

Evolution of the Ainu Language in Space and Time

Sean Lee*, Toshikazu Hasegawa

Department of Cognitive and Behavioral Science, Graduate School of Arts and Sciences, University of Tokyo, Tokyo, Japan

Abstract

Languages evolve over space and time. Illuminating the evolutionary history of language is important because it provides a unique opportunity to shed light on the population history of the speakers. Spatial and temporal aspects of language evolution are particularly crucial for understanding demographic history, as they allow us to identify when and where the languages originated, as well as how they spread across the globe. Here we apply Bayesian phylogeographic methods to reconstruct spatiotemporal evolution of the Ainu language: an endangered language spoken by an indigenous group that once thrived in northern Japan. The conventional dual-structure model has long argued that modern Ainu are direct descendants of a single, Pleistocene human lineage from Southeast Asia, namely the Jomon people. In contrast, recent evidence from archaeological, anthropological and genetic evidence suggest that the Ainu are an outcome of significant genetic and cultural contributions from Siberian hunter-gatherers, the Okhotsk, who migrated into northern Hokkaido around 900–1600 years ago. Estimating from 19 Ainu language varieties preserved five decades ago, our analysis shows that they are descendants of a common ancestor who spread from northern Hokkaido around 1300 years ago. In addition to several lines of emerging evidence, our phylogeographic analysis strongly supports the hypothesis that recent expansion of the Okhotsk to northern Hokkaido had a profound impact on the origins of the Ainu people and their culture, and hence calls for a refinement to the dual-structure model.

Citation: Lee S, Hasegawa T (2013) Evolution of the Ainu Language in Space and Time. *PLoS ONE* 8(4): e62243. doi:10.1371/journal.pone.0062243

Editor: Luísa Maria Sousa Mesquita Pereira, IPATIMUP (Institute of Molecular Pathology and Immunology of the University of Porto), Portugal

Received: January 17, 2013; **Accepted:** March 19, 2013; **Published:** April 26, 2013

Copyright: © 2013 Lee, Hasegawa. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) Fellows (248973). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: seanlee@darwin.c.u-tokyo.ac.jp

Introduction

Patterns of linguistic variation among individuals often carry the signature of a speech community's demographic past. Accumulating evidence indicates that languages evolve by a process of descent with modification and they form into distinct families in a manner similar to their speakers forming into different ethnic groups through evolutionary history [1–3]. The intertwined history between languages and their speakers appears most vividly in the areas that experienced large-scale population expansions, often driven by agricultural intensification and cultural innovation since the end of the last Ice Age [4]. Recent empirical evidence supporting this phenomenon includes a range of language phylogenies reconstructed with computational methods [5–8].

While the computational phylogenetic methods have been fruitful in shedding new light on language evolution and the speakers' prehistory, their application has been focused mainly on inferring temporal and sequential aspects. As a result, inferences about the homeland or geographic diffusion pattern often relied on heuristic approaches such as locating a monophyletic outgroup and formulating post-hoc diffusion scenarios from the branching order. Recent progress in phylogenetic methods is, however, producing innovative ways to embed phylogenetic inference in a geographical context, and allow us to explicitly estimate both temporal and spatial aspects of evolution while accounting for phylogenetic uncertainty [9–11]. In this paper, we adopt these methodological innovations and directly reconstruct spatiotemporal evolution of the Ainu language: a nearly extinct language spoken by indigenous people of Japan whose origins remain obscure.

Considerable debate surrounds the apparent incompatibility between the conventional model of human prehistory for the Japanese islands and the emerging evidence from modern archaeology, anthropology and genetics. For several decades, the dual-structure model [12] has posited that similarities in dental [13] and cranial features [14] between the Ainu people and Southeast Asians meant that the Ainu ancestry originated in Southeast Asia around 10700 years before present (BP) [15]. Similarly, reconstructed proto-Ainu lexicons have also been suggested to share some similarities with proto-Austroasiatic lexicons [16]. Therefore, the Ainu have long been thought to be direct descendants of a single ancient Southeast Asian lineage, the Jomon, and have remained isolated from neighboring populations throughout the Holocene. However, recent evidence from genetic [17,18], morphological [19,20], and cultural studies [21] are beginning to suggest that the Okhotsk people, a hunter-gatherer group from the Amur river basin, migrated into northern Hokkaido around 900–1600 BP bringing significant genetic and cultural contributions to the preexisting Jomon, and subsequently gave rise to modern Ainu people as well as their culture. In essence, this 'Okhotsk expansion scenario' suggests that, far from being direct descendants of a single ancient human lineage that had no contact with the rest of the world, the Ainu and their culture are the outcome of a recent population expansion into northern Hokkaido.

