PARENTAGE ASSIGNMENT AND GENE EXPRESSION OF GROWTH RELATED GENE IN SNAKEHEAD MURREL, *Channa striata*

FONG POOI HAR

UNIVERSITI SAINS MALAYSIA

2020

PARENTAGE ASSIGNMENT AND GENE EXPRESSION OF GROWTH RELATED GENE IN SNAKEHEAD MURREL, *Channa striata*

by

FONG POOI HAR

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

March 2020

ACKNOWLEDGEMENT

I am very thankful that God has bestowed to me His blessing and with all His grace the strength and resilience to finish my PhD journey. This may probably the single most challenging journey I have taken throughout my life. Throughout this journey I am very lucky to have the strong support and encouragement of Professor Dr. Siti Azizah Mohd Nor, she was very patient and her constant encouragement was among the main reasons I am able complete my PhD journey. I am also very thankful to Dr. Darlina Md Naim who later stepped up and became my main-supervisor in Universiti Sains Malaysia (USM) after Prof Siti's retirement along with Dr. Mohd Ghows Mohd Azzam who guided me in analysis and write up for this study. Both of them were integral in the analysis part and subsequently the write up of this thesis. I credit the work that I have done in this study to these three extraordinary individuals.

It has been a long journey full of ups and downs. I wish to thank my laboratory mates who I have the pleasure to work with and share 6 years of my PhD in USM, Adib, Ana, Ayesha, Danial, Evan, Abang Fadli, Fidaus, Hadi, Ili, Jamsari, Lia, Lim, Mel, Mira, Kak Nurul Norli, Nurul, Siti Hasmah, Siti Zuliana and others I have not mentioned above, all have been there sharing the joy when things were good as well as supportive ideas, advices, guidance, encouragements and your precious knowledge. These 6 years of friendship in USM are among the most valuable friendship I had the pleasure the experience and may our friendship last forever. I am also indebt to my close friend Dr Suganthi Appalasamy who took the time and effort to train me on many aspects of the RNA work even though she was busy with her own works.

I am also very thankful to the Ministry of Higher Education (Now Ministry of Education) for awarding me the MyPhD scholarship under the MyBrain15 scheme to pursue my PhD. I would also like to acknowledge the grant 1001.PBIOLOGI.855003 led by Prof. Dr Alex Chong and shared by Prof. Dr Siti Azizah, which was the source of funding for my research. All of this would not have happened in anywhere else except School of Biological Science, Universiti Sains Malaysia. Hence, I am thankful to my university for various facilities and administrative support.

I would also like to dedicate this PhD thesis to my late father Fong Yeng Wah who has always encouraged me to pursue my education. I hope he is in a better place and have attained peace in the afterlife. I am also thankful to my mother and siblings for their constant encouragement in this long and hard PhD journey.

Like the lyrics of the song “I will be right here waiting for you’ sung by Richard Marx, my husband has always been there for me through thick and thin. We have gone through many obstacles and here we are still together. In his previous love letter he said I complete his life, I am telling him now that my life was never complete before I met him. I don’t aim to be rich or famous, but I want to stay happy with you together along with our pets for the rest of my life.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

cDNA	Complementary DNA
CEMACS	Centre for Marine and Coastal Studies
C-TAB	cetyl trimethyl ammonium bromide
Cyt *b*	Cytochrome *b* gene
ddH_2O	distilled and deionised water
DIA1	Deleted in autism protein 1
DNA	Deoxyribonucleotide acid
dNTP	Dideoxynucleoside triphosphate
EDTA	Ethylenediaminetetraacetic acid
EST	expressed sequence tags
EtOH	ethanol
gDNA	Genomic Deoxyribonucleotide acid
GO	Gene Ontology
HPHL1	Hephaestin Like 1
LIFR	Leukemia inhibitory factor receptor
LOD	Logarithm of the likelihood ratio
$MgCl_2$	magnesium chloride
MYA	million years ago
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
PacBio	Pacific Biosciences
PANTHER	Protein ANalysis THrough Evolutionary Relationships

