LANGUAGE DIVERSITY OF THE JAPANESE ARCHIPELAGO AND ITS RELATIONSHIP WITH HUMAN DNA DIVERSITY

Naruya Saitou and Timothy A. Jinam

Introduction

The Japanese Archipelago stretches over 4000 km from north to south, and is the homeland of three human populations; Ainus, Mainlanders, and Okinawans. The origins of these people have been studied for a long time. The standard theory based on craniofacial data is the "dual structure model" propagated by Japanese physical anthropologists, notably Hanihara (1991). According to this model, first migrants to the Japanese Archipelago came from somewhere in Southeast Asia in Upper Paleolithic age more than 30,000 years ago, and were probably ancestors of the Jomon people. The second wave of migration is believed to have taken place later in the Yayoi period (3000-1700 BP), where the people arrived from Northeast Asia. Indigenous Jomon people and new migrants in and after Yayoi period gradually mixed with each other. This model provides a reasonable explanation for the morphological similarity between the Ainu people of Hokkaido at the northernmost main island of Japanese Archipelago, and the Okinawan people in Southwest Archipelago, despite of large geographical distance. Similarity of these peoples was already noticed more than one hundred years ago by von Baelz (1911) who proposed the Ainu-Okinawan common origin theory.

We recently determined genome-wide single-nucleotide polymorphisms (SNPs) for Ainus and Ryukyuans (see Figure 1), and analysed these data with existing data sets (Japanese Archipelago Human Population Genetics Consortium 2012; Jinam *et al*. 2015). Major findings were: (1) Recent admixture with Mainlanders was observed for more than one third of Ainu individuals; (2) Ainus and Okinawans are tightly clustered with a high statistical support followed by Mainlander and Korean in the phylogenetic tree of East Asian populations; (3) Ainu population probably experienced admixture with other populations distributed north of Hokkaido. We further analysed these data, and found that Ainu people are genetically different from Mainlanders living in Tohoku, northern part of the Honshu Island, and the estimated age of starting admixture between Ainu and Mainlander ancestral populations was ca. 6th centruy A.D. (Jinam *et al.* 2015). We also determined partial nuclear genome DNA sequences of Jomon period people from DNA extracts used for mitochondrial DNA haplotype determination (Kanzawa-Kiriyama *et al.* 2013), and found that the Ainu people are genetically closest to the Jomon people, and

Figure 1

the mainlanders are sandwiched between Jomon and East Eurasian continental populations (Kanzawa-Kiriyama *et al.,* 2016).

Another aspect is the genome diversity of Japanese Archipelago Mainlander populations. Yamaguchi-Kabata *et al.* (2008) reported genome-wide SNP data analysis of 7,000 Japanese, and they computed Fst distances among people from seven geographical areas (see Figure 1). Figure 2 is a phylogenetic network based on these genetic distances using Neighbour-net (Bryant and Morton 2004). The Okinawan population is far apart from the remaining six populations, as expected. Kyushu area, which is geographically closest to Okinawa, is also genetically closest to Okinawa. Interestingly, Tohoku area, which is
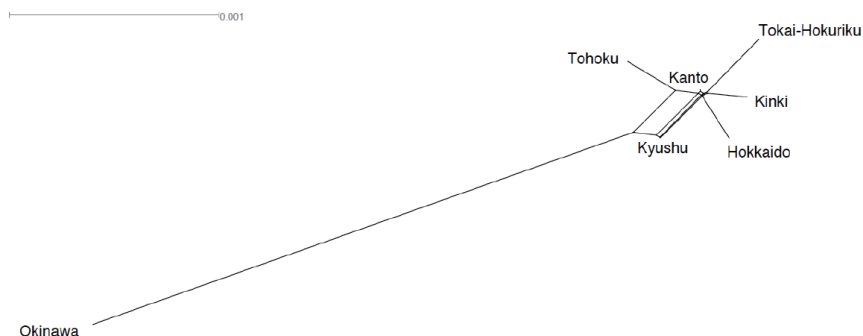
Figure 2

geographically far apart from Okinawa, shares a short split with the Okinawa population. If modern Tohoku people inherited DNAs of the Emishi who lived in Tohoku area more than 1,000 years ago, this pattern suggests a genetic affinity between the Emishi people and the ancestral Okinawans.

Nakaoka *et al.* (2013) analysed HLA data of ten regions in Japan, and found various interesting diversity among these regions. First of all, Okinawa was quite different from the remaining nine regions. Shikoku and Hokuriku were outliers and were different with each other. Shikoku was somewhat closer to Okinawa, while Hokuriku was most distant from Okinawa.

We recently examined the genome-wide SNP data of the Izumo population who are distributed relatively close to East Korea (see Figure 1). It should be noted that Izumo area is in the Chugoku-Shikoku District which was not examined by Yamaguchi-Kabata *et al.* (2008). Figure 3 is a Principal Component Analysis based on preliminary 21 Izumo individual data with four other East Asian populations: Tokyo (Nishida *et al.* 2008), Korea (Bae *et al.* 2011), North China (HapMap Consortium, 2005), and South China (Teo *et al.* 2009). The PCA coordinates of three continental populations (Koreans, Northern Chinese, and Southern Chinese) and Japanese Mainlanders in Tokyo shown in Figure 3 are in good agreement with their geographical relationship. In contrast, Izumo people are genetically more apart from Koreans than Japanese Mainlanders in Tokyo, although Izumo is geographically between Tokyo and Korea. The PCA plot of Figure 3 suggests the existence of some heterogeneous migrants to the Japanese Archipelago Mainland.
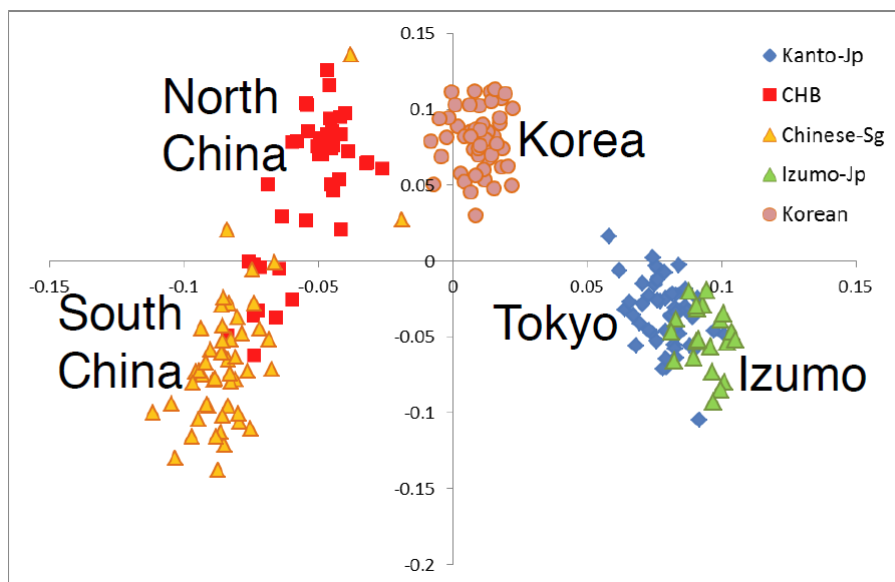
Figure 3

By combining these analyses with our new analysis on Ainu and other populations in the Japanese Archipelago, a new scenario on the evolutionary history of people on the Japanese Archipelago is emerging. This is shown in Figure 4, and we would like to name this as "the inner dual structure" model of the Japanese Mainlanders. Two phases of migrations are assumed to form the current Japanese Mainlanders in this model. Phase 1 is the earlier migrants who possibly arrived at the Japanese Archipelago in the late Jomon period (~4000 years ago), while phase 2 is later migrants who came to Northern Kyushu and later spread eastward. Phase 2 migrants initially brought paddy field rice agriculture around 3000 years ago, which started the Yayoi period. This phase 2 migration trend is assumed to continue to historical times of Japan. Saitou (2015) proposed "three migration waves model" based on this inner dual structure model for explaining the formation of Japanese Archipelago people.

