UNIVERSITI PUTRA MALAYSIA

VERIFICATION OF SEQUENCES DERIVED FROM *DE NOVO* ASSEMBLY OF mRNA SEQ. DATA OF *GRACILARIA CHANGII*

NUR AFIZA MOHD ZAINUDDIN

FBSB 2015 160

VERIFICATION OF SEQUENCES DERIVED FROM *DE NOVO* ASSEMBLY OF mRNA SEQ. DATA OF *GRACILARIA CHANGII*

NUR AFIZA BINTI MOHD ZAINUDDIN

BACHELOR OF SCIENCE (HONS.)

UNIVERSITI PUTRA MALAYSIA

2015

VERIFICATION OF SEQUENCES DERIVED FROM *DE NOVO* ASSEMBLY OF mRNA SEQ. DATA OF *GRACILARIA CHANGII*

By

NUR AFIZA BINTI MOHD ZAINUDDIN

Thesis Submitted to the Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Science.

Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Science (Hons) Cell and Molecular Biology

June 2015

Abstract of thesis presented to the Department of Cell and Molecular Biology in fulfilment of the requirement for the Degree of Science (Hons.) Cell and Molecular Biology

VERIFICATION OF SEQUENCES DERIVED FROM *DE NOVO* ASSEMBLY OF mRNA SEQ. DATA OF *GRACILARIA CHANGII*

By

NUR AFIZA BINTI MOHD ZAINUDDIN

JUNE 2015

Chair: Assoc. Prof. Dr. Ho Chai Ling

Faculty: Faculty of Biotechnology and Biomolecular Science.

De novo RNA –Seq assembly facilitates the study of transcriptomes without the need for a genome sequence especially for non-model organisms. However, assembly of such data may produce errors. One of these errors could be due to direct and inverted repeats in sequences. However, the presence of repeat contents could be due to either assembly problem or post transcriptional modification in the cells. This study was conducted to verify the sequences (1_CL2250Contig1 (Gc2250), CC_1_CL1944Contig1 (Gc1944), 1_CL8489Contig1 (Gc8489) and CC _Lc_4470 (Gc4470)) derived from *de novo* assembly of mRNA sequencing data of *Gracilaria changii* which contain direct and inverted repeats. The cDNA of control and sulphate deprivated *G.changii* samples were amplified by polymerase chain reaction (PCR) using specific primers. The PCR products were cloned into the TOPO TA vector and transformed into *Escherichia coli* DH5α. Colony PCR was performed for confirmation of positive transformants. The TOPO TA vector carrying the inserted sequence was extracted from the host. For each PCR products, plasmid DNA samples from three positive clones were sequenced using M13 universal primer. The sequencing results showed that, the repeats were due to the assembly errors.

Keywords: *De novo* assembly, repeat contents, sequencing

Abstrak tesis yang dikemukakan kepada Jabatan Biology Sel dan Molekul

sebagai memenuhi keperluan untuk ijazah Sains (Kepujian) Biologi Sel dan Molekul.

PENGESAHAN JUJUKAN YANG DIPEROLEHI DARIPADA *DE NOVO* HIMPUNAN DATA mRNA SEQ *GRACILARIA CHANGII*

Oleh

NUR AFIZA BINTI MOHD ZAINUDDIN

JUN 2015

Pengerusi: Prof. Madya Dr. Ho Chai Ling

Fakulti: Fakulti Bioteknologi dan Sains Biomolekul.