PAPA	Package for the Analysis of Parental Allocation
PASOS	Parental Allocation of Singles in Open Systems
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
PDLIM7	PDZ and LIM domain protein 7
PIT	Passive integrated transponder
qPCR	Quantitative Polymerase Chain Reaction
RAG1	Recombination Activation Gene-1
RIN	RNA Integrity Number
RNA	Ribonucleotide acid
RNA-seq	RNA sequencing
rRNA	Ribosomal Ribonucleotide acid
SNPs	single nucleotide polymorphisms
TBE	Tris/Borate/EDTA
TPM2	Tropomyosin beta chain

PENENTUAN INDUK DAN PENGEKSPRESIAN GEN BERKAITAN TUMBESARAN DALAM IKAN HARUAN, *Channa striata*

ABSTRAK

Ikan Haruan (*Channa striata*) ialah ikan air tawar yang terkenal dengan nilai perubatannya dan merupakan sumber protein yang penting di Asia Tenggara. Permintaan terhadap ikan ini masih tinggi; justeru, species ini mempunyai potensi pasaran yang baik dalam industri akuakultur di Malaysia. Tambahan pula, penternakan ikan haruan belum popular dan masih tiada program pembiakan terpilih dan penetasan yang sistematik dibangunkan di Malaysia. Oleh itu, objektif pertama bagi penyelidikan ini ialah untuk menyiasat kebolehgunaan primer mikrosatelit yang sedia ada dalam menentukan genotip *C. striata* untuk penentuan induk dan pengurusan pedigri. Objektif seterusnya iaitu menghasilkan rujukan bagi perpustakaan transkriptom untuk penyelidikan ekspresi gen. Dengan perpustakaan transkriptom ini, perbezaan ekspresi gen sebagai tindak balas terhadap tumbesaran pantas dan lambat boleh dikenalpasti dan diaplikasikan dalam program pembiakan terpilih ikan haruan. Penentuan induk yang tepat dalam pengurusan pedigri amat penting untuk kejayaan program penetasan yang sistematik terutamanya dalam industri akuakultur dan mencegah penurunan pembiakbakaan dalaman. Maka, kebolehgunaan tujuh pasangan primer yang dicirikan sebelum ini untuk kajian genetik populasi *C. striata* dinilai keberkesanannya dalam kajian ini. Sejumlah 211 anak ikan daripada 22 stok pembiak baka yang disimpan di dalam enam tangki berasingan telah ditentukan genotipnya berdasarkan tangki dan kemudiannya disusun semula mengikut 22 stok pembiak baka menggunakan lima perisian penentuan induk berlainan. Antara kelima-lima perisian ini, *Colony* adalah perisian

