Genetic analysis of Taishu horses on and off Tsushima Island: Implications for conservation

Ayumi TEZUKA1, *, Masaki TAKASU2, Teruaki TOZAKI2,3 and Atsushi J. NAGANO1
1Faculty of Agriculture, Ryukoku University, Shiga 520-2194, Japan
2Department of Veterinary Medicine, Faculty of Applied Biological Sciences, Gifu University, Gifu 501-1193, Japan
3Genetic Analysis Department, Laboratory of Racing Chemistry, Tochigi 320-0851, Japan

Taishu horses are a native Japanese breed, of which only 41 individuals remained on Tsushima Island in 2018. Their genetic diversity is considered lower than that of other Japanese native horse breeds; thus, it needs to be investigated for sustainable conservation of this breed. Historical records revealed that several Taishu individuals were released areas off-Tsushima Island in mid-1980s. At present, Taishu horses living outside of Tsushima Island, hereafter referred to as Non-Tsushima Taishus (NTTs), are tagged. However, the genetic structure of the NTT individuals remains unclear, and such individuals are not included in the current mating plans for Taishu horses. Herein, we examined the genetic structure of 18 NTT individuals by comparing their genomic (SNP) information with that of individuals on Tsushima Island (TT), four other native Japanese breeds, and one Anglo-Arabian breed by using ddRAD-seq. We found that all individuals related to the Taishu can be grouped in one cluster, which was separated from other horse breeds. Patterns of specific and shared SNPs in NTT individuals closely resembled those of TT individuals, suggesting very minor genetic differences. Meanwhile, the heterozygosity of NTT individuals was slightly higher than that of TT individuals, and many NTT individuals were of fertile age, suggesting that the pedigree of NTT individuals would be useful in breed conservation plans for Taishu horses. Based on their genomic information, we also reconstructed the pedigree structures of four NTT individuals with no family information. The inclusion of NTT individuals in future mating plans on Tsushima Island may be an effective and feasible method for conserving the Taishu horse breed in Japan.

Key words: ddRAD-Seq, genetic diversity, Japanese native horse, Taishu
been used in the mating and conservation plans of the Taishu in Japan. If NTTs were in the same genetic group as Taishu horses on Tsushima Island (hereafter referred to as Tsushima Taishus, or TTs), NTTs may be useful in future mating and conservation plans for the Taishu.

Our study aimed to examine the genetic structures of NTT individuals based on their genomic information relative to those of TTs, four other Japanese native breeds, and one Anglo-Arabian breed obtained in one of our previous studies (Ayumi Tezuka, unpublished results) [10] and to assess whether NTT individuals can be used in future mating and conservation plans for the Taishu in Japan. In this study, we used double digest restriction-site associated DNA sequencing (ddRAD-seq), which is favorable for examining the genetic structure of the Taishu for various reasons, including the fact that it is more cost-effective compared with whole-genome sequencing and single nucleotide polymorphism (SNP) chips. Although SNP chips designed for horses can provide tens of thousands of accurate SNPs, it is possible that a certain number of SNPs do not include those of the Taishu. Moreover, comparable data from a previous study (Ayumi Tezuka, unpublished results) were obtained by using ddRAD-seq.

Materials and Methods

Sample collection and DNA extraction

We collected fresh blood samples from 18 NTT individuals of so-called Japanese native Taishu horses (Equus caballus) living off Tsushima Island (Figs. 1 and 2). The blood samples were collected in a tube containing EDTA and kept at −20°C prior to DNA extraction. Total genomic DNA was extracted from whole blood using a Maxwell 16 Blood DNA Purification Kit (Promega, Madison, WI, U.S.A.). The blood samples of the other horse breeds were collected by the same method (Ayumi Tezuka, unpublished results). This research was conducted according to the regulations of the Animal Care and Use Committee of Ryukoku University, which provides Guidelines for Animal Care and Use in Research (https://www.ryukoku.ac.jp/research/ethics/animal.html).

The previous study used DNA samples from 9, 5, 5, 6, and 5 individuals of Miyako, Yonaguni, Kiso, Hokkaido, and Anglo-Arabian breeds, respectively (Ayumi Tezuka, unpublished results). The genetic similarity between the actual introgressed Anglo-Arabian individuals and the studied Anglo-Arabian individuals is not clear. Considering that there is only one existing Anglo Arabian breed in Japan, the studied breeds in this study were representative of all Japanese horse breeds. A total of 86 individuals were analyzed in the following analysis.

