COMPARATIVE GENOMICS AND PHYLOGENETIC ANALYSIS OF *Leptospira interrogans*

NAVAMUGANTHAN MURUGAN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

COMPARATIVE GENOMICS AND PHYLOGENETIC ANALYSIS OF *Leptospira interrogans*

NAVAMUGANTHAN MURUGAN

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE MASTER OF BIOTECHNOLOGY

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

2019

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **NAVAMUGANTHAN MURUGAN**

Matric No: **SGF 150003**

Name of Degree: **MASTER OF BIOTECHNOLOGY (BY MIX MODE)**

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

COMPARATIVE GENOMICS AND PHYLOGENETIC ANALYSIS OF *Leptospira interrogans*

Field of Study:

MICROBIAL BIOTECHNOLOGY

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:

COMPARATIVE GENOMICS AND PHYLOGENETIC ANALYSIS OF *Leptospira interrogans*

ABSTRACT

Leptospirosis is a re-emerging zoonotic disease worldwide and is caused by pathogenic strains of *Leptospira* spp. Most of the known serovars are isolated from their prime reservoirs of domestic and wild animals. Humans are infected either directly or in directly with exposure to the contaminated soil, water or urine of infected animals. However, very little is known about the genetics and virulence of *Leptospira interogans* that enable it to persist in environment and the reservoirs. Next Generation sequencing and advances in computational analysis have facilitated our understanding on the genomics of the causative agent of Leptospires. Three Malaysian strains of *L. interrogans* of different serovars isolated from three different hosts that is dog (D7), humans (L52) and rat (R123) were sequenced using the Illumina HiSeq 2000 platform. A5-miseq pipeline was used to perform *de novo* assembly. Genomes of 86 reference strains were obtained from National Center for Biotechnology Information (NCBI) database representing global sources. The protein coding sequences of the 89 genomes was determined independently using automated PROKKA annotation programs. Functional annotation of coding sequences of three Malaysian *L. interrogans* genomes were further determined using eggNOG-mapper. Conservation of the genes among 89 genomes was further examined using Roary program followed by pan-genome and evolutionary analysis using BPGA computational analysis. The draft genome sizes of D7, L52 and R123 are 4.70 Mbp, 4.88 Mbp and 4.80 Mbp, respectively. Comparative genomic analysis between Malaysian and global reference strains of *L. interrogans* have identified conserved functional genes involved in morphology determination (*mrdB*, *mreB* and *mreC*), adaptation (*cat, lexA* and *recA*), resistance (*katA*, *ccp*, *ccpA* and *tpx*), adhesion (*tlyC*, *ligA*, *inlA*, *slrP*, *sspH1* and *sspH2*), invasion (*colA* and *npr)*, gene

regulation (*degU*, *hupR1*, *pdtaR,* and *zraR)*, metabolism (*cbi*, *cob, sir, psel, legF*, *neuA* and *neuB*), chemotaxis (*cheA*, *cheB*, *cheC*, *cheD*, *cheR*, *cheW*, *cheX*, *cheY*, *mcp2*, *mcp3*, *mcp4*, *mcpA*, *mcpB*, *pctB*, *pctC*, *pomA*, *tar* and *tsr)*, motility (*cheD*, *cheR*, *cheW*, *flaAL*, *flaB*, *flgB*, *flgE*, *flgK*, *flgC flgG*, *fliD, fliE fliG*, *fliM, fliN*, *mcp4, motB*, *pomA* and *ycgR*) and virulence factors (*hlb*, *tlyAC*, *smc*L, *sph*, *cirA, htrB*, *kdsA*, *kdsB*, *kdsD*, *lpxA*, *lpxB*, *lpxC*, *lpxD*, *lpxK*, *waaA, cssS*, *liaS*, *mprA*, *pleD*, *rcp1*, *rcsC*, *tmoS*, *todS* and *uvrY)*. The pan genome is considered still open but maybe closed soon. *L. interrogans* is endowed with many genes that enable it to colonise different hosts. Phylogenetic study showed that strains of the same serovar isolated from different hosts and geographic locations are clustered together. The three Malaysian strains of *L. interrogans* D7, L52 and R123 appeared to be ancestrally related to the South America and Asia sub-clusters.

Keywords: *Leptospira interogans*, *de novo* assembly, comparative genomic analysis, pan genome, phylogenetic study

PERBANDINGAN GENOMIK DAN ANALISIS FILOGENETIK

Leptospira interrogans

ABSTRAK

Leptospirosis adalah penyakit zoonotik yang muncul semula di seluruh dunia dan disebabkan oleh strain patogen *Leptospira* spp. Kebanyakan serovar yang diketahui mempunyai perumah utama haiwan liar dan domestik. Manusia dijangkiti secara langsung dan tidak langsung menerusi pendedahan kepada tanah, air atau air kencing haiwan yang tercemar. Walau bagaimanapun, sangat sedikit yang diketahui mengenai maklumat genetik dan virulen *Leptospira interogans* yang membolehkannya kekal hidup di dalam pelbagai persekitaran dan perumah. Penjujukan Generasi Hadapan dan kemajuan analisis komputasi telah memudahkan pemahaman kita mengenai genomik agen penyebab Leptospirosis ini. Tiga strain *L. interrogans* Malaysia dengan serovar berbeza dipencilkan daripada tiga perumah yang berbeza iaitu; anjing (D7), manusia (L52) dan tikus (R123) telah dijujukan dengan menggunakan platform Illumina HiSeq 2000. Penjujukan *de novo* dilakukan menggunakan perisian A5-miseq. Genom dari 86 strain rujukan yang diperoleh daripada pangkalan data NCBI yang mewakili sumber global. Urutan pengekodan protein dari 89 genom ditentukan secara bebas menggunakan program anotasi automatik PROKKA. Anotasi fungsian bagi urutan pengekodan gen-gen tiga strain *L. interrogans* Malaysia telah ditentukan menggunakan aplikasi internet eggNOG-mapper. Pemuliharaan gen di antara 89 genom diperiksa menggunakan program Roary diikuti oleh analisis pan-genom dan evolusi dengan menggunakan perisian BPGA. Saiz draf genom D7, L52 dan R123 masing-masing adalah 4.70 Mbp, 4.88 Mbp dan 4.80 Mbp. Analisis perbandingan genomik diantara strain Malaysia dan rujukan global *L. interrogans* telah mengenal pasti pemuliharaan gen fungsian yang terlibat dalam penentuan morfologi (*mrdB*, *mreB* dan *mreC*), adaptasi (*cat, lexA* dan *recA*), rintangan (*katA*, *ccp*, *ccpA* dan *tpx*), lekatan (*tlyC*, *ligA*,

inlA, *slrP*, *sspH1* dan *sspH2*), pencerobohan sel (*colA* dan *npr)*, kawalatur gen (*degU*, *hupR1*, *pdtaR,* dan *zraR)*, metabolisma (*cbi*, *cob, sir, psel, legF*, *neuA* dan *neuB*), kemotaksis (*cheA*, *cheB*, *cheC*, *cheD*, *cheR*, *cheW*, *cheX*, *cheY*, *mcp2*, *mcp3*, *mcp4*, *mcpA*, *mcpB*, *pctB*, *pctC*, *pomA*, *tar* dan *tsr)*, motiliti (*cheD*, *cheR*, *cheW*, *flaAL*, *flaB*, *flgB*, *flgE*, *flgK*, *flgC flgG*, *fliD, fliE fliG*, *fliM, fliN*, *mcp4, motB*, *pomA* dan *ycgR*) dan faktor virulen (*hlb*, *tlyAC*, *smc*L, *sph*, *cirA, htrB*, *kdsA*, *kdsB*, *kdsD*, *lpxA*, *lpxB*, *lpxC*, *lpxD*, *lpxK*, *waaA, cssS*, *liaS*, *mprA*, *pleD*, *rcp1*, *rcsC*, *tmoS*, *todS* dan *uvrY)*. Pan genom dianggap masih terbuka tetapi bakal tertutup kelak. *L. interrogans* mempunyai banyak gen yang membolehkannya menjangkiti pelbagai perumah. Kajian filogenetik yang telah dijalankan menunjukkan bahawa strain dari serovar yang sama berada dalam kelompok yang sama walaupun dipencil dari perumah dan lokasi geografi yang berbeza. Strain *L. interrogans* D7, L52 dan R123 dari Malaysia kelihatannya berkait rapat dengan keturunan sub-kelompok dari Amerika Selatan dan Asia.

Kata kunci: *Leptospira interogans*, penjujukan *de novo*, analisis perbandingan genomik, pan genom, kajian filogenetik

ACKNOWLEDGEMENTS

I would first like to thank my supervisors Prof. Dr. Thong Kwai Lin and Dr. Saharuddin Bin Mohamad of the Institute of Biological Science at Faculty of Science of University of Malaya for their expertise who participate in the development and authentication of this research project. Without their passionate involvement, input and technical support the research and writing could not have been efficaciously conducted.

In addition, the doors to Prof. Thong and Dr. Saharuddin offices were always open whenever I ran into a trouble spot or had any question about my research or writing. Both of my supervisors do consistently permit this research and thesis to be my own work but steered me in the right direction whenever they thought I needed it and do keep on motivating me. My great appreciation to the trust and ample of time that were given for me to complete the research and writing.

I would also like to acknowledge Yousri Ab, Hamidah Ghani, Adib Wahab and Tony Yap as my lab and research mates, and I am gratefully indebted to their very valuable idea and comments on this research.

Finally, I must express my very profound gratitude to my parents, Murugan Perumal and Subahmal Muniandy and to my wife, Manggaiyarkarasi Velutham for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

Mbp	:	Mega base pair
B. burgdorferi	:	*Borrelia burgdorferi*
BLAST	:	Basic Local Alignment Search Tool
bp	:	Base Pairs
BPGA	:	Bacterial Pan Genome Analysis
CDS	:	Coding Sequences
COG	:	Cluster of Orthologous Group
DNA	:	Deoxyribonucleic Acid
ECM	:	Extracellular Matrix
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
L. alexanden	:	*Leptospira alexanden*
L. alstoni	:	*Leptospira alstoni*
L. biflexa	:	*Leptospira biflexa*
L. borgpetersenii	:	*Leptospira borgpetersenii*
L. broomii	:	*Leptospira broomii*
LERG	:	Leptospirosis Burden Epidemiology Reference Aggregate
L. fainei	:	*Leptospira fainei*
L. idonni	:	*Leptospira idonni*
L. inadai:	:	*Leptospira inadai*
L. interrogans	:	*Leptospira interrogans*
L. kirschneri	:	*Leptospira kirschneri*
L. kmetyi	:	*Leptospira kmetyi*
L. licerasiae	:	*Leptospira licerasiae*
L. meyeri	:	*Leptospira meyeri*

L. noguchii	:	*Leptospira noguchii*
L. santarosai	:	*Leptospira santarosai*
L. vanthielii	:	*Leptospira vanthielii*
L. weilii	:	*Leptospira weilii*
L. wolbachii	:	*Leptospira wolbachii*
L. wolfii	:	*Leptospira wolfii*
L. yanagawae	:	*Leptospira yanagawae*
L. tersptrae	:	*Leptospira tersptrae*
MAT	:	Microscopic Agglutination Test
MUSCLE	:	Multiple Sequence Comparison by Log-Expectation
NCBI	:	National Center for Biotechnology Information
NGS	:	Next-Generation Sequencing
PATRIC	:	Bacterial Bioinformatics Resource Center
qPCR	:	Real Time Polymerase Chain Reaction
QUAST	:	Quality Assessment Tool
rRNA	:	Ribosomal Ribonucleic Acid
spp.	:	species
T. pallidum	:	*Treponema pallidum*
tRNA	:	Transfer Ribonucleic Acid
WGS	:	Whole Genome Sequences
WHO	:	World Health Organization

LIST OF APPENDICES

CHAPTER 1

GENERAL INTRODUCTION

Leptospirosis, an important emerging zoonotic disease in Malaysia is caused by pathogenic spirochetes that is classified in the genus of *Leptospira* (Bharti et al., 2003). *L. interrogans* is the main causative agent of Leptospirosis disease in humans (Bourhy et al., 2007). Every year, more than 500,000 severe cases of Leptospirosis are reported by the Leptospirosis Burden Epidemiology Reference Aggregate (LERG) at the World Health Organization (WHO) (Lehmann et al., 2014).

The number of genomes that have been deposited in the databases has increased exponentially. Comparative genomics of bacterial pathogens isolated from various hosts, different sources and geographic locations have been reported. Comparative genetic analysis enables researchers to understand genetic materials of the pathogen and the knowledge has been applied for disease control, vaccine development, and molecular epidemiology.

Previous study conducted by using Pulsed-field Gel Electrophoresis (PFGE) technique showed genetic diversity between the strains of *Leptospira* spp. prevalent in the environment and rats (Benacer et al., 2013). Malaysian *L. interrogans* isolated from different hosts are hypothesised to carry genetic determinants that may contribute to various biological characteristic. Genomes of three Malaysian *L. interrogans* serovar *Canicola, L. interrogans* serovar *Batavie and L. interrogans* serovar *Ricardi* strains isolated from dog, rat and humans, respectively were sequenced through Next-Generation Sequencing (NGS) technology and subjected to appropriate bioinformatics analysis to comprehend their genetic make-up. This information might help to elucidate the strains' wide host adaptability, pathogenicity and virulence capability. In addition,

pan-genome analysis was conducted to determine genome plasticity and establish evolutionary relationship of the studied Malaysian strains to other reference strains isolated globally.

There is a paucity of scientific information about the status of Leptospirosis in Malaysia in terms of pathogenicity and transmission. Therefore, the research aims to address the following research questions and knowledge gap:

1. What are the genomic features, genetic similarities and variations between Malaysian *L. interrogans* strains isolated from different hosts of animal and human?
2. Which of the genes in genomes are conserved or shared (core genes) and dispensable genome that facilitated adaptation to wide number of hosts and responsible to virulence and pathogenicity of *L. interrogans* strains?
3. Whether the pan-genome of *L. interrogans* strains is open or closed?
4. The whole-genome based phylogenetic relationship of the studied Malaysian *L. interrogans* strains in relative to all other strains found globally?

The objectives of this study were;

1. To determine possible genes that facilitate adaptation of *L. interrogans* strains to wide number of hosts.
2. To postulate possible genes that are responsible for virulence and pathogenicity of *L. interrogans* strains.
3. To determine the pangenome of *L. interrogans* strains.
4. To define phylogenomic relationship of Malaysian *L. interrogans* strains isolated from rat, dog and human hosts in relation to global *L. interrogans* strains.

