1 **Short Communication for the Journal of Human Genetics**

2 **Title:**

3 mtDNA diversity of the Zapotec in Mexico suggests a population decline long before

4 the first contact with Europeans

5

6 **Contributors:**

7 Jun Gojobori[1] *, Fuzuki Mizuno[2], Li Wang[3] **, Keisuke Onishi[2], Julio Granados[4], Celta

8 Gomez-Trejo[5], Víctor Acuña-Alonzo[5], and Shintaroh Ueda[2]

9

10 **Affiliations:**

11 [1] School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced

12 Studies), Hayama 240-0193, Japan

13 [2] Department of Biological Sciences, Graduate School of Science, the University of

14 Tokyo, Tokyo 113-0033, Japan

15 [3] School of Medicine, Hangzhou Normal University, Zhejiang 311121, China

16 [4] Department of Transplants, Instituto Nacional de Ciencias Médicas y Nutrición

17 Salvador Zubirán, Mexico City 14000, Mexico

18 [5] Molecular Genetics Laboratory, National Institute of Anthropology and History,

19 Mexico City 14030, Mexico

20

21 Correspondence to: *gojobori_jun@soken.ac.jp (J.G.), **liwang@hznu.edu.cn (L.W.)

22

23 **Running title:**

24 Demographic history of indigenous Mesoamericans

25

26

1 **Abstract:**

2 The New World is the last continent colonized by anatomically modern

3 humans, *Homo sapiens*. The first migrants entered the New World from Asia

4 through Beringia. It is suggested that there were three streams of Asian gene

5 flow, one major and two additional minor gene flows. The first major

6 migrants took a Pacific coastal route and began spreading to the American

7 continent before the opening of the ice-free corridor. We investigated the

8 diversity of full-length mtDNA genomes of the Zapotec population, residing

9 in the Mesoamerican region, and reconstructed their demographic history

10 using Bayesian Skyline Plots. We estimated the initial date of gene flow into

11 the New World by Zapotec ancestors at around 17,000–19,000 years ago,

12 which is highly concordant with previous studies. We also show a population

13 decline after the initial expansion. This decline started 4,000 years ago, long

14 before European contact with Native Americans. This indicates that other

15 factors including climate change should be considered to explain the observed

16 demographic pattern.

17

1　**Description section:**

2　**Introduction**

3　　　　It is widely accepted that anatomically modern humans (*Homo sapiens*) from

4　Asia colonized the New World through Beringia during the ice age.[1] Many researchers

5　in archaeology, linguistics, and evolutionary genetics have investigated the route,

6　timing, and size of the first migration event to America.[2,3] The single migration

7　hypothesis was proposed based mainly on the analysis of uniparental genetic markers of

8　current Native Americans.[1,4] Genome-wide SNP data for Native Americans suggests

9　three streams of Asian gene flow, one major plus two additional minor gene flow

10　events.[5] These data also suggest a coastal route for the initial peopling of America, in

11　agreement with the results of mtDNA analyses.[4,5]

12　　　　The Mesoamerican region is adjacent to the Isthmus of Panama, a narrow strip

13　of land that links North and South America. It is an important region for understanding

14　the demography of Native Americans because the ancestors of South Americans

15　presumably passed through this region during their migration from North America. In a

16　previous study, we analyzed mtDNA haplotypes to estimate genetic diversity within the

1   Mesoamerican region.[6] Here, we report a more detailed demographic history of the

2   Zapotec inferred from Bayesian Skyline Plots (BSP) to gain better insight into their

3   genetic history. Of particular interest is whether a recent population decline corresponded

4   with European contact, as some studies have proposed, or other factors such as climatic

5   events.[7-11]

6

7   **Materials and Methods**

8   We used 88 complete Zapotec mtDNA genome sequences reported in Mizuno

9   *et al.*[6] The poly(C) tracts beginning at position 16,182 were not considered in this study.

10   Information regarding data collection and data availability can be found in

11   Mizuno *et al.*

12   To obtain BSP, we used BEAST v1.7 suite.[12] We applied the Tamura and Nei

13   TN93 mutation model using gamma-distributed rates and a proportion of invariant sites.

14   A clock model was used assuming a log-normal distribution. In the Skyline model, we

15   assumed a piecewise-constant model with ten groups. We set the Markov chain Monte

16   Carlo (MCMC) chain length to 1×10^8 with 1×10^7 burn-in steps to collect sufficient

1   samples for parameter estimation. We performed the MCMC simulation twice

2   independently to confirm that the simulation converged to the same state. We also

3   estimated the times to the most recent common ancestor (tMRCAs) by this MCMC

4   method for haplogroups A2, B2, C1, and D1 using Beauti.[12] The time was scaled by the