If we accept premises (i) population expansions often leave its signature in the patterns of linguistic variation and (ii) the cultural flow from the incoming Okhotsk people had a profound impact on the language, then we can reason that spatiotemporal reconstruc-

tion of the Ainu language evolution might allow us to test the plausibility of the Okhotsk expansion scenario for the Ainu origin, and examine whether or not the dual-structure model should be modified to accommodate the Okhotsk expansion scenario. Accordingly, we predicted that if the scenario were correct, then the estimated root age of the Ainu varieties should coincide with 900–1600 BP, and their geographic distribution should be the end result of expansion from northern Hokkaido, where the gene and cultural flows from the Okhotsk to Jomon is likely to have taken place (blue bar in Figure 1). It should be noted that even if the patterns of Okhotsk expansion were correct, the specific processes of language change could be interpreted in two ways (see Discussion for more details). Following the line of reasoning above, we also predicted that if the scenario were incorrect, then the Ainu language diffusion should conform to the conventional scenario and spread northward from southern Hokkaido with the root age being at least several thousand years older than 1600 BP but not beyond 10000 BP, which is the current methodological limit for tracing language ancestry.

Materials and Methods

The data consist of 19 geocoded lists of 200 basic vocabularies (Figure 1; for a full list of sites, see Figure S1) compiled by Hattori and Chiri during 1950s [22], when there was still a rich linguistic diversity among the Ainu people. The basic vocabularies are a set of words transmitted vertically from one generation to the next [23], thereby preserving evolutionary signal required for reconstructing phylogenetic history [3,24]. Nevertheless, one could argue that the 19 varieties that we analyze here are dialects of the Ainu language and if one supposes that only languages, not dialects, constitute representative units of analysis, then using these varieties implies that the resulting tree may potentially depict a confusing branching pattern with excessive detail, or even fail to recover the actual subdivisions of the speech community [25].

We do not, however, consider this to be a major obstacle for reconstructing Ainu language evolution for three reasons: (i) a natural model of language evolution that we use here is known to

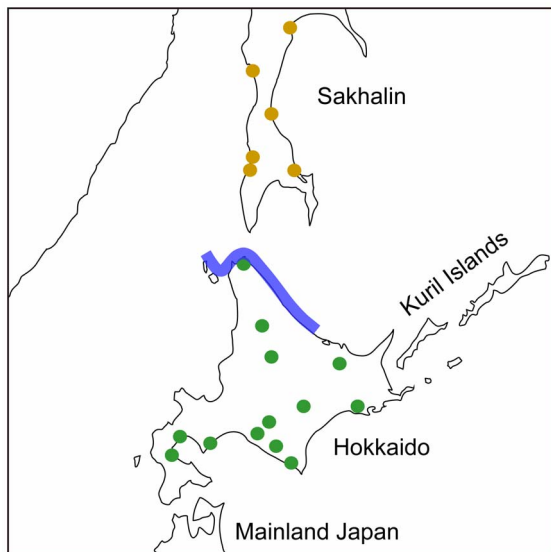


Figure 1. Map of the Ainu language varieties. Colored circles represent two major subgroups (Green-Hokkaido; Yellow-Sakhalin). Blue bar in the center indicates the main area of the Okhotsk settlement.

doi:10.1371/journal.pone.0062243.g001

be robust against reasonable levels of noise (i.e., up to 20% of horizontal transfer per 1000 years) [26], (ii) if we define languages as groups of tongues that are mutually unintelligible in a manner similar to biologists defining species as groups of animals that cannot interbreed [27], then Swadesh's criterion of mutual intelligibility (i.e., any two languages being mutually unintelligible if they share less than 90% of their basic vocabularies with each other [28]) and a matrix of pair-wise cognate similarities of the Ainu varieties [22] allow us to estimate that any one of the varieties would be able to communicate with the rest only about 18% at a time, meaning that the majority of the 19 varieties can actually be considered languages in their own right and (iii) we used SPLITSTREE4 [29] to estimate tree-likeness of the Ainu phylogeny [30,31] and obtained the average delta score = 0.25 and Q -residual score = 0.01, both indicating that the evolution of Ainu lexicons was reasonably tree-like, and hence suitable for phylogenetic analysis (to put this in perspective, the tree-likeness scores calculated from a subset of 12 Indo-European languages have similar scores as our 19 Ainu varieties with the average delta score = 0.23 and Q -residual score = 0.03 [31]). These observations provide us confidence that the data should carry robust evolutionary signal and the 19 Ainu varieties are appropriate units of analysis for the current purpose.