It may be interesting to compare this new scenario based on DNA data with linguistic data. Corresponding to three genetically defined human populations on this Archipelago, there are three languages; Ainu, Japanese, and Ryukyuan. The Ryukyu language is clearly close to the Japanese language, and they were once considered to be just dialect differences. Lee and Hasegawa (2011) analysed linguistic data of the Japanese and Ryukyuan that were compiled by Hirayama (1988, 1992), and concluded that the common ancestral language of the Japanese
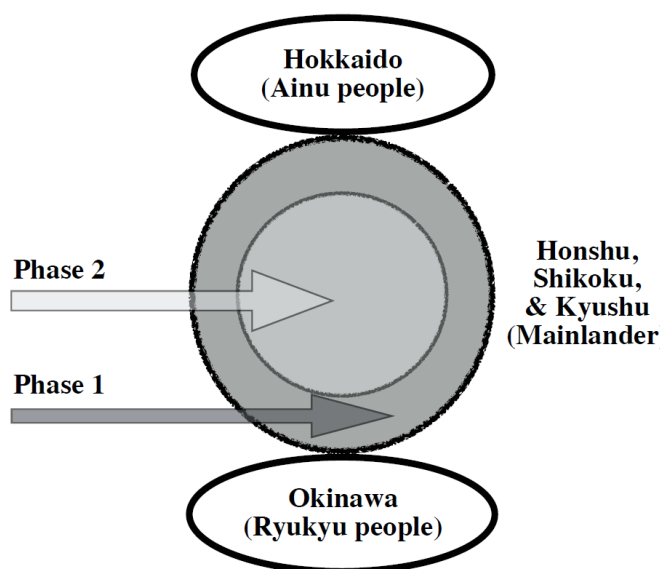
Figure 4

and Ryukyuan was introduced to the Japanese Archipelago with introduction of rice agriculture. This was based on their estimated divergence time of ~2,000 years ago between the Japanese and Ryukyuan languages. Lee and Hasegawa (2011) criticised other hypotheses such as affinity of the Japanese language with Austronesian and with Altaic languages. They never paid attention to the seminal work of Robbetts (2005) on the Altai hypothesis.

They used BEAST software (Drummond and Rambaut 2007) whose mathematical framework was similar to that used by Gray and Atkinson (2003) and Bouckaert *et al.* (2012), who both supported the Anatolian origin hypothesis of the Indo-European language. Recently, however, ancient DNA analysis of European individuals strongly supported the classic Steppe hypothesis (Haak *et al.* 2015). If this is true, there was something wrong in the mathematical model used by these studies on language phylogenies.

One clear drawback on the phylogenetic tree model is that there are often reticulations in linguistic data. In fact, Lee and Hasegawa (2011) showed phylogenetic networks of Ryukyuan and Japanese dialects using Neighbour-Net (Bryant and Moulton, 2004), and there were so many reticulations in Japanese dialects. We thus conduct more thorough phylogenetic network analysis of linguistic data used by Lee and Hasegawa (2011) in this paper.

Materials and Methods

We downloaded supplementary data of Lee and Hasegawa (2011), and computed Manhattan distances from their 0/1 data. When missing data (indicated as "?") were found in some words, that word was omitted from distance computation. In total, we used 159 word data for linguistic distances after omitting 52 words which included at least one piece of missing data. There were 49 words that had identical 0/1 patterns, meaning that the remaining 110 words contributed to the observed distances. There were 15 words which showed singleton patterns. Therefore, in terms of cladistic thinking, 95 words were phylogenetically informative; see Saitou (2014) for this concept. Distance matrices are available upon request. Phylogenetic networks were drawn by using SplitsTree4 (Huson and Bryant, 2006) based on the Neighbour-Net algorithm (Bryant and Moulton, 2004).

Results

Figure 5 shows the phylogenetic network of 10 Ryukyuan dialects. This figure is essentially similar to Figure 4B of Lee and Hasegawa (2011). Split *a* clusters dialects in Okinawa and Naze, and splits *b* and *c* cluster dialects of Miyako and Ishigaki Islands, respectively. The remaining two major splits *d* and *e* are reticulative. Split *d* clusters Hateruma with Taketomi and Ishigaki Islands, while split *e* clusters Hateruma with Yonaguni Island.

Figure 6 is a phylogenetic network for seven Kyushu prefectural dialects. Split *a* separates northern prefectures (Fukuoka, Saga, and Nagasaki) from southern four prefectures, and splits *b* and *c* cluster Miyazaki-Oita and Miyazaki-Kagoshima, respectively. Miyazaki Prefecture is geographically between Kagoshima and Oita Prefectures, and coexistence of splits *b* and *c* is reasonable. Split *d* clusters Oita and Fukuoka Prefectures which are geographically adjacent. The four major splits (*a-d*) are all consistent with geographical propinquity of prefectures.

Figure 7 is a phylogenetic network for seven Chugoku and Shikoku prefectural dialects. As we saw in Figure 6, this network also shows dialect clusters which are geographically close; split *a* clusters two San-in district prefectures (Tottori and Shimane), and split *c* further clusters these two prefectures with Yamaguchi prefecture which is adjacent to Shimane prefecture. Two trios of Shikoku district prefectures, one is Kagawa, Kochi, and Ehime and the other is Tokushima, Kagawa, and Kochi, are separated from the remaining prefectures with split *b* and *d*, respectively.