CHAPTER 2

LITERATURE REVIEW

2.1 *Leptospira*

Leptospira is a genetically diverse genus (Zuerner et al., 2000) containing at least 21 species which are further ordered into three large subgroups based on 16S rRNA phylogeny (Figure 2.1), DNA-DNA hybridization, pathogenicity, virulence, and *in vitro* growth characteristics (Lehmann et al., 2014). Figure 2.1 shows the clustering of *Leptospira* species into three major groups: infectious group, intermediate pathogen and non-pathogenic. The infectious group I is known as "pathogens" which consist of 9 species of *L. alstoni*, *L. kmetyi*, *L. santarosai*, *L. weilii*, *L. alexanden*, *L. borgpetersenii*, *L. noguchii*, *L. kirschneri* and *L. interrogans*. Group II is called "intermediate pathogens" which include 5 species; *L. fainei*, *L. inadai*, *L. broomii*, *L. wolfii* and *L. licerasiae*. *The* non-infectious group III is referred to as "saprophytes" is comprised of 7 species; *L. biflexa*, *L. meyeri*, *L. yanagawae*, *L. idonni*, *L. vanthielii*, *L.tersptrae and L. wolbachii,* and contain more than 60 serovars (Adler et al., 2010). Group I pathogens (Brenner et al., 1999; Slack et al., 2009) have been classified into 250 to 260 serovars (Adler et al., 2010; Lehmann et al., 2014) of distinct antigenic types and causes Leptospirosis with varying severity, ranging from subclinical infections to severe disease and death. Most severe infection is caused by serovars belonging to the evolutionarily-related species *L. interrogans*, *L. kirschneri*, and *L. noguchii* (Lehmann et al., 2014).

These Leptospires are antigenically diverse and classified into different serovars based on the expression and specificity of the surface-exposed epitopes. Mosaic of the lipopolysaccharide antigens depends on their sugar composition and orientation (Adler et al., 2010; Zuerner et al., 2000) making it more complicated to understand the Leptospires. In several cases, members of different species are serologically indistinguishable and belong to the same serovar. For example, strains of serovar Hardjo belong to the species *L.interrogans, L. borgpetersenii, and L. meyeri* (Brenner et al., 1999).

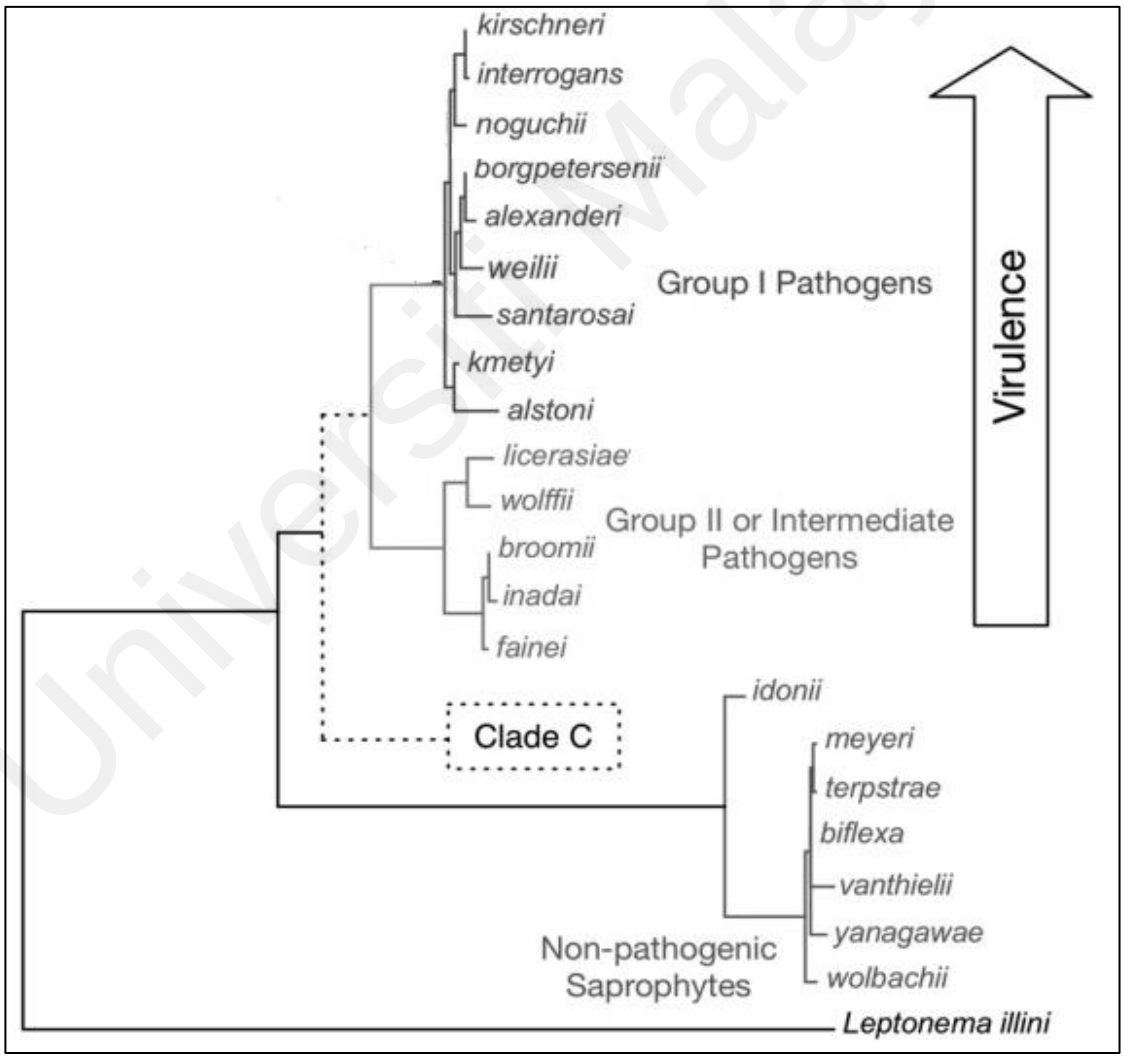


Figure 2.1 Phylogenetic tree of 21 known leptospiral species based on 16S rRNA sequences (Lehmann et al., 2014). Clade C referring to putative strain detected in Peruvian surface by qPCR with indefinite pathogenicity level (Ganoza et al., 2006). Genus name of *Leptospira* has been omitted for brevity purpose. Spirochete *Leptonema illini* which is a closely related species used as outgroup.

2.2 Leptospires

Leptospires were first observed in 1907 by Stimson (Stimson, 1907) in silver stained tissues from kidney sections of a patient who had been diagnosed incorrectly as a case of yellow fever (Feigin et al., 1975). At that time, the bacterial observed was identified as *Spirochaeta interrogans* (Cerqueira et al., 2009). The first valid description of saprophytic *Leptospira* was provided after it was discovered in a filtrate from stagnant water taken from the shores of a fresh water pond in the vicinity of Boston (Wolbach et al., 1914). In 1886, a pathogenic *Leptospira* named as *Spirochaeta icterohcemorrhagiae* was isolated from the blood of a Japanese patient causing Weil's disease (Inada et al., 1916). It appeared to be a unique clinical illness and accounted for cases of infectious jaundice led to symptoms of relapsing fever occasionally, jaundice, enlargement of the liver and spleen and the occurrence of hemorrhages (Davidson et al., 1934).

2.3 Leptospirosis

One of the common zoonotic diseases transmitted to humans by infected animals is Leptospirosis. Leptospirosis is a disease caused by pathogenic species of *Leptospira interrogans,* classified under pathogenic group of *Leptospira* genus. Leptospirosis spread through the urine of infected animals to environment and then to other animals and human. Leptospirosis disease is common in all countries including Malaysia and the implication depends on the severity of infection.

2.3.1 Sign and symptoms of Leptospirosis

Leptospirosis has an incubation period of 2 to 30 days and normally early symptoms are noticeable in the first 5 to 14 days of infection. Symptoms of this disease are not visible in the first 24 hours after infection (Terpstra, 2003). Anicteric Leptospirosis and icteric Leptospirosis are two types of infections with different symptoms. Anicteric Leptospirosis more common, affecting 90 percent of the patients. Patients who are in phase 1 of anicteric Leptospirosis may experience symptoms like flu, such as fever with temperature more than 39°C, muscle pain, tiredness, vomiting, heavy headaches, nausea, shivering cold and rashes. Patient is considered temporarily recovered if the above-mentioned symptoms last only for 3 to 5 days. Patients who are in phase 2 of anicteric Leptospirosis will develop anti-*Leptospira* antibodies due to the presence of bacterial in their urine. They get sick and this can last up to 30 days or longer. Inflammation of the meninges, membrane covering the brain and spinal cord may occur. In addition, inflammation of the iris, ciliary body, retina, choroid and uvea, the pigmented layer between the inner retina and the sclera and cornea may appear. However, vomiting, headache and fever are less severe compared to the septicemic stage and not all patients get into phase 2 of anicteric Leptospirosis. Aseptic meningitis is common in patients infected with anicteric Leptospirosis and fatality is uncertain. But this condition can cause pulmonary hemorrhage and can lead to death (Rao et al., 2003).

Icteric Leptospirosis or Weil's syndrome only affect 10 percent of hosts but is very dangerous compared to anicteric Leptospirosis. Symptoms of icteric Leptospirosis are the same as anicteric Leptospirosis, but probability of fatality is higher (Rao et al., 2003). It takes only 10 days to infect main organs such as heart, liver and kidney and brain (Sharma et al., 2008). Prime lesion in endothelium of little blood vessels prompting restricted ischemia in organs, bringing about renal tubular necrosis,

hepatocellular, myositis, pulmonary damage, meningitis and placentitis (Mohammed et al., 2011). Hemolysin is considered as phospholipases, destroys red blood cells (Thompson et al., 1986) and plasma membranes built by phospholipids, prompting cytolysis (Lee et al., 2002). Symptoms of icteric Leptospirosis are observed from third to ninth week of infection and causing death (Sharma et al., 2008).

2.3.2 Implications of Leptospirosis

Leptospirosis is labelled as the most omnipresent infectious disease in the world (Pappas et al., 2008). United States, Southeast Asia, Brazil, Malaysia, India and Nicaragua faced large outbreaks and Leptospirosis has been identified as threatening disease (Ratnam et al., 1993). This is mainly due to the characteristics of Leptospires that can survive in both humid and warm condition.

Leptospirosis normally affect adventurous travelers and its transmission is related to recreational activities and water sports. People involved in recreational activities such as rafting, kayaking, swimming and canoeing are at high risk of leptospirosis. In 2000, more than 150 athletes and participants in Expedition Race of EcoChallenge Sabah in Malaysian Borneo were infected with Leptospirosis. It is assumed that Leptospirosis infection happened during open water swim (Munoz et al., 1995).

Infection may happen because of bathing or immersion in water resources such as rivers and lakes contaminated with urine of livestock infected by pathogens of *L. interrogans* (Koutis, 2007). A rise of incidents related to Leptospirosis infection has been demonstrated during summer and rainy seasons in hot-weather areas (Guerreiro et al., 2001).

Widespread of Leptospirosis among livestock reduces the source of protein worldwide because of distorted growth, perinatal mortality, low milk production and abortion in pigs and calves. Indirectly, this led towards economical lost and affects main income of many countries. The main sources of human infection are rats and rodent (Terpstra, 2003). Species of rodents that are normally related to infectious diseases are *Mus musculus*, *Rattus norvegicus*, *Rattus rattus*, *Bandicota bengalensis*, and *Bandicota indica* (Gangadhar, 1999; Koutis, 2007; Matthias et al., 2002). Apart from rodents, *Leptospira* infect dogs even after vaccination. This phenomenon is an alarming sign because *Leptospira* may spread through the wastes of the infected domestic animals, causing infection on humans.

Initially, Leptospirosis was identified as one of the occupational diseases and the first group of workers highly risked to this disease was miners (Buchanan, 1925; Inada et al., 1916). Risk at work is undeniable for humans. Occupations which are directly or indirectly associated with the risk of Leptospirosis are veterinarians, rodent control workers, farmers, sewer workers, miners, fish farmers, soldiers and field farmers (Ko et al., 1999).

Leptospirosis is not only limited to warm and humid settings. Not long ago, it has been stated that Leptospirosis became a disease that affects metropolitan cities as rural population are moving to the cities (Johnson et al., 2004; Ko et al., 1999). Cities in many countries changed in ways where Leptospirosis can spread due to population shifts.

2.3.3 Epidemiology of Leptospirosis

Worldwide, more than 1 million Leptospirosis cases are reported annually (Adler et al., 2011). This re-emerging disease commonly affect people of tropical and subtropical countries (Bharti et al., 2003). Cases have been reported in Southeast Asian countries such as Indonesia, India, Thailand and Malaysia, South America and Central America (Mendoza, 2010; Victoriano et al., 2009). Due to the underreporting and poor documentation of human cases, widespread of Leptospirosis ranges from 0.1 to 10 per 100 000 population, with case fatality of 10% (Pappas et al., 2008). 100 or more 100 000 maybe infected in high risk countries during outbreak (Terpstra, 2003). However, these numbers might not be accurate due to wrong diagnosis and lack of surveillance facilities in many countries (Socolovschi et al., 2011).

In Malaysia, the rate of infection is estimated to be 2 to 5 cases per 100 000 populations. So far about 13% of Malaysians infected by Leptospirosis (El Jalii et al., 2000). The highest incidence and mortality rates were recorded, between the year 2004 until July 2015 with 30.2 per 100, 000 population and 0.31 per 100, 000 population respectively in nationwide (Wahab, 2015). The first case of Leptospirosis in Malaysia was discovered by Fletcher in 1925 and it is a common disease among Malaysian during that time (El et al., 2004). Leptospirosis cases in Malaysia increased fairly over the years. In year 2004, 263 cases with 20 deaths were reported. This increases up to 7 times in year 2010, whereby 1976 cases with 69 deaths were reported (Lim et al., 2011). In Sarawak, 49 cases were reported in year 2010 and increased to 186 cases in year 2011 (Thayaparan et al., 2013). Humidity of Malaysian environment assists the reproduction of pathogenic spirochetes (Abdullah et al., 2019). Urine of rodents that contaminate soil and water resources of Malaysia is the highest risk that Malaysians are encountering (Vke, 2011). Higher number of Leptospirosis cases were recorded after

flood (Hisham et al., 2009). Terengganu, Kelantan and Perak recorded about 110 cases in January 2015 due to massive flood and leaching of Leptospires from the earth (Yaakob et al., 2015). Kerala State Health Department officials have alarmed an increase in number of Leptospirosis cases been reported following a massive flood in Kerala India during August 2018. 570 and 18 persons confirmed to have contracted Leptospirosis and died respectively (James et al., 2018).