Cognate judgments, a process of revealing shared ancestry among lexicons, are carried out by identifying systematic correspondences in phonetic structure and meaning [25]. For our analyses, we adopted the cognate judgments made by the two linguists who compiled the data [22]. The cognate sets were encoded into binary states indicating presence ('1') or absence ('0') of a cognate, which resulted in 19×350 matrix.

We used BEAST [32] for all analyses because it allows us to reconstruct phylogenies without specifying an *a priori* outgroup. Continuous random walk model we use in this paper [9,33] is a Bayesian expansion of Brownian diffusion model developed in a maximum-likelihood framework [34]. In general, a Brownian diffusion model aims to estimate the vectors of latitudes and longitudes of internal nodes (i.e., common ancestors of extant languages) on a continuous surface, in which increments are independent and normally distributed with a mean centered on zero with variance that scales linearly in time, meaning that diffusion processes are assumed to be homogeneous over time and space. This can be unrealistic as many geographic features (e.g., mountains and rivers) can influence the rate of spread for each branch. Bayesian continuous diffusion model we adopt here effectively overcomes this limitation by relaxing the Brownian process: borrowing ideas from uncorrelated relaxed clock models [35], the method models branch-specific dispersal processes with the diffusion rate scalar in each branch being drawn independently and identically from a range of parametric distributions. Distributions used in our analyses are (i) Cauchy distribution that has fat tails accommodating long distance dispersals [36], (ii) gamma distribution that accommodates infinite variance in a manner similar to Lévy flight models [37] but without enforcing power-law tail behavior, and (iii) lognormal distribution that allows even greater degree of rate variability [35]. In order to make our geographic inference more realistic, we sampled the root and node locations only from the land by assigning a prior probability of zero to the water [11].

In addition, we compared the degree of model-fit between relaxed and strict clocks [35]. Temporal scale of phylogenies was calibrated using a probabilistic prior taken from well-attested evidence that modern Ainu expanded into Sakhalin around 15th century [38,39]: a normally-distributed prior with a mean of 500 BP with its 95% of the distribution incorporating 200 years of

uncertainty. For all analyses, we applied a stochastic Dollo model with a correction for ascertainment bias [40] and a Bayesian skyline tree prior [41]. We chose the best model by comparing Bayes Factors (BF) [42].

Results

Based on BF tests among diffusion models and evolutionary clock models, we chose the relaxed clock with gamma-distributed diffusion as the best model (Table S1). Figure 2 shows the summary of time-dated maximum clade credibility trees for 19 Ainu language varieties. Assuming that the patterns of linguistic diversity is shaped by the demographic dynamics of speakers, we predicted that if the recent evidence supporting the Okhotsk expansion scenario were correct, then the estimated root age should overlap with 900–1600 BP. The estimated root age of the Ainu language across post-burn-in trees has a median of 1288 BP [mean: 1323 BP; 95% Highest Posterior Density (HPD): 820–1862 BP], in strong agreement with the prediction. We also predicted that if the hypothesized scenario were correct, then the current distribution of 19 Ainu language varieties should be the end result of diffusion from northern Hokkaido; otherwise, we would observe northward expansion from southern Hokkaido, conforming to the conventional dual-structure model. Figure 3 (also in Animation S1) shows that the estimated diffusion pattern in natural time scale [43] is in clear agreement with the prediction, with the estimated homeland being in northern Hokkaido. Both the diffusion pattern and root time were consistent across all models we excluded based on BF tests.

In order to examine the robustness of our phylogeographic inferences, we carried out two additional tests. Firstly, we tested the strength of support for northern Hokkaido origin (i.e., the Okhotsk expansion scenario) over southern Hokkaido origin (i.e., the dual-structure model) by directly calculating BF: we divided Hokkaido into two broad regions of north and south at the centroid of Hokkaido, and estimated BF by comparing the posterior to prior odds ratio of observing potential homeland in either one of the two regions. In agreement with our results, we obtained substantial support (BF = 7.5) for northern Hokkaido being the homeland of the Ainu. Secondly, we investigated whether or not our results are statistical artifacts of the diffusion model falling into the center of language mass regardless of the data: we randomly reassigned the locations of 19 Ainu varieties to the data for fifty times, and then obtained 90% HPDs for all possible root locations (Figure S2). From this exercise, we observed that the absence of true signal could cause the estimated homeland to be as south as mainland Japan or as north as Sakhalin. This observation clearly demonstrates that our results are valid estimations based on true phylogeographical signal. Conversely, this also suggests that if the data contained signal indicating northward diffusion, or any other direction, our methods would have reconstructed it accordingly.