5   number of mutations for both BSP construction and the estimation of tMRCAs for each

6   haplogroup. To estimate divergence time based on the number of mutations,

7   we assumed a molecular clock of 1.665×10^{-8} and applied the correction method of

8   Soares *et al.*[13] To scale the female effective population size (N_{ef}), we assumed a

9   generation time of 25 years.[14] The simulation results were analyzed using Tracer v1.5

10   (http://tree.bio.ed.ac.uk/software/tracer/). For estimating tMRCAs, we also used another

11   common method, ρ statistics, to estimate the tMRCAs of the four haplogroups using

12   median joining calculation implemented in Network (http://www.fluxus-

13   engineering.com).[15]

14

15   **Results**

16   Zapotec mtDNA was classified as haplogroup A2, B2, C1, or D1, and these

1   groups included 59, 16, 10, and 3 sequences, respectively. In total, the data included 183

2   segregating sites (S) and was characterized by a nucleotide diversity per site (π) of

3   0.002850. Tajima's D was -1.464, and was significantly different from zero (p = 0.036).[16]

4   This excess of low frequency polymorphisms indicates that the Zapotec experienced a

5   past population expansion.   However, these summary statistics enable only a rough

6   estimation of the timing and magnitude of demographic events.

7   To obtain a more detailed demographic history of the Zapotec, we constructed

8   a BSP using BEAST v1.7 suite.[12] BSP showed that the ancestors of the Zapotec

9   experienced a population expansion around 19,000 to 17,000 years ago (Figure 1). The

10   N_{ef} was approximately 1,500 and reached approximately 20,000 after the expansion.

11   This observation is highly concordant with previous result[4]; if we assume that this time

12   of population expansion corresponds to the peopling of America, we can infer that the

13   ancestors of the Zapotec migrated from Beringia to the American continent at least

14   19,000 to 17,000 years ago. This time range predates the opening of the ice-free

15   corridor between the Laurentide and Cordilleran ice sheets, estimated to have occurred

16   roughly 12,000 years ago.[4,17] This suggests that the Zapotec ancestors migrated via a

7

1   coastal route rather than the ice-free corridor.[4,5] The BSP estimated from the mtDNA

2   coding regions and the full-length sequences showed the same results (data not shown).

3   Table 1 shows the estimated tMRCA for the four haplogroups determined

4   using BEAST[12]. Haplogroup A2 showed the oldest tMRCA, while Haplogroup D1

5   showed the youngest tMRCA. The tMRCAs of Haplogroups B2 and C1 were

6   intermediate. These results are not concordant with a previous study.[4] Because the

7   Zapotec people are considered to be the descendants of the first major migrants to the

8   New World[5], overlap between the tMRCAs of the four haplogroups was expected.[4]   To

9   verify our results, we estimated tMRCA for the four haplogroups using ρ

10   statistics.[15] Table 1 shows the estimates of ρ and the associated dates inferred from the

11   statistic, using the molecular clock of Soares *et al*.[13] The estimates based on the ρ

12   statistics were consistent with the results obtained using BEAST.

13   We carefully examined very recent demographic events estimated from BSP

14   to gain insights into the population decline of Native Americans. The BSP of the

15   Zapotec shows a population decline after an approximately 10-fold population

16   expansion (Figure 1). N_{ef} decreased to around 5,000 from a peak of around 20,000. If

1 we assume that this population decline was caused by the first contact with Europeans,

2 the time of this decline should correspond to the time of the contact event that occurred

3 500 years ago.[8] However, our BSP showed that the population decline started

4 approximately 4,000 years ago, which is 3,500 years earlier than the arrival of the first

5 Europeans to the New World.

6

7 **Discussion**

8 The tMRCAs estimated for the four haplogroups showed different result from

9 previous study[4] (Table 1). The sample size of Haplogroup D1 was only three; therefore,

10 it is quite possible that the tMRCA of D1 was underestimated. The sample sizes of

11 Haplogroups B2 and C1 were also small (16 and 10, respectively), which might also

12 result in an underestimate of the mutations and therefore can lead to more recent

13 coalescent time estimates.

14 Our BSP showed a population decline started approximately 4,000 years ago

15 (Figure 1). On the other hand, O'Fallon and Fehren-Schmitz detected a population

16 decline of Native Americans that occurred 500 years ago, which coincides with the first

1   European contact.[8] Unlike our analysis, O'Fallon and Fehren-Schmitz used extended

2   BSP (EBSP). We performed EBSP with our Zapotec data, applying the same

3   parameters used in O'Fallon and Fehren-Schmitz, and detected a population decline

4   starting 5,000 to 6,000 years ago. When we applied the molecular clock of Soares *et*

5   *al.*[13] in this EBSP analysis, the time of population decline was estimated at around 4,000

6   years ago, which matches our BSP results (data not shown).   Therefore, the difference

7   between our results and those of O'Fallon and Fehren-Schmitz was not due to

8   methodological differences. We assume that the difference may reflect the different

9   samples used in the studies.