There are only three major splits in Figure 8, in which the dialect network of seven Kinki district prefectures are shown. Split *a* separates the central part of
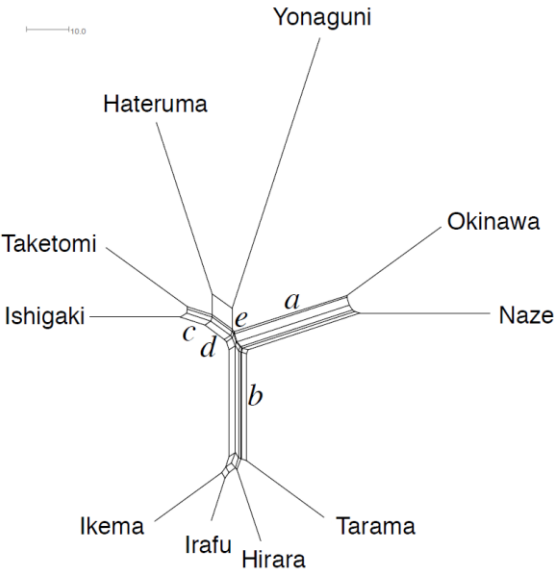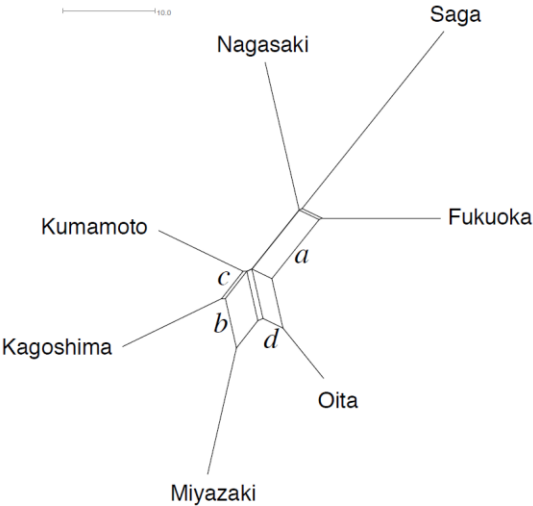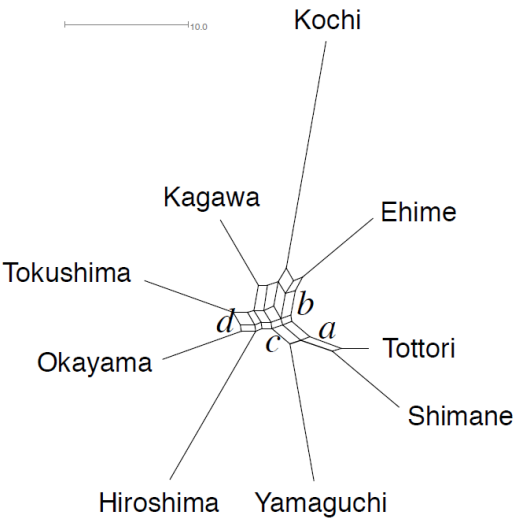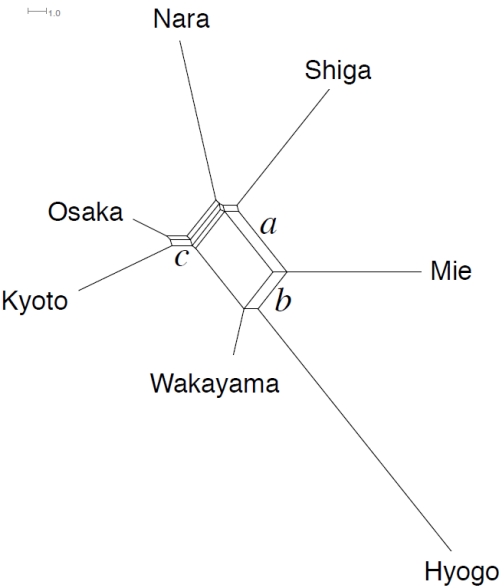
Figure 5

Figure 6

Figure 7

Figure 8

the Kinki district (Kyoto, Osaka, Nara, and Shiga) from peripheral ones (Wakayama, Mie, and Hyogo), while split *b* seems to correspond to west and east regions in the Kinki District. Finally, split c clusters dialects of Kyoto and Osaka, which were the cultural centers of Japan for a long time.

The phylogenetic network for the Chubu district dialects shown in Figure 9 is highly reticulated, and some splits show clusters in which their member prefectures are not geographically adjacent. The longest split *a* clusters three Hokuriku district prefectures (Fukui, Ishikawa, and Toyama), while the second longest split *b* clusters Yamanashi, Aichi, and Gifu. Although Aichi and Gifu prefectures are adjacent, Yamanashi is not. Among the remaining six splits (*c-h*), those compatible with geographical adjacency are splits *e*, *f*, and *h*. Split *c* clusters dialect of the Hachijo Island with that of Toyama prefecture. Hachijo Island dialect is well known to be quite distinct from all the other Japan Mainland dialects, but this network suggests that it shares some elements with Toyama dialect. It should be noted that Hachijo Island is on the Pacific, while Toyama is facing the Japan Sea. Fukui and Yamanashi dialects are clustered through split *d*, and Gifu and Shizuoka dialects are clustered through split *g*. If these splits are not statistical artefacts, some non-geographical factors are involved in forming these splits.

Phylogenetic networks of Kanto and Tohoku (including Hokkaido) districts are shown in Figures 10 and 11, respectively. Splits for these two district dialects are mostly representing geographical adjacency. However, there are some exceptions; split *a* of Figure 11 clusters Fukushima and Hokkaido, and split *d* clusters Miyagi and Aomori.

We now move to compare dialects of two geographically adjacent districts. Figure 12 is a phylogenetic network of Ryukyuan language dialects and Japanese language dialects in Kyushu district. There are two short but interesting splits in this figure. Yonaguni and Hateruma dialects are clustered with Saga prefecture dialect at split *a*, and these two Ryukyuan dialects are clustered with all Kyushu dialects at split *b*. This suggests some linguistic influence from Northern Kyushu to these two islands. Figure 13 is a phylogenetic network of Kyushu and Chugoku-Shikoku district dialects. Split *a* clusters Kyushu dialects except for Nagasaki, and Nagasaki dialect is clustered with Setouchi area dialects (Hiroshima, Okayama, and Tokushima) at split *b*. This suggests that Nagasaki dialect was influenced by Chugoku-Shikoku district dialects, especially by Setouchi dialects.