Library preparation and sequencing in ddRAD-Seq

We prepared a ddRAD-seq library, as described by Ayumi Tezuka (unpublished results). Accordingly, 50-bp sequences of DNA fragments were read using a HiSeq 2500 system (Illumina, San Diego, CA, U.S.A.), which was performed by Macrogen. The sequenced reads were demultiplexed by CASAVA 1.8.2 (Illumina). Fastq files were deposited into the National Center for Biotechnology Information Sequence Read Archive (accession no. PRJNA498901).

SNP calling

The genomic information of NTTs was analyzed and compared with those of TTs, four Japanese native breeds, and one Anglo-Arabian breed, which were obtained on one of our previous studies (DNA Data Bank of Japan Sequence Read Archive, accession no. DRA007047) (Ayumi Tezuka, unpublished results).

Poor-quality reads containing low-confidence bases and adapter sequences from the raw sequence reads were removed using the Trimmmomatic software, version 0.33 [3]. Next, the trimmed reads were mapped to the thoroughbred reference genome sequence (EquCab2) using the Bowtie2 software [8]. SNPs were called using the Stacks software, version 2.3 [5]. This process was performed according to the default settings of the pipeline ref_map.pl in Stacks (population analysis, 75% minimum percentage of individuals in a population required to process a locus [-r 0.75]). All SNPs obtained by Stacks were filtered by Plink ver. 1.9 [2]. All SNPs were screened according to the following criteria: SNPs with a minor allele frequency (MAF) below 0.05, those mapped on X and Y, and those on undetermined chromosomes were removed. A total of 10,471 SNPs were used for downstream analysis.

Estimating genetic structure and genetic status of the Taishu population

To clarify whether NTTs and TTs are grouped in the same cluster, we constructed a phylogenetic tree from the genome-wide ddRAD-Seq dataset using the maximum likelihood (ML) method in the SNPhylo software, version 20160204 [9]. SNPhylo keeps only one informative SNP in a linkage disequilibrium (LD) block. To reduce the times required to construct phylogeny, 4,532 SNPs were selected by SNPhylo (-r -b 1000 -A). We applied a rapid bootstrap algorithm with 1,000 replicates to each data set. The resulting tree was plotted using the MEGA software, version 7 [7].

We constructed multidimensional scaling (MDS) plots based on the genome-wide identity-by-state (IBS) pairwise distance matrix and genetic structure identified with a model-based clustering approach in Admixture [1]. We ran Admixture from K=2 to 7, and the optimal number of
clusters (K-values) showing the lowest cross-validation error was identified [1].

We chartered 10,471 SNPs to assess the genetic independence of NTT individuals. We defined SNPs that were found in one population as breed-specific SNPs and SNPs that were found in two or more populations as shared SNPs.

We also calculated the proportion of heterozygosity with the Tassel software [4] to clarify the results of adding NTT individuals to the TT population.

Reconstruction of pedigree relationships of NTT individuals

To reconstruct the pedigree relationships of NTT individuals that lacked family information, we calculated the identity-by-descent (IBD) scores of all Taishu individuals using the genome option in the Plink software (-chr-set 32, -nonfounders) [2]. In the reconstructions of the pedigrees, because theoretical estimation indicated that parents and full brings show a PI_HAT of 0.5, whereas half-siblings show a PI_HAT of 0.25, we regarded a proportional IBD (i.e., PI_HAT) of >0.375 as indicating parents or full siblings and a PI_HAT of 0.375–0.1825 as indicating half-siblings.
Results

Detection of SNPs detection in the Taishu, Japanese native breeds, and Anglo-Arabian breed

We obtained 388,717,865 primary alignments after removing poor reads. The median alignment number per sample was 4,933,673 (interquartile range: 930,783–13,833,796). Based on these reads, 1,027,741 loci that contained 0 or more than 1 SNP were found. The number of loci decreased as the number of matching samples increased, with or without SNPs. ddRAD-seq analysis included a large number of SNPs whose loci were shared among a small number of individuals and not suitable for population genetic analysis. For our study, we ultimately used 10,471 SNPs that appeared in over 75% of individuals (65/86 individuals showed these SNPs) yielded 1 SNP and 2 alleles with a minor allele frequency cutoff of 5% and mapped them to the autosomal chromosomes.