2.4 *Leptospira* cell biology

L. interrogans can live in the environment with low nutrient for long period. It can survive in fresh water sources and moist soil. It requires medium with suitable pH, salt concentration and viscosity to survive (Hartskeerl et al., 2011).

Leptospires are chemotrophs and they can grow in optimum temperature of 28 to 30 degree Celsius. Vitamin B12 and B1, ammonium salt and long fatty acids are crucial for Leptospires to grow. Long fatty acids, the only source of carbon is required for metabolism of Leptospires and is done via β-oxidation process (Faine et al., 1999).

Leptospira are normally 6 to 20 μm in length and spiracle coiled. This bacterium has hooked ends which give it a shape of question mark. The movement of *Leptospira* is assisted by two polar periplasmic flagella in periplasmic space (Picardeau et al., 2001).

Surface of *Leptospira* has characteristics of both Gram-negative and positive bacteria. One of the characteristics of Gram-negative bacteria present is the appearance of lipopolysaccharide (LPS) and two membrane layers, while characteristics of Gram-positive bacteria is the closeness of murein cell wall with cytoplasmic membrane (Haake, 2000).

2.5 Reservoir hosts and transmission

Rodents and rats are the primary reservoirs for pathogenic *Leptospira* and may contain pathogenic serovars of *L. interrogans* (Wangroongsarb et al., 2002). Furthermore, a lot of other mammalian and amphibians species could act as reservoirs of *Leptospira* (Athanazio et al., 2008; Gravekamp et al., 1991). It is believed that some species of hosts favor specific serovars such as rats harbor serovar Copenhageni and cattle harbor serovar Hardjo. However, this host and serovar association is not absolute (Lehmann et al., 2014). Leptospires colonises the convoluted tubules of the reservoir host kidneys and can be shed in the urine to the environment. Saprophytic species are normally present in natural water and soil and do not cause any infection most of time (Mohammed et al., 2011).

2.6 Humans as incidental hosts

Leptospirosis in human happens when pathogenic species are transmitted into the circulatory system by the mean of direct contact with tainted urine of animal repositories or in an indirect way via tainted water and soil (Bharti et al., 2003).

Leptospires penetrate the human host by ruptures on the skin surface, mucous membrane, conjunctiva, and genital tract. This requires chemotaxis ability for attachment and transmembrane entries. The microscopic organisms then need to pass through the vascular compartment to cause lesions. Endotoxin action has been accounted for in a few serovars of Leptospires (Mohammed et al., 2011).

2.7 Pathogenesis and virulence

L. interrogans is physiologically different from the other two spirochetes *Treponema pallidum* (Fraser et al., 1998) and *Borrelia burgdorferi* (Fraser et al., 1997). However, similarities exist in their genes, giving them similar form, shape and structure.

A detailed study of the ability of *L. interrogans* to move independently and synthesis of lipopolysaccharide provides information about the development of infection on human (Picardeau, 2017). Findings on cluster of genes related to adhesion, invasion and hematological effect provided valid reasons on the possibilities of *L. interrogans to* evolved as human pathogen (Fouts et al., 2016).

2.8 Molecular pathogenicity

Aspects at molecular level such as dynamics of cell motility, synthesis of surface proteins and virulence factors are important to be studied further as contributing factors of pathogenicity of *L. interrogans* (Saier Jr et al., 2001).

Movement of Leptospires is basically controlled by a structure which extended from each end of the bacterium, endoflagella or two periplasmic flagella. Pathogenic *L. interrogans* with the flagellar motor switch *fliY* mutant shows motion of an attenuated rotative movement pattern in semi solid and liquid mediums (Liao et al., 2009). Guinea pigs infected by *L. Interrogans* with *fliY* mutant exhibited a higher rate of survival against Leptospirosis, suggesting a major role of bacteria motility in pathogenicity of *Leptospira* infection. FlaA and Flab proteins make up flagellar sheath and core respectively. Electron microscopic view of *flaB* mutant show absence of endoflagella and fail to move (Picardeau et al., 2001).

The very first step of infection starts when Leptospires attach itself to the host. This attachment step is crucial for Leptospires to enter, spread and exist for prolong period in mammalian host tissues and extracellular matrix (ECM) components. Leptospires produce microbial surface components identified as adhesive matrix molecules that could help in colonization of host like other pathogens (Patti et al., 1994; Schwarz-Linek et al., 2004). It has been proven that *L. interrogans* infect different types of cells lines, such as monocytes, fibroblasts, macrophages, kidney epithelial and endothelial cells in vitro (Breiner et al., 2009).

Pathogenic leptospiral surface-exposed proteins can be identified via combination of experimental techniques and *in silico* analysis but their functions are still unknown (Pinne et al., 2009). Normally leptospiral attachment take place in outer surface proteins due to their bigger surface molecules that are exposed, hence assist attachment. Partially surface exposed proteins LigA, LigB and LigC contain bacterial immunoglobulin domain with redundant function are also important cause of Leptospirosis. However, a genetic knockout of *ligB* did not affect virulence or colonization in chronically infected hamsters or acutely infected rats (Croda et al., 2008). This shows that existence of other proteins like LigA are capable of similar interaction with host and causing infection.

Hemolysins are possible virulence factors due to its capability to lyse red blood cell and other cell membranes. Orthologs of hemolysin proteins Tly, that are recognized as virulence factor in the spirochete *Brachyspira hyodysteriae* (ter Huurne et al., 1994), are also found in *L. interrogan*. TlyB and TlyC surface exposed proteins did not possess hemolysin properties, but TlyC was found to attach with ECM components (Carvalho et al., 2009). The lysis of sheep red blood cells were caused by purified sphingomyelinase C from *L. interrogans* serovar Pomona (Bernheimer et al., 1986). *L. borgpetersenii* serovar Hardjo, another example of leptospire bacterium, contains sphingomyelinase

gene (*sphA*) and shows symptoms of sphingomyelinase activities (Segers et al., 1992). Hemolysin-encoding genes found in *L. interrogans* serovar Lai include a *sphA* homolog, *sphH*, coding a pore-forming protein without sphingomyelinase or phospholipase activities (Lee et al., 2000; Lee et al., 2002), and *sph2*, whose protein product induces membrane damage of endothelial cell and red blood cell (Artiushin et al., 2004). SphH and Sph2 are proteins causing human *Leptospira* infection (Artiushin et al., 2004) and have cytotoxic properties (Carvalho et al., 2010).

2.9 Comparative genomics of *L. interrogans* strains

Comparative genomics is the comparison study between DNA information of a species with another to find differences among them (Abby et al., 2007). The ultimate motive behind this study is to have better understanding regarding the genetic characteristics, size of the complete genome, repertoires sets of genes and establishment of pattern among living species (Sivashankari et al., 2007).

Comparative genomic study helps to identify the coding sequence of genes, functions of genes, presence and absence of genes between the same species and to understand the progressive events such as genome plasticity come out with phylogenetic relationships (Ogier et al., 2010; Rust et al., 2002). It is noticed that intracellular infectious pathogens such as *L. interrogans* are more vulnerable to loss of gene or reduction of gene development (Merhej et al., 2009).

Pan-genome, accessory genome and core genome are new terms developed in conjunction with the existence of thousands of infectious genomes in database and comparative genomics study (Lapierre et al., 2009).

2.9.1 Whole genome sequencing

Cost and time conservation of genome sequencing (Metzker, 2005) is possible with the technology of NGS. This tool enables us to observe a rapid increase in the number of genomes sequence that are available in databases.

The complete genome sequences of the pathogenic *L. interrogans* serovars Lai (NC_004342 & NC_004343) and Copenhageni (NC_005823 & NC_005824), are accessible via NCBI (National Center for Biotechnology Information) genome database (Nascimento et al., 2004; Ren et al., 2003). The leptospiral genome consists of approximately 3.9–4.6 Mbp, depend on the strain and species with G+C content of 35–41 mol% and is arranged in two circular chromosomes. Commonly, the *L. interrogans* genome (4,691,184 base pairs bp) (Ren et al., 2003) is much bigger than the other two Spirochaetes of 1,138,006 bp for *T. pallidum* (Fraser et al., 1998) and 1,519,857 bp for *B. burgdorferi* (Fraser et al., 1997), including plasmids.

2.9.2 Core and dispensable genome

The core genome is a part of genes present in all the genomes and can be identified via different genomes comparison (Muzzi et al., 2007). Based on Lapierre and Gogarten (2009), 250 genes specified as core genome and exists as evident of gene conservation. Basically, genes that act as core genomes are related to the function of organism's metabolism translation, replication and cellular homeostatic (Medini et al., 2005).

Dispensable genomes are a part of genes that are shared by some species but are not available in other organisms (Lapierre et al., 2009). Subset genes normally will have certain roles that are related with the ability of organism to live in different niches, resistance and pathogenicity of virulence factors (Mira et al., 2010; Read et al., 2006).

Dispensable genome are able to create new gene functions due to their gene sequence variations (Lapierre et al., 2009). Differences are only found in substrates specificity even though similarity observed in the nucleotide level. Dispensable gene emerge from gene duplication, horizontal gene transfer and mutation as it happens during evolution process cause divergence of same bacteria species to different strains (Croll et al., 2012).

2.9.3 Pan-genome characterisation

Pan-genome assessment involving same species of the different strains or genes enable researchers to understand the similarities and differences of genomic characteristics better (Snipen et al., 2009). Existence of large number of genomes enable the study of genomic characteristics of bacterial species via pan-genome assessment.

Pan-genomic investigation gives understanding of pathogens evolution, population design, niche characteristics and interactions with host (Carlos Guimaraes et al., 2015). This further enhanced the identification of virulence factors, vaccine and medicine design (Muzzi et al., 2007).

2.9.4 Genome-based phylogenetic analysis

Estimation of evolutionary relationships are known as phylogenetic analysis. The evolutionary history inferred from phylogenetic analysis is normally portrayed as branching, treelike diagrams that represent an estimated pedigree of the inherited relationship (Brinkman et al., 2001).

Moreover, strain-level relationships between base composition, serovar type, genome size and transmission in several microbial species over time were explored in recent analysis (Bohlin et al., 2014).

CHAPTER 3

MATERIAL AND METHODS

3.1 Strains background

Three Malaysian *L. interrogans* strains (D7, R123 and L52) were studied. The initial DNA fingerprinting analysis by Benacer et al., 2013 showed that all three strains were genetically distinct. These strains related to various epidemiological backgrounds that were isolated during 2011 in Kuala Lumpur. The D7 strain was isolated from urine sample of a 2 years male stray dog that showed skin lesions. Strain R123 was recovered from a rat kidney and strain L52 was isolated from human blood. Serological identification of all three isolates was performed previously using the microscopic agglutination test (MAT) as described by WHO (Terpstra, 2003). These three strains determined to be representing different serovars of Canicola (D7), Batavie (R123) and Ricardi (L52) (Benacer et al., 2016) (Table 3.1).

Table 3.1: Details of Malaysian *L. interrogans* strains used in this study.

Strain name	Serovar	Location of isolation	Host
D7	Canicola	Kuala Lumpur, Malaysia	Dog
R123	Batavie	Kuala Lumpur, Malaysia	Rat
L52	Ricardi	Malaysia	Human

The genome of these three isolates were compared with selected 86 reference genomes of *L. interrogans* that are available publicly in NCBI genome database at the

time of analysis. Information of all *L. interrogans* genomes currently available were obtained using web-tools, PATRIC (Pathosystems Resource Integration Center), Genome Viewer (https://www.patricbrc.org/). All these 86 reference strains represent *L. interrogans* from diverse sources of hosts, temporal and geographic backgrounds in association with variable epidemiological settings. The 86 references genome comprises of five complete genomes and 81 whole genome sequences (WGS) were manually downloaded using NCBI file transfer protocol service. The details of the bacterial strains are provided in Appendix A.

3.2 DNA sequencing and assembly

The three Malaysian genomes of each isolate were previously sequenced using the Illumina HiSeq 2000 (100-bp read length) with an insert size of 300 base pairs (bp). The reads were further assembled via *de novo* method using open A5-miseq software, an updated pipeline to assemble DNA sequence data generated on the Illumina sequencing platform. A5-miseq pipeline computerised processes involved adapter trimming, quality filtering, error correction, contig and scaffold generation and detection of misassembles with just one command line (Coil et al., 2014). The quality of the genomes assemblies was further determined using web-service of QUAST (Quality Assessment Tool for Genome Assemblies) (http://bioinf.spbau.ru/quast).

3.3 Genome and functional annotations

Genome annotation of three assembled Malaysian *L.interrogans* genomes and 86 *L. interrogans* reference genomes were performed using a tool of command-based line

software known as Prokka, rapid prokaryotic genome annotation, applied on Unix system. The annotations of all 89 genomes were performed simultaneously with the option of 'usegenus' of *Leptospira* using Prokka batch annotation script. Prokka depends on external application prediction tools of RNAmmer, Aragorn and Prodigal to identify the coordinates of genomic features of ribosomal RNA (rRNA), transfer RNA (tRNA) and coding sequences (CDS) genes respectively within contig (Seemann, 2014).

Functional annotation of CDS of three assembled Malaysian *Leptospira interrogans* genomes which were determined previously by Prokka software were performed using eggNOG-mapper (http://eggnogdb.embl.de/#/app/emapper), an application meant for quick useful explanation of novel sequences utilizing pre-figured profiles of sequences and assigned ortholog. The web-based application is intended for the explanation of substantial numbers of novel sequences, normally focusing on interpreted gene coding sequences. Utilitarian descriptors depend on the latest eggNOG fabricate, and recently incorporate curated GO terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Cluster of Orthologous Group (COG) useful classifications (Huerta-Cepas et al., 2017). The inputs of .faa file for the functional annotation step was obtained from the files generated by Prokka genome annotation.

3.4 Comparative genomic analysis

The presence and absence of genes among the first subset of three assembled Malaysian *L. interrogans* genomes and the second subset of 89 strains of *L.interrogans* representing global sources were determined using Roary, an application that quickly fabricates huge scale pan genomes, distinguishing the core and accessory genes (Page et al., 2015). The inputs of .gff file for the analysis was obtained from the files generated by Prokka genome annotation.