De novo RNA -Seq himpunan memudahkan kajian transkriptom tanpa memerlukan urutan genom terutama bagi organisma bukan model. Walau bagaimanapun, himpunan data tersebut boleh menghasilkan ralat. Salah satu ralat ini mungkin disebabkan jujukan langsung dan songsang. Walau bagaimanapun, kehadiran kandungan berulang boleh disebabkan sama ada masalah himpunan atau pengubahsuaian selepas transkripsi dalam sel-sel. Kajian ini dijalankan untuk mengesahkan jujukan (1_CL2250Contig1 (Gc2250), CC_1_CL1944Contig1 (Gc1944), 1_CL8489Contig1 (Gc8489) and CC _Lc_4470 (Gc4470)) berasal dari *de novo* himpunan data mRNA jujukan *Gracilaria changii* yang mengandungi ulangan langsung dan songsang. cDNA kawalan dan kekurangan sulfat sampel *G.changii* telah diamplifikasi melalui tindak balas berantai polimerase (PCR) menggunakan pencetus tertentu. Produk PCR telah diklonkan ke dalam vektor TOPO TA dan ditransformasi ke dalam *Escherichia coli* DH5α. Penyaringan koloni melalui PCR telah dilakukan untuk pengesahan transforman positif. Vektor TOPO TA yang membawa jujukan selitan telah diektsrak daripada perumah. Untuk setiap produk-produk PCR, sampel DNA plasmid dari tiga klon positif telah dijujuk menggunakan M13 pencetus umum. Keputusan penjujukan menunjukkan ulangan adalah disebabkan oleh kesilapan himpunan.

Kata kunci: *De novo* himpunan, kandungan berulang, penjujukan

Approval

This thesis was submitted to the Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences accepted as fulfilment of the requirement for the Degree of Science (Hons.) Cell and Molecular Biology. The member of the Supervisory Committee was as follows:

Name of Chairperson:

Assoc. Prof Dr. Ho Chai Ling,

Faculty of Biotechnology and Biomolecular Science,

Universiti Putra Malaysia.

……

(Prof. Madya Dr. Janna Ong Abdullah)

Head Department Cell and Molecular Biology,

Faculty of Biotechnology and Biomolecular Science,

Universiti Putra Malaysia

Date: .

Declaration

Declaration by undergraduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by the Department of Cell and Molecular Biology.
- written permission must be obtained from before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials
- there is no plagiarism or data falsification/fabrication in the thesis. The thesis has undergone plagiarism detection software (TURNITIN).

Signature:……. Date:……..

Name and Matric No:……..

Declaration by Supervisor

This is to confirm that:

- the research conducted and the writing of this thesis was under supervision;

Signature:…….

Name of Chairman of Supervisory:…….

ACKNOWLEDGEMENT

First and foremost, I would like to express my infinity gratitude to my beloved supervisor, Assoc. Prof. Dr. Ho Chai Ling for her guidance and patience to guide me throughout my final year project. She is very supportive and always gave any information that I need for my project.

Millions thanks go to Lee Wei Kang and other postgraduate students in Cell and Molecular laboratory for their knowledge, guidance, help and advice while I'm doing my project. I also would like to dedicate special thanks to all the laboratory assistants especially Pn. Noor Hanaliza Nandzori for their contribution either directly or indirectly during my final year project.

Next, I like to expand my special thanks to my friends who really helped me by giving moral support, ideas throughout my journey to finish this project. They are so supportive, friendly and helpful.

Not forgotten also, my beloved family especially my parents for their support, advice that they gave from the start of the project until I managed to finish writing of this thesis.

Lastly, thanks for those who involved in my project, directly or indirectly, your kindness and co-operation are appreciate. May Allah bless all of you. Thank you.

