MOLECULAR EPIDEMIOLOGY OF *Salmonella enterica* subspecies *enterica* serovar Typhi ISOLATES FROM KELANTAN USING RANDOM AMPLIFIED POLYMORPHIC DNA AND SINGLE NUCLEOTIDE POLYMORPHISM

JA'AFAR NUHU JA'AFAR

UNIVERSITI SAINS MALAYSIA
2014

MOLECULAR EPIDEMIOLOGY OF *Salmonella enterica* subspecies *enterica* serovar Typhi ISOLATES FROM KELANTAN USING RANDOM AMPLIFIED POLYMORPHIC DNA AND SINGLE NUCLEOTIDE POLYMORPHISM

by

JA'AFAR NUHU JA'AFAR

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

FEBRUARY 2014

ACKNOWLEDGEMENT

Alhamdulillah, all praise and glory be to Allah SWT for making things easier for me in my life. My family, without whose encouragement, I would not have been here. I thank them for their unflinching prayers and support. I acknowledge with thanks the invaluable assistance and advise rendered by Prof. Ribadu and his family.

My profound gratitude and appreciation goes to my able supervisor, Assoc. Prof. Dr. Phua Kia Kien. He has been very understanding, supportive and friendly. Thank you very much Prof. To my co-supervisors, Prof. Dr. Prabha Balaram and Assoc. Prof. Dr. Subhash Bhore, I say, may God continue to increase your knowledge and grant you good health.

So many people (both staff and students) are in the queue to be thanked and acknowledged. Trying to thank them individually would mean to risk leaving out someone by mistake. I therefore say "thank you" to all of you from my heart. However, my sincere thanks go to my group mates, Goay Yuan Xin *et al* and to the Institute for Research in Molecular Medicine (INFORMM) for providing the enabling working environment to carry out my research. Also, my appreciation goes to Dr. Eugene for his assistance in the abstract translation to Bahasa Melayu. Thank you Dr.

Finally, I wish to acknowledge the financial support provided by Universiti Sains Malaysia (USM) in the form of Postgraduate Research Grant (PRGS) (Grant No. 1001/CIPPM/844078), Graduate Assistant Fellowship (GA) and the Research

University Cluster (RUC) project (Grant No. 1001/PSKBP/86300111), without which this research would not have been possible. Thank you.

TABLE OF CONTENTS

ACKNOWLEDGEMENT……... ii

TABLE OF CONTENTS…..……..iv

LIST OF FIGURES……..…… xi

LIST OF TABLES…….....xv

LIST OF SYMBOLS AND ABBREVIATIONS…..…… xvi

GLOSSARY OF TERMS……..xvii

ABSTRAK…….....xxi

ABSTRACT……....…….……xxiv

CHAPTER 1: INTRODUCTION

1.1 Background…… 1

1.2 *Salmonella* and its Characteristics…… 3

1.3 *Salmonella enterica*……... 5

1.4 Genetic Diversity of *S.* Typhi…… 6

1.5 Typhoid Fever in Kelantan…… 8

1.6 Problem Statement……11

1.7 Significance of the Study…….. 13

1.8 Rationale of the Study……15

1.9 Objectives of the Study…… 17

1.9.1 General Objective…….. 18

1.9.2 Specific Objectives…… 18

CHAPTER 2: LITERATURE REVIEW

2.1 History of Typhoid Fever……...19

2.2 Typhoid Fever.....19

2.3 Epidemiology of Typhoid Fever……21

2.4 The Asymptomatic Typhoid Carrier……. 23

2.5 Evolutionary History of *Salmonella*…….. 24

2.6 Nomenclature of *Salmonella*……..26

2.7 Epidemiology of *Salmonella*……. 27

2.8 Identification of *Salmonella* Typhi…… 28

2.8.1 Morphological Identification…… 29

2.8.1(a) Xylose Lysine Deoxycholate (XLD) Agar…….29

2.8.1(b) MacConkey Agar……..... 30

2.8.2 Biochemical Identification……31

2.8.2(a) Triple Sugar Iron (TSI) Medium……..... 32

2.8.2(b) Urea Agar Base…….. 33

2.8.2(c) Simmons Citrate Agar…… ……... 33

2.8.2(d) Sulphide Indole Motility (SIM) Medium…….....34

2.8.2(e) Methyl Red-Voges Proskauer (MRVP) Medium……...34

2.8.3 Serological Identification……...35

2.9 Serological Classification of *Salmonella*……...36

2.9.1 Somatic Antigens……. 37

2.9.2 Capsular Antigens……..……...37

2.9.3 Flagella Antigens……..…… 38

2.10 *Salmonella* Typing Methods……..40

2.10.1 Phage Typing…… 46

2.10.2 Plasmid Typing…….. 46

2.10.3 Ribotyping…… 47

2.10.4 Insertion Sequence *200* (IS *200*) Typing……... 47

2.10.5 Restriction Fragment Length Polymorphism (RFLP)……... 48

2.10.6 Pulsed-Field Gel Electrophoresis (PFGE)……. 48

2.10.7 Amplified Fragment Length Polymorphism (AFLP)…… 49

2.10.8 Variable Number of Tandem Repeats (VNTR)…… 50

2.10.9 Multilocus Sequence Typing (MLST)……...51

2.10.10 Multilocus Enzyme Electrophoresis (MLEE)…….. 51

2.10.11 DNA Microarray……...52

2.10.12 Random Amplified Polymorphic DNA (RAPD)…… 53

2.10.12(a) Sequence Characterised Amplified Region (SCAR)… 56

2.10.12(b) RAPD Optimization……...57

2.10.12(c) Annealing Temperature Optimization……..... 58

2.10.13 Single Nucleotide Polymorphism (SNP)…….. 61

2.11 Pyrosequencing……..64

2.11.1 Principle of Pyrosequencing…….. 64

2.11.2 Applications of Pyrosequencing…… 65

2.12 Phylogeny…….. 66

2.12.1 Definition……...67

2.12.2 Uses of Phylogeny……. 67

2.12.3 Phylogenetic Tree Construction Methods……. 68

2.12.3(a) Distance Matrix…… 70

2.12.3(b) Maximum Likelihood……...72

2.12.3(c) Bayesian Inference……. 73

2.12.3(d) Maximum Parsimony……...73

CHAPTER 3: MATERIALS AND METHODS

3.1 Introduction……75

3.2 Overall Flow Chart of Research…… 76

3.3 Materials…… 77

3.3.1 Bacteria Isolates……77

3.3.2 Instruments……77

3.3.3 Reagents……77

3.3.4 Consumables……..77

3.3.5 RAPD Primers……... 78

3.3.6 Genes and Primers for Pyrosequencing……78

3.3.7 Softwares…… 78

3.4 Methods……. 79

3.4.1 Bacteria Confirmatory Tests…….....79

3.4.1(a) Microbiological Confirmation Tests…….. 79

3.4.1(b) Biochemical Confirmation Tests……..... 79

3.4.1(c) Serological Confirmation Tests……..80

3.4.2 Genomic DNA Extraction……. 80

3.4.3 RAPD-SCAR……81

3.4.3(a) Annealing Temperature Optimization……..... 81

3.4.3(b) Optimization of RAPD-PCR Master Mix……..82

3.4.3(c) Screening and Selection of RAPD Primers……..... 86

3.4.3(d) RAPD-PCR Assay……. 88

3.4.3(e) Gel Purification…….. 89

3.4.3(f) Cloning and Transformation of RAPD-PCR Product…….....89

3.4.3(g) Identification of Positive Clones and Plasmid Extraction……..... 91

3.4.3(h) Sequencing and Primer Design for SCAR Markers…….. 92

3.4.3(i) Validation of SCAR Primers…….. 92

3.4.4 SNP Typing by Pyrosequencing…… 92

3.4.4(a) Selection of Genes and Primer Design……. 93

3.4.4(b) PCR Amplification…….94

3.4.4(c) Sample Preparation…… 94

3.4.4(d) Pyrosequencing……. 97

3.4.4(e) SNP Analysis and Phylogenetic Tree Construction..... 97

3.4.4(f) SNP Typing by Pyrosequencing as an Epidemiological Marker……. 98

CHAPTER 4: RESULTS

4.1 *Salmonella* Typhi (*S.* Typhi) Confirmation…….……..... 99

4.2 RAPD-SCAR……. 100

4.2.1 Optimization of Annealing Temperature……....100

4.2.2 Optimization of RAPD-PCR Master Mix……...102

4.2.3 Selected RAPD Primers……113

4.2.4 RAPD-PCR Bands Selection……. 114

4.2.5 RAPD-PCR Product Cloning…….117

4.2.6 Primer Design for SCAR Markers…….118

4.2.7 Validation of SCAR Primers……. 119

4.3 SNP Typing by Pyrosequencing…… 122

4.3.1 PCR Products……122

4.3.2 Pyrosequencing Results……. 124

4.3.3 Phylogenetic Analyses……141

4.3.3(a) Phylogenetic Tree……141

4.3.3(b) Description of DNA Sequence Variation……...144

4.3.3(c) NETWORK Analysis…… 147

4.3.3(d) Geospatial Distribution of Haplotypes…….. 149

4.3.4 Possibility of Using SNP Typing for Routine Identification and as an Epidemiological Marker for *S.* Typhi…… 150

4.3.5 SNP Database…….158

CHAPTER 5: DISCUSSION…….. 163

5.1 Overview……163

5.2 Random Amplified Polymorphic DNA-PCR (RAPD-PCR)…… 163

5.3 SNP Studies……... 171

5.3.1 Global Mapping of Kelantan Isolates…… 180

5.3.2 Geographical Distribution of h-haplotypes in the State of Kelantan 182

5.4 Future Prospects in SNP Studies……... 183

5.5 Comparative Analysis of SNP and PFGE……. 184

CHAPTER 6: CONCLUSION, LIMITATIONS AND RECOMMENDATIONS

6.1 RAPD-SCAR Studies...……. 188

6.2 SNP Studies..……. 189

6.3 Limitations…….....191

6.4 Recommendations……..192

REFERENCES……..…….. 194

Appendix A List of Isolates Used in the Study…… 233

Appendix B List of Instruments and Suppliers…….. 234

Appendix C List of Media, Reagents, Chemicals, Kits, Suppliers and Chemical Preparation……..235

