Whole Genome Analysis of Yakushima macaque, *Macaca fuscata yakui*

Scientific Report of Genome Science Course

Presented to Kyoto University

by

Chung-Kun Lee[1]

Kei Hazehara[2]

Kota Kuroki[2]

Mayu Onoda[3]

1, Department of Botany, Kyoto University

2, Primate Research Institute, Kyoto University

3, Department of Biophysics, Kyoto University

1st July, 2016

Approved by _____

Dr. Takushi, KISHIDA

Abstract

Yakushima macaques (*Macaca fuscata yakui*) are endemic to Yakushima Island in Japan, and are the southernmost population of Japanese macaques (*Macaca fuscata*), which in turn the northernmost-distributed nonhuman primates. However, its position as a subspecies is not supported by genetic analysis despite morphological features. Their population history had not been analyzed before and divergence time estimation of Japanese macaque from Chinese population of rhesus macaques (*Macaca mulatta*) was still unsettled issue. Here, we conducted whole genome analysis of Yakushima macaques and 1) estimated divergence time from rhesus macaques and population history of Yakushima macaques through PSMC method, 2) investigated some of newly revealed genetic differences between the species by scanning single nucleotide polymorphisms (SNPs). As for PSMC analysis, we found Yakushima macaques diverged from Rhesus macaques about 0.5 million years ago, which is consistent to some previous research and the fossil records. We also estimated the divergence of Yakushima macaques from other Japanese macaques to be 0.1 - 0.2 mya, although sufficiently accurate estimation requires comparative analysis with other Japanese macaques. The decrease in Yakushima macaques' population size five thousand years ago suggests a dynamic environmental change, possibly immense volcanic

eruptions near the island around that time. Genetic analysis illustrated a number of significant mutations on multiple genes which appear relating to attributes in each species: for instance, first, a group of hair-density genes including *EDAR*, *FGFR2*, *DSG4*, *LIPH*, *LPAR6*, *JUP*, *DSP*, *DSC2*, *KANK2*, and *KRT83*, second: *Myo5B-ACAA2* expression complex, third: *ZNF37A* and fourth: mental disorder-related genes including *SCN1A*, *KCNQ2*, *CHRNB4*, *CACNA1C*, *VIPR2*, *EGR4* and *AUTS2*.

1. Introduction

Macaques have the widest distribution among primates except for human beings, and have contributed to medical studies as model animals close to human beings (Gibbs *et al.*, 2007; Fleagle,1988). Previous genetic studies of Asian macaques have revealed their intrageneric phylogenetic relationships and attempted to estimate their ancestral movement (Wu et al. 2013), speciation period, and even population dynamics history (Fan et al., 2013; Osada et al,. 2015).

Yakushima macaques (YM, Macaca fuscata yakui), endemic to Yakushima Island, Kyushu, Japan, are the southernmost population of Japanese macaque (JM, Macaca fuscata), the northernmost primates except for human beings. Yakushima Island has great environmental and biological diversity; therefore, it is conceivable that YMs have adaptively acquired unique features such as darker pelage and smaller body size (Fooden and Aimi, 2005). However, taxonomic treatment of YM has been debated due to insufficient genetic supports despite of distinctive morphological character states. Compared to RMs, JMs are known to have shorter tails and darker hairs with higher density (Fooden 1976, Inagaki *et al.*, 1985).

YM's position in radiation history of JM is still unclear (Marmi et al., 2004), and even the divergence time of JM from Chinese population of rhesus macaques (RM, Macaca mulatta) is not settled problem; genetic estimations varies from 0.18 mya of Chu et al. (2007) to 0.65-0.73 mya of Hayasaka (1996), and a fossil evidence indicates the first settlers appeared at least 0.43 mya (Aimi, 2002).

This study conducted the first whole genome analysis of YM for three purposes. First, to provide the first population history estimation of YM, we used PSMC (Pairwise Sequentially Markovian Coalescent) model invented by Li and Durbin (2009). Second, we obtained the most reliable divergence time estimation by virtual pseudo-diploid analyze with PSMC. Third, we looked for some genetic differences responsible for the characteristic unique to YM.

3. Materials and Methods

Our analyses focused on DNA sequence of Single YM in Japan Monkey Center. The donner was born in a large captive population, which has been in the center for several decades.