Figure 14 is a phylogenetic network of Chugoku-Shikoku and Kinki districts. Split *a* clusters five Kinki dialects, while Hyogo dialect clusters with Okayama dialect at split b. This clustering is understandable, for Hyogo and Okayama prefectures are geographically adjacent. However, clustering of Mie
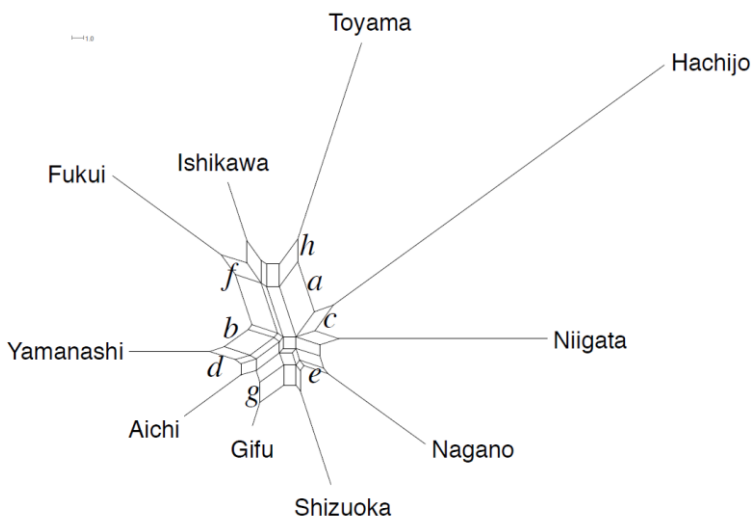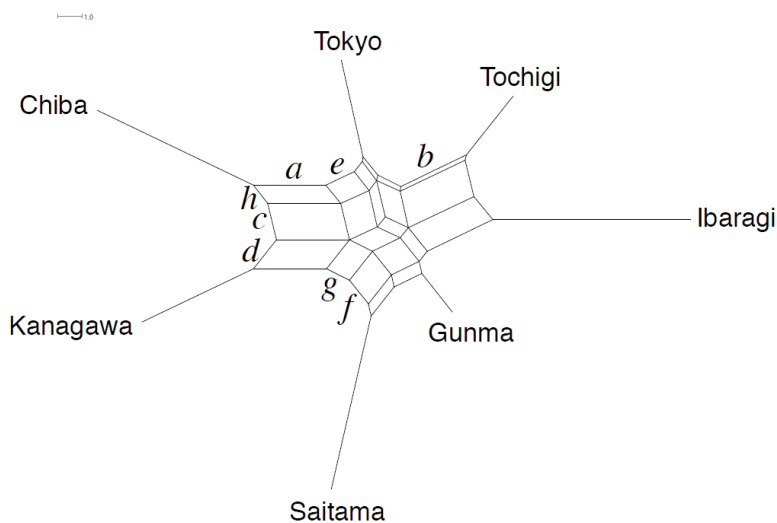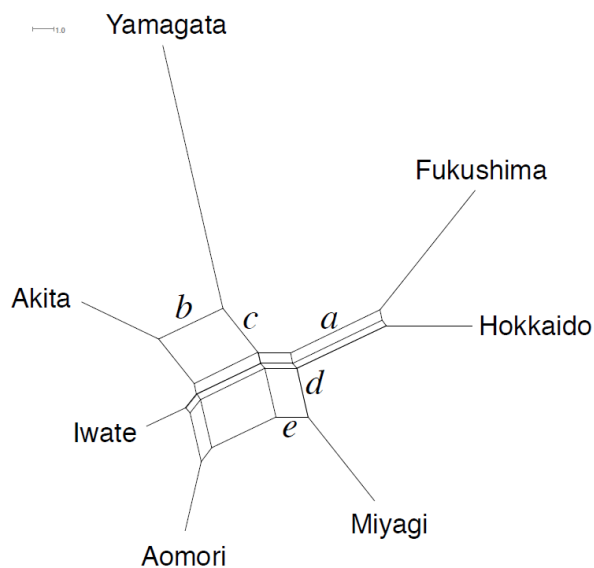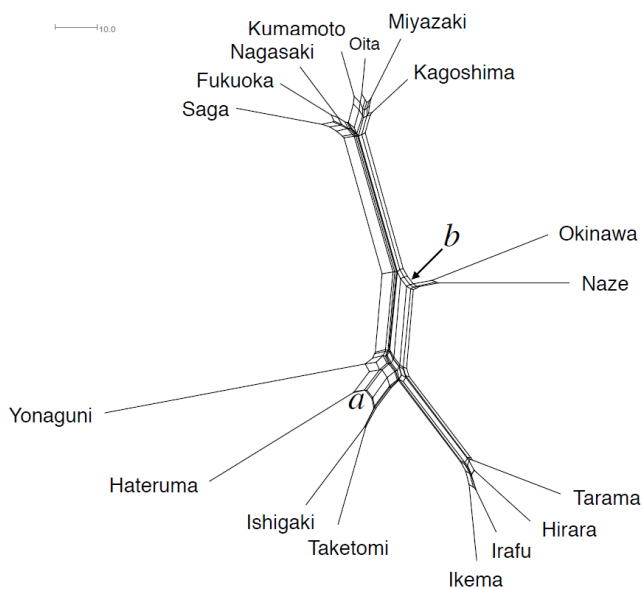
Figure 9

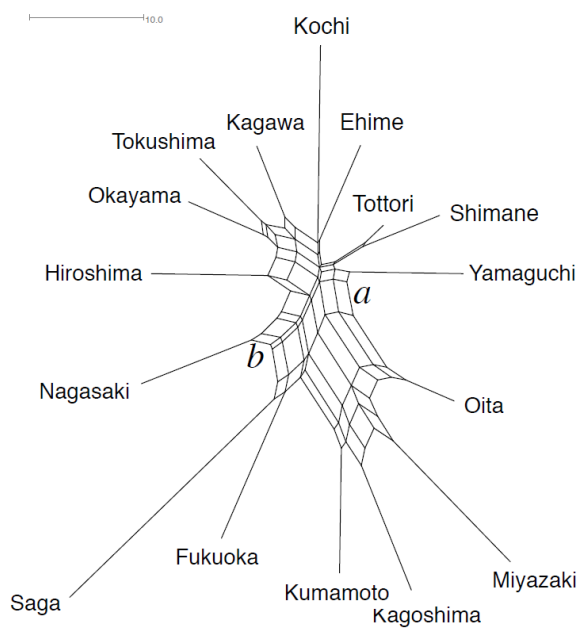**Figure 10**
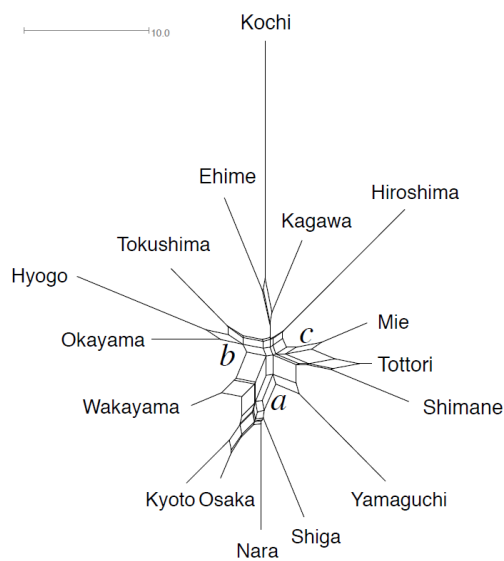
Figure 11

Figure 12
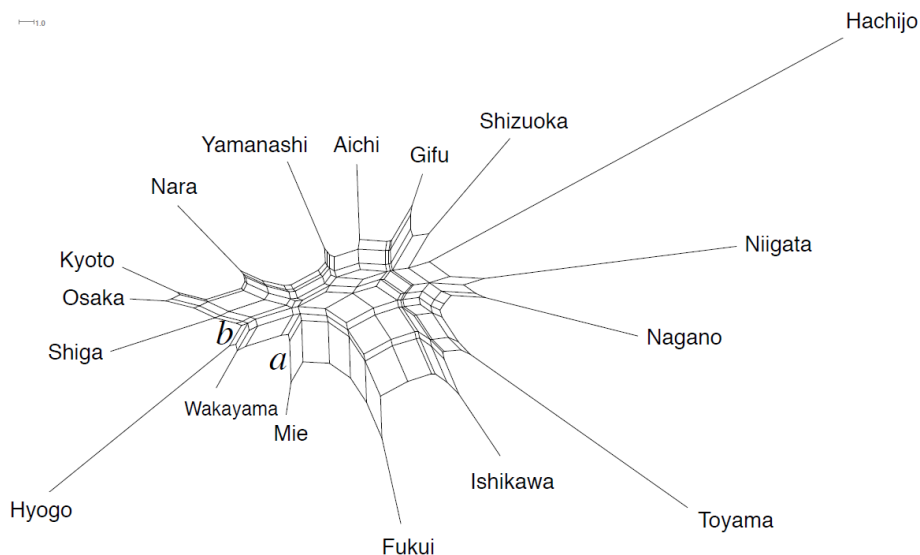
Figure 13

Figure 14

Figure 15

and Tottori dialects at split *c* is puzzling. The unique status of Mie dialect is also found in Figure 15 in which dialects of Kinki and Chubu were compared. Mie is now clustered with Fukui dialect at split *a*, while it also clustered with Wakayama and Hyogo dialects at split *b*.