terbaik yang berjaya menentukan anak ikan kepada induknya. Tambahan pula, hasil keputusan penyusunan anak ikan kepada induk menunjukkan bahawa tingkahlaku pengawanan spesies ini adalah secara poligami dan poliandri. Perpustakaan transkiptom *de-novo* untuk *C. striata* telah berjaya dibangunkan dengan tujuh data RNA-seq yang berkualitit tinggi dan menghasilkan 1045123 *contigs* selepas penyusunan transkrip dengan 155973 telah dianotasikan dengan sekurang-kurangnya satu pemadanan dalam pangkalan data. Terdapat 33556 transkrip yang unik kepada UniProt ID, 5313 transkrip unik kepada domain Pfam dan 14727 transkrip yang unik kepada istilah dalam *Gene Ontology*. Satu sampel daripada setiap kumpulan replikat telah dikecualikan daripada analisis pembezaan ekspresi sebab mempunyai korelasi yang rendah berbanding dengan replikat biologi lain dalam hasil keputusan penilaian kualiti replikasi. Dalam analisis pembezaan ekspresi menggunakan dua replikat dalam setiap kumpulan, terdapat 5113 transkrip yang berbeza ekpresinya secara ketara. Terdapat 1749 transkrip yang diekspresikan dalam kumpulan tumbesaran laju dan lambat yang mana 550 dan 1199 transkrip ini telah dinaikkan kadar regulasinya dalam kumpulan tumbesaran lambat dan laju masing-masing. Sejumlah dua jujukan RNA berkadar regulasinya dikurangkan dan tiga jujukan RNA berkadar regulasinya dinaikkan daripada kumpulan tumbesaran lambat dipilih secara rawak untuk pengesahan keputusan RNA-seq menggunakan qPCR. Gen 18S digunakan sebagai gen pengemasan dan rujukan. Keputusannya telah disahkan adalah sama dengan keputusan RNA-seq. Lima gen yang terpilih kemudiannya digunakan sebagai calon biopenanda untuk penskrinan anak ikan haruan yang cepat tumbesarannya. Keputusan perbandingan ukuran fizikal dan kadar ekspresi gen untuk kelima-lima biopenanda ini, menunjukkan bahawa LIFR, HPHL1 dan TPM2 adalah penanda berpotensi untuk mencari anak ikan yang cepat tumbesaran bagi *C. striata*.

PARENTAGE ASSIGNMENT AND GENE EXPRESSION OF GROWTH RELATED GENE IN SNAKEHEAD MURREL, *Channa striata*

ABSTRACT

Haruan (*Channa striata*) is a fresh water fish known for its medicinal value and is an important source of protein in Southeast Asia. The demand for this fish is still high; hence this species has good market potential for aquaculture industry in Malaysia. Furthermore, snakehead farming is not popular and there is yet a selective breeding as well as systematic hatchery programmes developed in Malaysia. Hence, the first objective of this study was to investigate the feasibility of using the existing microsatellite markers in genotyping *C. striata* for parentage identification and pedigree management. The next objective is to establish a reference for transcriptome library for gene expression study. With the transcriptome library, differentially expressed genes in response of fast and slow growth could be identified and applied in snakehead murrel selective breeding programmes. Accurate parentage identification for pedigree management is integral to a successful systematic hatchery programme especially in aquaculture industry to prevent inbreeding depression. Hence, the feasibility of seven pairs of microsatellite markers characterized previously for a population genetics study of *C. striata* was evaluated in this study. A total of 211 offspring from 22 brood stocks kept in six separated tanks were genotyped according to tanks and were later assigned back to these 22 brood stocks using five different parentage assignment software. Among these software, Colony was the best software to successfully allocate the offspring back to their parents. In addition, the results of the parentage allocations revealed that mating behaviour of this species is polygamous and polyandry. *De novo* transcriptome

library of *C. striata* was successfully constructed with seven high qualities of RNA-seq data and yielded 1045123 contigs after transcript assembly and 155973 were annotated with at least one data database hits. There were 33556 transcripts unique to UniProt IDs, 5313 transcripts unique to Pfam domains and 14727 unique transcripts to Gene Ontology terms. A sample from each group was excluded from the differential expression analysis due to low correlation to other biological replicates in their respective groups in replicates quality assessment. In the differential gene expression analysis with two biological replicates in each group, there were a total of 5113 transcripts were significant differentially expressed. There were 1749 transcripts were expressed in both fast and slow growing groups while 550 and 1199 were up-regulated in slow and fast growing group respectively. A total of two transcripts of the down-regulated and three transcripts of the up-regulated in the slow growing groups were pick randomly to validate the RNA-seq results using qPCR. The 18S gene was used as the housekeeping gene as well as the reference gene. The validation results were congruent with the RNA-seq results. The five selected genes were then used as the candidate of biomarkers to screen for fast growing *C. striata* fingerlings. The results of the fingerlings physical measurements and expression levels of these five candidate biomarkers showed that LIFR, HPHL1 and TPM2 were potential markers for screening of fast growing fingerlings for *C. striata*.