10   In the Mesoamerican area, agriculture began more than 5,000 years ago[18] and

11   the growth of civilizations occurred afterward.[19] Steady population expansion is

12   generally expected after the introduction of agriculture, but such steady growth was not

13   observed in Europe.[20] Our study is the first to uncover unexpected demographic

14   patterns in a population of Mesoamericans after the introduction of agriculture. During

15   the Holocene, millennial or multi-century periodical cycles of climate change have been

16   suggested.[9, 10] The timing of these cycles is related to Bond event no. 3 of the Holocene

1   Bond cycles.[10] An abrupt climate change occurred in Mexico around 4,000 years ago.[11]

2   We concluded that the N_{ef} of the ancestors of the Zapotec population decreased around

3   4,000 years ago, long before European contact, possibly owing to severe climate

4   change.[9-11]

1 **Conflict of Interest**

2 The authors declare no conflict of interest.

3

4

5 **Acknowledgments**

1   1. Fagan, BM. *The Great Journey: The Peopling of Ancient America. London.*

2   (University Press of Florida, Gainesville, FL, USA, 2004).

3   2. Perez, S. I., Bernal, V., Gonzalez, P. N., Sardi, M. & Politis, G. G. Discrepancy

4   between cranial and DNA data of early Americans: implications for American

5   peopling. *PLoS ONE* **4,** e5746 (2009).

6   3. Goebel, T., Waters, M. R. & O'Rourke, D. H. The late Pleistocene dispersal of

7   modern humans in the Americas. *Science* **319,** 1497–1502 (2008).

8   4. Fagundes, N. JR., Kantiz, R., Eckert, R., Valls, A. C. S., Bogo, M. R., Salzano F.

9   M. *et al.* Mitochondrial population genomics supports a single pre-Clovis origin

10   with a coastal route for the peopling of the Americas. *American journal of human*

11   *genetics* **82,** 583–592 (2008).

12 5. Reich, D., Patterson, N., Campell, D., Tandon, A., Mazieres, S., Ray, N. *et al.*

13   Reconstructing Native American population history. *Nature* **488,** 370–374 (2012).

14 6. Mizuno, F., Gojobori, J., Wang., L. Onishi, K., Sugiyama, S., Granados, J.*et al.*

15   Complete mitogenome analysis of indigenous populations in Mexico: its relevance

16   for the origin of Mesoamericans. *J Hum Genet* **59,** 359–367 (2014).

1   7.   Cook, ND *Born to Die: Disease and New World Conquest, 1492-1650.*

2   (Cambridge University Press, Cambridge, UK, 1998).

3   8.   O'Fallon, B. D. & Fehren-Schmitz, L. Native Americans experienced a strong

4   population bottleneck coincident with European contact. *Proc Natl Acad Sci USA*

5   **108,** 20444–20448 (2011).

6   9.   Denton, G. H. & Karlén, W. Holocene climatic variations—Their pattern and

7   possible cause. *Quaternary Research* **3,** 155–205 (1973).

8   10.   Bond, G. A Pervasive Millennial-Scale Cycle in North Atlantic Holocene and

9   Glacial Climates. *Science* **278,** 1257–1266 (1997).

10   11.   Wanner, H., Beer, J., Bütikofer, J. & Crowley, T. J. Mid- to Late Holocene climate

11   change: an overview. *Quaternary Science Reviews* **27,** 1791–1828 (2008).

12   12.   Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics

13   with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29,** 1969–1973 (2012).

14   13.   Soares, P. et al. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl,

15   A. *et al*. Correcting for Purifying Selection: An Improved Human Mitochondrial

16   Molecular Clock. *American journal of human genetics* **84**, 740–759 (2009).

1 14. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in

2   genetics-based population divergence studies. *Am J Phys Anthropol* **128,** 415–423

3   (2005).

4 15. Bandelt, H. J., Forster, P., Röhl, A. Median-joining networks for inferring

5   intraspecific phylogenies. *Mol Biol Evol* **16**, 37–48 (1999).

6 16. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA

7   polymorphism. *Genetics* **123**, 585–595 (1989).

8 17. Arnold, T. G. Radiocarbon dates from the ice-free corridor. *Radiocarbon* **44,** 437–

9   454 (2006).