Appendix D RAPD Primers……... 241

Appendix E List of Genes and Primers Used in the Study ……... 245

Appendix F Softwares……. 249

Appendix G Pyrosequencing Setup……... 254

Appendix H Phylogenetic Analyses (A Systematic Guide)…….....268

Appendix I List of Publications and Presentations……...286

Appendix J Permissions of Copyrighted Materials……...288

LIST OF FIGURES

		Page
Figure 1.1	*Salmonella* organism showing rod-shaped body and peritrichous flagella	4
Figure 1.2	Map of Kelantan state showing the different districts	8
Figure 1.3	Trend of typhoid fever in Kelantan state from 1990–2012 showing number of cases recorded	10
Figure 1.4	Incidence of typhoid fever in Kelantan state from 2000 – 2010 showing a decrease in number of new cases reported	11
Figure 2.1	Typhoid fever pathogenesis describing the developmental stages of the disease	20
Figure 2.2	Global geographic distribution of typhoid fever showing its incidence	22
Figure 2.3	Evolutionary model of *Salmonella*	25
Figure 2.4	Outer membrane structure of gram-negative bacteria	36
Figure 2.5	Schematic representation of RAPD-PCR reaction	54
Figure 2.6	Single nucleotide polymorphism showing mutated "T" allele in the second DNA sequence compared to the ancestral "C" allele in the first and third DNA sequences	62
Figure 2.7	Principle of pyrosequencing showing the sequential catalytic activities of the four enzymes; polymerase, sulphurylase, luciferase and apyrase	65
Figure 2.8	A simplified phylogenetic tree of three samples, A, B and C	67
Figure 2.9	Phylogenetic tree construction methods based on data types	69
Figure 3.1	Summarized layout of RAPD-PCR optimization	83
Figure 3.2	Map of pDrive cloning vector	90
Figure 3.3	Overall workflow of pyrosequencing	93
Figure 3.4	PyroMark® Q96 vacuum prep workstation	95

Figure 3.5	PyroMark reagent cartridge with labelled compartments for enzyme, substrate and dNTPs	97
Figure 4.1	Morphological and biochemical characteristics of *S.* Typhi	99
Figure 4.2(a)	Optimum annealing temperature for primer SBSD02 was 36.8°C based on maximum number of reproducible and sharp bands	101
Figure 4.2(b)	Optimum annealing temperature for primer SBSD05 was 34.8°C based on maximum number of reproducible and sharp bands	101
Figure 4.3(a)	Effect of PCR buffer and $MgCl_2$ concentrations on number of PCR bands obtained	107
Figure 4.3(b)	Effect of primer and $MgCl_2$ concentrations on number of PCR bands obtained	108
Figure 4.3(c)	Effect of dNTP and $MgCl_2$ concentrations on number of PCR bands obtained	108
Figure 4.3(d)	Effect of Taq DNA polymerase and $MgCl_2$ concentrations on number of PCR bands obtained	109
Figure 4.3(e)	Effect of Taq DNA polymerase and primer concentrations on number of PCR bands obtained	110
Figure 4.3(f)	Effect of dNTP and Taq DNA polymerase concentrations on number of PCR bands obtained	110
Figure 4.4	Result of RAPD-PCR assay on the ten optimum conditions resolved in 1.2 % agarose gel	112
Figure 4.5	Reproducibility assay of RAPD-PCR over a period of three days	113
Figure 4.6	Score ratios for three example primers	114
Figure 4.7(a)	Primer SBSA02 showing a 2 kbp band that was not found in any of the non-Typhi Isolates	115
Figure 4.7(b)	Primer SBSA03 showing a 900 bp band that was not found in the other isolates except *S.* Paratyphi C	115
Figure 4.7(c)	Primer SBSD08 showing a 950 bp band that was not found in the other isolates except *S.* Paratyphi A	116
Figure 4.7(d)	Primer SBSD11 showing a 1,400 bp band that was not found in the other isolates except *S.* Typhimurium	116

Figure 4.8	Blue/white screening of transformed competent cells	117
Figure 4.9	Gradient PCR gel resolution of SCAR primer pair showing the best annealing temperature, which was 56°C based on the intensity of the PCR band	120
Figure 4.10	Result of PCR assay on SCAR Plasmid at optimum PCR condition	121
Figure 4.11	Results of PCR assay using SCAR primers for *S.* Typhi and non-Typhi isolates	121
Figure 4.12	Preliminary PCR result of bcfD gene fragment with and without Q-Solution resolved in 2% agarose gel	122
Figure 4.13	PCR amplicon sizes for all 37 gene targets resolved in 2% agarose gel using a representative isolate	123
Figure 4.14	Pyrosequencing pyrogram highlighting polymorphic positions with yellow background	125
Figure 4.15	Phylogenetic tree of 282 *S.* Typhi isolates using all 36 SNPs positions	142
Figure 4.16	Phylogenetic tree of 288 *S.* Typhi isolates using 30 SNPs positions	143
Figure 4.17	An example of single nucleotide polymorphisms (SNPs) and the haplotypes observed in six example isolates	144
Figure 4.18	Genealogical relationship of 282 *S.* Typhi isolates showing their h-haplotype distribution and ancestral h-haplotype	147
Figure 4.19	Global haplotype map of 650 *S.* Typhi isolates showing the distribution of the Kelantan isolates	148
Figure 4.20	Google map of Kelantan state showing circulating h-haplotypes	150
Figure 4.21(a)	fimI (C139T) SNP position	154
Figure 4.21(b)	sdhB (G222A) SNP position	154
Figure 4.21(c)	STY1889 (C372A (172)) -ve strand SNP position	155
Figure 4.21(d)	secA (C187T) SNP position	155
Figure 4.21(e)	yehU(2) (G146A (1541)) -ve strand SNP position	156
Figure 4.21(f)	glpB (G612A) SNP position	156

Figure 4.21(g) STY2713 (T295C (750)) -ve strand SNP position 157

Figure 4.21(h) STY4529 (G500A) SNP position 157

Figure 4.22(a) Main menu page to access the SNP database of *S.* Typhi 158

Figure 4.22(b) Summary Information of genes used in the study accessible using the SNP database developed in this study 159

Figure 4.22(c) Form view of SNP database showing detailed isolate's description for 37-SNP alleles as well as three haplotyping methods, designated 'G', 'P' and 'h' 160

Figure 4.22(d) Concatenated sequence of alleles in 36 genes used in the study forming the haplotype string 161

Figure 4.22(e) Comparison of three haplotyping methods 162

LIST OF TABLES

		Page
Table 1.1	Host adaptation of *Salmonella* serovars	6
Table 2.1	Biochemical properties of *S.* Typhi and other *Salmonella* subspecies	33
Table 2.2	*Salmonella* classification and nomenclature	40
Table 2.3	Summary of different typing methods highlighting their principles, uses and limitations	43
Table 2.4	Discriminatory indices for different typing methods	45
Table 2.5	Summary of different RAPD-PCR conditions	59
Table 2.6	Strengths and weaknesses of phylogeny methods	70
Table 3.1	Master mix concentrations and factorial levels for RAPD-PCR optimization	83
Table 3.2	Experimental runs for RAPD-PCR optimization generated using Design Expert	84
Table 3.3	List of sorted discrimination indices of all RAPD primers tested	87
Table 4.1	Results of RAPD-PCR experimental runs showing the response as number of bands	103
Table 4.2	ANOVA results for response surface quadratic model	106
Table 4.3	Numeric solutions for RAPD-PCR optimization	111
Table 4.4	PCR amplicon sizes of gene targets used in the study	124
Table 4.5	Genes used in SNP typing all 271 *S.* Typhi isolates using pyrosequencing method showing their respective alleles at the SNP positions	126
Table 4.6	P′-SNP profile of *Salmonella* Typhi isolates from Kelantan using 37 SNP positions	139
Table 4.7	P-SNP profile of *Salmonella* Typhi isolates from Kelantan using 36 SNP positions	140
Table 4.8	h-haplotype distribution of 282 *S.* Typhi isolates	146
Table 4.9	Global H-haplotype designation of Kelantan isolates	149

LIST OF SYMBOLS AND ABBREVIATIONS

°C	degree Celsius
µl	micro litre
µm	micro metre
AFLP	amplified fragment length polymorphism
bp	base pair
CDC	Centres for Disease Control
ET	electrophoretic type
IS*200*	insertion sequence 200
kbp	kilo base pair
min(s)	minute(s)
ml	milli litre
MLEE	multi locus enzyme electrophoresis
MLST	multi locus sequence type
mM	milli molar
OFAT	one factor at a time
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
RSM	response surface methodology
S. Typhi	*Salmonella* Typhi
SCAR	sequence characterized amplified region
sec(s)	second(s)
SNP	single nucleotide polymorphism
SPI	*Salmonella* pathogenicity island
ST	sequence type
T3SS	type three secretion system
Tm	melting temperature
U/l	units per litre
VNTR	variable number of tandem repeats
WHO	World Health Organization
Γ	gamma (capital case)
γ	gamma (small case)

GLOSSARY OF TERMS

Some common terminologies used in molecular epidemiology for prokaryotic organisms (Riley, 2004; Tenover *et al.*, 1997; Struelens, 1996) include:

***Allele*:** A nucleotide sequence variant of a gene at a particular locus.

***Biotype*:** A taxonomic unit of a microorganism belonging to one species based on a panel of biochemical tests.

***Clone*:** An isolate or a group of isolates descending from a common precursor strain by non-sexual reproduction exhibiting identical phenotypic or genotypic traits characterized by a strain typing method to belong to the same group.