3.5 Pan genome profile and functional analysis

Pan genome profile and functional analysis between all the 89 *Leptospira interrogans* genome sequences used in the study was conducted using ultra-fast computational pipeline, Bacterial Pan Genome Analysis (BPGA) tool performed five functional modules of (i) Pan Genome Profile Analysis, (ii) Pan Genome Sequence Extraction, (iii) Exclusive Gene Family Analysis, (iv) Species Phylogenetic Analysis and (v) Pan Genome Functional Analysis (Chaudhari et al., 2016). For clustering of gene families, CD-HIT clustering tools (Li et al., 2006) was used with identity cut off = 95%. The inputs of .faa file for the analysis was obtained from the files generated by Prokka genome annotation and converted to .fas format to be processed by the BPGA pipeline. The pan genome curve was produced by plotting the aggregate number of dissimilar gene families against the quantity of 89 genomes considered. In addition, core genome plot was produced by the number of shared gene families plotted against the quantity of 89 genomes to portray the pattern of constriction in the core genome size with consecutive expansion of more genomes. 20 random permutations of genomes were performed as it has been seen that highest number of permutations which did not modify the median value of pan genome and core genome sizes meaningfully. KEGG and COG identifiers were allocated for core and accessory genes considering the protein BLAST against reference KEGG and COG databases and utilizing gnuplot. Results were represented in the form of histograms.

3.6 Phylogenetic analysis

Concatenated core gene alignments and binary (presence/absence) pan-matrix concepts were used by BPGA pipeline for phylogeny analysis. Calculation of gene matrix was done using similarity or dissimilarity in contribution of genes to orthologous

gene clusters. BPGA first extracts the protein sequences (excluding paralogs) from 20 random orthologous gene clusters to generate core genome phylogeny tree for core genome based phylogenetic tree. Multiple Sequence Comparison by Log-Expectation (MUSCLE) application was used for BPGA automated multiplication of sequence alignments. Neighbor-joining phylogenetic tree was constructed once all alignments were concatenated (Chaudhari et al., 2016).

CHAPTER 4

RESULT AND DISCUSSION

4.1 General genome characteristics of Malaysian *L. interrogans* strains isolated from animal and human hosts

The genome sequencing and assembly of Malaysian strain of *L. interrogans* serovar Canicola D7 isolated from the dog host produced a draft genome of 344 contigs with minimum contig length of more than 200 bp. A total of 315 contigs contains more than 5,000 bp. The predicted genome size is approximately 4,704,388 bp, with a mean GC content of 35.04%. The genome annotation revealed approximately 3,778 CDS, 37 tRNA and 3 rRNA genes (Table 4.1).

For the Malaysian *Leptospira* strain *L. interrogans* serovar Batavie R123 isolated from the rat, a draft genome of an approximately 4,804,613bp was generated with 353 contigs, where 218 contigs consist of more than 5,000 bp. The draft genome also revealed a GC content of 35.15% and is composed of 3,883 CDS, 37 tRNA and 3 rRNA genes. On the other hand, the size of draft genome of the *Leptospira* strain *L. interrogans* serovar Ricardi L52 isolated from human of was approximately 4,884,364 bp with 232 contigs and 127 contigs consist of more than 5,000 bp, and a G+C content of 35%. There were 3,938 protein coding sequence found. The draft genome revealed 37 tRNA and 3 rRNA genes. The three new assembled Malaysian *Leptospira interrogans* genomes were then deposited on the NCBI WGS database (Table 4.1).

Table 4.1: Malaysian *L. interrogans* strains genome assembly and annotation results.

Feature	Malaysian *Leptospira interrogans* strains		
Strain	D7	R123	L52
Serovar	Canicola	Batavie	Ricardi
Host	Dog	Rat	Human
Level	Draft genome	Draft genome	Draft genome
Accession Number	MCLU00000000	MCLW00000000	MCLV00000000
Contigs	344	353	232
Genome size (Mbp)	4.70	4.80	4.88
GC Content (%)	35.04%	35.15%	35%
Number of tRNAs	37	37	37
Number of rRNAs	3	3	3
Coding Sequences	3,778	3,883	3,938

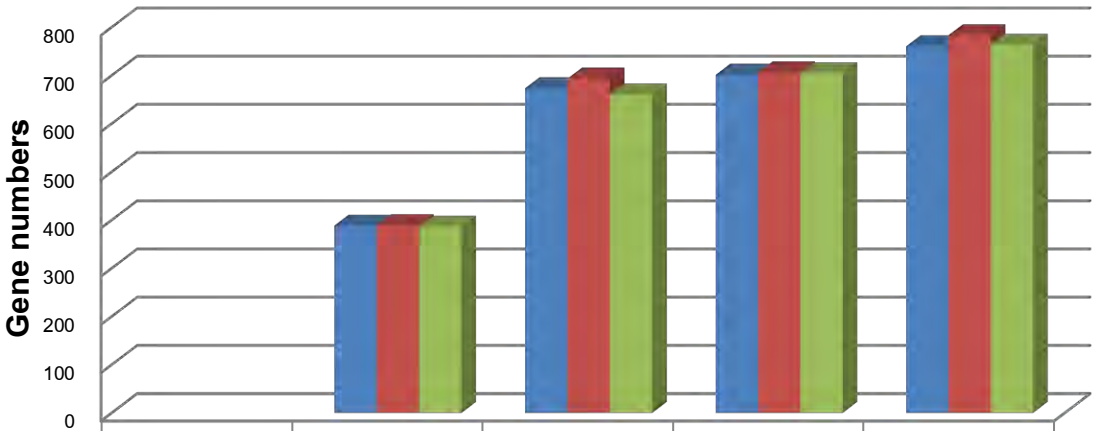
The genome size, GC content and the number of coding sequences for D7, R123 and L52 obtained were almost the same as the average values that have been reported for *L. interrogans* (Genome ID: 179) in NCBI genome database of 4.62Mbp genome size, 35% GC content and 3780 coding sequences. The differences in terms of the genome size for D7, R123 and L52 compared to the reference genome was probably due to the genome assembly procedure that was only performed at scaffold level and the presence of gaps between the contigs that cannot be resolved due to lack of overlapping read. In addition, loss of sequences during the library preparation and presence of repetitive region in *L. interrogans* could be the reason for high number of D7, R123 and L52 assembled contigs (Ekblom et al., 2014).

4.2 Comparative genomics of Malaysian *L. interrogans* strains

The major functional categories that stood out with the greatest number of genes for all three Malaysian strains of *L. interrogans* was "information storage and processing", "cellular processes and signaling" and "metabolism" (Table 4.2 & Figure 4.1). However, the features of many genes rmain with unknown function and warrants for further study.

Table 4.2: EggNOG functional categories for the predicted genes of Malaysian *L. interrogans* strains (D7, L52 and R123) isolated from animal and human hosts.

Category	D7	L52	R123
INFORMATION STORAGE AND PROCESSING	388	389	386
[J] Translation, ribosomal structure and biogenesis	141	140	141
[K] Transcription	88	97	90
[L] Replication, recombination and repair	157	150	153
[B] Chromatin structure and dynamics	2	2	2
CELLULAR PROCESSES AND SIGNALING	673	693	660
[D] Cell cycle control, cell division, chromosome partitioning	21	24	21
[V] Defense mechanisms	44	39	40
[T] Signal transduction mechanisms	208	211	202
[M] Cell wall/membrane/envelope biogenesis	201	219	200
[N] Cell motility	64	65	65
[Z] Cytoskeleton	1	1	1
[U] Intracellular trafficking, secretion, and vesicular transport	34	32	32
[O] Posttranslational modification, protein turnover, chaperones	100	102	99
METABOLISM	701	706	704
[C] Energy production and conversion	124	125	125
[G] Carbohydrate transport and metabolism	87	83	79
[E] Amino acid transport and metabolism	139	143	147
[F] Nucleotide transport and metabolism	53	53	54
[H] Coenzyme transport and metabolism	94	93	96
[I] Lipid transport and metabolism	87	89	90
[P] Inorganic ion transport and metabolism	89	91	87
[Q] Secondary metabolites biosynthesis, transport and catabolism	28	29	26
POORLY CHARACTERIZED	760	783	763
[S] Function unknown	760	783	763



		INFORMATION STORAGE AND PROCESSING	CELLULAR PROCESSES AND SIGNALING	METABOLISM	POORLY CHARACTERIZED
D7		388	673	701	760
L52		389	693	706	783
R123		386	660	704	763

Figure 4.1: EggNOG-mapper major functional categories for the predicted genes of Malaysian *L. interrogans* strains (D7, L52 and R123) isolated from animal and human hosts.

The pathogenic *L. interrogans* can survive and persists in natural habitats such as surface water for extended period (Bulach et al., 2006) before been transmitted into the host and became more sensitive towards UV light compared to saprophytic strains of *Leptospira* (Stamm et al., 1988). Large selection of genes involved in DNA replication, recombination and repair, and transcription regulatory genes, under the category of information storage and processing were observed in all three Malaysian strains of *L. interrogans* used in the study. These genes allow fast adaptative changes of *L. interrogans* for the existence under various range of environments and conditions.

The cell envelope of *Leptospira,* made up of lipoproteins, provides *L. interrogans* the ability to adhere towards various cell types and plays a major role in pathogenesis. In addition, cell wall act as an interface between the pathogen and host cell (Xue et al., 2009). Large number of genes were involved in biogenesis of cell wall, cell membrane

and envelope, under the category of cellular processes and signaling were annotated in D7, L52 and R123 contributing towards host-pathogen interaction.

Almost 64-65 and 202-211 genes involved in motility and signal transduction mechanism or chemotaxis under the category of cellular processes and signaling, were recorded respectively in all three newly assembled Malaysian genome of *L. interrogans*. Synthesis of motility and chemotaxis proteins may indicate how the pathogenic *L. interrogans* gain its ability to penetrate skin, mucous membrane and translocate cells immediately after infection (Barocchi et al., 2002) . Chemotaxis and motility have a major role in pathogenesis of *Leptospira*, suggested by the respond of *L. interrogans* towards hemoglobin via swimming action (Yuri et al., 1993).

Most of the genes involved in carbohydrate, amino acid, lipid and coenzyme metabolisms and transports under the category of metabolism are highly observed among all the three Malaysian strains of *L. interrogans,* indicating their role as essential genes to survive in different hosts.

The percentage of functional annotation is presented in the form of pie chart for visualization purpose according to categories for each D7, L52 & R123 strains respectively (Figure 4.2, 4.3 & 4.4).

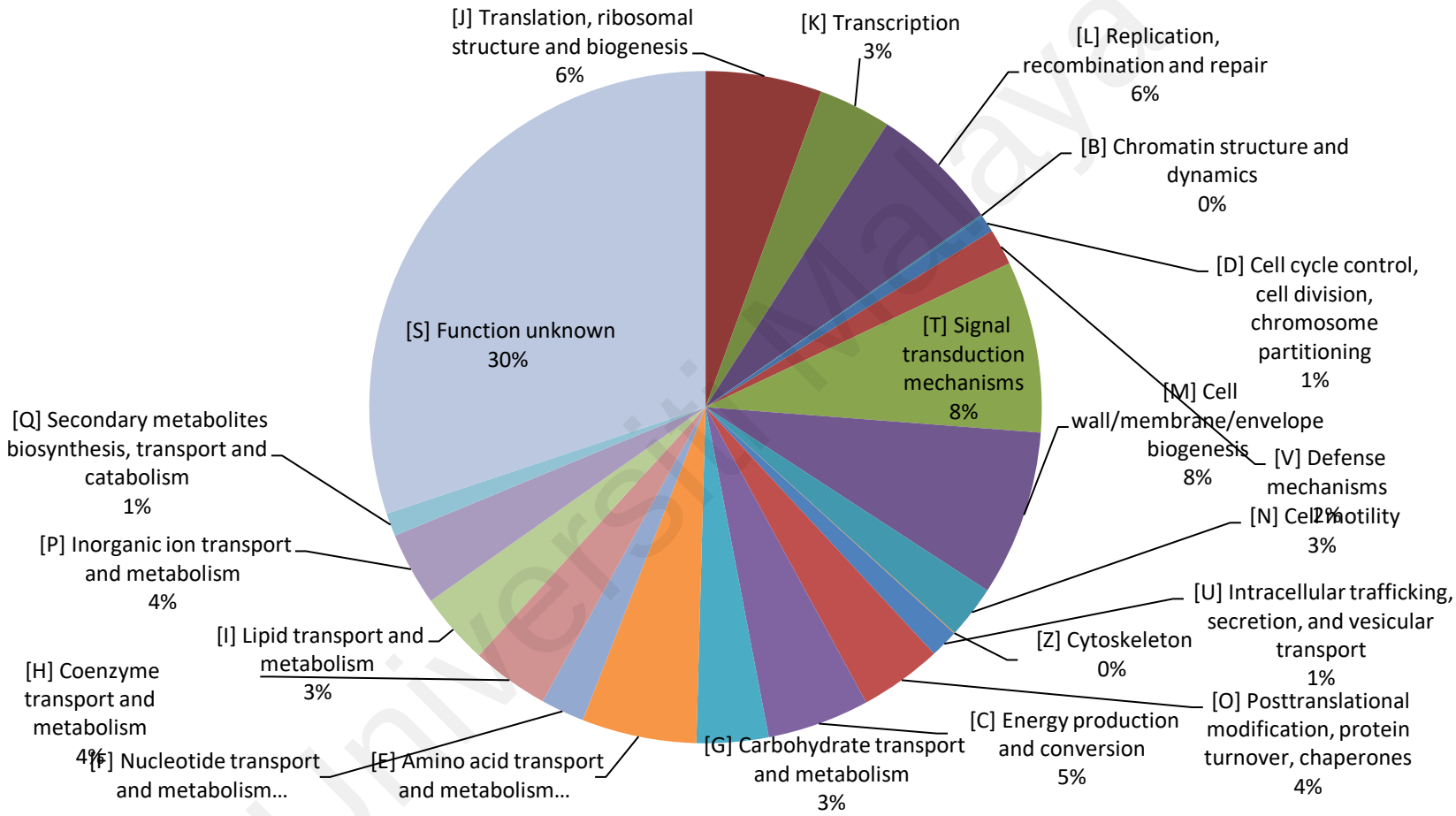


Figure 4.2: EggNOG-mapper functional categories percentage for the predicted genes of Malaysian *L. interrogans* serovar Canicola D7.