We acknowledge, however, that a well-established subgroup of the Ainu language, namely the Kuril, is absent from our data. This is because the Kuril had become extinct by the time the data were collected, and the Kuril lexicons seem to be available only through sketchy records scattered around the literature. For this reason, we currently have little information about the Kuril. If the point in

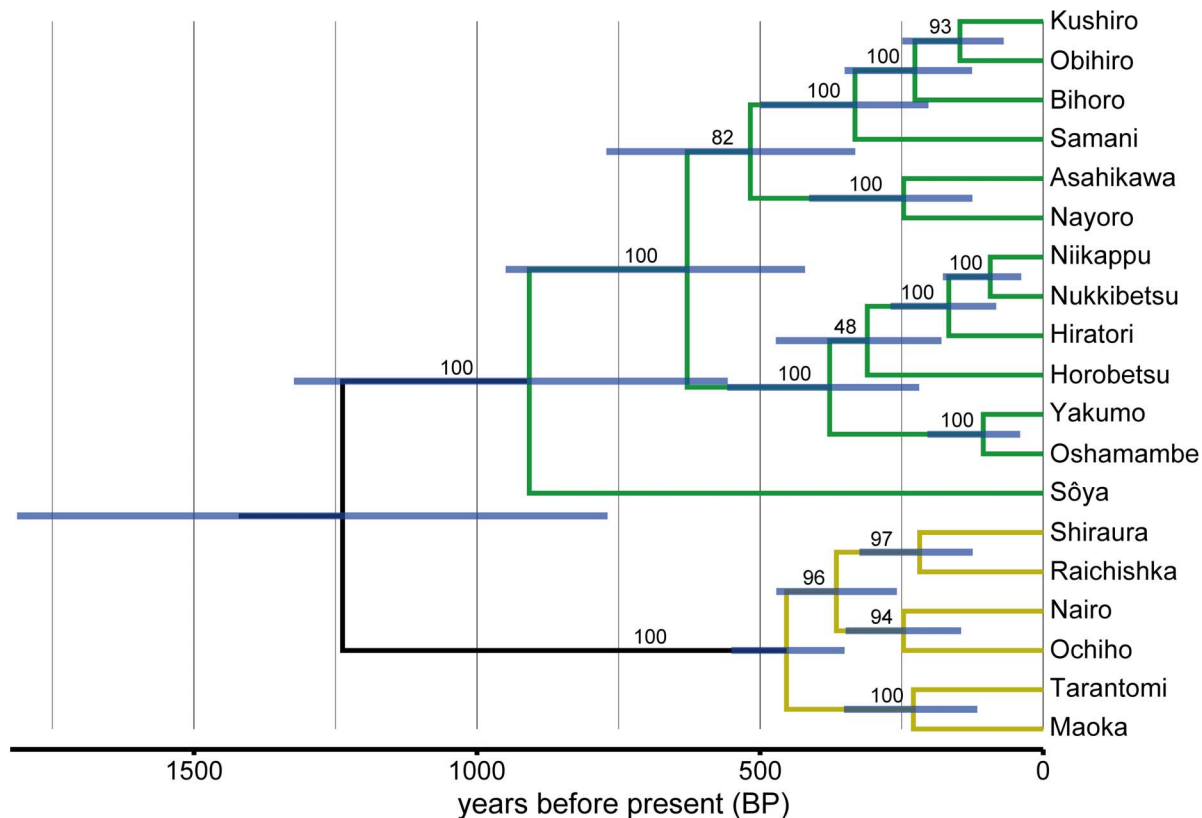


Figure 2. Maximum clade credibility tree of 19 Ainu language varieties. Colored branches represent two major subgroups (Green-Hokkaido; Yellow-Sakhalin). All node heights are scaled to match the posterior median node heights with bars indicating 95% HPD intervals of the estimated ages. The value on each branch is the posterior probability, showing the percentage support for the following node. doi:10.1371/journal.pone.0062243.g002

