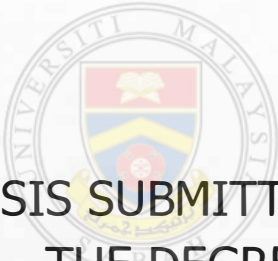


GENOME ASSEMBLY OF *Acropora tenuis* FROM
SHORT READ SEQUENCING DATA

SARAH HANISAH BT MOHD ZAIFULLIZAN

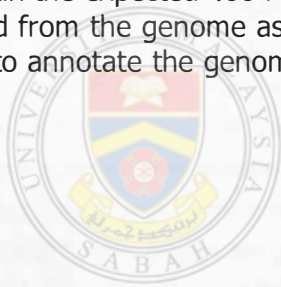
 PERPUSTAKAAN
UNIVERSITI MALAYSIA SABAH
UMS
UNIVERSITI MALAYSIA SABAH
THESIS SUBMITTED IN PARTIAL FULFILMENT FOR
THE DEGREE OF MASTER OF SCIENCE

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH

2018

ABSTRACT

Acropora tenuis is staghorn coral that is important globally to marine diversity as it makes up the majority of coral reefs in the world. Though much studies have been done to the species in particular, genomic studies are still lacking. The coral (branching fragment) specimen where DNA was extracted from for sequencing was a direct environmental sample. This in itself presents problems as though morphologically identified as *A. tenuis*, there is still the possibility of hybridization between other species of *Acropora* corals as they known to mass spawn and this produces problems such as the hybrid formation of the members of the same genus, as well as their symbionts such as dinoflagellate algae *Symbiodinium* spp. To address these unknowns, this project aimed to broaden the understanding of this species and to enrich the public genomic data available for future research by assembling a draft of *A. tenuis* genome *de novo* from reads generated from Illumina HiSeq3000. Quality control and trimming of low quality reads were performed using FastQC and Trimmomatic. Sequence reads that have passed quality control were binned and classified taxonomically against NCBI's GenBank genome database. Sequence reads were binned using the software Kraken, followed by *de novo* assembly using CLC Genomics Workbench 11.0. The binning results found no sequences belonging to the *Symbiodinium* spp. The assembled genome size was 340,895,668 bp, which was lower than the expected 400 Mbp genome size. The number of contigs and scaffolds produced from the genome assembly was 347,155. Further downstream analysis is needed to annotate the genome and to screen for genes that can be of commercial value.



UMS
UNIVERSITI MALAYSIA SABAH

ABSTRAK

Acropora tenuis merupakan sejenis batu karang staghorn yang penting untuk biodiversiti marin kerana ia membentuk sebahagian besar terumbu karang di dunia. Walaupun banyak kajian biologi yang telah dijalankan terhadap spesies ini secara khususnya, namun tidak banyak kajian genomik yang telah dijalankan. DNA yang diekstrak dari spesimen batu karang bagi tujuan penjujukan ini merupakan sampel yang dikumpul dari alam sekitar. Ini boleh menyebabkan masalah kerana walaupun pada luarannya sampel telah dikenalpasti sebagai *A. tenuis*, namun ada kemungkinan berlakunya hibridisasi antara spesies batu karang *Acropora* yang lain kerana batu karang *Acropora* membiak secara serentak pada satu masa. Batu karang ini juga boleh membuat formasi hybrid bersama ahli genusnya yang lain dan bersama alga dinoflagellate *Symbiodinium spp.* Bagi mengatasi masalah ini, kajian ini bertujuan untuk memperluaskan pemahaman tentang spesies ini dan memperkayakan data genomik awam untuk penyelidikan masa hadapan dengan memasang genom draf *A. tenuis* secara *de novo* daripada bacaan yang dijana dari mesin penjujukan Illumina HiSeq3000. Kawalan kualiti telah dijalankan dan bacaan yang rendah kualitinya telah diasingkan menggunakan FastQC dan Trimmomatic. Bacaan yang telah lulus kawalan kualiti telah diasingkan mengikut kumpulan dan dikelaskan secara taxonomi terhadap database genom NCBI GenBank. Peranti Kraken telah digunakan untuk mengasingkan bacaan mengikut kumpulan, dan kemudian dipasang menggunakan CLC Genomics Workbench. Hasil dari pengasingan mendapati bahawa tiada bacaan dari spesies-spesies *Symbiodinium*. Saiz genom yang telah dipasang adalah sebesar 340,895,668 bp, iaitu lebih kecil dari saiz 400 Mbp yang dianggarkan, manakala jumlah kontig dan perancah yang terhasil adalah sebanyak 347,155. Analisis perlu dijalankan lagi bagi menganotasi genom ini untuk mendapatkan gen-gen yang mempunyai nilai komersil.