598

IDENTIFICATION AND VALIDATION OF DNA SEQUENCE VARIATION (SNPS) IN RHESUS MACAQUES USING NEXT-GEN SEQUENCING

G.L. Fawcett, M. Raveendran, D. Rio Deiros, D. Chen, R.A. Harris, M. Bainbridge, D.M. Muzny, J. Reid, D.A. Wheeler, K.C. Worley, A. Milosavljevic, R.A. Gibbs, J. Rogers

Human Genome Sequencing Center and Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA

Presenter's Email: jr13@bcm.tmc.edu

The goal of this study was to identify and validate new DNA sequence polymorphisms (SNPs) in rhesus macaques (*Macaca mulatta*). Rhesus macaques are among the most widely studied of nonhuman primates. The existing rhesus whole genome sequence assembly is remarkably useful, but little is known about DNA sequence variation in this species. Furthermore, it is advantageous to develop an efficient strategy for identifying SNPs in other primate species that will be sequenced in the future. Using the original rhesus genome data produced using the Sanger sequencing method, we identified more than 4 million potentially heterozygous nucleotide positions in the original genome assembly. We validated specific SNPs by re-sequencing the original animal using the SOLiD sequencing method (Life Technologies). Both 50bp fragment and 50bp paired-end reads were generated totaling 15X genome coverage, and mapped to the reference genome. Next, we sequenced two more animals using SOLiD, and compared their genome sequences (5X coverage each) to the original animal. Third, we used Nimblegen DNA capture and Roche 454 sequencing to screen 2834 genes in three unrelated animals. Any basepair position that is heterozygous in two different animals, or is called heterozygous in one animal using two different sequencing methods, is considered a validated SNP. To date, SOLiD re-sequencing has validated more than 2 million SNPs. The Nimblegen capture-454 analyses found approximately 24,000 gene-specific SNPs. Our results show that rhesus macaques are highly polymorphic, and define the genomic locations for more than 2 million DNA sequence polymorphisms.

Keywords: genomics, next generation sequencing, genetic variation, single nucleotide polymorphisms