We mapped the donner sequence to reference genome of RM mapped DNA of single The reference genome is from Indian origin RM, presenting twenty chromosomes and 3.23 Gbps (GenBank Accession GCA_0007728753). Details are available from NCBI (http://www.ncbi.nlm.nih.gov/assembly/GCF_000772875.2/). The quality of sequenced DNA was checked by FastQC (Andrews, 2010) and then quality control was applied before genome mapping using Trimmomatic (version 0.33, LEADING: 15 TRAILING: 15 SLIDINGWINDOW: 5:15 CROP: 100 MINLEN: 50; Bolger *et al*, 2014) to remove adaptors, segmental repeats and low quality regions. Mapping was performed by BWA (Li H. 2009) and depth of coverage was estimated, SNPs were called using Genome Analysis TK (McKenna *et al*., 2010). We used snpEff (Cingolani *et al*., 2012) to annotate mutations in the coding regions for the genetic analysis. We followed the PSMC method (Li and Durbin 2011) to infer the demographic history of YM. The following parameters were used: mutation rate per generation per site = $1.2 * 10^{-8}$, and generation time = 12 [years]. PSMC was also applied for pseudo-diploid with parameters above in order to

detect their divergence time (Prado-Martinez *et al*. 2013).

Sex chromosomes were excluded for the PSMC calculation.

4. Result

Population Size History (PSMC analysis)

Before conducting PSMC, we examined depth of autosomes as well as sex chromosomes shown in Table 1. Coverage in total was 21.3, representing the suitable value for analysis (Foote *et al*. 2016, Nature Communications). In addition, it was confirmed the YM donor was a female by calculation of depth of sex chromosome.

To estimate the divergence period of YM from RM, we initially produced pseudo-diploids of YM and RM and analyzed it by PSMC. As Fig. 1 indicates, it is estimated to be around 0.5 million years ago (mya) that YM branched from RM. We also investigated population history of YM with larger scale in order to correlate its dynamics with environmental events in Yakushima (Fig. 2). Approximately 0.2 mya ago, the population size of YM was once boosted and decreased. It is gently decreased during the last ice age (approximately $7*10^4$ - $1*10^4$ years ago) and after that this decrease was mended until approximately five thousand years ago in which a strong downturn could be observed.

Genetic Analysis

To research the level of genetic differentiation between YM and RM, we performed annotation of mutations by snpEff. Transcripts with missense mutations were found to be 27122 out of 70539 in

total. Subsequently, we sorted different genes by type of variants its effect such as: nonsense, missense and silence mutation either in the exon/intron, splice region, or UTR region. Particularly, 41 missense mutations were found in *Myo5B*, and 2 were in *ACAA2*, neither of which showed nonsense mutation. All of them were confirmed to be heterozygotic diversity in *ACAA2* as well as in *Myo5B*. *ZNF37A* mutation had 11 stop codons that were the most in all of the genes. *ZNF37A* codes a protein correlated to myogenesis.

Mental disorders-related genes were found to have missense mutations: *SCN1A*, *KCNQ2*, *CHRNB4*, *CACNA1C*, *VIPR2*, *EGR4* and *AUTS2*.Brain development-related genes were also found to have missense mutations: *NDE1* and *WDR62*. Ten hair density-related genes are found to have missense mutations: *EDAR*, *FGFR2*, *DSG4*, *LIPH*, *LPAR6*, *JUP*, *DSP*, *DSC2*, *KANK2*, and *KRT83*. There was no missense mutation on *EDA2R* and *EDARADD*. mutations in hair color-related genes are also found:

5. Discussion

Population history of Yakushima macaques

There has been a significant inconstancy of time estimations between recent RM-focused studies and JM- focused studies. Pseudo-diploid PSMC indicates divergence between YM and Indian-origin RM at around 0.5 mya, which is significantly older than the results of some recent studies on RM; Chu *et al.* (2007) estimates speciation between JM and Chinese RM 0.18 mya, and Wu *et al*. (2013) estimates divergence between Indian RM and Chinese RM 0.16 mya. But our results are in agreement with studies on JM; JM immigration to Japan at least 0.43 mya based on fossil records (Aimi, 2002), and JM-RM speciation between 0.31 and 0.88 mya based on mtDNA analysis with variable molecular divergence rate (Marmi *et al*., 2004).

The population decline between 0.1 and 0.2 mya indicates possibly the divergence of YM from other JM; it is in accordance with Ohshima (1990), geologically estimating that the isolation of Yakushima Island from mainland Japan dates back to around 0.1mya. This decline also can be interpreted as just a divergence within JM not related to Yakushima Island, because there is another theory that Yakushima island had been connected to mainland Japan until about 0.02mya (Davison *et al*., 2005). However, no steep decline is detectable between 0.1 mya and 0.02 mya, which would be a footprint of other divergences, thus we strongly suspect that the separation of

YM from other JM populations occurred about 0.1 mya. Whole genome analysis of other JMs will provide more accurate estimation of divergence time of YM.