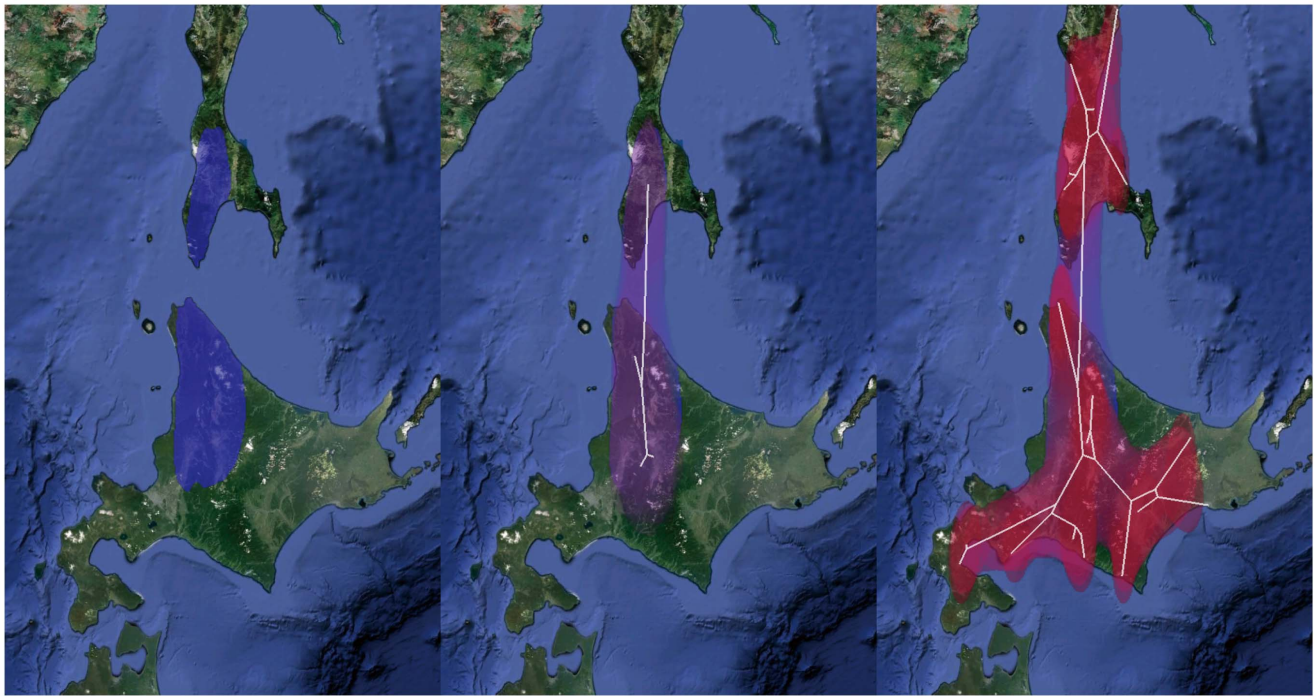


Figure 3. Inferred origin and diffusion of the Ainu language varieties in natural time scale. Color gradient of the polygons (80% HPD) indicates relevant age of the diffusion [Blue-older (1288 BP); Red-more recent (50 BP)]. White lines represent the phylogeny projected onto the surface. Image sources: © 2012 Google Earth; © 2012 Cnes/Spot Image; © 2012 TerraMetrics.
doi:10.1371/journal.pone.0062243.g003

time that the Kuril diverged from other varieties turns out to be much deeper, then the resulting divergence time and diffusion pattern may differ significantly from the current results. The search for a more complete set of data is, therefore, a direction that should be prioritized for further evaluation of our conclusion.

Discussion

In this paper, we reconstructed spatiotemporal evolution of 19 Ainu language varieties, and the results are in strong agreement with the hypothesis that a recent population expansion of the Okhotsk people played a critical role in shaping the Ainu people and their culture. Together with the recent archaeological, biological and cultural evidence, our phylogeographic reconstruction of the Ainu language strongly suggests that the conventional dual-structure model must be refined to explain these new bodies of evidence. The case of the Ainu language origin we report here also contributes additional detail to the global pattern of language evolution, and our language phylogeny might also provide a basis for making further inferences about the cultural dynamics of the Ainu speakers [44,45].

We recognize that there are also some evidence that the Jomon people, one of the two ancestral populations of the Ainu, may have descended from Northeast Asia rather than Southeast [46,47], thereby questioning the validity of dual-structure model on a greater time scale. Unfortunately, the scope of our results presented here have little bearing on the larger question of the Jomon prehistory because the linguistic traces of this process may have been wiped out by the recent rise of the Ainu as our results indicate. Regardless of what further research reveals about the Jomon ancestry, however, we argue that the evidence for the Okhotsk expansion scenario should remain valid, and therefore any future models of deeper historical process for the Japanese

islands must properly account for the recent northern Hokkaido origin of the Ainu. With this respect, we suggest that the most effective way of shedding light on the deeper history of the Jomon, or historical processes of any other regions, is to synthesize different lines of evidence from archaeology, biology and culture, and triangulate them to obtain a rigorous analytic framework [48] rather than relying on a single line of evidence.