***Discriminatory power*:** The ability of a laboratory test to generate distinct and discrete units of information from different isolates, usually at a subspecies level.

***Genotype*:** A genetic description of a cell or an organism according to its entire genome or a specific set of genetic loci (allele).

***Isolate*:** A population of microbial cells in pure culture derived from a single colony on an isolation plate and identified to the species level.

***Pathotype (Pathovar)*:** A taxonomic unit of a microorganism that describes a pathogenic variant of a species or subspecies of microorganisms that normally colonize a host.

***Phenotype*:** An observable characteristic expressed by a cell or an organism, such as drug resistance, virulence, and morphology.

***Serotype (Serovar)*:** A taxonomic unit of a microorganism based on antigenic properties of a distinct set of surface structures recognized by an antiserum.

***Strain*:** An isolate or a group of isolates exhibiting phenotypic and/or genotypic traits belonging to the same lineage, distinct from those of other isolates of the same species.

Systematics: The inference of phylogenetic relationships among species and the use of such information to classify species.

Taxonomy: The description, classification and naming of species.

Cluster algorithm: An algorithm of assigning a set of individuals to groups (or clusters) so that objects of the same cluster are more similar to each other than those from different clusters.

Markov chain: A stochastic sequence (or chain) of states with the property that the probabilities for the next state do not depend on the past states.

Transitions: Substitutions between the two pyrimidines (T↔C) or between the two purines (A↔G).

Transversions: Substitutions between a pyrimidine and a purine (T or C↔A or G)

Long-branch attraction: The phenomenon of inferring an incorrect tree with long branches grouped together by parsimony or by model-based methods under simplistic models.

Molecular clock: The hypothesis or observation that the evolutionary rate is constant over time or across lineages

Prior distribution: The statistical distribution assigned to parameters before the analysis of the data. That is the probability that describes our expectation of what trees look like, before observing the data.

Posterior distribution: Posterior probability is the measure of confidence for Bayesian phylogeny. The statistical distribution of the parameters (or models) is

conditional on the data. Also, it is the probability that a particular monophyletic group is correct.

Markov chain Monte Carlo algorithms (MCMC): A Monte Carlo simulation is a computer simulation of a biological process using random numbers. An MCMC algorithm is a Monte Carlo simulation algorithm that generates a sample from a target distribution (often a Bayesian posterior distribution).

***Bootstrapping*:** The idea is to generate a fake data set of the same size as the real one, and each site corresponds to a randomly chosen site of the real one. The fake data set is analysed using the same method as the real one repeatedly (at least 1000 times). Percentage of each cluster from the real data set that appears in the trees made from fake data sets is referred to as the bootstrap support of that cluster.

Epidemiologically related strains: These isolates are cultured from specimens collected from different patients or their excretions (faeces, urine), or from the environment during a discrete time frame, or from a well-defined (geographical) area as part of an epidemiologic investigation that suggests that the isolates may be derived from a common source.

Genetically related strains (clones): These isolates are (almost) indistinguishable from each other by one or a variety of genetic tests, supporting the suggestion that they are derived from a common ancestor.

***Cluster or clonal complex*:** The term "cluster" is used to indicate isolates with identical or highly similar DNA typing results (fingerprints), but it is also used to indicate a group of persons (patients) from whom these isolates were derived.

Outbreak strain: Outbreak strains are isolates of the same species that are epidemiologically related *and* genetically related. Such isolates are presumed to be clonally related since they have common genotypes and phenotypes.

Epidemic strain: Isolates that are frequently recovered from infected patients in a particular health-care setting or community and that are genetically closely related, but for which no direct or epidemiologic relation can be established. Their common origin may be more temporally distant from those of outbreak strains.

KAJIAN EPIDEMIOLOGI MOLEKULAR TERHADAP *Salmonella enterica* subspecies *enterica* serovar Typhi DI KELANTAN DENGAN MENGGUNAKAN KAEDAH DNA POLIMORFIK TERAMPLIFIKASI SECARA RAWAK DAN POLIMORFISME NUKLEOTIDA TUNGGAL

ABSTRAK

Tifoid ialah satu jangkitan akut sistemik manusia yang disebabkan oleh bakteria *Salmonella enterica* subspesies *enterica* serovar Typhi (*S.* Typhi). Tifoid merupakan penyakit yang endemik di Kelantan, iaitu sebuah negeri di Timur Laut Semenanjung Malaysia. Walaupun Jabatan Kesihatan Awam negeri telah mengambil tindakan ekstensif untuk mengawal penyebaran tifoid, wabak tifoid masih berlaku dengan kewujudan strain rintang pelbagai-drug (multi-drug resistant strains). Satu strategi penting untuk membasmi tifoid adalah untuk mengesan sumber dan penyebaran jangkitan. Oleh itu, satu kaedah pencapjarian *S.* Typhi yang piawai amat diperlukan. Pengesahan bakteria ini dengan menggunakan teknik-teknik fenotip rumit kerana bilangan serovar yang besar (>1,500) dalam subspecies *enterica*. Walau bagaimanapun, kaedah pencapjarian molekul terkini boleh memberikan resolusi yang tinggi; tetapi kebanyakannya terhad dalam pengenalpastian hubungan filogenetik, aliran evolusi atau ciri-ciri genotip. Maka, objektif utama kajian ini adalah untuk menyiasat kesan variasi-variasi jujukan nukleotida dalam bakteria *S.* Typhi yang berhubungan dengan kepelbagaian filogenetiknya, dan untuk mengenalpasti penanda-penanda epidemiologi untuk mewujudkan satu kaedah mudah untuk pengenalpastian dan pengesahan patogen ini. Dua kaedah telah digunakan dalam kajian ini.

Dalam kajian DNA Polimorfik Teramplifikasi Secara Rawak (RAPD), pengoptimuman ujian, pengklonan dan penjujukan kumpulan-kumpulan jalur RAPD khusus dilakukan untuk tujuan membangunkan satu penanda rantau teramplifikasi dengan jujukan terciri (sequence characterized amplified region (SCAR)). Suhu optimum penyepuhlindapan primer RAPD dan campuran ulung optimum RAPD-PCR ditentukan dengan kaedah matematik untuk memperolehi satu kaedah boleh ulang semula pentaipan RAPD-PCR untuk *S.* Typhi. Walaupun penanda SCAR untuk mengenalpasti *S.* Typhi secara spesifik berjaya ditentukan, tetapi serovar lain untuk subspesies *enterica* turut teramplifikasi dan ini menghadkan penggunaan penanda-penanda SCAR ini untuk pengenalpastian organisma homolog.

Untuk kajian-kajian SNP, satu kaedah pentaipan SNP dengan menggunakan kaedah 'pyrosequencing' telah digunakan untuk menyiasat 37 kepelbagaian SNP yang diketahui dalam 271 isolat *S.* Typhi berasingan dari negeri Kelantan. Dua kaedah penghaplotipan yang dilabel sebagai "P" and "h" telah dibangunkan dengan mencantumkan alel SNP-36 dan SNP-30 dan semua isolat telah diuji. Sebagai tambahan, skim ini telah dipetakan kepada koleksi 400 *S.* Typhi global yang telah diterbitkan. Analisis filogenetik klaster mendedahkan tiga kelompok utama menunjukkan kesamaan isolat tersebut. Kebanyakan isolat didapati tergolong dalam haplotip h11. Haplotip ini dipetakan kepada haplotip global H50, yang telah dilaporkan wujud sejak tahun 1986, justeru mencadangkan haplotip ini amat kekal di Kelantan.

Kelebihan kaedah yang telah dibangunkan dalam kajian ini adalah pengautomatan yang membolehkan integrasi mudah ke dalam pangkalan data.