Dialects of Chubu and Kanto districts are compared in Figure 16. Dialects of these two districts are separated at split *a*, while Shizuoka is clustered with Ibaraki at split *b* and Tokyo, Yamanashi, and Aichi are clustered at split *c*. Figure 17 is a phylogenetic network of Kanto, Tohoku, and Hokkaido dialects. All Kanto dialects are clustered at split *a*, while Tokyo is outside of the cluster defined by split *b* and Fukushima is now included in the cluster. Split *c* defines Tohoku dialects except for Fukushima, and Hokkaido dialect is clustered with dialects of Kanto district and Fukushima prefecture at split *c*. This intermediate location of Hokkaido dialect between Kanto and Tohoku districts probably reflects the cultural influence to Hokkaido both from Tohoku and from Kanto.

So far, only modern dialects were compared. We would like to add Old Japanese and Middle Japanese to our comparison. Figures 18-24 are phylogenetic networks of these two historical Japanese and dialects of seven districts. In all cases, Old Japanese and Middle Japanese were clustered. Some dialects showed closer relationship with these historical Japanese. Among Ryukyuan dialects, Okinawa and Naze dialects are clustered with historical
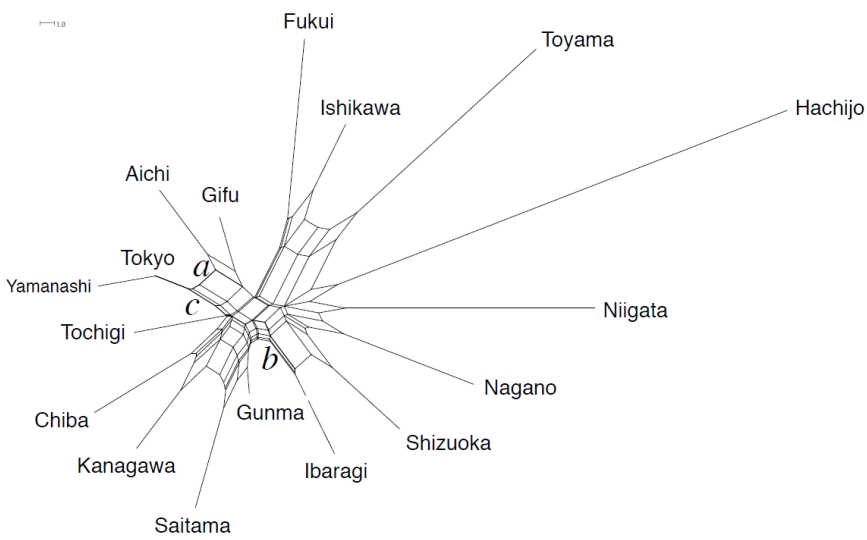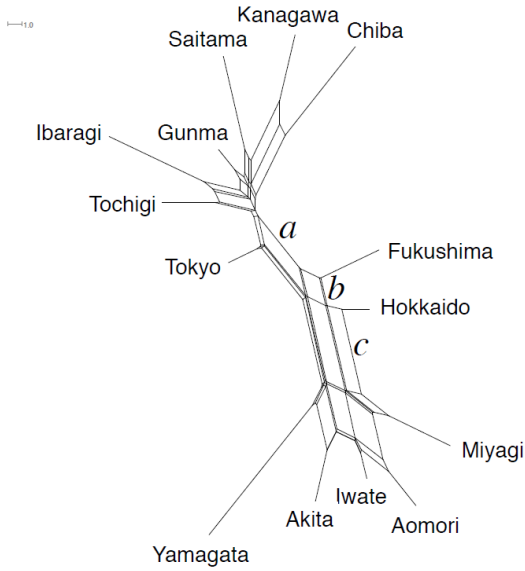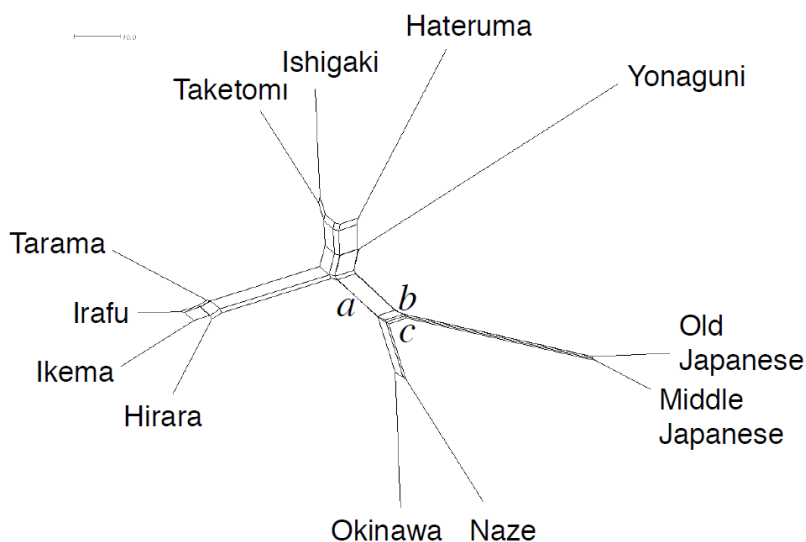
Figure 16

Figure 17

Figure 18

Japanese at split *a* of Figure 18. Naze dialect is slightly closer to historical Japanese, for it shares a short split *b* with them. Historical Japanese also cluster with Hateruma and Yonaguni dialect at split *c*. This split probably corresponds to split *b* of Figure 12.

Nagasaki dialect is clustered with historical Japanese at split *a* of Figure 19, and Miyazaki and Oita dialects cluster with Old Japanese at split *b*. Figure 20 shows that historical Japanese are quite different from Chugoku-Shikoku dialects, yet Kochi and Ehime dialects showed a weak similarity with them as shown at split *a*. Kinki dialects are also quite different from historical Japanese as shown in Figure 21, yet Mie and Nara dialects are somewhat similar to these historical Japanese at split *a*.

Hachijo Island dialect is well known among those who study Japanese dialects to be quite unique and keeps old words. Split *a* of Figure 22 confirms this pattern. Yamanashi dialect is also somewhat similar to historical Japanese as shown at split *b*. The phylogenetic network of Kanto dialects and historical Japanese (Figure 23) shows that Tokyo dialect is somewhat similar to historical Japanese at split *a*. The same situation can be found for Hokkaido dialects at Figure 24, in which Tohoku and Hokkaido dialects were compared with historical Japanese.

