Research Activity Report
Supported by "Leading Graduate Program in Primatology and Wildlife Science"
(Please be sure to submit this report after the trip that supported by PWS.)

	2015.06.25
Affiliation/Position	CES, IISc, Bangalore, India / Associate professor
Name	Praveen Karanth

1. Country/location of visit

Kyoto, Japan

2. Research project

Genome Science Course

3. Date (departing from/returning to Japan)

2015. 06. 01 – 2015. 06. 15 (15 days)

4. Main host researcher and affiliation

Dr. Takushi Kishida, Wildlife Research Center of Kyoto University

5. Progress and results of your research/activity(You can attach extra pages if needed)

Please insert one or more pictures (to be publicly released). Below each picture, please provide abrief description.

I participated in the Genomic Science Course that was offered by Dr. Kishida and Dr. Hayakawa. The course dealt with various bioinformatics tools that one has to use for analyzing genomic data. In this course we analyzed Japanese macaque whole genome. We started with checking the various quality control parameters of the sequence file using the program *FastQC*. The raw data in Fastq format was then "cleaned up" using the program *Trimmomatic*. This program was used to trim the sequences to 100 bp, remove adapter, discard sequences <50bp and to remove low quality base from the ends. The program *FastQC* was run again to check if the quality of the reads improved.

The next step was mapping Japanese macaque genome to Rhesus reference genome. For this we first retrieved Rhesus reference seq from Ensemble database. The downloaded reference genome was decompressed using *gunzip* and the reference sequence was indexed using the program Burrows-Wheeler Alignment Tool (*BWA*). The file format was changed from sam to bam format using *samtools*. Mapping of Japanese macaque genome on to Rhesus reference genome was undertaken using *BWA*. Index and sorting of bam files done using *samtool*. The alignment was viewed in IGV (Integrative Genomic Viewer-Broad Institute).

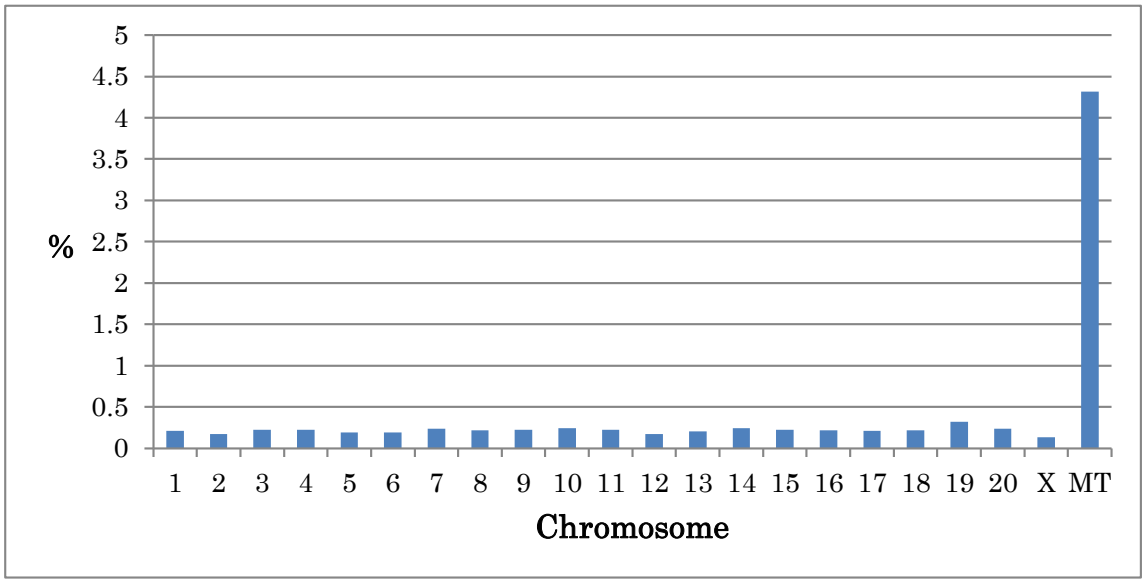
Genome Analysis Toolkit (*GATK*) was used for depth calculation, variant calling and for annotation of mutation in coding sequence. The output file in Variant Call Format (VCF) was used for further analysis. The VCF file were analyzed further using grep command in Linux and the following variations were characterized: Non synonymous mutations, heterozygotes, Indels, number of variants/chromosome. The coding regions of Japanese and Rhesus macaques were compared to determine if the degree of divergence between these two species varied across chromosomes.

Results
Overall a total of 69087 differences were observed between the two species in their coding regions (31175181 bp) suggesting that the percent sequence divergence between the two species is 0.22%. Interestingly among the 69087 changes 33758 were nonsynonymous substitutions. Separate comparisons were done for each autosomal chromosome (1-20), X-chromosome and the mitochondrial DNA to determine how this variation is distributed across chromosomes. The highest sequence divergence between these two species in the autosomal region was 0.32% for chromosome 19 and the lowest was 0.17% for chromosome 2. The average across the autosomal chromosomes was 0.22%. The X-chromosome exhibited the lowest overall sequence divergence (0.13%) for a coding region and the mtDNA the highest (4.3%).

Conclusion:
Overall the Japanese macaques is genetically very similar to Rhesus, this is not a surprise give they

are sister species. The overall percent sequence divergence between them in the coding region is less than 0.5% (0.22), which is comparable to human-chimp divergence. However it would be interesting to compare the non-coding regions between these two species. Also chromosome 19 might harbour some interesting genes that have undergone positive selection, given the higher than average divergence in this region between the two species. Additionally it would be very interesting to determine the genes that harbour the 33758 nonsynonymous substitutions. The divergence in mitochondrial regions was very high (4.3%), this is expected given 10-20 fold higher mutation rate in mammalian mtDNA when compared to nuclear DNA (0.22x20=4.4). Additionally the divergence in X-chromosome was the least (0.13) which is approximately 2/3 the divergence in autosomal genes due to lower mutation rate in the female germline.

6. Others



Japanese vs. Rhesus macaque genetic divergence (p-distance) in the coding region across chromosomes.