Genetic structure of the NTT population

To clarify whether NTTs and TTs could be grouped in the same cluster, we constructed an ML phylogenetic tree using 4,532 SNPs, which were filtered by LD using the SNPhylo software. Each breed belonged to a separate clade, and NTTs and TTs were assigned in a single clade (Fig. 3).

To investigate the introgression of Anglo-Arabian to NTT individuals and genetic differences among TTs and NTTs, we performed MDS analysis using 10,471 SNPs. MDS plots indicated that TTs and NTTs were separated, whereas Anglo-Arabian individuals, TTs, and NTTs were not separated. The first dimension largely separated all Taishu and Anglo-Arabian individuals, whereas the first and second demotions did not separate TTs and NTTs. The MDS results indicated that dimension 3 separated Miyako and Yonaguni individuals, whereas dimension 4 separated Kiso, Hokkaido, and Anglo-Arabian individuals (data not shown). To identify the genetic structure of NTT individuals, we used Admixture analysis. Based on the cross-validation error of each K, the optimal value of K was 4 (Fig. 4B). Admixture results of K=4 showed that the genetic structure of Taishu individuals had multiple clusters, but the clusters were not derived from TTs or NTTs. When K=5 and 6, Taishu, Miyako-Yonaguni, and Kiso-Hokkaido-Anglo

Fig. 3. Molecular phylogenetic analysis by the maximum-likelihood method. Maximum-likelihood phylogeny was constructed by using 4,517 screened single nucleotide polymorphisms. Colored boxes represent different breeds. Green, purple, yellow, orange, blue, and gray represent the Kiso, Hokkaido, Yonaguni, Miyako, Taishu, and Anglo-Arabian, respectively. Black circles indicate Non-Tsushima Taishu individuals.
Arabian individuals were separated. Miyako and Yonaguni individuals were separated when K=7. Sample No. 65 was separately plotted among Taishu individuals (Fig. 4A and 4B) because it had already been recognized as a result of crossing with other Japanese breeds.

**Taishu-specific SNPs**

We counted breed-specific SNPs and shared SNPs among the breeds to assess the genetic independence of NTT individuals. Four hundred and eighty-eight Taishu-specific SNPs were present in NTT individuals, 258 of which were present in TT individuals (Fig. 5).

**Genetic diversity of the Taishu**

We calculated proportion of heterozygosity for TT, NTT, and all Taishu individuals (TT + NTT) as an index of genetic diversity in each population (Fig. 6). The proportions of heterozygosity of the NTT population and all Taishu individuals were slightly higher than that of TTs; however, the difference was not significant (Steel-Dwass test; for TTs-NTT, t=0.38, P=0.92; for TTs-All Taishu=0.166, P=0.98)

**Reconstruction of pedigree structures**

We reconstructed the pedigree structures of NTT and TT individuals without pedigree data. Among the 18 NTTs, 8 individuals had no or incomplete pedigree information. Among the 38 TTs, one individual had no pedigree information. We calculated IBD and proportion IBD (PI_HAT) by plink (Supplementary Table 1: online only). Based on the PI_HAT, the pedigree structures of four NTTs and one TT were reconstructed (Fig. 7).

---

![Fig. 4. Genetic Structure of Non-Tsushima Taishus. A. Genetic clustering using Admixture with an admixture model. Admixture results for the Tsushima Taishu, Non-Tsushima Taishu, four other Japanese horse breeds, and an Anglo-Arabian breed from K=2 to 7. B, Multi-dimensional scoring of Japanese breeds. C, Cross validation error of Admixture results.](image-url)
Age status of TT and NTT individuals

We investigated age information for all Taishu horses included in this study and a previous study (Fig. 8). The results showed that 61% of TTs and 50% of NTTs were under 15 years old.

Discussion

The phylogenetic tree (Fig. 3) showed that each breed consisted of a single clade, with NTTs and TTs belonging to the same cluster, consistent with the results of previous studies (Ayumi Tezuka, unpublished results) [10]. In addition, both Admixture results and MDS plots indicated that both TTs and NTTs individuals were separated from Anglo-Arabians. These results suggested that the NTTs in this study were genuinely Taishu horses, with few or very minor genetic introgressions from Anglo-Arabian horses. One of the NTTs (sample no. 65) was clarified to have a rather specific genetic structure containing clusters of other Japanese native breeds, suggesting that this individual may have recently been introgressed from other Japanese breeds.