Drastic decline of the population is shown approximately 5000 years ago. It is near to disastrous eruption of Kikai caldera 7300 years ago (Okuno, 2002), which drove ancient Yakushima people into extinction. Otherwise, this decline is possibly due to influence of incest, which is not uncommon in captive-bred.

Genetic analysis

There are significant numbers of mutations in several genes, such as *MYO5B* and *ZNF37A*. 41 missense mutations in *MYO5B*, and 2 missense mutations in *ACAA2* were found in this study. None of nonsense mutations were observed in either of them. The genetic differentiation in this study did not elucidate phenotypic features of Indian RM, JM, and Chinese RM, respectively. However, RNA expression profile of *MYO5B* and *ACAA2* slightly differs in tissues compared with other primates including human according to Nonhuman Primate Reference Transcriptome Resource (NHPRTR) database (available from http://www.ncbi.nlm.nih.gov/ieb/research/acembly/av.cgi?db=human&term=MYO5B&submit=Go). JM and RM India (RMI) present certain expression in entire tissues but RM China (RMC)

indicates highest expression level in liver only.

MYO5B and ACAA2, both closely located in the locus on Chromosome 18, forms an expression complex. *MYO5B* encodes Myosin VB, which is a type of processive myosin motors, representing involvement in transport of cargo like sugars, organelles, and mRNAs toward the periphery of cells along actin filaments rather than in muscle contraction (Hammer, J. A. and Sellers, J. R. 2012). *ACAA2*, encodes a catalyst in the last step of beta oxidation in mitochondria. Therefore, the high-level expression in liver is supposedly because; 1) *ACAA2* is in need for the process of beta oxidation to obtain energy, and 2) *Myo5B* plays a role of carrying vesicles containing a variety of proteins produced within the cell. Indeed, *LIPG* coding lipase is in the upstream region of both of them. This will support *ACAA2-MYO5B* expression complex in liver. The reason for in RMC liver-specific expression of the complex remains unclear. Thus, further genomic analysis in RMC and JM in addition to this study will be conducted in the future, which leads not only to investigate expression profile but also historical genetic changes. It is conceivable RMC will have more genetic variations on these genes.

ZNF37A codes Zinc finger protein 37A, a transcription factor involved in myogenesis in human, and its defection in expression contributes to altered myogenesis in myotonic dystrophy type 1

(Gauthier *et al*. 2003). Such mutation-rich genes do not seem to play normal roles, but YM obviously avoid fatal consequences.

Some missense mutations on hair-related and behavior-related genes were detected, and some of them could be responsible for distinctive features of JM such as high hair density (Inagaki *et al*., 1985) or possibly of YM such as pelage color (Kuroda 1940). We did not determined the causative genes, and it is also possible that causative genes have other kinds of mutation such as insertion and deletion.

6. Acknowledgments

This study was conducted as one of the curriculums of Genome Science Couse in Kyoto University. Takushi Kishida and Kei Matsumura (Wildlife Research Center of Kyoto University) are thanked for direct instruction. Naoki Agetsuma (Hokkaido University) is thanked for comments on the manuscript. PWS Leading Graduate Program is thanked for funding and support for Genome Science Course.

7. References

Aimi, M. (2002). The Oldest Fossil Macaque from Japan. *Primate research,* **18(2)**, 239-245.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30,** 2114-2120.

Cingolani, P., Platts, A., Wang, L. L., Coon, M.., Nguyen, T., Wang L., Land, S. J., Lu, X., and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain 1118; iso-2; iso-3. *Fly (Austin),* **6,** 80-92.

Chu, J. H., Lin, Y. S., and Wu, H. Y. (2007). Evolution and dispersal of three closely related macaque species, *Macaca mulatta, M. cyclopis, and M. fuscata,* in the eastern Asia. *Mol. Phylogenet. Evol.,* **43,** 418-429.

Davison, A., Chiba, S., Barton, N. H., and Clarke, B. (2005). Speciation and gene flow between snails of opposite chirality. *PLoS Biol*, **3(9)**, e282.