Figure 19

Figure 20
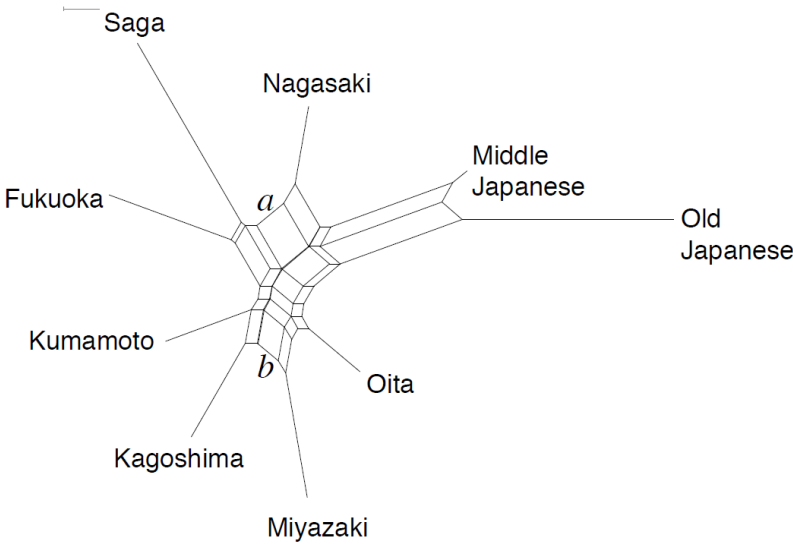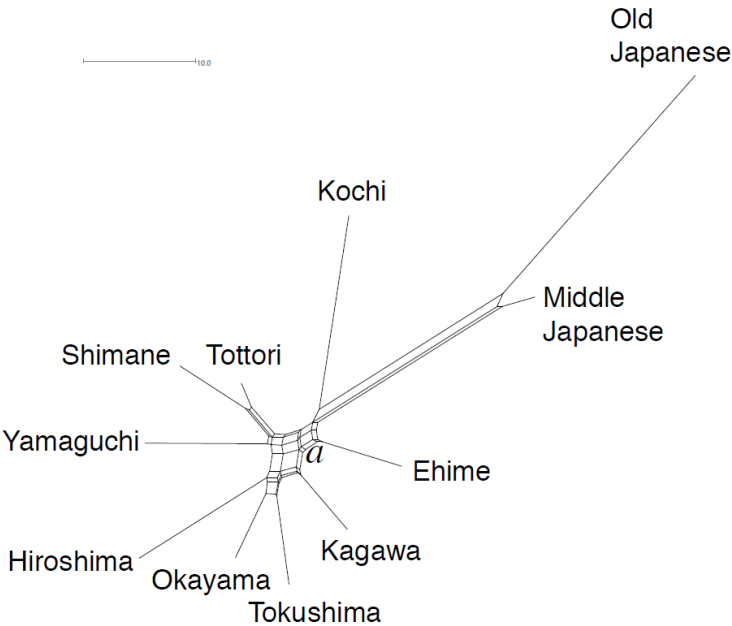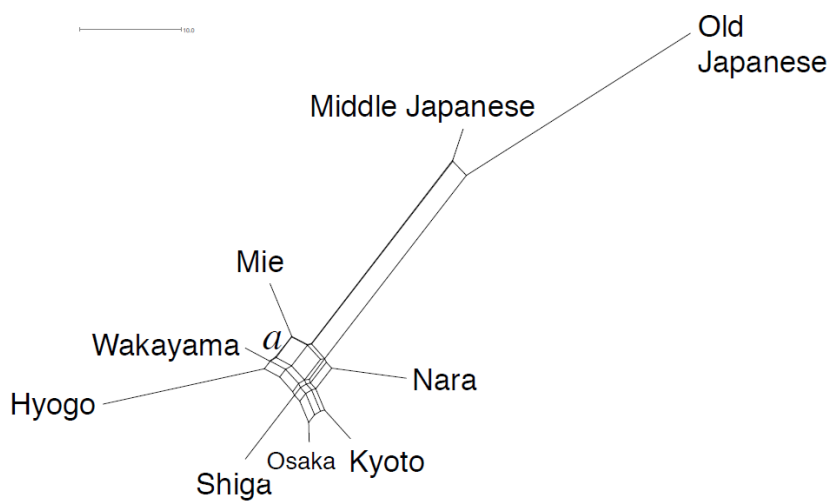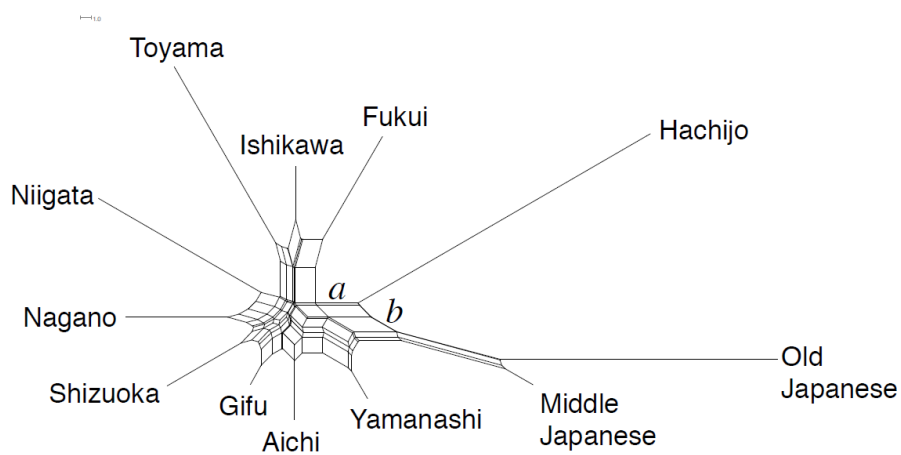
Figure 21

Figure 22

Saitama
Ibaragi
Gunma
Tochigi
Kanagawa
a
Chiba
Tokyo

Middle
Japanese
Old
Japanese

Figure 23

Middle
Japanese
Old
Japanese

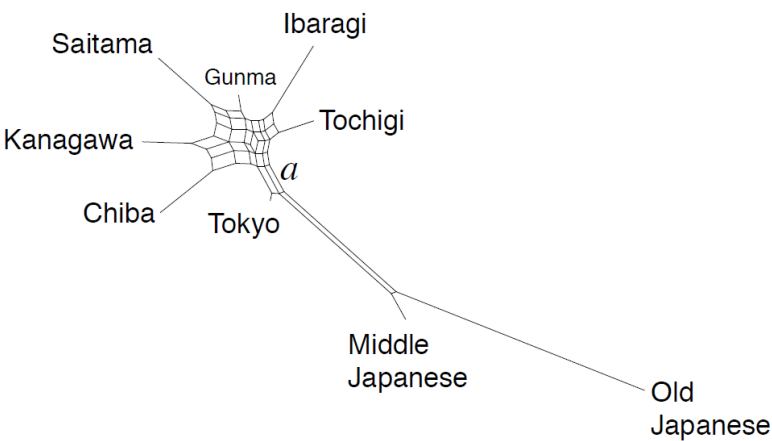Fukushima   Hokkaido
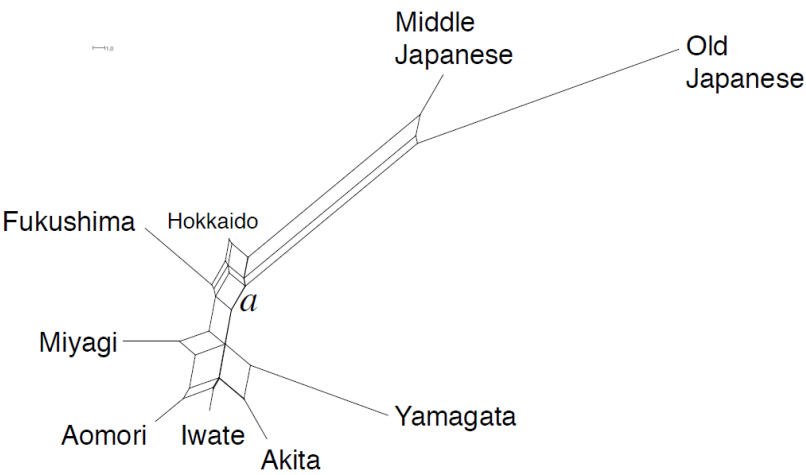
a