If our inferences are correct, then the recent Okhotsk expansion scenario for the Ainu origin leads us to a new question: what historical factors drove the Okhotsk people to migrate from the Amur river basin to Hokkaido and give rise to the Ainu? It is now clear that early farming populations went through similar processes due to agricultural intensification and cultural innovation [4] but the Okhotsk people were hunter-gatherers, not farmers. While not resolving this question directly, Hudson [49] provides a comprehensive model of the Okhotsk socio-environmental conditions that allows us to sketch out a possible scenario: (i) the diet of the Okhotsk people relied heavily on marine mammal products and (ii) the time in which the Okhotsk expansion occurred seems to be characterized by dramatic climate changes, beginning with a cold sea-ice stage between 1300–1800 BP followed by a warmer open-ocean stage. Based on these observations, we speculate that the Okhotsk expansion may have been opportunistic in nature: the sea-ice condition in the early stage probably resulted in increased area for exploiting marine mammals as well as convenient routes for exploring new territory, thereby leading to the migration into Hokkaido. The drastic climate change in the later stage, however, may have deteriorated the hunting conditions for the Okhotsk with rapid break up of sea-ice, which may also have necessitated increased reliance on other types of food source, and hence causing a greater degree of niche overlap with the preexisting Jomon population. The end result was

probably the admixture of the two populations, followed by the rise of a new ethnolinguistic group, namely the Ainu.

If we accept a view that transmission of language may be gender-specific [50–52], then we are able to formulate at least two hypotheses for the specific processes of the Ainu language origin. Because Y-chromosome haplogroup D is thought to represent Jomon male ancestry, the predominance of that particular haplogroup in the Ainu (75–87.5%) implies that the majority of Ainu male ancestry is from the Jomon [53,54], whereas a heavy mixture of mtDNA haplogroups indicates that a significant proportion of the Ainu female ancestry is from the Okhotsk (excluding 35.3% of mtDNA haplogroups that the Ainu share with other neighboring populations, 39.4% of the remaining female heritage is shared exclusively with the Okhotsk and the rest is a mixture of both Jomon and Okhotsk [18,47,54]). If we thus assume male-specific language transmission for the Ainu, the first hypothesis for the processes behind the Ainu language origin could be that proto-Ainu arose from a large number of Jomon males who intermarried with Okhotsk females in northern Hokkaido, and subsequently spread to the rest of region. Similarly, if we assume that the transmission of Ainu language corresponds with female ancestry, the second hypothesis could be that proto-Ainu was spoken by the incoming Okhotsk females who merged with the preexisting Jomon males. Based on these observations, we propose that one potential way of understanding how language change occurred for the Ainu is to estimate which gender was more influential when early Ainu people established family membership. This may be carried out indirectly by revealing the signature of historical post-marital residence pattern via estimating the degrees of genetic variation in their Y-chromosome and mtDNA [55] as well as reconstructing ancestral post-marital residence rules from regional cultural variation [56]. Investigating which model of language change [57] is relevant to the Ainu is a direction that deserves more attention, and acquiring an accurate description of how language change occurred for the Ainu would allow us to make further inferences about the deeper history of the human lineage that once thrived in northern Japan.

Languages rise and fall, and so do the communities who speak them. Although significant progress has been made in recent years, we are still far from thoroughly understanding why languages are so deeply related to the fates of their speakers or how the process unfolds through evolutionary history. These are perhaps some of the most challenging questions in human sciences, and a complete understanding of this complex phenomenon might thus be reached only with further methodological innovations as well as more language data from around the world. But as we demonstrate in this paper, a combination of spatiotemporal

reconstruction of language evolution and synthesis of several different historical evidences is probably one of the most promising methodologies that can further illuminate the process and consequence of this fascinating phenomenon.

Supporting Information

Figure S1 Full list of the Ainu language varieties. Colored circles represent subgrouping (Green-Hokkaido; Yellow-Sakhalin). (TIF)

Figure S2 Ninety percent highest probability density obtained from fifty random reassignments of location coordinates to the tips of phylogeny. This demonstrates that our results are not statistical artifacts of the diffusion model returning to the center of language mass. For all analyses, we applied an arbitrary root calibration consisting of a normal distribution with the mean of 1500 BP and the standard deviation of 400 years. (TIFF)

Table S1 Log-marginal likelihoods estimated from all models fitted to data. The model with a relaxed clock and gamma-distributed random walk model shows the best fit with the highest log-marginal likelihood. (DOCX)

Animation S1 Animated origin and diffusion of the Ainu language varieties in natural time scale. Color gradient of the polygons (80% HPD) indicates relevant age of the diffusion [Blue-older (1288 BP); Red-more recent (50 BP)]. White lines represent the phylogeny projected onto the surface. Image sources: © 2012 Google Earth; © 2012 Cnes/Spot Image; © 2012 TerraMetrics. (MOV)

Analysis S1 NEXUS and BEAST input files for full details of the analysis. (ZIP)

Acknowledgments

We thank Russell Gray, Simon Greenhill and Quentin Atkinson for their insightful comments and help with technical issues. We also thank two thoughtful reviewers for their constructive comments.

