COMPARATIVE GENOMIC ANALYSIS OF *VIBRIO CHOLERAE* AND ITS COLONIZATION FACTORS

CINDY TEH SHUAN JU

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILISOPHY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2011

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Cindy Teh Shuan Ju I.C/Passport No.: 820129-01-5654

Registration /Matric No: SHC080021

Name of Degree: Doctor of Philosophy (PhD)

Title of Project Paper/Research Report/ Dissertation/ Thesis ("this Work"):

Comparative Genomic Analysis of *Vibrio cholerae* and its Colonization Factors

Field of Study: Molecular Microbiology

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

Abstract

Vibrio cholerae, the causative agent of cholera, is endemic in Malaysia and is spread through the ingestion of contaminated food or water. The aims of this study were to develop a multiple PCR assay for differentiation of *V. cholerae* from other *Vibrio* species; to compare the efficiency of PCR assay with conventional biochemical tests and API 20E.; to characterize the strains based on their biotypes, serogroups, and virulotype by using a multiplex PCR assay; to determine the genetic relatedness of strains by using genotyping methods such as RAPD-PCR, ERIC-PCR, REP-PCR, MLVA, PFGE, MLST and MVLST; to study the virulence factors which cause colonization in different serogroups of *V. cholerae*; and the influence of host environment for colonization. Four pairs of primers were designed for differential detection of sister groups of *Vibrio* species Strains tested were differentiated into *V. cholerae* (493/338 bp), *V. parahaemolyticus* (493/409 bp), *V. vulnificus* (493/656 bp), *Vibrio* species (493 bp), and non-*Vibrio* (no amplification) based on *pntA* and *gyrB* genes. This multiplex PCR assay was more sensitivie and specific than API 20E identification assay. Another multiplex PCR assay based on *ompW*, *hlyA*, *orf* complex, *toxR*, *ctxA*, *tcpI* for *V. cholerae* biotyping, serogrouping and virulotyping was developed and tested on 43 *V. cholerae* strains. A total of 22 El Tor O1 and one O139 *V. cholerae* that harboured all virulence genes were identified. One El Tor O1 *V. cholerae* presented identical virulotype to 17 other non-O1/non-O139 *V. cholerae*, while the *tcpI* gene was detected in two non-O1/non-O139 *V. cholerae*. The 43 strains were also subtyped into 38, 40, 35, 30, 35, 38, 29 and 27 profiles by RAPD-PCR, ERIC-PCR, REP-PCR, VCR-PCR, PFGE, MLVA, MLST and MVLST, respectively with discriminatory power ranging from 0.910 to 0.996. Overall, genetic diversity of non-O1/non-O139 *V. cholerae* strains was high while some of the O1 strains were indistinguishable. However,

the unrelated strains which shared the same profiles were distinguished based on the combined analyses of the eight genotyping methods. However, each method possesses its own limitations. MLST and MVLST gave precise description of point mutation but were expensive. Overall, MLVA developed in this study remains the most suitable genotyping methods based on discriminatory ability, ease of operation, cost, timeline and data management. However, a combination of several genotyping methods may overcome the inefficiency of each single method and therefore able to distinguish unrelated strains. Finally, clinical and environmental O1 strains could colonize the mouse intestines but prolonged colonization was only observed with environmental strains which showed upregulated expression of *rtxA* and *hlyA* genes. The *tcpI*$^{+}$ non-O1/non-O139 *V. cholerae* strain was a more efficient colonizer while the non-toxigenic O1 *V. cholerae* could colonize the mouse intestine once the virulence genes were favourably enriched and 'turn-on' in the host environment. In conclusion, this study provided alternative approaches for rapid differentiation of *V. cholerae* from other pathogenic *Vibrio* species, as well as to biotype, serogroup and virulotype the *V. cholerae* strains. Regardless of serogroups, year, source and location of isolation, all the unrelated strains were distinguishable and therefore suggests a high diversity of *V. cholerae* population in Malaysia. Different traits of strains posses different colonization ability and *tcpI* gene might be the key regulator for colonization in non-O1/non-O139 *V. cholerae*. However, colonization in non-toxigenic O1 *V. cholerae* might be facilitated once the virulence genes were 'enriched' in the host environment.