Maklumat yang dimasukkan ini kemudiannya boleh digunakan untuk perbandingan dengan keputusan makmal lain pada masa hadapan. Kaedah ini juga boleh digunakan untuk menentukan aliran evolusi *S.* Typhi di dalam negeri Kelantan dan boleh digunakan sebagai sistem pemantauan awal untuk penyebaran strain rintang pelbagai-drug. Kesimpulannya, kaedah pentaipan SNP yang dibangunkan dalam kajian ini boleh digunakan untuk pengesahan, analisis filogenetik dan juga pemantauan perubahan evolusi isolat *S.* Typhi di Kelantan. Akhir sekali, analisis filogenetik untuk *S.* Typhi yang dipencilkan dalam tempoh 11 tahun kebelakangan ini (2002 – 2012) menunjukkan bahawa kepelbagaian genetik bakteria tersebut terhad untuk hampir 30 tahun sejak pengenalpastian pertama haplotip H50 di rantau ini. Tambahan pula, kelompok-kelompok kecil yang diperhatikan menunjukkan bahawa penyebaran tifoid di negeri Kelantan berlaku secara dalaman dan bukan dari sumber luaran.

MOLECULAR EPIDEMIOLOGY OF *Salmonella enterica* subspecies *enterica* serovar Typhi ISOLATES FROM KELANTAN USING RANDOM AMPLIFIED POLYMORPHIC DNA AND SINGLE NUCLEOTIDE POLYMORPHISM

ABSTRACT

Typhoid fever is an acute systemic infection of humans caused by *Salmonella enterica* subspecies *enterica* serovar Typhi (*S.* Typhi). Typhoid fever is endemic in Kelantan, a North-eastern state in Peninsular Malaysia. In spite of extensive intervention by the state's Public Health Department, outbreaks of typhoid fever continue to occur with the emergence of multi-drug resistant strains. An important strategy to eradicate this persistent disease is to identify the sources and distribution of infection; and for this, a reliable *S.* Typhi fingerprinting method is needed. Confirmation of the bacteria by phenotypic techniques is difficult due to the vast number of serovars (>1,500) within the subspecies *enterica*. Current molecular typing methods offer very high discriminatory resolutions; however, the majority are limited in identifying phylogenetic relationships, evolutionary trends or genotypic characteristics. Thus, the main objective of this study was to investigate the effect of nucleotide sequence variations in *S.* Typhi in relation to its phylogenetic diversity, and to identify possible epidemiological markers for easy identification and confirmation of this pathogen. Two methods were used in this study.

In random amplified polymorphic DNA-PCR (RAPD-PCR) studies, optimization of the assay, cloning and sequencing of specific RAPD bands were

performed with a view to develop a sequence-characterized amplified region (SCAR) marker. Optimum RAPD-primer annealing temperatures and an optimum master mix were established using a mathematical approach to obtain a reproducible RAPD-PCR typing method for *S.* Typhi. Although the SCAR marker developed to specifically identify *S.* Typhi was successful, yet other serovars of the subspecies *enterica* were also amplified, suggesting the limited functionality of SCAR markers for identification of homologous organisms.

For the SNP studies, a robust and reproducible SNP typing method developed using pyrosequencing was used to investigate 37 known SNP variations in 271 individual *S.* Typhi isolates in the state of Kelantan. Two haplotyping methods, designated "P" and "h", were developed by concatenating 36- and 30-SNP alleles, and all isolates were tested. In addition, this scheme was mapped with those published for a 400 global *S.* Typhi collection. Phylogenetic analyses of the isolates revealed three major clusters suggesting clonality of the isolates. It was also found that the majority of the isolates belonged to h11. This haplotype was mapped to the global haplotype H50, which was reported to be in existence as far back as 1986 in Malaysia, suggesting the persistence of this haplotype in the country.

The advantage of the method developed in this study is its automation, which allows for easy integration into a database. This would provide future comparisons with other laboratories. The method could also be used to determine the evolutionary trend of *S.* Typhi within the state and can serve as an early monitoring system for the spread of multi-drug resistant strains. In conclusion, the SNP typing assay developed in this study may be used for confirmation, phylogenetic analyses as well as for

monitoring the evolutionary changes of *S.* Typhi isolates in Kelantan. Finally, the phylogenetic analyses of *S.* Typhi from this 11-year period (2002 - 2012) showed its limited genetic diversity for almost 30 years since the first identification of the H50 haplotype in this region. Furthermore, the sub-clusters observed indicate that typhoid transmission in the state of Kelantan is within and not from an outside source.

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

The *Salmonella* serovars that can cause enteric fever in humans include *Salmonella enterica* subspecies *enterica* serovar Typhi (*S.* Typhi), the causative agent of typhoid fever, and *Salmonella enterica* subspecies *enterica* serovars Paratyphi A, Paratyphi B, Paratyphi C and Sendai, which cause paratyphoid fever.

Typhoid fever is a generalized acute infection of the intestinal lymphoid tissue, reticuloendothelial system, and gallbladder caused by *S.* Typhi. The disease is restricted to humans and human chronic carriers serve as reservoirs of the disease (Levine, 2009). In addition, the pathogen has the ability to survive for several months in soil and water (Tran *et al*., 2005).

While typhoid fever is regarded as the disease of history in most industrialized countries, it is still a devastating disease and is endemic in countries where hygiene and sanitation are poor. In industrialized countries however, typhoid fever is linked to persons traveling abroad (Ziprin and Hume, 2001). The low incidence of typhoid fever in industrialized countries could be associated with improvements in waste disposal, food handling, personal hygiene, patient care and availability of clean water supply (Molbak *et al*., 2006).

It is very difficult to get a true estimate of typhoid fever prevalence since only severe cases are reported. Subclinical cases are usually not reported. To obtain

an estimate for typhoid fever prevalence, it is assumed that for each case confirmed by a laboratory test, there are 38 other unconfirmed cases (Voetsch *et al*., 2004). In addition, for every case of typhoid fever detected it is assumed that there is 0.25 case of paratyphoid fever occurring (Crump *et al*., 2004). According to the World Health Organization (WHO), typhoid fever was globally estimated at 16.6 million and averaging 600,000 deaths. However, from a more recent study, typhoid fever cases were estimated to be 21.6 million and averaging 200,000 deaths annually (Crump *et al*., 2004). The authors suggest that the difference of 5 million more cases between the two studies could have been a 20% increase in global population over the period. This claim may be due to human-to-human transmission. However, direct human-to-human transmission is hard to prove, because epidemiologically unrelated *S.* Typhi isolates are often so similar and look identical using most typing techniques (Roumagnac *et al.*, 2006). Coincidentally, advances in DNA sequence analysis have offered new prospects of studying microbes at the molecular level. Hence, this study aims to explore the nucleotide sequence of several *S.* Typhi isolates as a new and more efficient way of identifying the organism.

Epidemiology as defined by Evans (2009) is "the study of the distribution and determinants of health-related states, conditions, or events in specified populations and the application of the results of this study to the control of health problems." In contrast, molecular epidemiology is the study of genetic factors that determine and regulate an organism's specific mode of transmission among hosts within an environmental context (Riley, 2009). Molecular epidemiology uses new molecular biology tools and techniques to subtype microorganisms, to characterize

disease occurrence, distribution, or determinants of disease distribution (Riley, 2009).

1.2 *Salmonella* AND ITS CHARACTERISTICS

Lignieres coined the name *Salmonella* in 1900 after Daniel Elmer Salmon, the bacteriologist who identified *Salmonella Choleraesuis* (*S. Choleraesuis*) in 1885 from pigs (Threlfall and Frost, 1990). The genus *Salmonella* is phylogenetically clustered in the family of *Enterobacteriaceae* (Chart, 2003). It is characterized as gram-negative, ubiquitous, non-encapsulated, straight-rod shaped, facultative, non-spore forming, and generally motile with peritrichous flagella (Fig. 1.1) (Gray and Fedorka-Cray, 2002; Molbak *et al*., 2006), with the exception of the poultry-specific serovars, *Salmonella* Gallinarum (*S.* Gallinarum) and *Salmonella* Pullorum (*S.* Pullorum) (Grimont *et al*., 2000). The bacterium has a width of 0.7 to 1.5 μm and a length of 2.0 to 5.0 μm (Holt *et al*., 1994). It can grow within a wide range of temperatures from 8 to 45°C (Hanes, 2003), but the optimal growth temperature is within the range of 35 to 40°C (Dickson, 2000). Typically, *Salmonella's* pH growth range lies within 4.5 to 9.0 (D'Aoust, 1989). However, the most favourable pH for growth is within 6.5 to 7.5 (Garcia-Del Portillo, 1999).