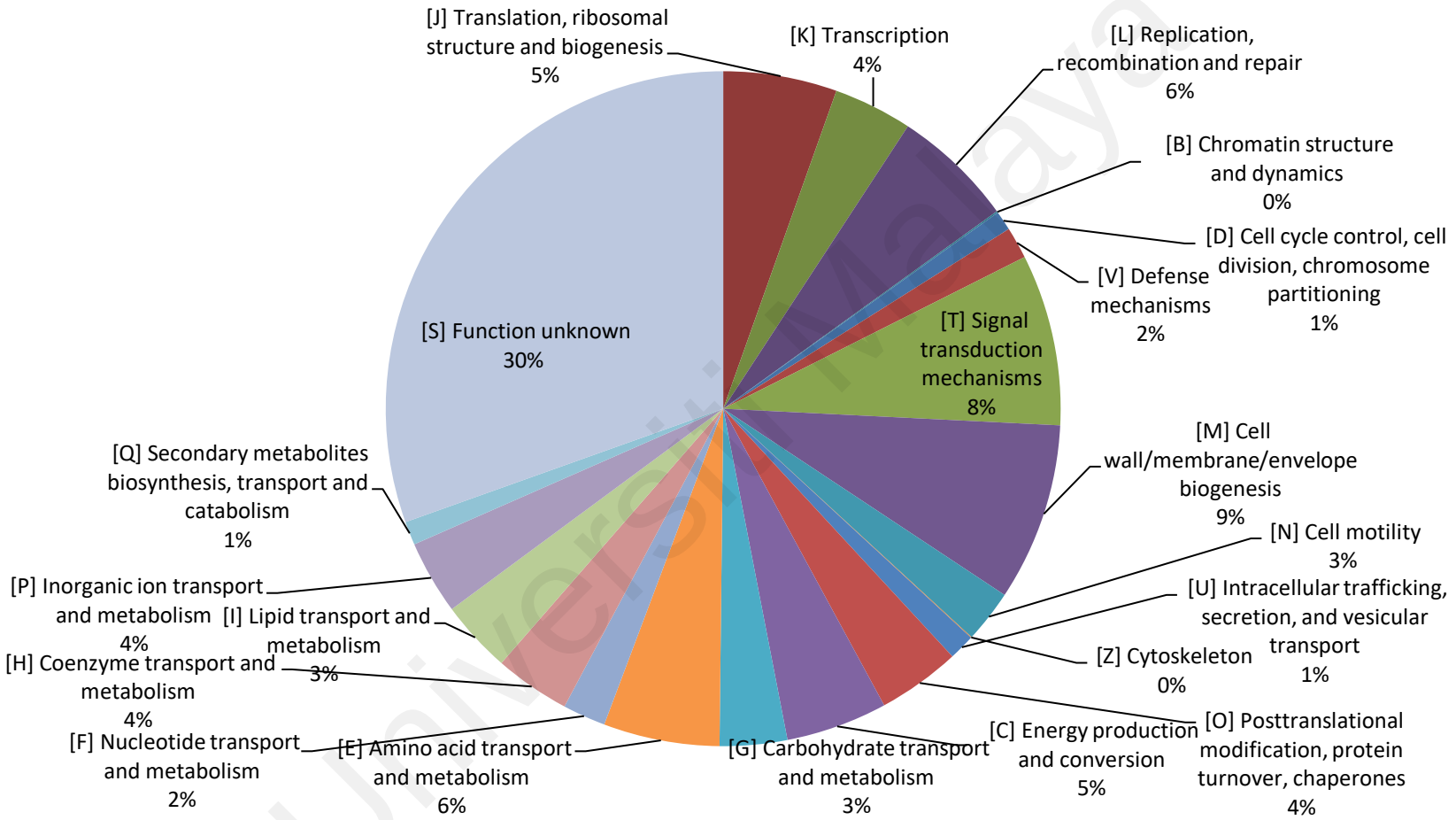


Figure 4.3: EggNOG-mapper functional categories percentage for the predicted genes of Malaysian *L. interrogans* serovar Ricardi L52.

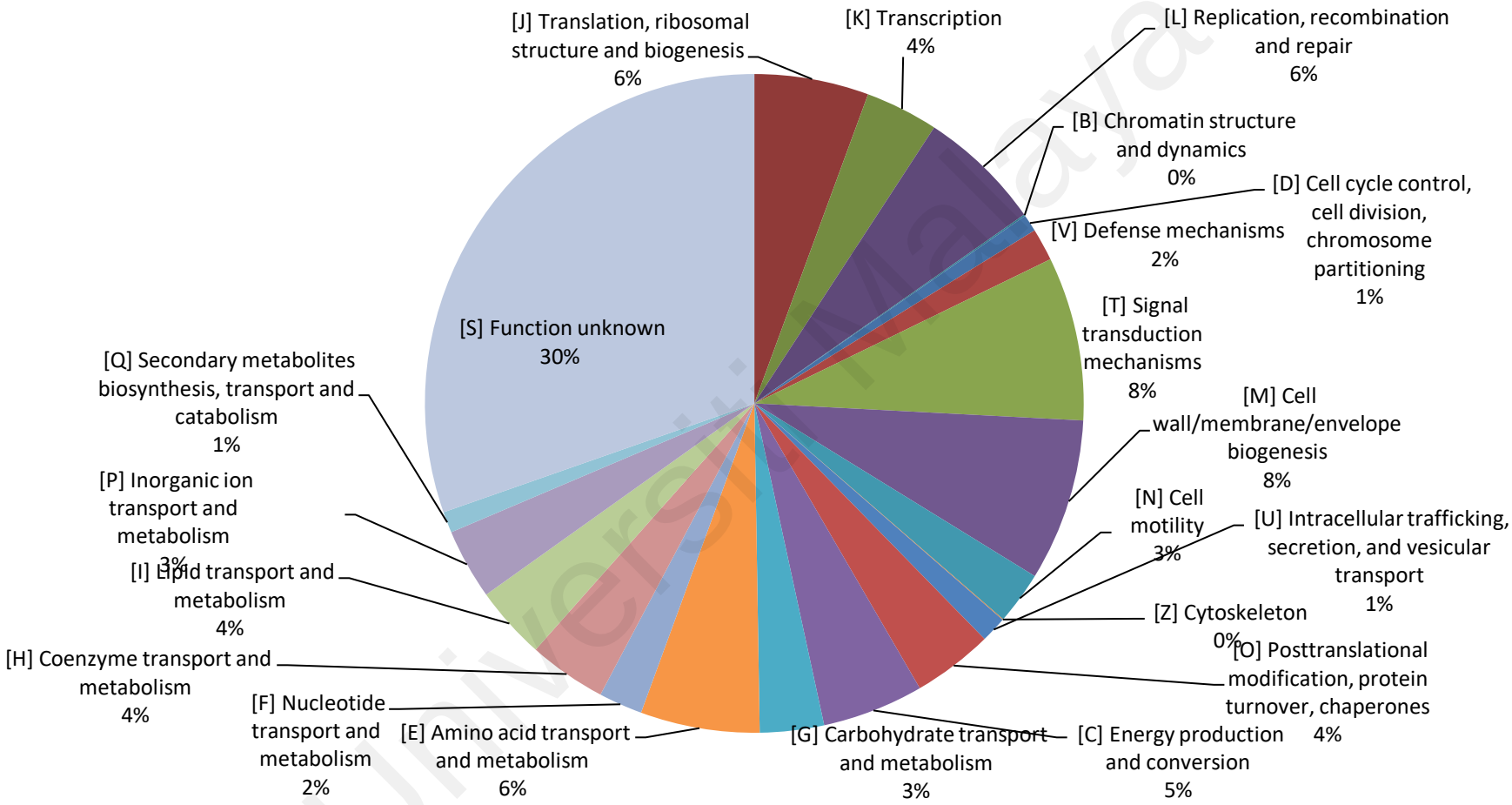


Figure 4.4: EggNOG-mapper functional categories percentage for the predicted genes of Malaysian *L. interrogans* serovar Batavie R123.

4.3 Shared and dispensable genomes among Malaysian and global *L. interrogans* strains

The genomes of three Malaysian *L. interrogans* strains have been analysed to determine the shared and dispensable genes. A total of 3,150 genes were detected as shared genes among all three Malaysian strains and 1720 genes were detected as accessory or dispensable genes (Figure 4.5).

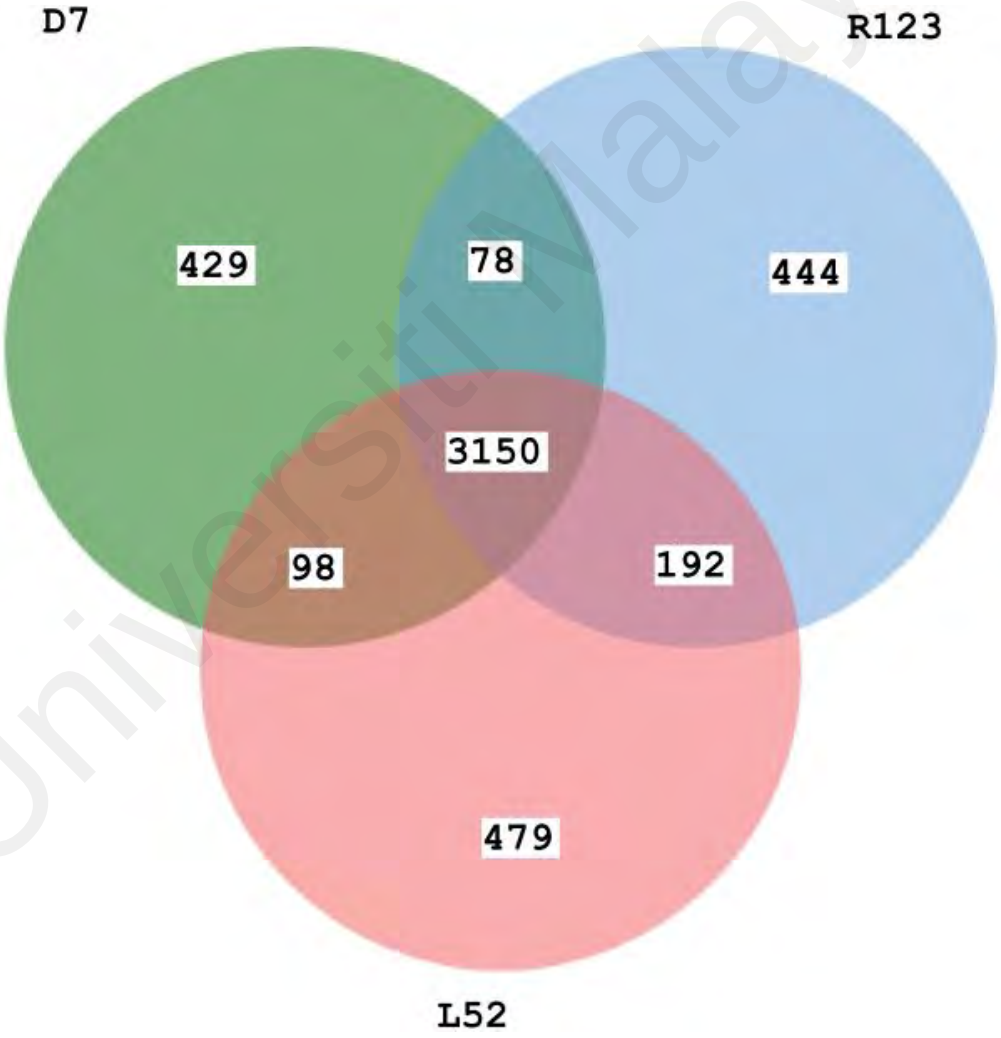


Figure 4.5: The shared and dispensable genomes of Malaysian *L. interrogans* strains (D7, L52 and R123). The Venn diagram illustrates the number of shared and unique genes based on clustering analysis with a percentage sequence identity of 90% using Roary.

Further analysis on the 3,150 core genes, shows that 2,207 CDS were inferred with functional categories by eggNOG-mapper. Out of the 2,207 CDS around 2, 166 CDS were inferred with one functional category, 40 CDS with two functional categories and 1 CDS with three functional categories. A total of 2249 functional annotation were obtained from the analysis (Table 4.3).

Table 4.3: EggNOG functional categories for the core genes of Malaysian *L. interrogans* strains (D7, L52 and R123) isolated from animal and human hosts.

Functional Categories	Count of core genes
[J] Translation, ribosomal structure and biogenesis	139
[K] Transcription	80
[L] Replication, recombination and repair	111
[B] Chromatin structure and dynamics	2
[D] Cell cycle control, cell division, chromosome partitioning	19
[V] Defense mechanisms	35
[T] Signal transduction mechanisms	198
[M] Cell wall/membrane/envelope biogenesis	165
[N] Cell motility	62
[Z] Cytoskeleton	1
[U] Intracellular trafficking, secretion, and vesicular transport	31
[O] Posttranslational modification, protein turnover, chaperones	98
[C] Energy production and conversion	121
[G] Carbohydrate transport and metabolism	78
[E] Amino acid transport and metabolism	131
[F] Nucleotide transport and metabolism	53
[H] Coenzyme transport and metabolism	93
[I] Lipid transport and metabolism	85
[P] Inorganic ion transport and metabolism	87
[Q] Secondary metabolites biosynthesis, transport and catabolism	24
[S] Function unknown	636
Total	**2249**

The percentage of functional annotation of core genes is presented as pie chart according to categories for visualization purpose (Figure 4.6).

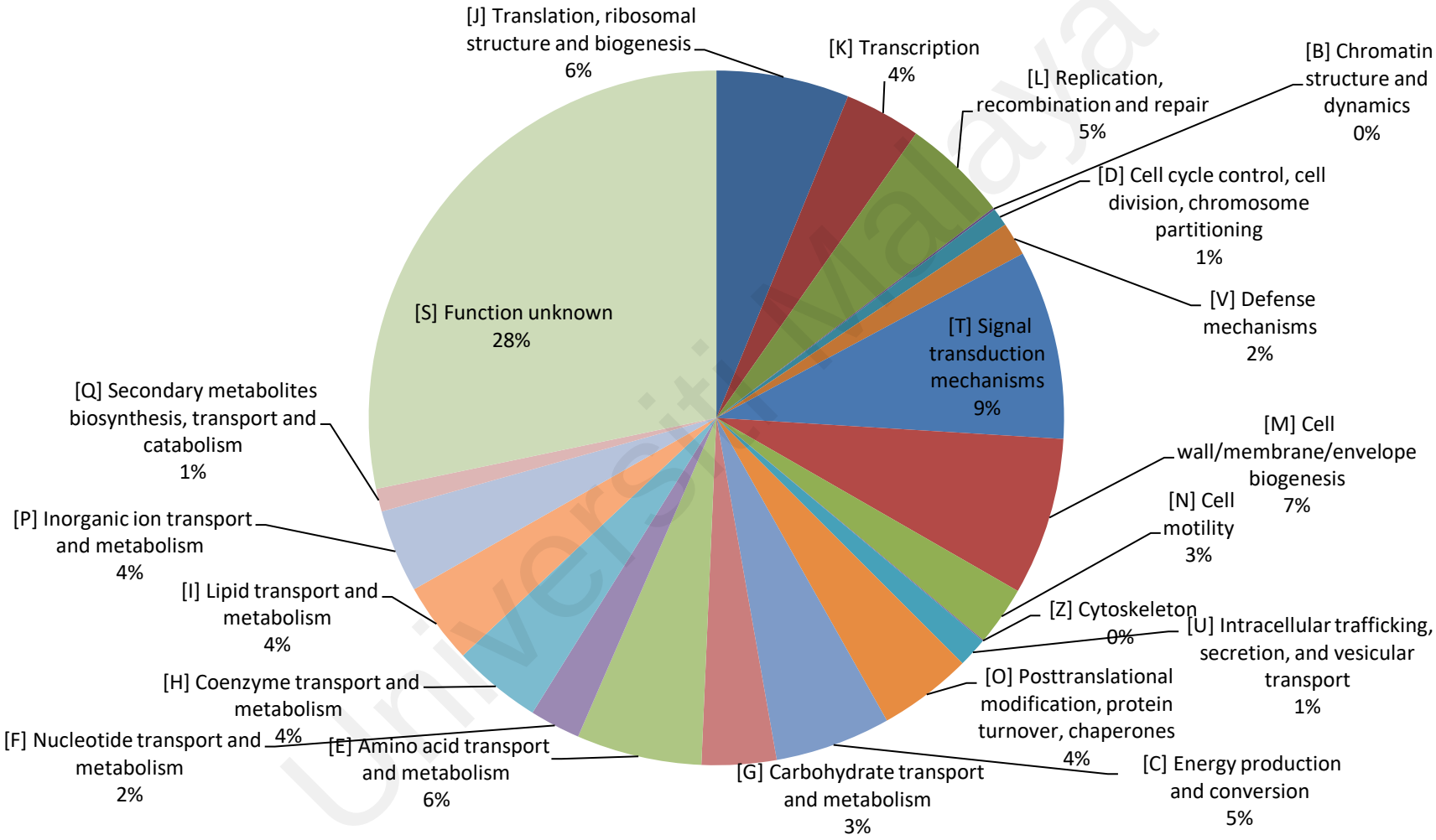


Figure 4.6: Functional categories of Malaysian *L. interrogans* strains (D7, L52 and R123) core genes inferred by EggNOG-mapper.