10 18. Brown CH (2010) *Pre-Columbian Foodways*. (Springer, New York, NY, USA,

11   2010).

12 19. Hendon JA, Joyce R. *A Mesoamerican archaeology: theory and practice*.

13   (Blackwell, Malden, MA, USA, 2004).

14 20. Shennan, S., Downey, S. S., Timpson, A., Edinborough. K., Colledge, S., Kerig, T.

15   *et al.* Regional population collapse followed initial agriculture booms in mid-

16   Holocene Europe. *Nat Comms* **4,** 2486 (2013).

1   Figure legend

2

3   Figure 1. Bayesian Skyline Plot (BSP) of the Zapotec.

4   N_{ef} (Y-axis, log-scale) was plotted against time (X-axis, years ago). A mutation rate of

5   1.664×10^{-8} /site/year and a generation time of 25 years were assumed.[14, 15] Solid line

6   shows the mean N_{ef} from the posterior distribution. Dashed lines show the 95%

7   posterior density intervals.
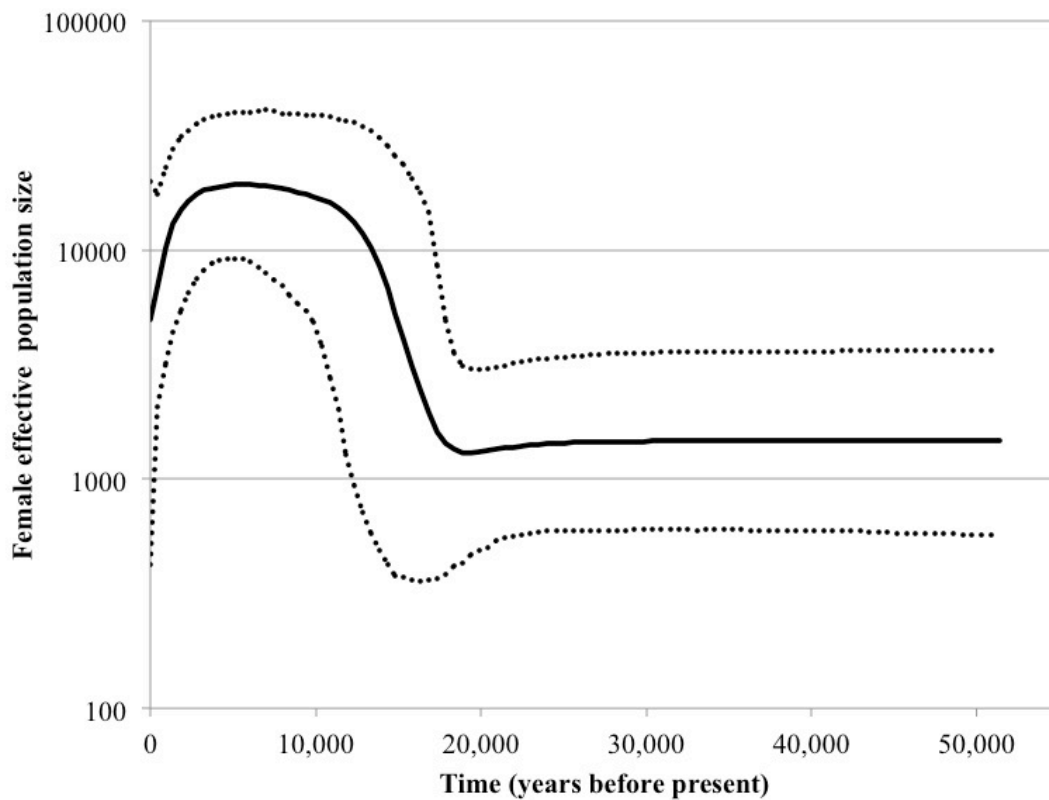
8

1   Table 1. Estimated tMRCAs in years for Haplogroups A2, B2, C1, and D1 from the

2   Zapotec using BEAST and ρ statistics.

Haplogroups	BEAST tMRCA in years (95% c.i.)	ρ (s.d.)	ρ statistics tMRCA in years (95% c.i.)
A2	23126.9 (16888.6–30542.3)	8.31 (1.54)	24432.6 (15727.1–33472.3)
B2	17769.5 (12896.4–23625.8)	7.75 (1.67)	21168.5 (11908.1–30833.5)
C1	18048.6 (12457.8–24820.1)	6.90 (1.52)	18726.7 (10393.8–27402.3)
D1	15137.7 (8636.2–20663.4)	5.00 (1.37)	13370.5 (6030.8–21006.2)

3   s.d. = Standard Deviation, c.i. = Confidential Interval.