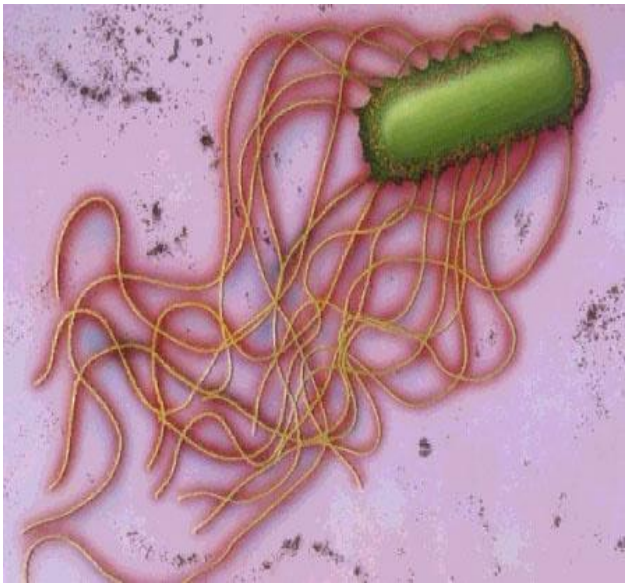

Figure 1.1: *Salmonella* organism showing rod-shaped body and peritrichous flagella http://sakittifus.com/sakit-tifus-abdominalis/ Accessed 15th June 2011

Salmonella are usually aerogenic - producing gas from glucose, and can utilize citrate as their sole carbon source. The failure to ferment lactose and the ability to produce hydrogen sulphide (H_2S) gas from sulphur-containing amino acids are features used to identify colonies on primary culture media (Ryan and Falkow, 1994). Nevertheless, *S.* Typhi is a facultative anaerobe, incapable of producing gas from fermentation of glucose (Hanes, 2003), and does not utilize citrate. The *Salmonella* bacterium is non-tolerant to oxidase and can convert nitrate to nitrite (Hanes, 2003). In addition, lysine and ornithine are decarboxylated, they do not hydrolyse urea nor deaminate tryptophan or phenylalanine (Anderson and Ziprin, 2001). These unique characteristics provide a form of demarcation when identifying *Salmonella* from other closely related organisms.

1.3 *Salmonella enterica*

As bacteria reproduce asexually by binary fission, their populations are considered groups of clones deriving from a common ancestor. This implies that genetic transmission occurs vertically from parent to daughter cell and any genetic differences between them are a result of mutation. Recombination in the form of horizontal gene transfer and lateral gene transfer could also occur and contribute to the diversification of bacterial clones. The proportion of mutational forces versus recombination affects the population structure of bacteria (Octavia, 2008).

The genus *Salmonella* has been assigned into more than 2,500 different serovars (Popoff *et al*., 2004). Initial classification and identification of these serovars were based on the Kauffman-White serotyping scheme whereby the type and arrangement of sugar residues and the flagella proteins determine antigenic variation, and about 60 types of the former have been recognized (Popoff and Minor, 1997). Most serovars of *Salmonella enterica* (*S. enterica*) can undergo phase variation as they have alternative expression of the two-flagellum genes (diphasic). However, *S.* Typhi has only one active H1 flagella gene (monophasic) while *S.* Gallinarum and *S.* Pullorum are aflagellate.

DNA hybridization and ribotyping have classified the *S. enterica* serovars into seven subspecies: I, II, IIIa, IIIb, IV, V, and VI (Le Minor *et al.*, 1986). Multilocus enzyme electrophoresis (MLEE) has defined an eighth group, designated as subspecies VII, consisting of few isolates that were initially grouped to subspecies IV (Reeves *et al*., 1989; Shelobolina *et al*., 2004). In addition, MLEE clustered subspecies I, IIIa, IIIb and VI separately from subspecies II, IV and VII. Subspecies

V was clustered separately from the other subspecies. This subspecies was considered the most divergent *S. enterica* subspecies, and hence classified as a different species of *Salmonella*, termed *S. bongori* (Reeves *et al.*, 1989).

The *S. enterica* subspecies *enterica* serovars can further be divided into those that are host adapted, host restricted and generalists (Table 1.1) (Uzzau *et al*., 2000; Ziprin and Hume, 2001; Edwards *et al*., 2002; Wallis, 2006). Other *Salmonella* infections not adaptive to humans may result in self-limiting diarrhoea to life threatening systemic diseases (Chart, 2003).

Table 1.1
Host adaptation of *Salmonella* serovars (Uzzau *et al*., 2000; Edwards *et al*., 2002)

Classification	**Serovar**	**Natural host**	**Rare hosts**
Host restricted	Typhi	Humans	None
	Paratyphi A and C	Humans	None
	Sendai	Humans	None
	Abortusovis	Ovines	None
	Gallinarum	Poultry	None
	Typhisuis	Swine	None
	Abortusequi	Equines	None
Host adapted	Choleraesuis	Swine	Humans
	Dublin	Bovines	Human and ovines
Generalist	Typhimurium	Humans, poultry, swine, bovines, and rodents.	None
	Enteritidis	Humans, poultry, and rodents	Swine and bovines

1.4 GENETIC DIVERSITY OF *S*. Typhi

Population structure studies of a pathogen integrate its epidemiological, phylogenetic and evolutionary relationships to provide a better understanding of its behaviour. *S.* Typhi shows limited genetic variation (Holt *et al*., 2008) based on two population structure studies conducted, MLEE (Selander *et al*., 1990) and multilocus sequence typing (MLST) (Kidgell *et al*., 2002). These methods have been widely accepted as

powerful tools to examine the population genetics of bacterial species. An MLEE study carried out on 334 global *S.* Typhi isolates to analyse the electrophoretic mobilities of 24 metabolic enzymes identified only two electrophoretic types (ET), ET1 and ET2 (Selander *et al*., 1990). The two ETs differed by only two of the 24 enzyme loci analysed with ET1 representing 82% of the isolates which are globally distributed, while ET2 is restricted to Senegal and Togo in Africa (Selander *et al*., 1990). MLST, which is a more sensitive method, analyses isolates based on sequence diversity of the seven housekeeping genes namely aroC, dnaA, hemD, hisD, purE, sucA and thrA. In contrast to MLEE, MLST could identify all the sequence changes, including synonymous changes that do not result in amino acid replacements. When MLST was performed on the seven housekeeping genes from 26 global *S.* Typhi isolates, only 3 base substitutions in the 3336 bp analysed were found, and they divided the 26 isolates into four sequence types (ST) designated as ST1, ST2, ST3 and ST8 (Kidgell *et al*., 2002). Similarly, the two completely sequenced *S.* Typhi strains, Ty2, commonly used in experimentation and vaccine development (Deng *et al.*, 2003), and CT18, the multi-drug resistant isolate from Vietnam (Parkhill *et al.*, 2001), belonged to different ST (Kidgell *et al.*, 2002). One of the differences between CT18 and Ty2 is 282 single point mutations (Deng *et al*., 2003). Recently, Holt *et al* (2008) has sequenced 17 additional global Typhi genomes and found nine different haplotypes. In a separate study in France, Ramisse *et al* (2004) used variable number of tandem repeats (VNTRs) to differentiate 27 *S.* Typhi isolates into 25 genotypes. These studies demonstrate the different approaches used in determining the diversity of this pathogen.

1.5 TYPHOID FEVER IN KELANTAN

Kelantan, a North-eastern state in Peninsular Malaysia with co-ordinates 5° 25' 0" North and 101° 55' 0" East (Fig. 1.2), is endemic for typhoid fever (Fig. 1.3). Under Act 342 of the Laws of Malaysia (Prevention and Control of Infectious Diseases Act, 1988), government hospitals, health care centers, outpatient departments and private hospitals are required to notify all cases of typhoid fever to individual states' Public Health Departments. Because of this legal requirement, a typhoid population-based registry was set up by the Kelantan State Public Health Department, which can be used to study the association of typhoid fever with various epidemiological parameters amongst Kelantanese in the state.

Figure 1.2: Map of Kelantan state showing the different districts

Epidemiological surveillance of *S.* Typhi is essential for public health management to identify origins of outbreaks, transmission patterns, and risk factors contributing to the persistence and spread of typhoid fever in this endemic area. Several studies have been conducted on the risk factors for contracting typhoid in several countries including Malaysia (Black *et al*., 1985; Velema *et al*., 1997; Zain & Naing, 2002; Vollaard *et al*., 2004; Tran *et al*., 2005; Hosoglu *et al*., 2006). Similarly, an association study has been conducted to link the incidence of typhoid to its spatial distribution in the state of Kelantan (Safian *et al*., 2008). This study reported that the majority of the population in Kelantan state are Malays, comprising 96.2% of the total population, compared to 5.35% Chinese and 0.33% Indians. Hence, no statistically significant association between ethnic group and typhoid fever was observed (Safian *et al*., 2008). In Malaysia, typhoid fever affects all classes of the society (Malik & Malik, 2001). Other reported risk factors include a history of contact with other patients prior to illness, poor housing, not using soap for washing hands, and previous infection with *Helicobacter pylori* were established in a study in India (Bhan *et al*., 2005).

