yang, C. C., Szathmary, E. J., Williams, R. C., Schanfield, M. S. *et al.* Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130,** 153–162 (1992).

55 Batista, O., Kolman, C. J. & Bermingham, E. Mitochondrial DNA diversity in the Kuna Amerinds of Panamá. *Hum. Mol. Genet.* **4,** 921–929 (1995).

56 Kolman, C. J., Bermingham, E., Cooke, R., Ward, R. H. & Arias, T. D. Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. *Genetics* **140,** 275–283 (1995).

57 Torres, M. M., Bravi, C. M., Bortolini, M. C., Duque, C., Callegari-Jacques, S., Ortiz, D. *et al.* A revertant of the major founder Native American haplogroup C common in populations from northern South America. *Am. J. Hum. Biol.* **18,** 59–65 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)