Around 50 coding sequences annotated as functional and structural protein for bacterial cell motility has been reported in *B. burgdorferi* and *T. pallidum* (Fraser et al., 1997; Fraser et al., 1998). In contrast, chemotaxis mechanism of *L. interrogans* is more complicated as 65 genes associated with motility were annotated in D7, R123 and L52 genomes. Sixty-two out of 65 genes are found to be present in all three strains (Appendix B) and highly conserved. The high number of genes involved in motility are accounted by many copies of protein coding sequences associated to chemoreceptor glutamine deamidase, chemotaxis, flagellar filament, flagellar basal-body rod, flagellar biosynthesis, flagellar hook, flagellar motor switch, methyl-accepting chemotaxis and flagellar filament outer layer. All three strains contain 2 *cheD* genes, 2 *cheR* genes, 3 *cheW* genes, 2 *flaAL* genes, 5 *flaB* genes, 2 *flgG* genes, 3 *fliG* genes, 2 *fliN* genes, 2 *mcp4* genes and 2 *pomA* genes. The high number of genes related to motility or chemotaxis as core genes suggesting the flexibility of pathogenic *L. interrogans* to adapt and survive in wide range of environment and invading different nature of hosts.

Hemolysin plays a major role in exotoxin process, destroys red blood cell membranes and caused cell rupture and pathogenesis process. Primary lesion or damage towards endothelium blood vessels caused localised ischemia and hemorrhage in multiple organs. Core genes responsible for virulence such as sphingomyelinase (*hlb* and *tlyAC*) and phospholipase (*smc*L and *sph*) were identified in all three Malaysian *L. interrogans* strains of D7, R123 and L52 (Appendix C). Red blood cells are a major target for these enzymes since erythrocytes are rich with antigen that determine ABO blood groups such as glycosphingolipids (Xue et al., 2009). In addition, genes involved in coding of TonB dependent outer membrane receptor for iron accession may play a major role as virulence factor and survival factor in various environmental condition. *cirA* gene is highly conserved among D7, R123 and L52 strains (Appendix D), which explain its

function as core gene that is involved in the transport of iron containing molecules such as heme group (Louvel et al., 2006).

Survival of Leptospira in the environment and adhesion to host cells as key steps during infection has been inducted by the presence of conserved genes such as *cat, lexA* and *recA* among D7, R123 and L52 strains (Appendix E). *lexA* and *recA* genes are involved in SOS response and autocatalytic cleavage respectively. These genes are important for DNA repair and enable *L. interrogans* to survive in different types of natural habitats even though *L. interrogans* is highly sensitive towards UV light compared to other saprophytic bacteria (Stamm et al., 1988). In addition, conserved catalase genes provide a mechanisms for *L. interrogans* to counter act upon oxidative stress present in specific environments (Xue et al., 2009). Moreover, the presence of regulatory genes (Appendix F) as repertoire core genes provides an ability for *L. interrogans* to rapidly change and survive under various environmental and host conditions. Repetitive number of genes involved in transcription regulation were observed such as 2 *degU* genes, 5 *hupR1* genes, 2 *pdtaR* genes and 2 *zraR* genes suggested major role of these genes as response regulator towards environmental change and host conditions.

A second set of analysis using Roary was performed on 89 strains of *L. interrogans* representing global sources to obtain comprehensive understanding of genetic diversity among worldwide strains related to survival on diverse ecological niches, adhesion to host cells, invasion of various number of hosts, virulence and pathogenicity factors.

Peptidoglycan sacculi made of peptidoglycan macromolecules is important to maintain morphology of *L. interrogans* with intact helical shape. Rod or cell shape determining protein is essential for maintenance of cell shape in most bacteria including *Leptospira* to survive in various environmented conditions and during intracellular

infection (Slamti et al., 2011). Three genes, *mrdB*, *mreB* and *mreC* (Appendix G) are highly conserved among all the 89 strains, suggesting their role in determining helical morphology of *L. interrogans*.

Adhesion related protein promote interaction with macromolecules and host cell membrane receptors during infection started and the colonisation of internal organs such as kidney. *tlyC* gene coding for surface exposed adhesins (Fouts et al., 2016) were identified with highest nucleotide sequence identity among all the 89 strains of *L. interrogans,* (Appendix H) suggested it's primary role to code for adhesion protein. Presence of adhesion protein contributes towards the binding of *L. interrogans* on ECM such as fibronectin, laminin and collagen. Hemolytic activity of tlyC hemolysins remain unknown in *Leptospira* (Carvalho et al., 2009). Moreover, based on the comparative analysis between 89 strains of *L. interrogans* showed that *ligA* gene was present in almost all the strains (Appendix H). This gene encodes surface exposed immunoglobulin-like (Lig) protein (Matsunaga et al., 2003). High level of nucleotide similarity of *ligA* gene between all the strains studied suggested function of the Lig protein as an important factor contributing towards cell adhesion regardless of hosts type (Ptak et al., 2014). In addition, several copies of genes (*inlA*, *slrP*, *sspH1* and *sspH2*) coding for leucine-rich repeat containing protein were identified in 89 strains of pathogenic *L. interrogans* studied (Appendix I). Leucine-rich repeat containing protein has been reported to be involved in pathogen and host interaction (Bierne et al., 2007).

The presence of genes coding for protease of thermolysin (Fraga et al., 2013) and collagenase which function (Kassegne et al., 2013) to degrade complement factors in hosts as immune evasion mechanism and have been suggested as one of the important virulence factors for invasion and transmission of *L. interrogans*. The sequence identity of the *colA* and *npr* genes encoding for thermolysin and collagenase respectively is found to be among low 89 strains of *L. interrogans* (Appendix J). It may suggest that

the genes have distinct roles in producing protease with different protein structures complement to various complement factors in different hosts. In addition, the presence and absence of variable type of lipid A biosynthesis genes (Appendix K) among 89 strains of *L. interrogans* studied suggested how it can complement endotoxinogenic activity (Fouts et al., 2016) of *L. interrogans* in various hosts. Endotoxinogenic activity cause toxic damage towards vascular walls and endothelial cells (Böhm, 1982), and is one of the important virulence mechanism to be consider. A total of 11 known enzymes involved in lipid A biosynthesis pathway, coded by genes *htrB*, *kdsA*, *kdsB*, *kdsD*, *lpxA*, *lpxB*, *lpxC*, *lpxD*, *lpxK* and *waaA* were identified.

Motility factor is highly required by pathogenic *L. interrogans* to invade and colonise wide range of hosts. Phylum of Spirochaetes showed the presence of endoflagellum with unique characteristics that consist of basal body, filament with inner and outer core layers, flexible hook and larger and more complex flagellar motor (Picardeau, 2017). Repetitive number of genes coding for proteins to assemble structure of endoflagellum; basal body (*flgB*, *flgC* and *flgG*), filament (*flaB*), hook (*flgE*, *flgK*, *fliD* and *fliE*) and switch motor (*fliG, fliM* and *fliN)* were identified present in all the 89 strains studied with high percentage of sequence identity (Appendix L). Synthesis of switch motor proteins may enable *L. interrogans* to swim by changing their direction within certain hundreds of milliseconds upon sensation of substrates (Goldstein et al., 1990). Additional genes to fasten the rotation of flagellar motor (*motB*) and regulate swimming and swarming (*ycgR*) (Appendix L) are required for motility, invasion of hosts and pathogenesis (Kearns et al., 2004; Paul et al., 2010).

Acquisition of coding sequences through gene duplications and horizontal genes transfer (Xu et al., 2016) encode for chemotaxis proteins, two-component systems (TCSs), catalase enzyme and vitamin B12 may have contributed towards evolution of *L.*

interrogans with the ability to colonise their hosts and survive under low nutrient condition (Fouts et al., 2016).

Coding sequences producing chemotaxis protein are highly conserved among pathogenic bacteria species (Fouts et al., 2016). Among the genes discovered in the genomes of 89 *L. interrogans* studied, majority were coding for methyl-accepting chemotaxis and chemotaxis regulator proteins (Appendix M). Series of genes identified were *cheA*, *cheB*, *cheC*, *cheD*, *cheR*, *cheW*, *cheX* and *cheY* synthesising chemotaxis regulator proteins while *mcp2*, *mcp3*, *mcp4*, *mcpA*, *mcpB*, *pctB*, *pctC*, *pomA*, *tar* and *tsr* were coded for methyl-accepting chemotaxis proteins.

TCSs are uttermost switches at molecular level involved in controlling of bacteria signaling event for colonisation of host, which may determine pathogenicity mechanism of pathogenic *L. interrogans* (Fouts et al., 2016). In general, TCSs consist of response regulator protein and enzyme known as effector response regulator and sensor histidine kinase respectively. In the study, repetitive number of 9 genes annotated as *cssS*, *liaS*, *mprA*, *pleD*, *rcp1*, *rcsC*, *tmoS*, *todS* and *uvrY* were conserved among the 89 strains with low level of nucleotide sequence similarity, coding for different types of effector response regulators and sensor histidine kinase proteins (Appendix N).

Catalase action in pathogenic *L. interrogans* such as oxidative rupture-facilitated killing may provide intracellular resistance towards pathogens killing mechanism of host cell (Fouts et al., 2016). Catalase enable *L. interrogans* to live and continually present in subcellular compartments (Fouts et al., 2016). Gene of *katA* coding for catalase protein were detected in genomes of all 89 strains of *L. interrogans*. In addition, the presence of *katA* suggested an alternative way of *L. interrogans* to survive in mammalian hosts and detoxify oxygen instead of superoxide mutase which is absent in *L. interrogans*. Moreover, the presence of two enzyme systems of cytochrome c551

peroxidase and thiol peroxidase provided the clue about ability of the *L. interrogans* to survive against host-derived peroxidases action (Appendix O). Cytochrome c551 peroxidase coded by *ccp* and *ccpA* functions by catalyzing the peroxidative oxidation of cytochrome c551 and azurin (Karlsen et al., 2005). Thiol specific peroxidase encoded by *tpx* catalyzes the reduction of organic hydroperoxides and hydrogen peroxide to alcohols and water, respectively (Cha et al., 2004). Both enzymes play a major role in detoxifying peroxides to protect cell against oxidative stress.

Cluster of *cbi* and *cob* genes and one *sir* gene were discovered among 89 strains of *L. interrogans* (Appendix P). These genes have been predicted to code for proteins involved in the biosynthesis of Cobalamin (vitamin B12) for growth in mammalian hosts with limited nutrient amount. Vitamin B12 autotrophy of *L. interrogans* is hyphotised as the survival mechanism towards mammalian system that is known to deprive iron from pathogens (Fouts et al., 2016).

Genes encoding transferase and synthase proteins related to Sialic acid biosynthesis of Pseudaminic acid (*psel*) and Legionamic acid (*legF*, *neuA* and *neuB*) were discovered among the 89 strains of pathogenic *L. interrogans* studied (Appendix Q). Pseudaminic acid is shown to be involved in biogenesis of flagella in *Helicobacter* spp. and *Campylobacter* spp. (McNally et al., 2008). While Legionamic acid located on the cell surface of pathogenic bacteria such as *Camplyobacter coli* and *Legionella pneumophila* seems to function in cell adhesion and immune evasion (Schoenhofen et al., 2009). Both of the similar mechanisms are expected to happen in pathogenic *L. interrogans* and may determine pathogenicity and function as a virulence factor (Fouts et al., 2016).

4.4 Pangenome-global wide and molecular genome function analysis of the *L. interrogans* species

A total of 86 reference genomes of *L. interrogans* strains which have been isolated from wide numbers of hosts and geographic locations (Appendix A) were obtained through the public NCBI database. Obtained 86 reference genomes were analysed together with three assembled genomes of Malaysian *L. interrogans* strains to define the pangenome of *L. interrogans.* The preparatory analysis of the *L. interrogans* genome size inconstancy is shown in Table 4.4.

Table 4.4: Main features of the *L. interrogans*.

Features analysed	***L. interrogans* pangenome**
Genomes	89
Total genes	331, 114
Average genome size	3720
Pangenome size (non-redundant genes)	21, 826
Core genome	1008
Accessory genome	7060
Unique genes	6698

The *L. interrogans* contain 3720 genes on average ranging from 3266 to 4460 genes per genome (Appendix R), indicating a variation of gene content between 12 – 19 % among all the 89 strains analysed. A total of 21, 826 non-redundant genes were obtained from all strains analysed. Core genome or genes present in all genomes are highly likely associated with maintenance of the essential biology processes (Medini et al., 2005). Based on the analysis, 1008 genes were identified as core genes that present in all 89 *L. interrogans* genomes studied and these accounted for 4.6 % of the pangenome. Low percentage of core genome explained the high diversity among *L. interrogans* (Lawrence et al., 2005). Almost one third of the full set of genes constituting the *L. interrogans* pangenome, 7060 genes or 32% were found as accessory or dispensable genome, composed of genes that absent in any one of the 89 strains. This subset of

genes could be related to the adaptation in different environments and hosts and virulence or pathogenicity factors (Read et al., 2006).