Author Contributions

Conceived and designed the experiments: SL TH. Analyzed the data: SL. Wrote the paper: SL TH.

References

- Darwin C (1871) The descent of man. London: Murray.
- Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci* 85: 6002–6006. doi:10.1073/pnas.85.16.6002.
- Pagel M (2009) Human language as a culturally transmitted replicator. *Nat Rev Genet* 10: 405–415. doi:10.1038/nrg2560.
- Diamond J, Bellwood P (2003) Farmers and their languages: the first expansions. *Science* 300: 597–603. doi:10.1126/science.1078208.
- Holden CJ (2002) Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc R Soc B* 269: 793–799. doi:10.1098/rspb.2002.1955.
- Gray RD, Atkinson QD (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426: 435–439. doi:10.1038/nature02029.
- Gray RD, Drummond AJ, Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323: 479–483. doi:10.1126/science.1166838.
- Lee S, Hasegawa T (2011) Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proc R Soc B* 278: 3662–3669. doi:10.1098/rspb.2011.0518.
- Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 27: 1877–1885. doi:10.1093/molbev/msq067.
- Walker RS, Ribeiro LA (2011) Bayesian phylogeography of the Arawak expansion in lowland South America. *Proc R Soc B* 278: 2562–2567. doi:10.1098/rspb.2010.2579.
- Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, et al. (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337: 957–960. doi:10.1126/science.1219669.
- Hanihara K (1991) Dual structure model for the population history of Japanese. *Japan Review* 2: 1–33.
- Turner CG II (1990) Major features of Sundadonty and Sinodonty, including suggestions about East Asian microevolution, population history, and late Pleistocene relationships with Australian aboriginals. *Am J Phys Anthropol* 82: 295–317. doi:10.1002/ajpa.1330820308.