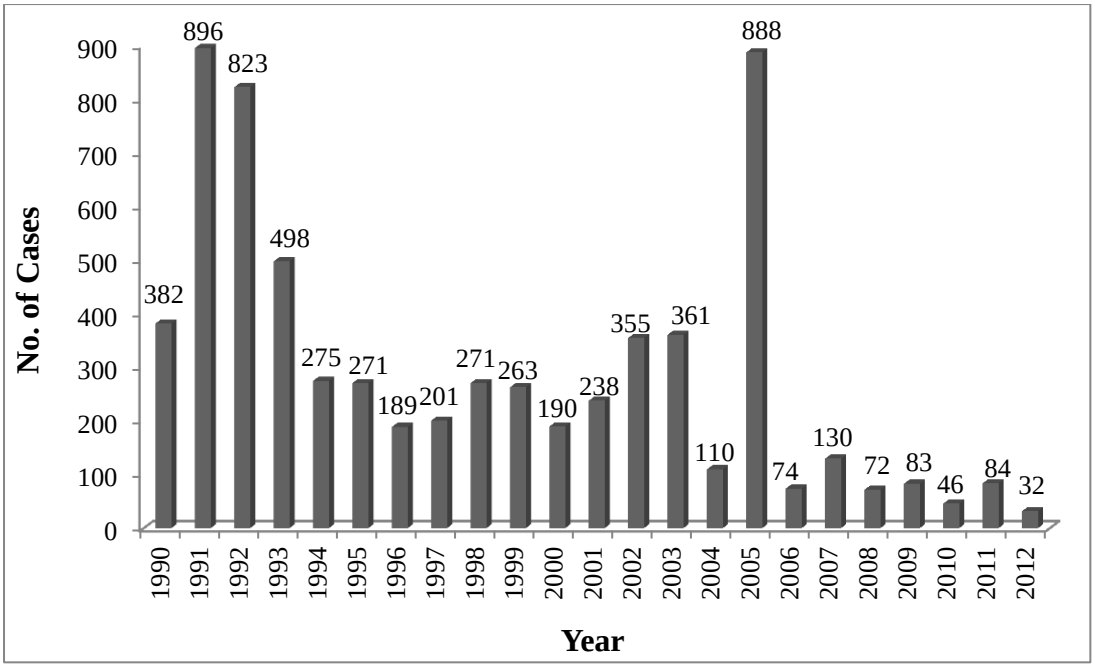

Figure 1.3: Trend of typhoid fever in Kelantan state from 1990–2012 showing number of cases recorded (Hamzah *et al*., 2011; Kelantan State Health Department, Malaysia).

In endemic areas, the incidence of typhoid fever peaks in the youngest age group of 1-19 years (Lin *et al*., 2000; Ja'afar *et al*., 2013). Malik & Malik (2001) reported from the Ministry of Health, Malaysia that the highest number of typhoid fever cases reported in 1998 and 1999 was from Kelantan, and the majority of the patients were children. A study conducted by Choo *et al* (1988) reported that the average age incidence of typhoid fever patients admitted to Hospital Universiti Sains Malaysia (HUSM) was 7.3 years, which was comparable to the study by Malik & Malik (2001) who reported 7.5 years from the same hospital. These data, however, contrasted with the study by Levine *et al* (1982) in Santiago, Chile. This often cited report highlighted the well-known "4Fs" risk factors amongst Chilean people; *viz.* 1) Female, 2) Fat, 3) Fertile, and 4) Over forty years old. However, there has been no similar study conducted amongst Asian people as to whether such risk factors prevail.

Yap and Puthucheary (1998) cited the annual incidence of typhoid fever in Malaysia as 10.2-17.9 cases per 100,000 population between 1978 to 1990, and as high as 50.3 cases per 100,000 population in the state of Kelantan. However, a significant improvement was achieved in reducing the incidence of typhoid fever in Kelantan from 14.7 cases per 100,000 population in 2000 to 2.8 cases per 100,000 population in 2010 (Fig. 1.4) (Hamzah *et al*., 2011).

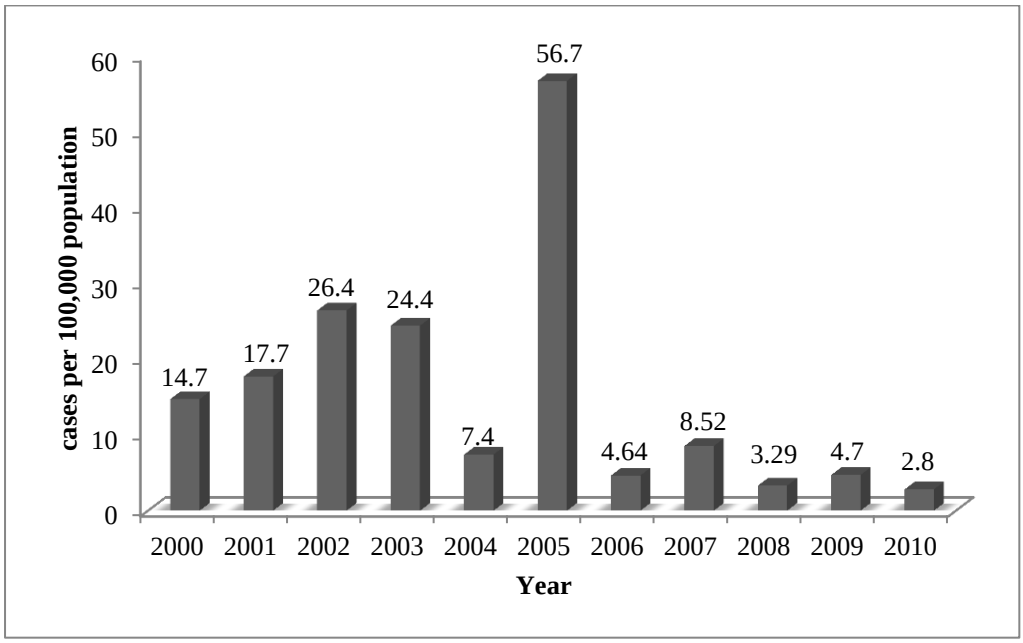

Figure 1.4: Incidence of typhoid fever in Kelantan state from 2000 – 2010 showing a decrease in number of new cases reported (Hamzah *et al*., 2011).

1.6 PROBLEM STATEMENT

Typhoid fever is a global problem and it claims thousands of lives every year (World Health Organization, 2005). Infants, children, pregnant women, elderly people and immunocompromised patients are at high risk (Hemsworth and Pizer, 2006; Pasmans *et al*., 2008). It is very difficult to get a true estimate of typhoid fever since only severe cases are ever reported. In the most recent study, typhoid fever was estimated to affect more than 21 million people annually (Crump *et al*., 2004).

The complete eradication of typhoid fever in Kelantan remains a challenge despite extensive efforts from the State's Public Health Department. Generally, understanding the epidemiology of an endemic disease such as typhoid fever in a given population could sometimes be complex, especially when multiple factors are involved in perpetuating the disease. Updated epidemiological databases will not only be of help to the epidemiologists and public health officers in rectifying and re-defining more effective and specific control measures, but will also help in elucidating and characterizing host-pathogen interactions, and provide better understanding of the disease pathogenesis at the cellular level. In addition, it will influence health-policy decision-making.

The occurrence of typhoid fever outbreaks and their spread is often associated with the asymptomatic typhoid carriers working as food-handlers. The occurrence of chronic asymptomatic carriers arising after acute infection with *S.* Typhi is about 1-5% (Gupta *et al*., 2006). Therefore, an important strategy in the eradication of typhoid fever is the detection of these asymptomatic carriers by screening food-handlers for *S.* Typhi, and preventing them from being involved in the food-handling industry and subjecting them to intense treatment with antibiotics to prevent the unintentional transmission of typhoid fever in the state. However, there is currently no reliable laboratory marker for identification of typhoid carriers. Stool culture remains the gold standard for identification of carriers, but it is unreliable and frequently gives false-negative results.

Genotypic studies of *S.* Typhi have revealed that it is highly homogeneous and estimated to be about 10,000 - 50,000 years old. This suggests that *S.* Typhi is

relatively young to accumulate significant variations (Kidgell *et al*., 2002). Due to this homogeneity, it is often challenging to distinguish *S.* Typhi isolates responsible for outbreaks or sporadic cases of typhoid fever, and even harder to establish the epidemiological and evolutionary relationships between these isolates.

Current molecular typing methods offer very high discriminatory resolution and are used to fingerprint isolates which are isolated over short periods. DNA-based molecular typing methods such as pulsed-field gel electrophoresis (PFGE) and ribotyping have demonstrated their effectiveness to differentiate isolates for epidemiological studies of *S.* Typhi. However, these molecular approaches are limited in identifying evolutionary trends or genotypic characteristics (Baker *et al*., 2007).

Although the Kelantan state public health authorities have a stringent policy of surveillance and reporting of every outbreak, no detailed molecular studies of the isolates have been performed to determine the predominant haplotypes circulating in the state.

1.7 SIGNIFICANCE OF THE STUDY

Typhoid fever is still of public health concern in Southeast Asia considering the spread of multi-drug resistant strains across the region (Roumagnac *et al*., 2006; Baker *et al*., 2008; Holt *et al*., 2012). One of the main approaches for combating typhoid fever today is controlling the dissemination of the causative agent. Therefore, investigating the molecular epidemiology of *S.* Typhi is of major importance to effectively control the disease. A robust molecular method such as

SNP typing is reproducible, and the results from such method can easily be integrated as databases so that global accessibility and comparisons can be made in contrast to many other typing methods. Investigation of the nucleotide sequence variation of *S.* Typhi may provide epidemiological markers that can aid in the identification, confirmation and prognosis of *S.* Typhi infection. However, this is only at a laboratory scale. These markers, when combined with other relevant epidemiological investigations, could assist in tracing the source(s) of an outbreak and monitoring trends in *S.* Typhi's diversification and its influence on treatment of the disease, and to prevent carrier states.

Furthermore, because *S.* Typhi is cosmopolitan and Malaysia is fast becoming a tourist attraction that is associated with rapid economic growth, there is a high possibility of increased microbial diversity. This could lead to the introduction of new strains that could be more pathogenic (virulent) to the local population but otherwise non-pathogenic to the expatriates. In view of this, this study is a stepping-stone for public health officers and epidemiologists to curtail the prevalence of virulent strains of the pathogen in the country.