TABLE OF CONTENTS

ABSTRACT Ii

ABSTRAK iii

APPROVAL iv

DECLARATION v

ACKNOWLEDGEMENT vi

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF ABBREVIATIONS xii

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW

2.1 *Gracilaria Changii* 4

2.2 Sequencing 5

2.3 Next generation sequencing (NGS) and applications 6

2.4 De novo assembly of NGS 8

2.4.1 Principle of *de novo* assembly 9

2.4.2 Limitations of *de novo* assembly 10

2.4.3 Verification of *de novo* assembled sequences 11

3 MATERIAL AND METHODOLOGY

3.1 Sequence analysis and primer design 12

3.2 Polymerase Chain Reaction (PCR) 13

3.3 Purification of PCR product 14

3.4 A –tailing and cloning of the PCR products 16

3.5 Transformation 16

3.6 Colony PCR 17

3.7 Conventional plasmid isolation 17

3.8 Plasmid isolation using kit 18

3.9 Sequencing 19

4 RESULTS AND DISCUSSION

4.1 PCR amplification 20

4.2 Purification of PCR products 23

4.3 Transformation 27

4.4 Screening for positive transformants 27

4.5 Plasmid isolation

4.5.1 Conventional plasmid isolation 31

4.5.2 Isolation of plasmid using kits 32

4.6 Sequencing analysis 35

5 CONCLUSION AND RECOMMENDATION 41

REFERENCES 42

APPENDICES

LIST OF TABLES

Table	Title	Page
Table 1	Applications of NGS	8
Table 2	Primer sequences	13
Table 3	Analysis of purified PCR products	26

LIST OF FIGURES

Figure		Page
Figure 1	PCR amplification of Gc2250 (A) and Gc8489 (B) using gene specific primers with Taq polymerase	21
Figure 2	PCR amplification of Gc1944(A) and Gc4770 (B) using gene specific primers with Taq polymerase	22
Figure 3	Optimization of annealing temperature of Gc4770 using gradient PCR	24
Figure 4	PCR amplification of Gc2250 and Gc8489 using gene specific primer and KOD Hot Start Polymerase	24
Figure 5	PCR amplification of Gc1944 (A) and Gc 4770 (B) using gene specific primer and KOD Hot Start Polymerase.	25
Figure 6	Screening of positive transformants for Gc8489 (A) and Gc2250 (B) using colony PCR	29
Figure 7	Screening of positive transformants for Gc4770(A) and Gc1944(B) using colony PCR	30
Figure 8	Plasmid extraction of Gc8489 and Gc2250 using commercial kits	33
Figure 9	Plasmid extraction of Gc1944 (A) and G4770 using commercial kits	34
Figure 10	Multiple alignment of sequence derived from de novo assembly of RNA-Seq reads (1_CL8489Contig1 /Gc8489), sequences of Gc8489 which were PCR-amplified from control (C) and treated (T) samples, and sequence derived from re-assembly of RNA-Seq reads (with Oases omitted from the assembly pipeline)(1_CL4721Contig1)	37
Figure 11	Multiple alignment of sequence derived from de novo assembly of RNA-Seq reads (1_CL2250Contig1 /Gc2250), sequences of Gc2250 which were PCR-	38

amplified from control (C) and treated (T) samples, and sequence derived from re-assembly of RNA-Seq reads (with Oases omitted from the assembly pipeline)(1_CL552Contig1).

Figure 12 Multiple alignment of sequence derived from de novo assembly of RNA-Seq reads (CC_LC_4770/Gc4470), sequences of Gc4470 which were PCR-amplified from control (C) and treated (T) samples, and sequence derived from re-assembly of RNA-Seq reads (with Oases omitted from the assembly pipeline)(1_CL951Contig1). 39

Figure 13 Multiple alignment of sequence derived from de novo assembly of RNA-Seq reads (CC_1_CL1944Contig1 /Gc1944), sequences of Gc1944 which were PCR-amplified from control (C) and treated (T) samples, and sequence derived from re-assembly of RNA-Seq reads (with Oases omitted from the assembly pipeline)(1_CL1171Contig1) 40

LIST OF ABBREVIATIONS

bp	Base pair
cDNA	Complementary deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
dATP	Deoxyadenosine triphosphate
dNTP	Deoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
IPTG	Isopropyl-beta-D-thiogalactopyranoside
LB	Luria bertani
min	minute
mM	Milimolar
mg	Miligram
ml	Mililitre
ng	Nanogram
NaoAC	Sodium Acetate
NaOH	Sodium hydroxide
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
s	second
SDS	Sodium dodecyl sulphate
TE	Tris-EDTA (Ethylenediamine Tetraacetic Acid)
Tris	Tris(hydroxymethyl)amino methane
µl	Microlitre
µM	Micromolar
V	Volt
w/v	Weight/Volume
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside

CHAPTER 1

INTRODUCTION

1.1 Research background

DNA sequencing techniques are important tools in many fields such as in genetic study, biotechnology, molecular biology and forensic sciences. The most common approach used to sequencing DNA was introduced by Frederick Sanger in 1977 which is known as the Sanger sequencing method. The design of his method is still very popular today. Sanger's method is also known as dideoxy sequencing or chain termination method. This technique uses dideoxynucleotide triphosphates (ddNTPs) labelled with radioactive labelling such as dideoxyadenine (ddATP) and dideoxyguanine (ddGTP). These molecules resemble normal nucleotides but lack the normal -OH group. The replication process will stop every time the corresponding ddNTP is incorporated. The products are then separated through four lanes of a polyacrylamide gel and scored according to their molecular masses (Ji *et al.,* 2008). In 1986, the Caltech team developed an automated platform using fluorescent detection of termination which is less harzadous. The former method is a radioactive method which requires four gel lanes, one for each reaction vessel while the later method uses nonradioactive fluorescence and requires only one gel lane because each nucleotide has a distinct colour. The first generation sequencing is laborious and costly (Ulucan 2014).

After years of improvement, the Next-Generation Sequencing (NGS) has been developed. The NGS enables researchers to study biological systems at a level never before possible. The NGS usually produces a large amount of data in a short time with reduced cost (Medini *et al*., 2008). There are three major sequencing platforms

for NGS (454 sequencing, SOLID System and Illumina sequencing) and they do not rely on the traditional method (Haridas *et al*., 2011). These platform produce short reads approximately 50 to 400 nucleotides depending on the selected platform. Among them, Illumina has been commonly used for sequencing purpose such as for model and non-model organisms (De Donato *et al*., 2013).

De novo assembly is widely used for assembly of RNA sequencing (RNA-seq) data of non-model organisms and metagenomic data (Zhang *et al*., 2012). According to Paszkiewicz *et al* (2010), *de novo* sequence assembly is a process whereby individual sequence reads were merged to form long contiguous sequences ('contigs') sharing the same nucleotide sequence as the original template DNA from which the sequence reads were derived. It allows the study of transcriptomes without the needs of genome sequences. RNA has become a cost effective way to quickly obtain sequence information for non –model systems. RNA-seq reads represent short pieces of all the mRNA present in the tissue at the time of sampling. RNA-seq is a robust tool for simultaneous transcriptome characterization and differential gene expression (DGE) analysis in a single cell type, tissue type, or entire organism under defined conditions (Peterson *et al*., 2012).

Although *de novo* assembly offers many advantages in several ways compared to reference based strategy, it also has several drawbacks. The sequences derived from *de novo* assembly may contain repeats content, contamination from other organisms or new insertions and segmental duplication (Alkan *et al*., 2011). The repeat content of repeats is one of the biggest challenges with *de novo* assembly resulting in reduced or lost genomic complexity. Besides that, *de novo* plant genome sequencing and assembly are challenging due to their large complex plant genomes (Pellicer *et al*., 2010). Extraction of large quantities of RNA from plant sample with high quality

is difficult, this would result difficulty in preparation good libraries for sequencing purpose (Schatz *et al.,* 2012).

In this study, the sequences derived from *de novo* assembly of RNA seq data of *Gracilaria changii* will be verified. In order to identify the type of errors produced a few steps are required. RNA will be extracted from fresh weed sea and converted into complementary DNA (cDNA). Both forward and reverse primers will be designed based on the *de novo* assembly sequencing result which contains directed repeats and inverted repeats. Next, PCR amplification will be conducted, follow with A-tailing. The products will be cloned into TOPO TA vector and transformed into DH5α. The plasmid DNA will be extracted and sequenced.

1.2 Problem statement

De novo assembly of RNA-Seq data allows us to study transcriptomes without the need for a genome sequence especially for non-model organisms. However, assembly of such data into high-quality data may have some errors. Therefore, the focus of this study was to verify the sequences derived from *de novo* assembly of mRNA sequencing data of *Gracilaria changii* .