Fan, Z., Zhao, G., Li, P., Osada, N., Xing, J., Yi, Y., Du, L., Silva, P., Wang, H., Sakate, R.,

Zhang, X., Xu, H., Yue, B., and Li, J. (2014). Whole-Genome Sequencing of Tibetan Macaque (Macaca thibetana) Provides New Insight into the Macaque Evolutionary History. *Molecular biology and evolution*, **31(6)**, 1475-1489

Fleagle J., G. (1988). Primate Adaptation and Evolution. Academic Press, New York.

Fooden, J. (1976). Provisional classifications and key to living species of macaques (primates: Macaca). *Folia Primatol (Basel).* **1976;25(2-3)** *225-36*

Fooden, J., and Aimi, M. (2005). Systematic review of Japanese macaques, Macaca fuscata (Gray, 1870). *Fieldiana zoology*, 1-198

Foote, A., Vijay N., Avila-Arcos, M., Baird, R., Durban, J., Fumagalli, M., Gibbs, R., Hanson, M., Kurneliussen, T., Martin, M., Robertson, K., Sousa, V., Vieira, F., Vinar, T., Wade P., Worley, K., Excoffier, L., Morin, P., Gilbert, M., and Wolf, J. (2016) Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature communications*, **7**, 1-12.

Gauthier, M., Marteyn, A., Denis, J, A., Cailleret, M., Giraud-Triboult, K., Aubert, S., Lecuyer, C., Marie, J., Furling, D., Vernet, R., Yanguas, C., Baldeschi, C., Pietu, G., Peschanski, M., and Martinat, C. (2003) A defective Krab-domain zinc-finger transcription factor contributes to

altered myogenesis in myotonic dystrophy type 1. *Human Molecular Genetics,* **22(25)**, 5188-5198

Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., *et al.* (2007). Evolutionary and biomedical insights from the rhesus macaque genome. science, **316(5822)**, 222-234.

Hammer, J. A., and Sellers, J. R. (2011). Walking to work: roles for class V myosins as cargo transporters. Nature.Rev. *Mol. Cell Biol.,* **13**,13–26.

Hayasaka, K., Fujii, K., and Horai, S. (1996). Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-bp region of mitochondrial DNA. *Mol. Biol. Evol.*, **13**, 1044– 1053.

Inagaki, H., and Hamada, Y. (1985). Differences in Hair Density of Japanese Monkeys (Macaca fuscata fuscata) with Locality and Age. *Primates*, **26(1)**, 85-90

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* **25,** 1754-1760.

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature,* **475,** 493-496.

Marmi, J., Bertranpetit, J., Terradas, J., Takenaka, O., and Domingo-Roura, X. (2004) Radiation and Phylogeography in the Japanese macaque, *Macaca fuscata*. *Mol. Phylogenet. Evol.,* **30,** 676-685.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.,* **20,** 1297-1303.

Ohshima, K. (1990). The History of Straits around the Japanese Islands in the Late-Quaternary. *The Quaternary Research*, **29(3)**, 193-208.

Okuno, M. (2002). Chronology of Tephra Layers in Southern Kyushu, SW Japan, for the last 30,000 Years. *The Quaternary Research*, **41(4)**, 225-236.

Osada, N., Hettiarachchi, N., Babarinde, I. A., Saitou, N., and Blancher, A. (2015) Whole-genome sequencing of six mauritian cynomolgus macaques (*Macaca fascicularis*) Reveals a genome-wide pattern of polymorphisms under extreme population bottleneck. *Genome biology and evolution*, **7(3)**, 821-830.

Wu, S. J., Luo, J., Li, Q. Q., Wang, Y. Q., Murphy, R. W., Blair, C., Wu, S. F., Yue, B. S. and

Zhang, Y. P. (2013). Ecological genetics of Chinese rhesus macaque in response to mountain building: all things are not equal. *PloS one*, **8(2)**, e55315.

7. Appendix

7.1. Tables

chromosome	depth
chr3	19.71
chrX	20.26
chrY	0.986

Table 1. Chromosome Depth of Coverage

Coverage depth was evaluated in each chromosome: Chromosome 3, X, and Y. The values for chromosome 3 and X were nearly 20.