In addition, the paucity of genotype population studies of *S.* Typhi in Malaysia has resulted in the lack of epidemiological, phylogenetic and evolutionary relationship data of the pathogen that could be used to facilitate formulation of control policies of the disease.

1.8 RATIONALE OF THE STUDY

Typhoid fever remains a major global health problem but could be controlled by implementing adequate food-handling practices and proper management of safe water supplies (Octavia, 2008). However, the immigration of foreign workers, and the emergence, spread and persistence of the multidrug resistant *S.* Typhi haplotype H58 (Baker *et al*., 2008; Holt *et al*., 2011), especially around the neighbouring countries, pose a threat to the preventive measures of this country.

After recovery from acute typhoid fever, 1-5% of the patients continue to harbour the bacteria in their hepatobiliary system for over a year and are designated as chronic carriers (Gupta *et al.*, 2006). These asymptomatic carriers may be the major source of *S.* Typhi infection, especially those carriers that are associated with food-handling (Kariuki *et al.*, 2006). The carriers shed the bacteria in their stools for years and impose a potential threat to the healthy population. Because *S.* Typhi's persistence in the body of carriers will influence antibiotic treatment, in-depth understanding of the functional genomics of the bacteria becomes imperative.

Although the isolation and identification of *S.* Typhi in acute patients of typhoid fever is well established, it is, however, still a problem with asymptomatic carriers. Traditional methods used in the laboratory to identify the bacteria have limited differentiating ability because they mostly target phenotypic characteristics of the pathogen. In contrast, DNA-based methodologies have the potential for greater discriminatory power. In fact, molecular methods have enabled the epidemiologist to determine the genetic relatedness of similar organisms more accurately and faster. Furthermore, genotyping is beneficial in addressing epidemiological concerns related to bacterial infectious diseases in the following

ways: characterizing the distribution and dynamics of disease transmission in geographically widespread areas; identifying and quantifying risk in sporadic occurrence and outbreaks of infectious disease. Others are for distinguishing pathovars from non-pathovars; addressing hospital and institutional infectious disease problems; and identifying the genetic determinants of disease transmission (Riley, 2009). However, none of the molecular techniques is optimum for all infectious disease causing pathogens. Low-throughput procedures such as restriction fragment length polymorphism (RFLP) and allele-specific PCR (Rose *et al*., 2003) are time consuming and laborious when multiple samples and/or multiple assays are to be used. PFGE is widely accepted as the gold standard method for typing and has facilitated epidemiological tracing of isolates from sporadic and outbreak cases. However, its discriminatory power has been reported to be too low to establish evolutionary relationships of relatively new pathogens (Roumagnac *et al*., 2006). For example, PFGE profiling of four carrier isolates isolated between 2007 and 2009 in Kelantan grouped them into three pulsotypes. But, it is still yet to be determined how these strains are related to one another or if they resulted from converging evolution in multiple phylogenetic ancestries. Furthermore, the requirement for specialized skill to carry out PFGE has resulted in a protocol that is routinely practiced in only a few reference laboratories (Nauerby *et al*., 2000).

In contrast, high throughput procedures utilizing chips or bead arrays (Suarez *et al*., 2005) are cheap per variant but are limited by the large amounts of samples needed and to studies such as whole-genome scanning where data of up to 500,000 polymorphisms are required (Marsh, 2007). Microarray technique has been used to detect genome variations, like single nucleotide polymorphisms (SNPs) or structural

variations (Smith *et al*., 2008), but it also has limitations such as high cost and need for specialized equipment.

Therefore, for epidemiological investigation and understanding the evolution of *S.* Typhi, a method that can fulfil both purposes will be ideal. RAPD-SCAR and SNP are molecular methods that can detect polymorphisms in the genome. SNPs are increasingly used for the phylogenetic analysis of bacteria, particularly monomorphic bacteria (Achtman, 2008). Although SNP typing is more specific in polymorphism detection, the low cost and simplicity of RAPD typing calls for its utilization. While high throughput technologies are currently used in big organizations for detecting polymorphisms as mentioned above, a range of medium-throughput technologies have emerged to bridge the gap, each with their merits and demerits (Freimuth *et al*., 2004). Pyrosequencing is one such medium-throughput technology that allows rapid and reliable genotyping to be performed in a 96-well plate format that facilitates handling.

1.9 OBJECTIVES OF THE STUDY

Population structure studies have shown that serovar Typhi is highly homogeneous, suggesting that it is relatively young to accumulate significant variation (Kidgell *et al*., 2002). However, previous PFGE studies performed on serovar Typhi isolates in our laboratory demonstrated high diversity. This suggests that comparison of these serovar Typhi isolates at nucleotide level is necessary to confirm the population structure studies. Relatively, little work has been done on typing serovar Typhi isolates in this endemic region using their nucleotide sequence information.

1.9.1 General Objective:

To perform phylogenetic analysis of *S.* Typhi isolates from Kelantan using RAPD and SNP methods.

1.9.2 Specific Objectives:

1. To use SNP technique in genotyping *S.* Typhi isolates isolated from Kelantan.
2. To improve the discriminatory power and reproducibility of RAPD by a mathematical approach.
3. To identify possible epidemiological marker(s) either from RAPD or SNP for easy identification and differentiation of *S.* Typhi isolates from other serovars.
4. To construct and examine the phylogenetic relationship of *S.* Typhi isolates isolated from Kelantan.

CHAPTER TWO

LITERATURE REVIEW

2.1 HISTORY OF TYPHOID FEVER

French physician, Pierre Charles Alexandre Louis (1787–1872) first used the term "typhoid fever" in 1829. William Budd (1811–1880) concluded in 1873 that typhoid fever was spread by the faecal-oral route. Karl Eberth (1835–1926) observed in 1880 rod-shaped organisms in the spleens and lymph nodes of typhoid patients, and he was credited with discovering the causative organism *Salmonella* Typhi (*S.* Typhi). Similarly, Lignieres attributed the name *Salmonella* to Daniel Elmer Salmon following his identification of *Salmonella Choleraesuis* (*S. Choleraesuis*) from pigs in 1885. Georg Gaffky (1850–1918) first successfully cultured *S.* Typhi from patients in Germany in 1884. Fritz Kauffmann (1899–1978) extends the work of Phillip Bruce White (White, 1926) and established a serological classification of *Salmonella* in the 1940s. In 1948, Theodore E. Woodward (1914–2005) and colleagues (Woodward and Smadel, 1948; Woodward *et al.*, 1950) successfully treated patients in Malaya with chloromycetin (chloramphenicol) (Ellermeier and Slauch, 2006).

2.2 TYPHOID FEVER

Typhoid fever is a serious systemic disease involving the reticuloendothelial system and the gall bladder, usually characterised by extended fever, abdominal discomfort, malaise, headache, constipation, hepatomegaly and splenomegaly. Infection develops following ingestion of food or water that is contaminated with 10^3-10^6 cfu/ml of *S.* Typhi (Hornick *et al.*, 1970). *S.* Typhi has remarkable mechanisms for

persistence in the host that are yet to be fully elucidated (Merrell and Falkow, 2004). Upon ingestion, *S.* Typhi initially traverses through the gastric acid-rich stomach to reach and colonize the intestine; a process that requires two type III secretion systems (T3SS), T3SS-1 and T3SS-2, which promotes uptake by intestinal epithelial cells and survival in host macrophages, respectively (Tischler and Mckinney, 2010). The bacteria then adhere to and invade the epithelial cells of the small intestine before they are phagocytosed by macrophages. Inside the macrophages, the bacteria survives the phago-lysosome system and gains free passage to the lymphatic and reticuloendothelial systems in the small intestine, liver and spleen and remain there for a few days before being transported back to the bloodstream (Fig. 2.1) (Everest *et al*., 2001; House *et al*., 2001; Parry *et al*., 2002).

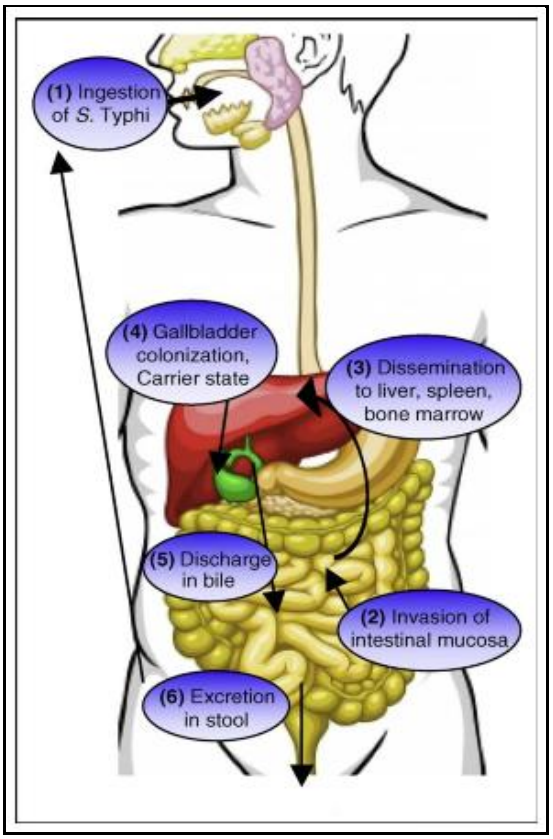

Figure 2.1: Typhoid fever pathogenesis describing the developmental stages of the disease. Each stage is labelled in Arabic numerals (Tischler and Mckinney, 2010).