Miyagi

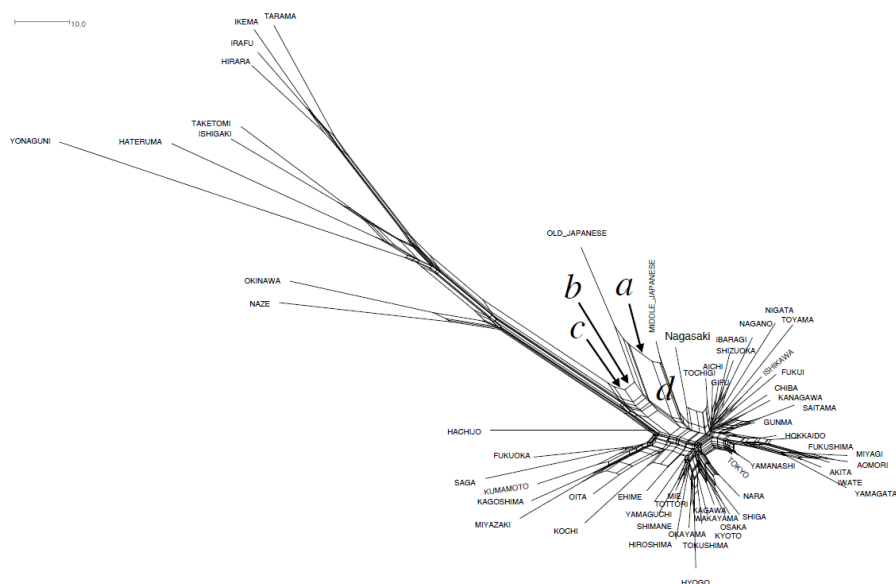Aomori   Iwate
Yamagata
Akita

Figure 24

Figure 25

Finally, we constructed a phylogenetic network of Ryukyuan dialects, modern Japanese dialects, and two historical Japanese, as shown in Figure 25. Ryukyuan dialects are clearly more diverged with each other than Japanese dialects. Hachijo island dialect is located at a unique position among Japanese dialects; it is located between Ryukyuan and Japanese. Old Japanese is more closely related to Ryukyuan language than Middle Japanese as depicted by split *a*. Split *b* separates historical Japanese, Nagasaki dialect, and dialects of eastern Japan (Chubu, Kanto, Tohoku, and Hokkaido) from the rest, while split *c* separates Ryukyu dialects, Hachijo island dialect, and Kyushu dialects except for Nagasaki dialect from the rest.

Discussion

The Japanese Mainland and Okinawa Islands has been populated for thousands of years. Therefore, if the original languages spoken in the Jomon period evolved to Japanese and Ryukyuan languages, they are expected to have diverged more than 3,000 years ago. However, Japanese and Ryukyu languages are much more closely related. Hattori (1959) estimated the divergence of these two languages (he used the term "dialects") to be around 1500 years ago (5th Century A.D.) by applying the glottochronological technique which was popular in 1950s. Lee and Hasegawa (2011) estimated that divergence to be around 2200 years ago.

Although their new estimate, which applied a new mathematical framework and linguistic data, is more than 500 years older than Hattori's estimate, its 95% confidence interval was 1200 - 4200 years ago, and Hattori's estimate is within this range.

Glottochronological studies were heavily criticised in linguistics, and a new Bayesian method seems to be also problematic. When Drummond and Rambaut (2007) intoduced a new software called BEAST, they used 17 dengue virus env protein amino acid sequences reported by Lanciotti *et al.* (1997) as an example. The env protein data consisted of 495 amino acid sequences with no gap, and those 17 viral strains were collected between 1956 and 1994. The genome of dengue virus is RNA, and the evolutionary rates of RNA viruses are usually very high, so only scores of years are enough for viral proteins to accumulate amino acid changes. One of us once conducted influenza virus (also RNA genome) sequence analysis and estimated its evolutionary rates (Saitou and Nei 1986). Therefore, basic assumptions used in BEAST is understandable. Lee and Hasegawa (2011) applied BEAST to linguistic data of only 110 words, and the divergence time estimation was based on only one time period; between Old Japanese and Middle Japanese. Linguistic data used by Lee and Hasegawa (2011) were quite scanty compared to molecular data which are expected for BEAST applications.

When Hattori (1959) estimated the divergence time between Ryukyu and Japanese languages, he implicitly assumed the constancy of the word changing rate. If we use the neighbor-joining method (Saitou and Nei 1987) to the same distance matrix used for the phylogenetic network shown in Figure 25, a tree shown in Figure 26 was produced. MEGA6 (Tamura *et al.* 2013) was used to draw that tree. If the root of this tree is between Japanese dialect cluster and Ryukyuan dialect cluster, the rate of word change is clearly faster in Ryukyu dialects than Japanese dialects.

The branching pattern of this unrooted tree is different from that of Figure 2 tree of Lee and Hasegawa (2011) in many respects. Middle Japanese and modern Japanese dialects were clustered in their tree (corresponding to split *a* of Figure 25), while middle Japanese clustered with Old Japanese in the neighbour-joining tree. The latter clustering was also observed in split *d* of Figure 25. Another noteworthy difference is the placement of Hachijo island dialect. It is outside of all modern Japanese dialects and historical Japanese in Figure 26, and this pattern is consistent with the unique and old feature of this dialect. In contrast, Hachijo island dialect was clustered with Shizuoka dialect in the tree shown by Lee and Hasegawa (2011). Another puzzling pattern of their BEAST-produced tree is the clustering of Hokkaido, Yamanashi, and Tokyo dialects which are basal to all the other modern Japanese dialects. In the neighbour-joining tree, Hokkaido dialect is within the Tohoku cluster, and Tokyo and Yamanashi dialect
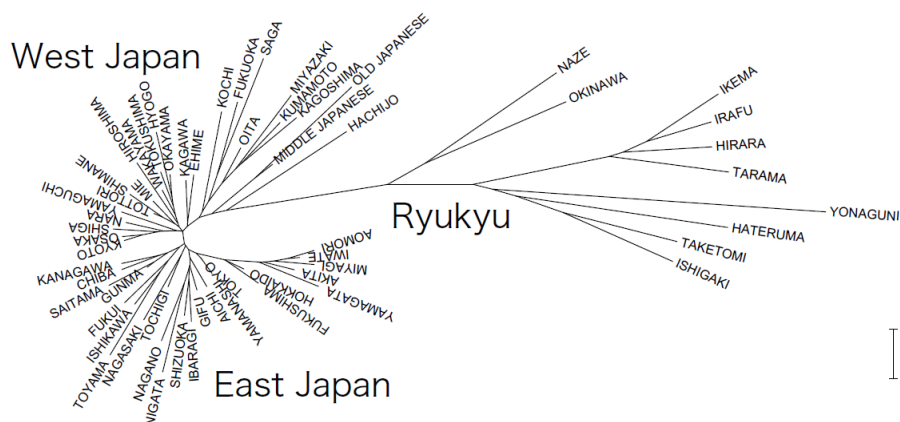
Figure 26

cluster is just outside of this Tohoku-Hokkaido cluster. They further cluster with other Kanto and Chubu dialects and formed East Japan dialect cluster.