Using power fit and exponential decay model, fluctuations of pan and core genome after the addition of new *L. interrogans* sequences can be exploited mathematically. For pan genome, the expected size of gene repertoire was 14, 766 genes. Meanwhile for core genome, the extrapolation of exponential decay curve indicates no changes on the number of core genes with the addition of new sequences. According to power law equation (f(x)= a.x^b), the parameter b= 0.282507 estimated that *L. interrogans* pan genome is still open but maybe be closed soon (Figure 4.7) .

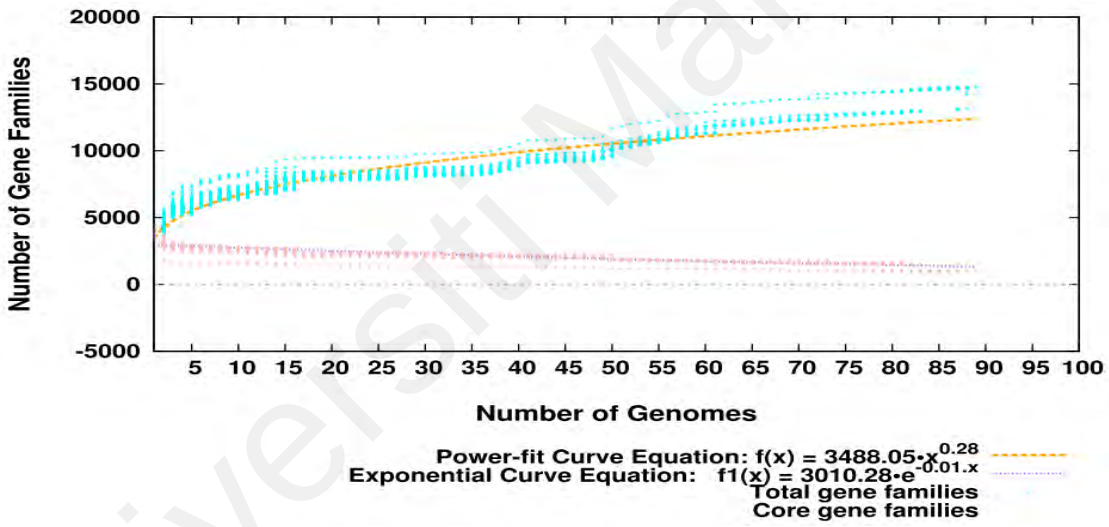


Figure 4.7: Total number of pan genome and core genome according to the number of genomes sequentially added. The blue and pink curves was indicating pan genome and core genome respectively as number of genomes added.

Functional analysis based on KEGG classification for *L. interrogans* gene repertoire showed that the genes were mainly assigned to categories of carbohydrate, amino acid and energy metabolism for core genes and carbohydrate metabolism, amino acid metabolism and metabolism of cofactor and vitamins for accessory genes (Figure 4.8). In addition, COG classification had postulated major categories of signal transduction mechanism, translational, ribosomal structural biogenesis, energy production conversion

and amino acid transport metabolism for core genes and cell wall/ membrane/ envelope biogenesis, replication, recombination, repair and amino acid transport metabolism for dispensable genes (Figure 4.9).

A total of 89 strains of *L. interrogans* including Malaysian strains of D7, L52 and R123 were selected for this pan-genome analysis to obtain better understanding and comprehensive idea of genetic conservation and diversity that facilitated the adaptation of these spirochetes to different hosts and responsible to virulence and pathogenicity. Of these strains, 43 isolates were isolated from human and remaining genomes from various animal hosts such as cattle, swine, horse, dog, mouse, rat, frog and eared seal.

In divergence to the pan-genome (21, 826 pan genes) of 89 *L. interrogans* strains, the number of core-genome was kept comparatively persistent at 1008 genes. From the functional analysis, high number of core genes were found involved in fundamental metabolism pathway such as amino acid metabolism, carbohydrate metabolism, cell motility, energy metabolism, glycan biosynthesis and metabolism, lipid metabolism, membrane transport, metabolism of cofactors and vitamins and signal transduction (Appendix S).

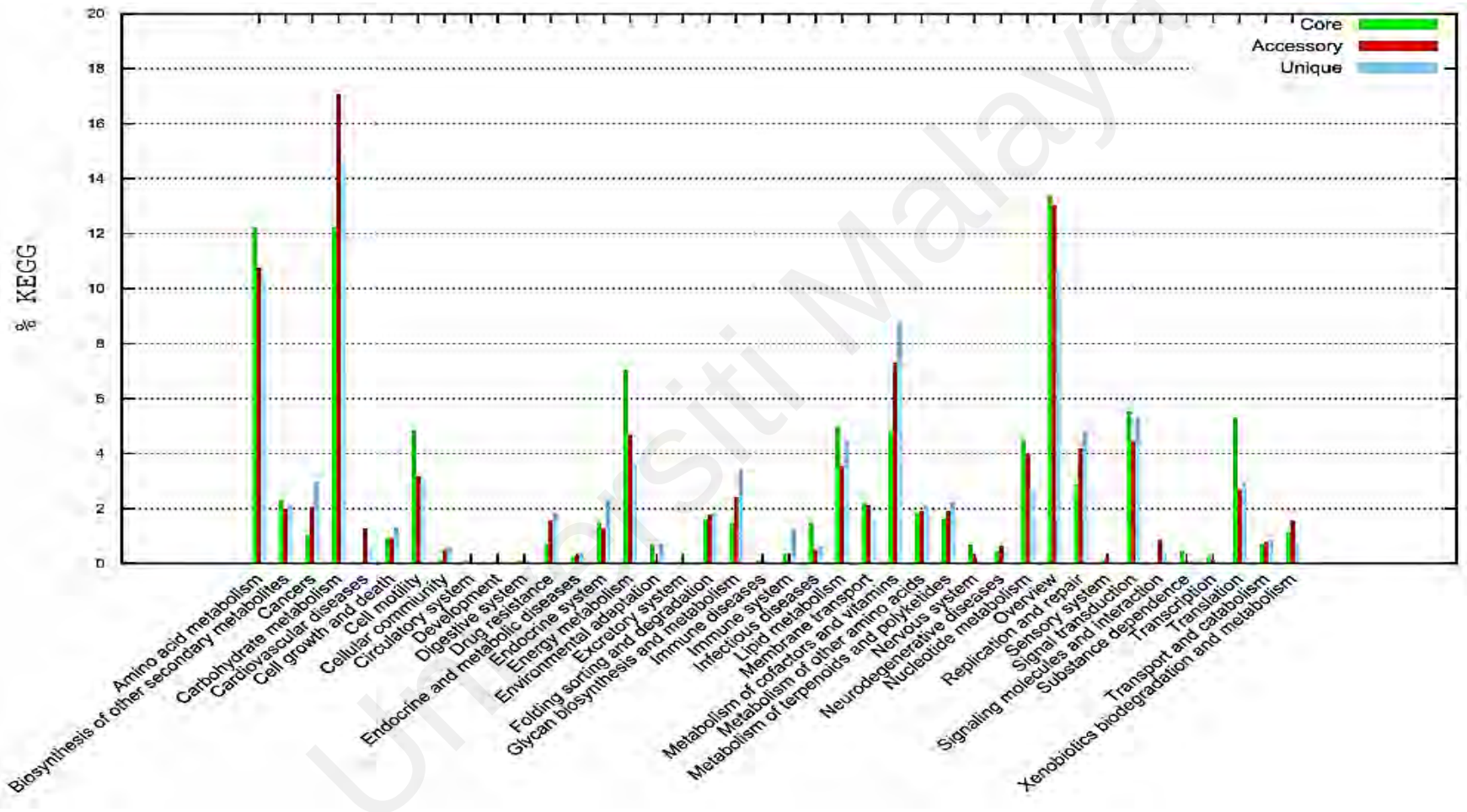


Figure 4.8: Percentage of distribution of KEGG functional categories of pan, accessory and unique genes in 89 isolates of *L. interrogans*.

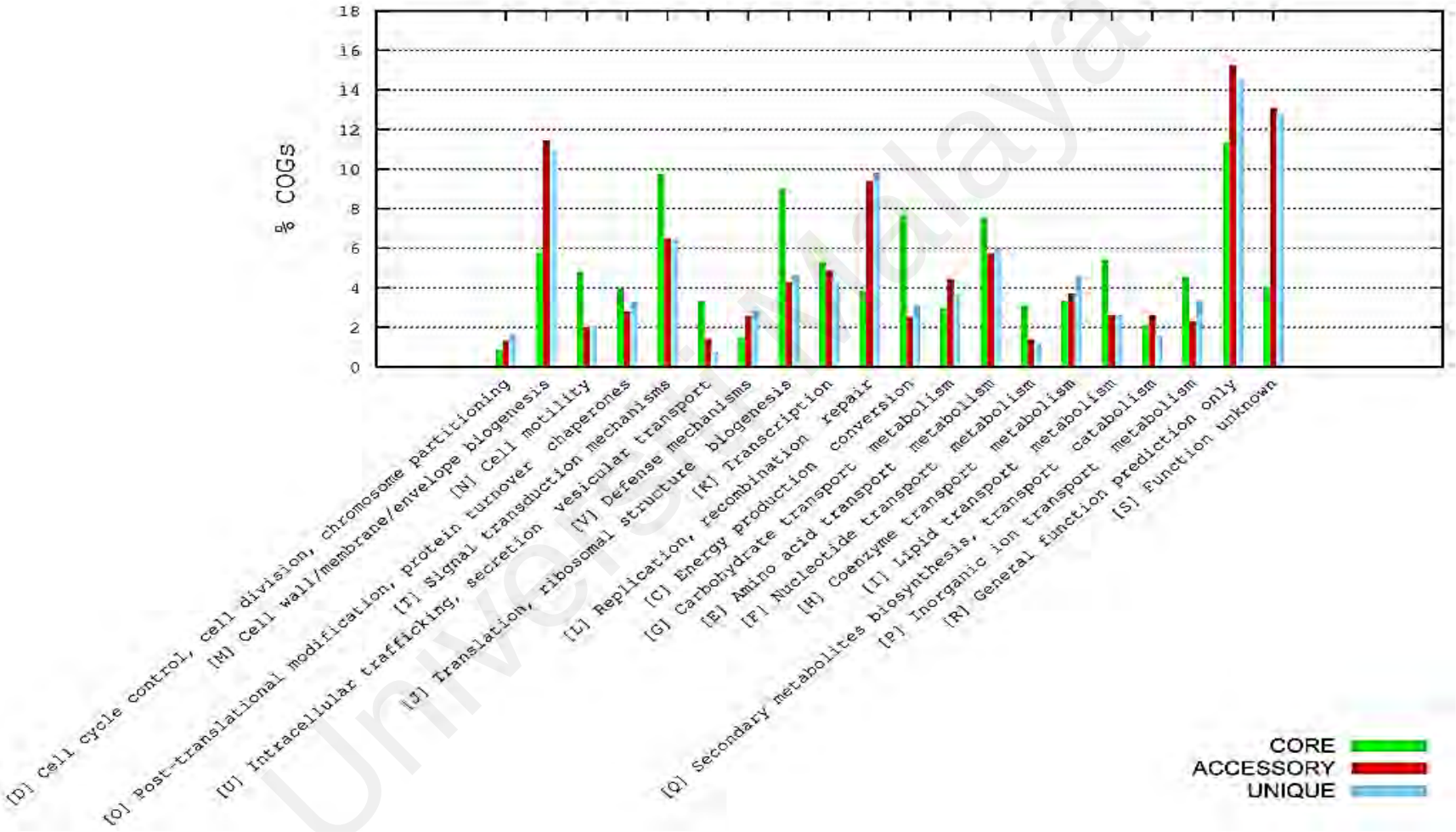


Figure 4.9: Percentage of distribution of COG functional categories of pan, accessory and unique genes in 89 isolates of *L. interrogans*.

4.5 Malaysian *L. interrogans* strains phylogenetic analysis relative to global reference strains

Two phylogenetic trees were built from concatenated core gene alignments and binary (presence/absence) pan-matrix concepts (Figure 4.10 & Figure 4.11). Interestingly serovar based clustering were detected in both phylogenetic trees. Most of the sub-clusters were consists of *L. interrogans* with the same serovar type despite being separated by year, geographic locations and hosts. This could highlight the development of phylogenetic analysis method to identify *L. interrogans* at the serovar level as result of NGS technology developments and availability of whole genome sequence.

The results presented by both phylogenetic trees also support the idea that *L. interrogans* to be spread via secondary transmission method of human migration and international travel across nations borders as the clusterings observed were not according to countries or continents. Secondary transmission of Leptospirosis is re-shaping the landscape of disease occurrence and prevalence worldwide (Bandara et al., 2014).

Various sort of activities involving human and animal allow infectious bacteria such as *L. interrogans* to be transmitted and adapt more effectively to various geographical locations by their host. Horizontal gene transfer and gene duplications is a common mechanism contributing towards gene diversity of *L. interrogans* due to its wide number of intracellular hosts. Hence by taking into consideration of core gene conversation and presence and absence of genes, a strong phylogenetic link of Malaysian *L. interrogans* D7, L52 and R123 strains was demonstrated by both phylogenetic analysis relative to *L. interrogans* distinctive topographical origins.

Notably, the Malaysian *L. interrogans* strains of D7, L52 and R123 appeared to be ancestrally related to the South America and Asia sub-clusters, in which it diverged from common ancestor to form a distinct cluster. *L. interrogans* strain D7 shows close was

phylogenetic relation with strains isolated from Brazil, China and Vietnam, while L52 strain phylogenetically related with strains originated from China, India, Indonesia, Japan, Philippines. Meanwhile *L. interrogans* strains R123 share close phylogenetic relationship with other strains isolated from China, Indonesia, Malaysia and Peru.