1.3 Objectives:

To verify de *novo* assemblies of RNA-seq from *Gracilaria changii.*

REFERENCES

Alkan C, Sajjadian S, & Eichler E E (2011). Limitations of next-generation genome sequence assembly. *Nature Methods* 8 ,61-65.

Ansorge J W (2009). Next-generation DNA sequencing techniques. *New Biotechnology* 25 ,195-202.

Bautista R, Claros M G, Darío G F, Benzerki H, Seoane P, & Noe F P (2012). Why Assembling Plant Genome Sequences Is So Challenging. *Biology 2012* 1 ,439-459.

Egger G, Liang G, Aparicio A, & Jones P A (2004). Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429 ,457-463.

Capobianchi M R, Giombini E, & Rozera G (2013). Next-generation sequencing technology in clinical virology National Institute for Infectious Diseases. *Clinical Microbiology and Infection* 19 ,15-22.

Chan C X, Teo S S, Ho C L, Othman R Y, & Phang S M (2004). Optimisation of RNA extraction for marine red alga, *Gracilaria changii* (Gracilariales, Rhodophyta). *Journal of Applied Phycology* 16 ,297-301.

Claros M G, Bautista R, Fernández D G, Benzerki H, Seoane P, & Pozo F N (2012). Why Assembling Plant Genome Sequences Is So Challenging. *Biology 2012* 1 ,439-459.

Darling A E, Andrew T, Jonathan A E & Marc T F (2011). Genome analysis. *Advanced access publication* 27 ,2756-2757.

De Donato M, Peters S O, Mitchell S E, Hussain T, Imumorin I G (2013). Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *Public library of Science* 8.

De Wit P, Pespeni M H, Ladner J T, Barshis D J, Seneca F, Jaris H, & Palumbi S R (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* 12 ,1058-1067.

Destombe C & Kain J M (1995). A review on the life history, reproduction and phenology of Gracilaria. *Journal of Applied Phycology* 7: 269-281.

Erwin L, Auger H, Jaszczyszyn Y, & Thermes C (2014). Ten years of next-generation sequencing technology. *Cell Press* 30 ,418-423.

Garibyan L (2013). Polymerase Chain Reaction. *Journal of Investigated Dermatology* 6 ,1-4.

Grada A, & Weinbrecht K (2013). Next-Generation Sequencing: Methodology and Application. *Journal of Investigative Dermatology* ,1-4.

Grabherr M G, Hass B J, Yassour M, & Regev A (2011). Full length transcriptome assembly from RNA-Seq data without a reference genome. *Nature America* 29 ,644-652.

Haridas S, Breuill C, Bohlmann J, & Hsiang T (2011). A biologist's guide to *de novo* genome assembly using next-generation sequence data: A test with fungal genomes. *Journal of Microbiology Methods* 86 ,368-375.

Hecker K H, & Roux K H (1996). High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *National Library of Medicine National Institutes of Health* 20 ,478-485.

Imelfort l, & Edwards D (2009). *De novo* sequencing of plant genomes using second generation technologies. *Briefing In Bioinformatics* 10 ,609-618.

Jackson S A, Varshney K R, Nayak N S, & May D G (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Cell Press* 27 ,522-530.

Jarvie T (2005). Next generation sequencing technologies: Drug discoveries Today. *Emerging Technologies* 2 ,255-260.

Ji H & Shendure J (2008). Next generation DNA sequencing. *Natural Biotechnology* 26 ,1135-1143.

Kitabayashi M, Nishiya Y, Esaka M, Itakura M, & Imanaka T (2002). Gene cloning and polymerase chain reaction with proliferating cell nuclear antigen *from Thermococcus kodakaraensis*, KOD1. *Bioscience Biotechnology Biochemistry* 10 ,2194-2200.

Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, & Bai Y (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* 463 ,311-317.