It is generally accepted in molecular phylogenetic studies that tree topologies may be acceptable while the divergence time estimates frequently have various problems even when we have relatively large molecular sequence data. If the original molecular data are the outcome of a tree-like evolutionary history, reconstructed trees are often concordant irrespective of methods used. The existence of many discrepancies between the neighbour-joining tree (Figure 26) and the BEAST tree (Figure 2 of Lee and Hasegawa 2011) suggests that the word data used include many reticulations. This feature is in fact shown in this paper. Therefore, we should be careful of any divergence time estimates not only for molecular data but for linguistic data. When tree topology is quite different among various tree-making methods, tree topology is also not easy to infer.

Language evolution on the Japanaese Archipelago can be interpreted by applying the elite dominance model in linguistics; people on the Ryukyu Islands once spoke language(s) totally different from the Japanese language. Later, their original language changed to the Japanese language by Mainlanders who influenced the Ryukyu people genetically and culturally. If we accept this assumption, the language shift possibly occurred when Kinki area was the

political center and when the government expanded their territory to Ryukyu Island. This may have occurred between 5th and 10th century A.D., which is much later than the divergence estimate of Lee and Hasegawa (2011).

In conclusion, linguistic difference between Japanese and Ryukyu languages may not reflect DNA difference between Mainlanders and Okinawans, for the language change in Okinawa probably occurred after the Phase 2 migration started. However, this language change coincided with gene flow from the Mainland to Ryukyu Archipelago that resulted in close genetic relationship between them, although the Okinawans still retained more Jomon genetic components.

Address for communication

Naruya Saitou and *Timothy A. Jinam,* National Institute of Genetics, Mishima, Japan.

References

Bae J. S., Cheong H. S., Kim J. H., Park B. L., Kim J. H., Park T. J. *et al.* (2011) The genetic effect of copy number variations on the risk of type 2 diabetes in a Korean population. PLoS One 6: e19091.

Bouckaert R. *et al.* (2012) Mapping the origins and expansion of the Indo-European language family. Science*,* 337: 957-960.

Bryant D. and Moulton V. (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution, 21: 255-265.

Drummond A. J. and Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7: 214.

Gray R. D. and Atkinson Q. D. (2003) Language-tree diver- gence times support the Anatolian theory of Indo- European origin. Nature, 426: 435 – 439.

Haak W. *et al.* (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature, 522: 207–211.

Hanihara K. (1991) Dual structure model for the population history of the Japanese. Japan Review, 2: 1–33.

Hattori S. (1959) *Nihongo no keitou*. Iwanami Shoten, Tokyo.

Hirayama T. (1988) *Minami Ryukyu no hogenkisogoi*. Ohfusya, Tokyo.

Hirayama T. (1992) *Gendai Nihongo daijiten*. Meiji Shoin, Tokyo.

Huson D. H. and Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution, 23: 254–267.

The International HapMap Consortium (2005) A haplotype map of the human genome. Nature, 437: 1299–1320.

Japanese Archipelago Human Population Genetics Consortium [Jinam T., Nishida N., Hirai M., Kawamura S., Oota H., Umetsu K., Kimura R., Ohashi J., Tajima A., Yamamoto T.,

Tanabe H., Mano S., Suto Y., Kaname T., Naritomi K., Yanagi K., Niikawa N., Omoto K., Tokunaga K.,   and Saitou N.] (2012) The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. Journal of Human Genetics, 57: 787-795.

Jinam T. A., Kanzawa-Kiriyama H., Inoue I., Tokunaga K., Omoto K., and Saitou N. (2015). Unique characteristics of the Ainu population in Northern Japan. Journal of Human Genetics, 60: 565-571.

Jinam T. A., Kanzawa-Kiriyama H., and Saitou N. (2015) Human genetic diversity in the Japanese Archipelago: dual structure and beyond. Genes and Genetic Systems, 90: 147–152.

Kanzawa-Kiriyama H., Kryukov K., Jinam T. A., Hosomichi K., Saso A., Suwa G., et al. (2016) A partial nuclear genome of the Jomons who lived 3,000 years ago in Fukushima, Japan. Journal of Human Genetics, advance online publication.

Kanzawa-Kiriyama H., Saso A., Suwa G., and Saitou N. (2013) Ancient mitochondrial DNA sequences of Jomon teeth samples from Sanganji, Tohoku district, Japan. Anthropological Science, 121: 89-103.

Lanciotti R. S., Gubler D. J., and Trent D. W. (1997) Molecular evolution and phylogeny of dengue-4 viruses. Journal of General Virology, 78: 2279-2284.

Lee S. and Hasegawa T. (2011) Bayesian phylogenetic analysis supports an agricultural origin. Proceedings of Royal Society B, 278: 3662-3669.

Nakaoka H., Mitsunaga S., Hosomichi K., Shyh-Yuh L., Sawamoto T., Fujiwara T., Tsutsui N., Suematsu K., Shinagawa A., Inoko H. *et al.* (2013) Detection of ancestry informative HLA alleles confirms the admixed origins of Japanese population. PLoS One, 8: e60793.

Nishida, N., Koike, A., Tajima, A., Ogasawara, Y., Ishibashi, Y., Uehara, Y. *et al.* (2008) Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. BMC Genomics, 9: 431–431.

Robbetts M. I. (2005) *Is Japanese related to Korean, Tunguisic, Mongolic and Turkic?* Harrossowitz Verlag, Wisbaden.

Saitou N. (2008) in *Simulations, Genetics and Human Prehistory* (eds Matsumura, S., Forster, P., Renfrew, C.) 89-92. (McDonald Institute Monographs, Cambridge.

Saitou N. (2014) *Introduction to Evolutionary Genomics.* Springer, London.

Saitou N. (2015) *History of Japanese Archiplelago people.* Iwanami Shoten, Tokyo.

Saitou N. and Nei M. (1986) Polymorphism and evolution of influenza A virus genes. Molecular Biology and Evolution, 3: 57-74.

Saitou N. and Nei M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4: 406-425.

Tamura K. *et al.* (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution, 30: 2725-2729.

Teo Y.Y., Sim X., Ong R.T., Tan A.K., Chen J., Tantoso E. *et al.* (2009) Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. Genome Research, 19: 2154-2162.

von Baelz E. (1911) Die Riu-Kiu-Insulaner, die Aino und andere kaukasier-ahnliche Reste in Ostasien. Korres. Blatt. Dtsch. Ges. Anthrop. Ethnol. Urgesch., 42: 187-191.

Yamaguchi-Kabata Y., Nakazono K., Takahashi A., Saito S., Hosono N., Kubo M., Nakamura Y., and Kamatani N. (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. American Journal of Human Genetics, 83: 445–456.