7.2. Figures

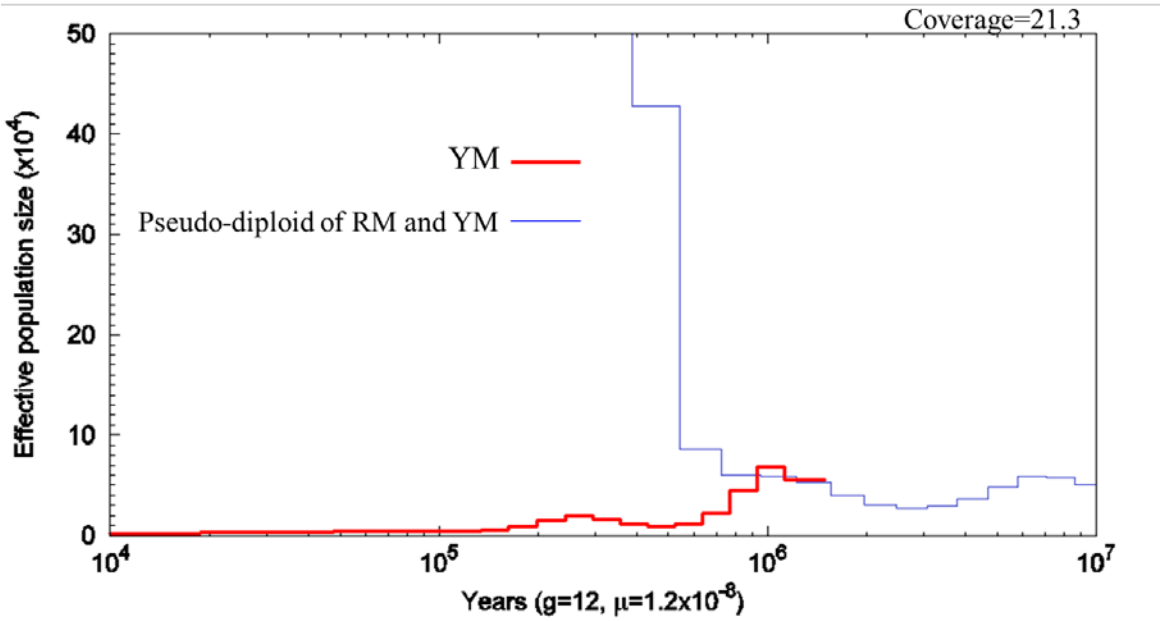

Fig.1. PSMC result of Yakushima macaque and pseudo-diploid

YM and RM means Yakushima macaque and Rhesus macaque respectively.

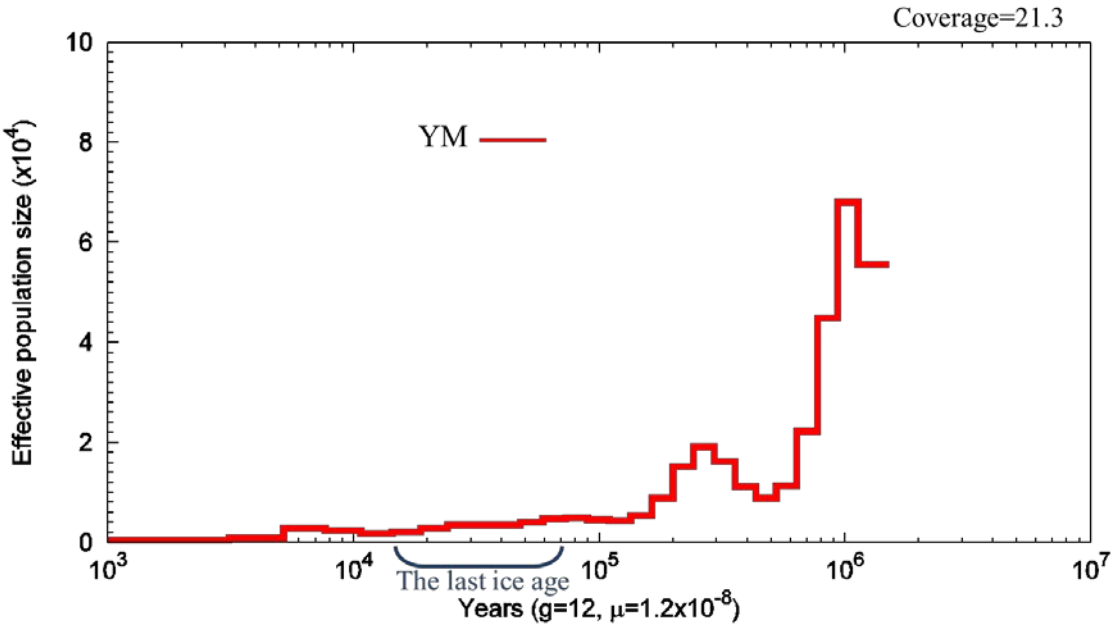

Fig.2. PSMC result of Yakushima macaque in detail

This is the result of population size of YM (Fig.1 shows) in which the scale of effective population size is scaled up to 10 x10^4 at most.