14. Dodo Y, Kawakubo Y (2002) Cranial affinities of the Epi-Jomon inhabitants in Hokkaido, Japan. *Anthro Sci* 110: 1–32. doi:10.1537/ase.110.1.
15. Turner CG II (1986) Dentochronological separation estimates for Pacific Rim populations. *Science* 232: 1140. doi:10.1126/science.232.4754.1140.
16. Vovin A (1993) A reconstruction of proto-Ainu. Leiden: Brill.
17. Sato T, Amano T, Ono H, Ishida H, Kodera H, et al. (2007) Origins and genetic features of the Okhotsk people, revealed by ancient mitochondrial DNA analysis. *J Hum Genet* 52: 618–627. doi:10.1007/s10038-007-0164-z.
18. Sato T, Amano T, Ono H, Ishida H, Kodera H, et al. (2009) Mitochondrial DNA haplogrouping of the Okhotsk people based on analysis of ancient DNA: an intermediate of gene flow from the continental Sakhalin people to the Ainu. *Anthro Sci*: 905260063. doi:10.1537/ase081202.
19. Ishida H, Hanihara T, Kondo O, Fukumine T (2009) Craniometric divergence history of the Japanese populations. *Anthro Sci* 117: 147–156. doi:10.1537/ase.081219.
20. Hanihara T (2010) Metric and nonmetric dental variation and the population structure of the Ainu. *Am J Hum Biol* 22: 163–171. doi:10.1002/ajhb.20969.
21. Masuda R, Amano T, Ono H (2001) Ancient DNA analysis of brown bear (*Ursus arctos*) remains from the archeological site of Rebum Island, Hokkaido, Japan. *Zool Sci* 18: 741–751. doi:10.2108/zsj.18.741.
22. Hattori S, Chiri M (2011) A lexicostatistic study on the Ainu dialects. *The Japanese Journal of Ethnology* 24: 31–66. Available: <http://ci.nii.ac.jp/naid/110001838079/en/>.
23. Embleton SM (1986) *Statistics in historical linguistics*. Bochum, Germany: Brockmeyer.
24. Greenhill SJ, Blust R, Gray RD (2008) The Austronesian basic vocabulary database: from bioinformatics to lexicomics. *Evol Bioinfo* 4: 271.
25. Crowley T, Bowern C (2009) *An introduction to historical linguistics*. Oxford, London: Oxford University Press.
26. Greenhill SJ, Currie TE, Gray RD (2009) Does horizontal transmission invalidate cultural phylogenies? *Proc R Soc B* 276: 2299–2306. doi:10.1098/rspb.2008.1944.
27. Pagel M, Mace R (2004) The cultural wealth of nations. *Nature* 428: 275–278. doi:10.1038/428275a.
28. Swadesh M (2006) *The origin and diversification of language*. Sherzer JF, editor. New Brunswick: Transaction Publishers.
29. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254–267. doi:10.1093/molbev/msj030.
30. Holland BR, Huber KT, Dress A, Moulton V (2002) δ plots: A tool for analyzing phylogenetic distance data. *Mol Biol Evol* 19: 2051–2059. doi:10.1093/oxfordjournals.molbev.a004030.
31. Gray RD, Bryant D, Greenhill SJ (2010) On the shape and fabric of human history. *Phil Trans R Soc B* 365: 3923–3933. doi:10.1098/rstb.2010.0162.
32. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*: 1–4. doi:10.1093/molbev/mss075.
33. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, et al. (2012) Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci*: 1–12. doi:10.1073/pnas.1206598109.
34. Lemmon A, Lemmon EM (2008) A likelihood framework for estimating phylogeographic history on a continuous landscape. *Syst Biol* 57: 544–561. doi:10.1080/10635150802304761.
35. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4: e88. doi:10.1371/journal.pbio.0040088.sd001.
36. Paradis E, Baillie SR, Sutherland WJ (2002) Modeling large-scale dispersal distances. *Ecol Model* 151: 279–292. doi:10.1016/S0304-3800(01)00487-2.
37. Reynolds AM, Rhodes CJ (2009) The Lévy flight paradigm: random search patterns and mechanisms. *Ecology* 90: 877–887. doi:10.1890/08-0153.1.
38. Ohyi H (1985) On the process of crystallization of Sakhalin Ainu. *Bulletin of the Institute for the Study of North Eurasian Cultures, Hokkaido University* 17: 165–192.
39. Ishida H, Kida M (1991) An anthropological investigation of the Sakhalin Ainu with special reference to nonmetric cranial traits. *J Anthropol Soc Nippon* 99: 23–32.
40. Alekseyenko AV, Lee CJ, Suchard MA (2008) Wagner and Dollo: a stochastic duet by composing two parsimonious solos. *Syst Biol* 57: 772–784. doi:10.1080/10635150802434394.
41. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22: 1185–1192. doi:10.1093/molbev/msi103.
42. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. doi:10.1093/molbev/mss084.
43. Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 27: 2910–2912. doi:10.1093/bioinformatics/btr481.
44. Currie TE, Greenhill SJ, Gray RD, Hasegawa T, Mace R (2010) Rise and fall of political complexity in island South-East Asia and the Pacific. *Nature* 467: 801–804. doi:10.1038/nature09461.
45. Mace R, Jordan FM (2011) Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Phil Trans R Soc B* 366: 402–411. doi:10.1098/rstb.2010.0238.
46. Hanihara T, Ishida H (2009) Regional differences in craniofacial diversity and the population history of Jomon Japan. *Am J Phys Anthropol* 139: 311–322. doi:10.1002/ajpa.20985.
47. Adachi N, Shinoda K-I, Umetsu K, Kitano T, Matsumura H, et al. (2011) Mitochondrial DNA analysis of Hokkaido Jomon skeletons: Remnants of archaic maternal lineages at the southwestern edge of former Beringia. *Am J Phys Anthropol* 146: 346–360. doi:10.1002/ajpa.21561.
48. Gray RD, Greenhill SJ, Ross RM (2007) The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory* 2: 360–375. doi:10.1162/biot.2007.2.4.360.
49. Hudson MJ (2004) The perverse realities of change: world system incorporation and the Okhotsk culture of Hokkaido. *J Anthropol Archaeol* 23: 290–308. doi:10.1016/j.jaa.2004.05.002.
50. Forster P, Renfrew C (2011) Mother tongue and Y chromosomes. *Science* 333: 1390–1391. doi:10.1126/science.1205331.
51. Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, et al. (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 68: 537–542. doi:10.1086/318200.
52. Wen B, Li H, Lu D, Song X, Zhang F, et al. (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* 431: 302–305. doi:10.1038/nature02878.
53. Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, et al. (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 51: 47–58. doi:10.1007/s10038-005-0322-0.
54. Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, et al. (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet* 49: 187–193. doi:10.1007/s10038-004-0131-x.
55. Oota H, Setteetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet* 29: 20–21. doi:10.1038/ng711.
56. Jordan FM, Gray RD, Greenhill SJ, Mace R (2009) Matrilineal residence is ancestral in Austronesian societies. *Proc R Soc B* 276: 1957–1964. doi:10.1098/rspb.2009.0088.
57. Renfrew C (1989) Models of change in language and archaeology. *Transactions of the Philological Society* 87: 103–155. doi:10.1111/j.1467-968X.1989.tb00622.x.