After an incubation period of 5 - 21 days, some patients will have diarrhoea and abdominal pain that may last several days. Individuals with or without diarrhoea then develop fever that rises over 2 - 3 days to 39 - 40°C. Other symptoms can include chills, myalgias, cough, sore throat and rose-coloured spots, a maculopapular rash seen on the chest of 30% of patients. Psychosis and confusion occur in 5 - 10% of patients. Perforation of the terminal ileum or proximal colon occurs in 3 - 10% of patients and it is a significant cause of mortality (Ellermeier and Slauch, 2006). For those untreated patients that survive, symptoms last approximately 4 weeks, although weakness may be evident for months. Ten percent of individuals will suffer a relapse (Ellermeier and Slauch, 2006). Confirmatory diagnosis is based on isolation and identification of the bacteria, usually from blood or stool. Antibiotic treatment significantly reduces the mortality rate and the duration of fever. Vaccination against typhoid fever is recommended for travellers to endemic areas and for laboratory personnel who work with *S.* Typhi (Miller and Pegues, 2000).

2.3 EPIDEMIOLOGY OF TYPHOID FEVER

Typhoid fever is global in distribution but it is more prevalent in Oceania, Africa, Latin America and Asia with prevalence rate of 15.4, 49.8, 53.1, and 274.3 per 100,000 population, respectively (Fig. 2.2) (Crump *et al*., 2004). Studies suggest that more than 70% of typhoid cases in developed countries are from people who had a history of visitation to a typhoid endemic area (Mead *et al*., 1999; Ackers *et al*., 2000; Reller *et al*., 2003; Connor and Schwartz, 2005; Ekdahl *et al.*, 2005; Lynch *et al*., 2009). However, outbreaks do occur, for example, from imported contaminated fruit (Katz *et al*., 2002), or from food service workers that are asymptomatic carriers of the organism (Greig *et al*., 2007).

Infection is reportedly highest in 1 - 19 year-olds (Lin *et al*., 2000; Merrell and Falkow, 2004). However, more recent data indicate children of less than 5 years of age to constitute 44 - 54% of infections (Graham, 2002; Siddiqui *et al*., 2006; Ja'afar *et al*., 2013). The case fatality rate for untreated typhoid fever is 10 - 20%, but the advent of antibiotic treatment has reduced the rate to less than 1% (Miller and Pegues, 2000). Chloramphenicol has historically been the drug of choice, but resistance to it and others, such as ampicillin and trimethoprim/sulfamethoxazole has been rising since the 1980s (Ellermeier and Slauch, 2006). Fluoroquinolones are very effective (Parry *et al*., 2002), but resistance to these drugs is also reported to be on the rise (Threlfall, 2002). Chloramphenicol resistance in *S.* Typhi is plasmid mediated whereas nalidixic acid resistance is chromosomal (Ray *et al*., 2006).

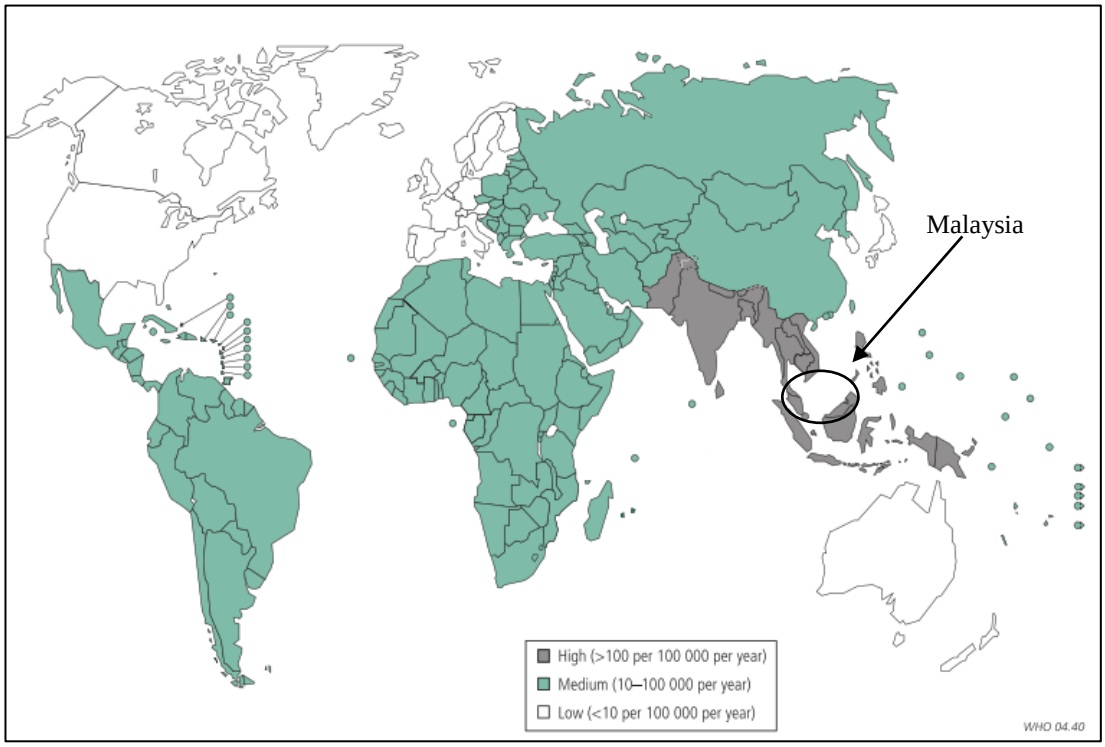

Figure 2.2: Global geographic distribution of typhoid fever showing its incidence (Crump *et al*., 2004).

2.4 THE ASYMPTOMATIC TYPHOID CARRIER

After clinical recovery, patients often continue to excrete *S.* Typhi for 12 days to 5 weeks (Sirinavin *et al*., 2004). Approximately 1% of such patients become chronic asymptomatic carriers and excretes 10^6 *S.* Typhi cells per gram of feces (D'Aoust, 1991; Selander *et al*., 1991), which can cause typhoid outbreaks (Pradier *et al*., 2000; Olsen *et al*., 2001).

In untreated patients, typhoid fever rapidly progress to bacteremia, particularly in those with underlying diseases such as cancer (Upadhye *et al*., 2005) and immunocompromised subjects (Manfredi *et al*., 1998). Bacteremia could subsequently cause localized tissue infections, such as endocarditis, meningitis, renal failure and pneumonia (Malik, 2002; Wani *et al*., 2004). Ninety percent of untreated typhoid patients who survive will completely clear the infection within several weeks. However, 10% will continue to shed the bacteria for up to 3 months, and 1 - 4% become long-term carriers, producing infective bacteria in their stools for over a year (Miller and Pegues, 2000; Parry *et al*., 2002; WHO, 2005). The chronic asymptomatic carrier stage is characterized by long-term colonization of the bacilli in the hepatobiliary system (Gupta *et al*., 2006; Nath *et al*., 2010). Bacteria replicating in the liver are shed through the gallbladder into the intestine and, thereby, the faeces. It has been suggested that *S.* Typhi can form biofilms on the gallstones, providing a more permanent habitat (Prouty *et al*., 2002). Both acute (cholecystits) and chronic infection of the gallbladder can develop. In addition, progress to gall bladder cancer is approximately 8.5 times more likely in typhoid fever carriers (Shukla *et al*., 2000; Kumar *et al*., 2006). These chronic infections are

more common in women and the elderly and are associated with gallstones (Ellermeier and Slauch, 2006).

Carriers are typically asymptomatic, and many have no apparent history of typhoid fever, but provide a constant source of infectious bacteria in the population (Bhan *et al*., 2005). As such, the occurrence of a typhoid fever outbreaks and the spread of this illness are often associated with asymptomatic typhoid carriers working as food-handlers. Lack of hygiene during food preparation by carrier food-handlers often result in transmission of *S.* Typhi to other individuals on a mass scale. Therefore, an important strategy in the eradication of typhoid fever is the detection of carriers, screening of food-handlers for *S.* Typhi carriers, preventing them from being involved in the food-handling industry, and subjecting them to intense antibiotics treatment in order to cure them and to prevent repeated outbreaks of typhoid fever. Unfortunately, there is currently no reliable laboratory marker for identification of asymptomatic typhoid carriers. Stool culture remains the gold standard for identification of carriers, but it is unreliable and frequently gives false-negative results since only less than 5% of stool samples show positive result by bacterial culture due to the intermittent release of the bacteria from the gall bladder (Koplan *et al.*, 1980).

2.5 EVOLUTIONARY HISTORY OF *Salmonella*

The genus *Salmonella* are members of the *Enterobacteriaceae* family of Gamma-proteobacteria subdivision. Two species of *Salmonella* are recognized: *Salmonella bongori* and *Salmonella enterica. Salmonella enterica* (*S. enterica*) can further be divided into several subspecies based on their biochemical characteristics (biotype),