



SCIENTIFIC REPORTS



OPEN

Discovery of novel genic-SSR markers from transcriptome dataset of an important non-human primate, *Macaca fascicularis*

Wendy Chang¹, J. EE-ULI², W. L. NG ³, Jeffrine J. Rovie-Ryan⁴, S. G. Tan² & Christina S. Y. Yong ¹

Macaca fascicularis, also known as the cynomolgus macaque, is an important non-human primate animal model used in biomedical research. It is an Old-World primate widely distributed in Southeast Asia and is one of the most abundant macaque species in Malaysia. However, the genetic structure of wild cynomolgus macaque populations in Malaysia has not been thoroughly elucidated. In this study, we developed genic-simple sequence repeat (genic-SSR) markers from an in-house transcriptome dataset generated from the Malaysian cynomolgus macaque via RNA sequencing, and applied these markers on 26 cynomolgus macaque individuals. A collection of 14,751 genic-SSRs were identified, where 13,709 were perfect SSRs. Dinucleotide repeats were the most common repeat motifs with a frequency of 65.05%, followed by trinucleotide repeats (20.55%). Subsequently, we designed 300 pairs of primers based on perfect di- and trinucleotide SSRs, in which 105 SSRs were associated with functional genes. A subset of 30 SSR markers were randomly selected and validated, yielding 19 polymorphic markers with an average polymorphism information content value of 0.431. The development of genic-SSR markers in this study is indeed timely to provide useful markers for functional and population genetic studies of the cynomolgus macaque and other related non-human primate species.

Macaca fascicularis (Raffles 1821), also known as the 'long-tailed macaque' or the 'cynomolgus macaque', is a macaque species that is native to Southeast Asia and widely distributed in Malaysia, Thailand, Myanmar, Laos, Cambodia, Vietnam, Indonesia, Timor Leste, and the Philippines¹. Despite being one of the predominant macaque species in Malaysia, the genetic structure of their wild populations remains unclear. In Malaysia, most studies were conducted to examine the distribution², behaviour³, human-macaque conflict^{4,5} and their association with zoonotic diseases^{6,7}. Only a few genetic studies involving phylogeography and population genetics of cynomolgus macaques were conducted thus far. Most genetic studies conducted were based on the maternally inherited mtDNA marker⁸⁻¹⁰, and a few reports were based on the Y-chromosome¹¹ and genomic SSR markers^{12,13}.

Simple sequence repeats (SSRs) are repetitive DNA sequences, generally with motifs of 2–6 bp long, and present abundantly in eukaryotic genome. Its codominant and multi-allelic properties are highly valued by geneticist and evolutionary biologist, and are commonly used as DNA markers in genetic studies. Despite the recent thriving of single nucleotide polymorphism (SNP) markers, SSR markers are still relevant in many applications¹⁴⁻¹⁸. SSRs can be broadly categorized into genomic SSR and genic-SSR, depending on their locations in the genome. SSRs sited in the transcribed region are generally known as genic-SSRs. As more SSRs associated with protein coding genes are found, it is more evident now that the previously presumed junk-DNA possibly play a crucial role in adaptive evolution¹⁹. While genic-SSR is not as abundant and as polymorphic as genomic SSR, it offers several advantages over genomic SSR marker – higher probability of finding association with functional gene,

¹Department of Biology, Faculty of Science, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia.

²Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia. ³China-ASEAN Institute of Marine Sciences, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900, Sepang, Selangor Darul Ehsan, Malaysia. ⁴National Wildlife Forensic Laboratory (NWFL), Department of Wildlife and National Parks, KM 10, Jalan Cheras, 56100, Kuala Lumpur, Malaysia. Correspondence and requests for materials should be addressed to C.S.Y.Y. (email: chrisyong@upm.edu.my)