Li R , Zhu H, Jue R, Qian W, Fang X, Shi Z, Li Y, Li S, Gao S, Kristiansen K, Li S, Yang Hu, Wang J, & Wang J (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Biology* 20 ,265–272.

Li B, Fillmore N, Bai Y, Collins M, Thomson A J, Stewart R, & Dewey N C (2014). Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. *Genome Biology* ,1-21.

Liu S T, LU B X, & Bing Z Z (2012). Comparative study of *de novo* assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Science China Life Sciences* 56 ,143-155.

Maxam A M, & Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of United State America* 74 ,560-564.

Morozova O, &. Marra M A (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92 ,255-264.

Munshi A (2012). DNA Sequencing – Methods and Applications. Intech books and journal online, 13-21.

Paszkiewicz K, & Studholme S J (2010). *De novo* assembly of short sequence reads. *Briefing In Bioinformatics* 11 ,457-472.

Paterson A H, Freeling M, Tang H, & Wang X (2010) . Insights from the comparison of plant genome sequences. *Annual Review Plant Biology* 61 ,349-372.

Palacios G, Druce J, & Du L (2008). A new arenavirus in a cluster of fatal transplant associated diseases. *New England Journal Medicine* 358 ,991–998.

Pellicer J, Fay M F, & Leitch I J (2010). The largest eukaryotic genome of them all? *Botanical Journal of the Linnean* 164 ,10-15.

Phang M S, Shaharuddin S, Hashim Noraishah, & Sasekumar A (1996). Studies on *Gracilaria changii* (Gracilariales, Rhodophyta) from Malaysian mangroves. *Kluwer Academic Publishers* 347-352.

Phang S M, & Lewmanomont K (2001). *Gracilaria changii. Plant Resources of South-East Asia,* 178-180.

Rozera G, Capobianchi M R, & Giombini E (2012). Next-generation sequencing technology in clinical virology. *Clinical Microbiol Infect 2013* 19 ,15-22.

Russell P (2002). *iGenetics.* Pearson Education, Incorporation San Francisco, pp 187-189

Sambrook J, Fritsch E F, & Maniatis T (1989). Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S, & Coulson A R (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74 ,5463–5467.

Schatz M C, Witkowski J, & McCombie W R (2012). Current challenges in *de novo* plant genome sequencing and assembly.*Genome Biology* 13 ,1-7.

See H C, & Wynne M J (1985). Introduction to the algae: Structure and reproduction. Englewood Cliffs, New Jersey: Prentice-Hall.

Sekvenciranja I (2013). The history of dna sequencing. *Journal of Medical Biochemistry* 32 ,301–312.

Shendure J, & Ji H (2008). Next-generation DNA sequencing. *Nature Biotechnology* 26 ,1135-1145.

Schuster S C (2008) Next-generation sequencing transforms today's biology. *Nature Methods* 5 ,16–18.

Towner J S, Sealy T K, Khristova M L, Downing R, Albarino C G, Conlan S, Reeder S A, Quan P, Lipkin W I, Tappero J W, & Okware S (2008). Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *Public library of Science Pathogen* 4 ,1-7.

Tritt A, Eisen A J, Facciotti T M, & Darling E A (2012). An Integrated Pipeline for de Novo Assembly of Microbial Genomes. *Public library of Science One* 7 ,1-9.

Ulucan K (2014). DNA Sequencing, Sanger and Next-Generation Sequencing. *Applications of Molecular genetics in Personalized Medicine* ,3-7. OMICS Group eBooks.

Xie L L, Clarke K, Yi Y, Ronald M, & Zhang K K (2013). Comparative analysis of *de novo* transcriptome assembly. *Science China Life Sciences* 56 ,156-162.

Yadav P, Yadav A, Garg V, Datta K T, Goswami L S, & De S (2011). A novel method of plasmid isolation using laundry detergent. *Indian Journal of Experimental Biology* 49 ,558-560.

Yang Y, & Smith A S (2013). Optimizing *de novo* assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics* 14 ,1-11.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, & Qi H (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490 ,49–54.