Abstrak

Vibrio cholerae merupakan agen penyebab kolera, adalah endemik di Malaysia dan teserbar melalui pengambilan makanan atau air yang tercemar. Terdapat beberapa tujuan dalam menjalankan kajian ini, terutamanya menghasilkan beberapa kaedah PCR untuk membezakan *V. cholerae* daripada spesis *Vibrio* yang lain serta membandingkan kecekapan PCR dengan ujian biokimia konvensional dan API 20E (suatu kit untuk pengenalan bakteria). Kajian ini juga bertujuan untuk menghasilkan suatu PCR multipleks yang boleh membezakan strain-strain dari berbagai ciri biokimia, jenis kumpulan serologi, dan gen kevirulenan. Seterusnya, perkaitan genetik antara strain disiasat dengan menggunakan kaedah genotip seperti RAPD-PCR, ERIC-PCR, REP-PCR, MLVA, PFGE, MLST dan MVLST. Selain itu, faktor kevirulenan yang menyebabkan gastroenteritidis bagi strain-strain *V. cholerae* yang berbeza kumpulan serologi juga dikaji. Keseluruhannya, empat pasang primer telah direka untuk mengesan perbezaan spesies *Vibrio* yang berlainan. PCR multipleks ini mampu membezakan *V. cholerae* (493/338 bp), *V. parahaemolyticus* (493/409 bp), *V. vulnificus* (493/656 bp), *Vibrio* species (493 bp), dan juga organisma bukan-*Vibrio* (tiada amplifikasi) berdasarkan gen *pntA* dan *gyrB*. PCR multipleks ini menunjukkan kepekaan dan kekhususan sebanyak 1.0 jika dibandingkan dengan ujian biokimia konvensional dan API 20E dalam pengesanan *V. cholerae*. Ujian biokimia konvensional juga didapati lebih berkesan dan khusus berbanding dengan API 20E. Di samping itu, PCR multipleks berdasarkan gen-gen *ompW*, *hlyA*, kompleks *orf*, *toxR*, *ctxA*, dan *tcpI* untuk membezakan *V. cholerae* telah dihasilkan. Sebanyak 22 El Tor O1 dan satu *V. cholerae* O139 yang mengandungi semua gen virulens telah dikenalpasti. Satu El Tor V. O1 cholerae telah menunjukkan jenis kevirulenan yang sama dengan 17 strain non-O1/non-O139. Gen *tcpI* dikesan dalam dua *V. cholerae* non-O1/non-O139. Empat puluh

tiga strain juga dibahagikan kepada 38, 40, 35, 30, 35, 38, 29 dan 27 corak dengan menggunakan RAPD-PCR, ERIC-PCR, REP-PCR, VCR-PCR, PFGE, MLVA, MLST dan MVLST, masing-masing. Kuasa pembezaan bernilai antara 0.910-0.996. Secara keseluruhan, kepelbagaian genetik strain *V. cholerae* non-O1/non-O139 adalah tinggi sementara terdapat beberapa strain O1 yang tidak dapat dibezakan. Strain-strain tertentu yang berkongsi jenis corak tersebut kemudiannya dibezakan berdasarkan analisis gabungan dari lapan kaedah genotip. Selain itu, setiap kaedah mempunyai kelemahannya. MLST dan MVLST memberikan gambaran tepat dari segi mutasi titik tetapi berkos tinggi. Secara keseluruhan, MLVA yang dihasilkan dalam kajian ini masih merupakan kaedah yang paling sesuai untuk menaipkan genotip *V. cholerae* berdasarkan kemampuan diskriminasi, kemudahan operasi, kos, masa dan pengurusan data. Walaubagaimanapun, penggabungan beberapa kaedah menaip genotipik dapat mengatasi ketidakcekapan daripada sesuatu kaedah dan strain yang tidak berkaitan dari segi genetik dapat dibezakan. Sementara itu, strain O1 dari klinikal dan persekitaran menunjukkan penjajahan usus tikus tapi penjajahan berpanjangan hanya didapati pada strain persekitaran yang menunjukkan pengekspresan gen *rtxA* dan *hlyA* yang tinggi. Strain *V. cholerae* bukan serogroup O1/O139 yang mengandungi *tcpI* adalah agen gastroenteritidis yang lebih kuat sedangkan strain O1 yang tidak toksik dapat menjangkit usus tikus selepas gen kevirulenan diperkayakan dan dimulakan dalam persekitaran perumah. Sebagai kesimpulan, kajian ini telah memberikan pendekatan alternatif untuk membezakan *V. cholerae* daripada *Vibrio* spesies lain yang berpatogenik selain daripada membezakan strain-strain *V. cholerae* yang berlainan ciri biokimia, kumpulan serologi dan kevirulenan dengan cekap. Semua strain yang tidak berkaitan dapat dibezakan dalam kajian ini dan menyarankan genetik kepelbagaian yang tinggi dalam populasi *V. cholerae* di Malaysia. Strain yang bersifat berlainan mempunyai kemampuan penjangkitan usus yang berbeza dan gen *tcpI* diyakinkan adalah salah satu