Regarding specific SNPs between breeds, we found 263 Taishu-specific SNPs limited to only TTs and NTTs. Additionally, the number of SNPs shared by NTTs with other breeds was almost the same as that shared with TTs. Moreover, TTs and NTTs were not separated in MDS plots. These results suggested that the genetic structure of almost all NTTs was close to that of TT individuals. On the other hand, the finding that K=4 was optimal in Admixture was inconsistent with the breed information. We considered that these results occurred because of the divided genetic structure of the Taishu, which was due to a drastic decrease in population size and unequal sample size between the Taishu and other breeds. The Taishu showed a greatly reduced population size, and this kind of bottleneck effect may cause a divided genetic structure. Furthermore, there were more Taishu samples than samples from other breeds in the current study. Unequal sample sizes between breeds are known to lead to analysis bias [6]. Because the phylogenetic tree, dimensions 3 and 4 in MDS plots, and K=5 and 6 in Admixture results separated the other breeds, we considered these results reliable. Because this study aimed to assess the genetic condition of NTTs, we did not re-run the admixture analysis with an equal sample size.

To confirm whether the addition of NTTs to TTs improves the genetic diversity of the Taishu population, we calculated the proportion of heterozygosity in TTs, NTTs, and both pooled. The heterozygosity of all Taishu individuals was higher than that of TT individuals, but the difference was
However, many NTT individuals, which were the most appropriate candidates for introduction, were revealed to be of under fertile age.

Pedigree structure information is important in mating plans and conservation of NTTs. Here, we reconstructed the pedigree structures of individuals with unknown parents using proportion of IBD and known pedigree structures of other individuals with known parents. In many cases, parental information was not available because, generally, the parents had died. On the other hand, information on full or half-siblings was generally available for the current populations. This suggests the usefulness of reconstruction of pedigree structures based on genomic information and comparison with full or half-siblings with known parents.

Ordinarily, pedigrees of horses require information on three paternal generations and the mother. That is, eight great-grandfathers, four grandfathers, and the father and mother. The pedigree reconstruction approach based on genomic information and partially known pedigree structures, as proposed in this study, could be an alternative approach to estimate the familial information of individuals with unknown parents. With this approach, inclusion of individuals with pedigrees known for several previous generations would be very useful to precisely reconstruct
pedigree structures for individuals of unknown parents. In the future mating and conservation plans for the Taishu, the inclusion of individuals from other populations than that of Tsushima Island may allow sustainable conservation of the Japanese Taishu horse breed.

Acknowledgments

We thank the conservation organizations of Taishu horse, Tsushima-shi, Nagasaki, Japan, Dr. Teruyuki Miyazaki, Mr. Sakita Hiromi, Mr. Tatsuya Ejima, Mr. Kenichi Okajima, Tokachi Station (National Livestock Breeding Center), Kujishima Zoo and Botanical Garden, Yagiyama Zoological Park, Toyama Municipal Family Park, Isahaya Yuyu Land “Kantaku-no-Sato”, and the riding club of Wakamatsu in Kitakyusyu for providing samples; Fumie Kobayashi and Satoko Kondo for their help with the experiments; Naomi Niwa for handling the paperwork; and Yumie Shinohara for assistance with the research. This work was supported by the Research Institute for Food and Agriculture of Ryukoku University.

References

1. Alexander, D.H., Novembre, J., and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genom Res. 19: 1655–1664. [Medline] [CrossRef]

2. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575. [Medline] [CrossRef]

3. Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120. [Medline] [CrossRef]

4. Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635. [Medline] [CrossRef]

5. Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22: 3124–3140. [Medline] [CrossRef]

6. Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., and Andrew, R.L. 2017. The K = 2 conundrum. Mol. Ecol. 26: 3594–3602. [Medline] [CrossRef]

7. Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870–1874. [Medline] [CrossRef]

8. Langmead, B., and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359. [Medline] [CrossRef]

9. Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15: 162. [Medline] [CrossRef]

10. Tozaki, T., Takezaki, N., Hasegawa, T., Ishida, N., Kurosawa, M., Tomita, M., Saitou, N., and Mukoyama, H. 2003. Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse outgroup. J. Hered. 94: 374–380. [Medline] [CrossRef]