

ABSTRACT

Sequencing the genome is an important step towards detailed understanding of the biology and evolution of Proboscis Monkey. It helps in finding the genes that are related to its unique characteristics especially that influenced its diet preferences. Presence-absence and alteration of any gene may influence the functionality of the gene. Moving in line with the advent of modern technology in genomics, the whole-genome of Proboscis Monkey was successfully sequenced using next-generation sequencing (NGS) approaches, namely, paired-end and mate pair (Illumina) and 454 Shotgun (Roche). The assembly using ABySS, SGA and PHRAP yielded 290 fold coverage and the genome size is 2.67 Gbp comprises of 20 + X chromosomes. Proboscis Monkey genome has the highest coverage compared to other primates and it was proved that NGS methods are more efficient than the previous method of sequencing (Sanger sequencing). From the genome, a total of 14,635 protein-coding genes were predicted (improved by transcriptome) and compared between 10 other primates. There were 738 genes that were specific to Proboscis Monkey including the glycogen synthase 1 (*GYS1*) and expanded SusE outer membrane protein (PF14292). These genes are related to starch utilization that is needed by Proboscis Monkey to digest leaves. Subsequently from the genome, sweet taste receptor *Tas1r2* gene was mapped and characterised using blastn and Geneious. The gene is 20,596 bp in length including the exons and introns, encodes an approximately 800 amino acid peptides and has seven transmembrane domains. There was approximately 210 bp deletion in exon 4 of *Tas1r2* gene in Proboscis Monkey after comparing with other taxa representing omnivore, herbivore and carnivore. There was also 4 nucleotides deletion in exon 6 that resulted in multiple premature

stop codons. These deletions alter the gene structure and function to be inactivated as the protein was truncated. Next, phylogenetic trees of *Tas1r2* gene in anthropoid primates were constructed using four different methods (neighbor-joining, maximum parsimony, maximum likelihood and Bayesian). The phylogeny successfully elucidated the evolution and phylogenetic relationship of *Tas1r2* gene in primates. Although the gene can be considered as conserved gene, the variable nucleotides give enough information to separate the primates into their taxonomic group as well as their dietary preferences. Furthermore, the reviewed time of divergence estimation by BEAST in this study change the current geological time of primate's speciation pattern to be younger than previously thought (Miocene epoch). It was speculated that during Miocene epoch, series of climate changes took place and caused vast alteration in the environment, forest type and fruit seasonality was the trigger behind the evolution of sweet taste receptor and dietary shift of Proboscis Monkey. The Proboscis Monkey whole-genome sequence reported here has already moved the iconic Proboscis Monkey to a whole-organism system that can be interrogated at the level of the single DNA base. The Proboscis Monkey genome simply provides a starting point for further investigation.

Keywords: Proboscis Monkey, primates, diet, genomics, next-generation sequencing, Illumina, Roche, gene prediction, phylogenetics, time of divergence, sweet taste receptor, *Tas1r2* gene

***Jujukan seluruh genom dan penyambungan Monyet Belanda (Nasalis larvatus)
(Primat: Cercopithecidae) dan sejarah evolusi bagi reseptor rasa manis Tas1r2
gen dalam Simiiformes (Primat: Haplorhini)***

ABSTRAK

Penjujukan genom adalah satu langkah yang penting ke arah pemahaman terperinci tentang biologi dan evolusi Monyet Belanda. Ia membantu dalam mencari gen yang berkaitan dengan ciri-ciri yang unik terutamanya yang mempengaruhi pemilihan makanan. Kehadiran atau ketiadaan dan perubahan mana-mana gen akan mempengaruhi fungsi gen. Bergerak seiring dengan perkembangan teknologi moden dalam bidang genomik, tesis ini berjaya menghasilkan draf pertama seluruh genom Monyet Belanda menggunakan pendekatan penjujukan 'next generation sequencing' (NGS), iaitu, 'paired-end' dan 'mate pair' (Illumina) dan 454 Shotgun (Roche). Penyambungan jujukan menggunakan Abyss, SGA dan PHRAP menghasilkan 290 liputan kali ganda dan saiz genom adalah 2.67 Gbp, terdiri daripada 20 + X kromosom. Genom Monyet belanda mempunyai liputan yang tertinggi berbanding dengan primat lain dan ini telah membuktikan bahawa kaedah NGS adalah lebih berkesan daripada cara sebelumnya iaitu penjujukan dari kaedah Sanger. Dari genom, sejumlah 14,635 gen protein pengkodan telah diramalkan oleh AUGUSTUS 2.5.5 dan Blat (diubahsuai oleh transcriptom) dan perbandingan dibuat antara 10 primat lain. Terdapat 738 gen yang khusus kepada Monyet Belanda termasuk glikogen synthase 1 (GYS1) dan perkembangan SUSE membran luar protein (PF14292). Gen ini berkaitan dengan penggunaan kanji dan diperlukan oleh Monyet Belanda untuk mencerna daun.

Seterusnya, gen reseptor rasa manis *Tas1r2* telah dipetakan dan dicirikan menggunakan *blastn* dan *Geneious*. Panjang gen ini ialah 20,596 bp termasuk 'exons' dan 'introns', ia mengkod 800 peptida asid amino dan mempunyai tujuh domain transmembran. Analisa dengan taksa lain yang mewakili omnivor, herbivor dan karnivor menunjukkan terdapat penghapusan kira-kira 210 bp dalam exon 4 *Tas1r2* gen dalam Monyet Belanda. Terdapat juga penghapusan 4 nukleotida dalam exon 6 yang mengakibatkan perubahan struktur dan fungsi gen. Seterusnya, pokok filogenetik primat antropoid dibina dengan menggunakan *Tas1r2* gen menggunakan empat kaedah yang berbeza ('neighbor-joining', 'maximum parsimony', 'maximum likelihood' dan Bayesian). Pokok filogenetik berjaya menjelaskan evolusi dan hubungan filogenetik *Tas1r2* gen dalam primat. Walaupun gen boleh dianggap sebagai gen dipulihara, pembolehubah nukleotida memberi maklumat yang cukup untuk memisahkan primat ke dalam kumpulan taksonomi mereka serta kumpulan pilihan pemakanan mereka. Tambahan pula, anggaran tafsiran perbezaan masa dikaji oleh BEAST dalam kajian ini mengubah masa geologi semasa corak penspesiesan primat menjadi lebih awal dari yang difikirkan. Spekulasi telah dibuat iaitu ketika zaman 'Miocene', siri perubahan iklim menyebabkan perubahan pada alam sekitar, jenis hutan dan musim buah-buahan dan ia adalah pencetus di sebalik evolusi reseptor rasa manis dan perubahan diet Monyet Belanda. Genom Monyet Belanda yang diperolehi dalam kajian ini telah pun membawa ikonik Monyet Belanda kepada sebuah sistem-organisma yang boleh disoal siasat di peringkat asas DNA tunggal. Genom Monyet Belanda telah menyediakan titik permulaan untuk siasatan lanjut.

Kata Kunci: Monyet Belanda, primat, diet, genomik, 'next generation sequencing', Illumina, Roche, gen ramalan, filogenetik, masa kecapahan, reseptor rasa manis, gen *Tas1r2*