faktor di kalangan *V. cholerae* yang non-O1/non-O139 untuk menyebabkan gastroenteritidis. Sementara itu, penjajahan usus bagi *V. cholerae* O1 yang tidak berpatogenik dapat diamati selepas gen-gen kevirulenan 'diperkayakan' dalam persekitaran tuan rumah.

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude to my supervisor, Professor Dr. Thong Kwai Lin for her valuable guidance, supervision throughout the whole research study and providing all necessary facilities to carry out this study. Her encouragement and guidance are very much appreciated. Besides, I would like to thank my co-supervisor, Dr. Chua Kek Heng for his supervision, valuable comments and helpful advises during the study.

I would like to thank my fellow lab mates, King Ting, for her suggestion, assistance, and motivation in writing this thesis; Boon Hong, Junie, Shabnam, Soo Tein, Wing Sze, Xiu Pei, Lai Kuan, Javad, Hosein and Omar for their kind helps and supports.

This thesis is dedicated to my beloved husband, father and mother. I would like to express my appreciation to my husband, Chun Wie for his persistent support, suggestions, and never-ending encouragement. A special thank to my father who provided me the inspiration to pursue my study, his all-time support, and pushing me to complete this thesis in time. I am grateful to my mother for her kind help, and being the great listener and my motivator in doing research and writing this thesis. To my father and mother-in-law, thank you for taking care of my son, so that I could concentrate in my research and thesis writing.

Finally, I would like to acknowledge IMR for contributing strains in this study. This study is financed by Science & Technology Research Grant from Malaysia Toray Science Foundation (MTSF) 2009, and Universiti Malaya for the scholarship and PPP grant (PS250/2008A).

TABLE OF CONTENTS

CHAPTER TWO: LITERATURE REVIEW

CHAPTER THREE: DEVELOPMENT OF A MULTIPLEX PCR ASSAY FOR *VIBRIO CHOLERAE*, *VIBRIO PARAHAEMOLYTICUS*, *VIBRIO VULNIFICUS* AND *VIBRIO* SPECIES DIFFERENTIATION

CHAPTER FOUR: COMPARISON OF PCR, BIOCHEMICAL TESTS AND API 20E FOR DETECTION OF AQUATIC *VIBRIO CHOLERAE* IN MALAYSIA

CHAPTER FIVE: DEVELOPMENT OF A MULTIPLEX PCR FOR BIOTYPING, SEROGROUPING AND VIRULOTYPING OF *VIBRIO CHOLERAE*

LIST OF FIGURES

LIST OF TABLE

LIST OF SYMBOLS AND ABBREVIATIONS

bp	base pair
cfu	colony forming unit
ddH_2O	double distilled water
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
EDTA	ethylenediamine-tetraacetic acid
ERIC	enterobacterial repetitive intergenic consensus
IMR	Institute for Medical Research
kb	kilobase
LB	Luria Bertani
M	molar
$MgCl_2$	magnesium chloride
min	minute
ml	mililitre
MLVA	multiple-locus variable-number of tandem repeat analysis
mM	milimolar
NaCl	sodium chloride
ng	nanogram
OD	optical density
PCR	polymerase chain reaction
PFGE	pulse-field gel electrophoresis
RAPD	random amplified polymorphic DNA
REP	repetitive extragenic palindromic

rpm	revolutions per minute
RT	room temperature
Taq DNA polymerase	Thermus aquaticus DNA polymerase
TBE	Tris-borate EDTA
TCP	toxin-coregulated pili
TE	Tris-EDTA
Tris-HCl	Tris-hydrochloric acid
Tm	melting temperature
U	unit
UV	ultraviolet
V	volt
VCR	*V. cholerae r*epeats
VPI	*Vibrio cholerae* pathogenicity island
WHO	World Health Organization
w/v	weight/volume
~	approximately
°C	degree Celcius
=	equals to
<	less than
>	more than
%	percent
µl	microlitre
µg	microgram
µM	micromolar