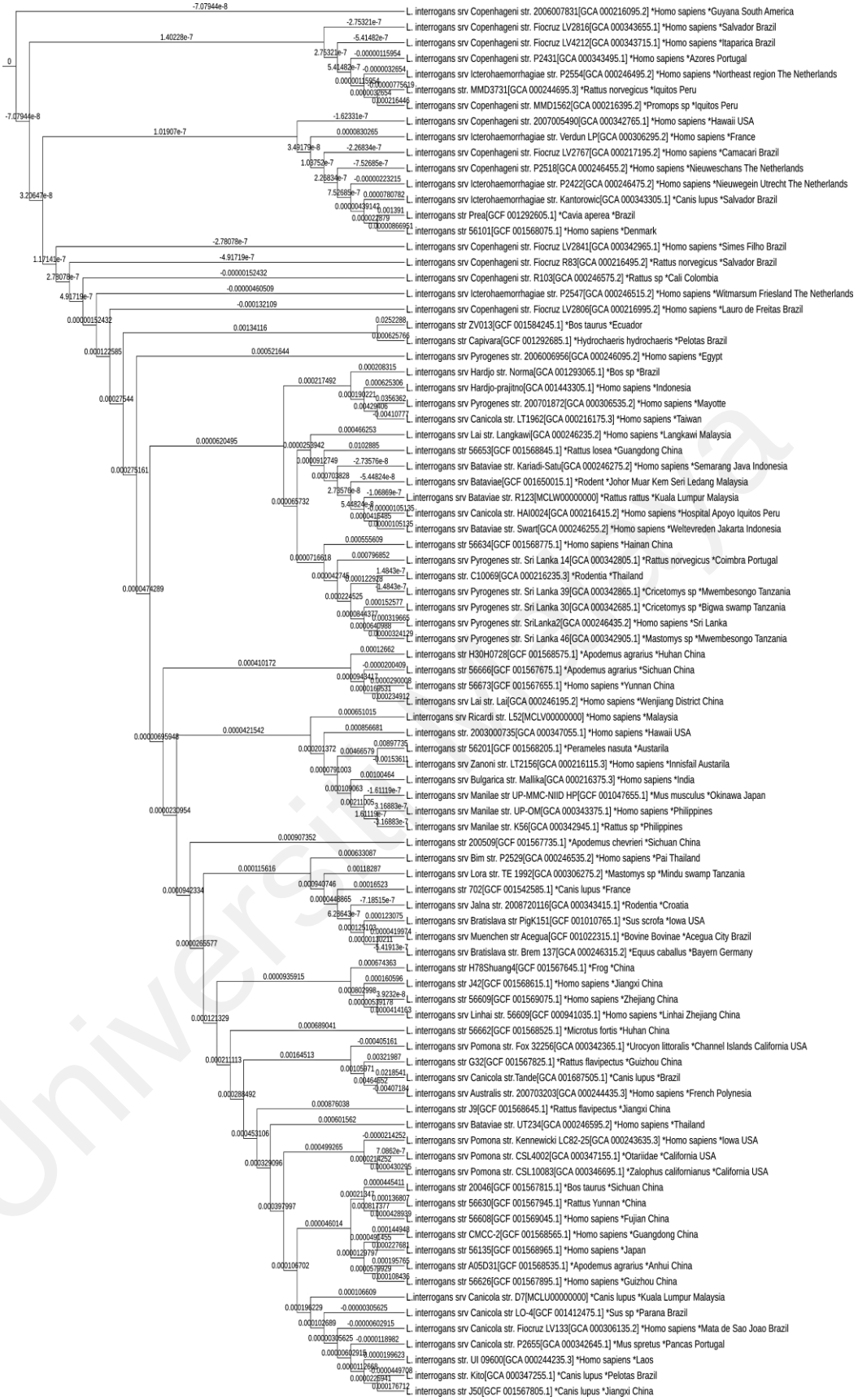


Figure 4.10: Phylogenetic tree generated based on concatenated core gene alignments of 89 strains of L. interrogans studied representing global source.

*As the diagram consist of multiple branching lines, the clearer and detailed version of the image is not included in the text. The original file is attached together with the thesis in softcopy version for better visualization.

Figure 4.11: Phylogenetic tree generated based on binary pan-matrix concept of 89 strains of L. interrogans studied representing global source.

*As the diagram consist of multiple branching lines, the clearer and detailed version of the image is not included in the text. The original file is attached together with the thesis in softcopy version for better visualization.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

In conclusion, large scale comparative analysis between 89 strains of *L. interrogans* isolated from various hosts and different geographic location has provided a good understanding regarding genetic variations that enable *L. interrogans* to survive and adapt as infectious member of *Leptospira* genus. In addition, several virulence or pathogenicity factors causing Leptospirosis infection were successfully eluciaded. The ability of *L. interrogans* to survive as intracellular pathogen enlightened by their spiral shape and immune evasion mechanisms. Synthesis of surface exposed adhesins are responsible for host-pathogen interaction during early stage of infection. In addition, endoflagellar structure of *L. interrogans* enabled the penetration through barrier of host connective tissues causing significant infection in human.

Pan genome of *L. interrogans* is considered open as the number of new genes predicted increase upon added genome numbers, indicating diverse biological properties of *L. interrogans* for survival and pathogenecity. The use of pan genome and pan-matrix concepts for phylogenetic analysis revealed strong sign of serovar based clustering to be applied for identification of *L. interrogans* at the serovar level.

Even though the work provided an important foundation regarding *L. interrogans* pathogenicity or virulence factors, it is also important to perform biochemical and

functional analysis on series of proteins to explain possible interactions and mechanisms that determine the severity of Leptospirosis infection. Eventually, the knowledge and understanding of genetic content and biological mechanism of *L. interrogans* obtained from this study can be applied for disease control, vaccine development, and molecular epidemiology.

In a time of globalisation with expanding worldwide travel and movement, illnesses that were believed to be detached to tropical locales can never again be considered as 'firmly contained' static disease anymore (Pappas et al., 2008). It is crucial to see new global patterns in Leptospirosis transmission as an immediate effect of globalisation. Such maladies are not rising dangers, but instead have been dangers from the start without human realisation. Leptospirosis has for a long time gone to a great extent undetected and dismissed. Even though Leptospirosis acknowledged as the most widespread zoonotic disease globally, the greatest problem of the disease still falls upon downgraded rice farming and fishing communities. They confront regular work-related exposure risks in order to earn their incomes. The new patterns in worldwide Leptospirosis transmission remains as a significant update that it is important to design and implement an effective and efficient disease detection method, treatment plan and prevention strategy starting first at every community level.

REFERENCES

Abby, S., & Daubin, V. (2007). Comparative genomics and the evolution of prokaryotes. *TRENDS in Microbiology, 15*(3), 135-141.

Abdullah, N. M., Mohammad, W. M. Z. W., Shafei, M. N., Sukeri, S., Idris, Z., Arifin, W. N., . . . Zainudin, A.-W. (2019). Leptospirosis and its prevention: Knowledge, attitude and practice of urban community in Selangor, Malaysia. *BMC Public Health, 19*(1), 628.

Adler, B., & de la Peña Moctezuma, A. (2010). Leptospira and leptospirosis. *Veterinary Microbiology, 140*(3), 287-296.

Adler, B., Lo, M., Seemann, T., & Murray, G. L. (2011). Pathogenesis of leptospirosis: The influence of genomics. *Veterinary Microbiology, 153*(1-2), 73-81.

Artiushin, S., Timoney, J., Nally, J., & Verma, A. (2004). Host-inducible immunogenic sphingomyelinase-like protein, Lk73. 5, of Leptospira interrogans. *Infection and Immunity, 72*(2), 742-749.

Athanazio, D. A., Silva, E. F., Santos, C. S., Rocha, G. M., Vannier-Santos, M. A., McBride, A. J., . . . Reis, M. G. (2008). Rattus norvegicus as a model for persistent renal colonization by pathogenic Leptospira interrogans. *Acta Tropica, 105*(2), 176-180.

Bandara, M., Ananda, M., Wickramage, K., Berger, E., & Agampodi, S. (2014). Globalization of leptospirosis through travel and migration. *Globalization and Health, 10*(1), 61.

Barocchi, M. A., Ko, A. I., Reis, M. G., McDonald, K. L., & Riley, L. W. (2002). Rapid translocation of polarized MDCK cell monolayers by Leptospira interrogans, an invasive but nonintracellular pathogen. *Infection and Immunity, 70*(12), 6926-6932.

Benacer, D., Woh, P. Y., Mohd Zain, S. N., Amran, F., & Thong, K. L. (2013). Pathogenic and saprophytic Leptospira species in water and soils from selected urban sites in peninsular Malaysia. *Microbes and Environments, 28*(1), 135-140.

Benacer, D., Zain, S. N. M., Ahmed, A. A., Khalid, M. K. N. M., Hartskeerl, R. A., & Thong, K. L. (2016). Predominance of the ST143 and ST50 Leptospira clones in the urban rat populations of peninsular Malaysia. *Journal of Medical Microbiology, 65*(6), 574-577.

Bernheimer, A. W., & Bey, R. F. (1986). Copurification of Leptospira interrogans serovar pomona hemolysin and sphingomyelinase C. *Infection and Immunity, 54*(1), 262-264.

Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., . . . Gotuzzo, E. (2003). Leptospirosis: A zoonotic disease of global importance. *The Lancet Infectious Diseases, 3*(12), 757-771.

Bierne, H., Sabet, C., Personnic, N., & Cossart, P. (2007). Internalins: A complex family of leucine-rich repeat-containing proteins in Listeria monocytogenes. *Microbes and Infection, 9*(10), 1156-1166.

Bohlin, J., Sekse, C., Skjerve, E., & Brynildsrud, O. (2014). Positive correlations between genomic% AT and genome size within strains of bacterial species. *Environmental Microbiology Reports, 6*(3), 278-286.

Böhm, N. (1982). Adrenal, cutaneous and myocardial lesions in fulminating endotoxinemia:(Waterhouse-Friderichsen Syndrome). *Pathology-Research and Practice, 174*(1-2), 92-105.

Bourhy, P., Salaün, L., Lajus, A., Médigue, C., Boursaux-Eude, C., & Picardeau, M. (2007). A genomic island of the pathogen Leptospira interrogans serovar Lai can excise from its chromosome. *Infection and Immunity, 75*(2), 677-683.

Breiner, D. D., Fahey, M., Salvador, R., Novakova, J., & Coburn, J. (2009). Leptospira interrogans binds to human cell surface receptors including proteoglycans. *Infection and Immunity, 77*(12), 5528-5536.

Brenner, D. J., Kaufmann, A. F., Sulzer, K. R., Steigerwalt, A. G., Rogers, F. C., & Weyant, R. S. (1999). Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for Leptospira alexanderi sp. nov. and four new Leptospira genomospecies. *International Journal of Systematic and Evolutionary Microbiology, 49*(2), 839-858.

Brinkman, F. S., & Leipe, D. D. (2001). Phylogenetic analysis. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, 2*, 349.

Buchanan, G. (1925). Spirochaetal jaundice. *Special Reports Series of Medical Research Council, London, 113*, 101.

Bulach, D. M., Zuerner, R. L., Wilson, P., Seemann, T., McGrath, A., Cullen, P. A., . . . Alt, D. P. (2006). Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. *Proceedings of the National Academy of Sciences, 103*(39), 14560-14565.

Carlos Guimaraes, L., Benevides de Jesus, L., Vinicius Canario Viana, M., Silva, A., Thiago Juca Ramos, R., de Castro Soares, S., & Azevedo, V. (2015). Inside the pan-genome-methods and software overview. *Current Genomics, 16*(4), 245-252.

Carvalho, E., Barbosa, A. S., Gómez, R. M., Cianciarullo, A. M., Hauk, P., Abreu, P. A., . . . Gonçales, A. P. (2009). Leptospiral TlyC is an extracellular matrix-binding protein and does not present hemolysin activity. *FEBS Letters, 583*(8), 1381-1385.

Carvalho, E., Barbosa, A. S., Gómez, R. M., Oliveira, M. L., Romero, E. C., Gonçales, A. P., . . . Ho, P. L. (2010). Evaluation of the expression and protective potential of leptospiral sphingomyelinases. *Current Microbiology, 60*(2), 134-142.

Cerqueira, G. M., & Picardeau, M. (2009). A century of Leptospira strain typing. *Infection, Genetics and Evolution, 9*(5), 760-768.

Cha, M.-K., Kim, W.-C., Lim, C.-J., Kim, K., & Kim, I.-H. (2004). Escherichia coli periplasmic thiol peroxidase acts as lipid hydroperoxide peroxidase and the principal antioxidative function during anaerobic growth. *Journal of Biological Chemistry, 279*(10), 8769-8778.

Chaudhari, N. M., Gupta, V. K., & Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Scientific Reports, 6*, 24373.

Coil, D., Jospin, G., & Darling, A. E. (2014). A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics, 31*(4), 587-589.

Croda, J., Figueira, C. P., Wunder, E. A., Santos, C. S., Reis, M. G., Ko, A. I., & Picardeau, M. (2008). Targeted mutagenesis in pathogenic Leptospira species: Disruption of the LigB gene does not affect virulence in animal models of leptospirosis. *Infection and Immunity, 76*(12), 5826-5833.

Croll, D., & McDonald, B. A. (2012). The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens, 8*(4), e1002608.

Davidson, L., Campbell, R., Rae, H., & Smith, J. (1934). Weil's Disease (Leptospirosis): A Clinical And Bacteriological Study Of Nineteen Cases Occurring Chiefly Among Fish Workers. *British Medical Journal, 2*(3859), 1137.

Ekblom, R., & Wolf, J. B. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications, 7*(9), 1026-1042.

El, I. J., & Bahaman, A. (2004). A review of human leptospirosis in Malaysia. *Tropical Biomedicine, 21*(2), 113-119.

El Jalii, I., Bahaman, A., Mohd-Azmi, M., & Mutalib, A. (2000). Occurrence of human leptospirosis in Malaysia: A retrospective study. *Tropical Biomedicine, 16*, 1-5.

Faine, S., Adler, B., Bolin, C., & Perolat, P. (1999). Leptospira and leptospirosis. *Australia MediSci*, 259.

Feigin, R. D., Anderson, D. C., & Heath, C. W. (1975). Human leptospirosis. *Critical Reviews in Clinical Laboratory Sciences, 5*(4), 413-467.

Feigin, R. D., Lobes, L. A., Anderson, D., & Pickering, L. (1973). Human leptospirosis from immunized dogs. *Annals of Internal Medicine, 79*(6), 777-785.

Fouts, D. E., Matthias, M. A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D. E., . . . Galloway, R. L. (2016). What makes a bacterial species pathogenic?: Comparative genomic analysis of the genus Leptospira. *PLoS Neglected Tropical Diseases, 10*(2), e0004403.

Fraga, T. R., Courrol, D. d. S., Castiblanco-Valencia, M. M., Hirata, I. Y., Vasconcellos, S. A., Juliano, L., . . . Isaac, L. (2013). Immune evasion by pathogenic Leptospira strains: The secretion of proteases that directly cleave complement proteins. *The Journal of Infectious Diseases, 209*(6), 876-886.

Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., . . . Hickey, E. K. (1997). Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. *Nature, 390*(6660), 580.

Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., . . . Ketchum, K. A. (1998). Complete genome sequence of Treponema pallidum, the syphilis spirochete. *Science, 281*(5375), 375-388.

Gangadhar, N. (1999). Rodents and leptospirosis: A global perspective. *Proceedings of 1st National Leptospirosis Conference* (pp. 7-14). Bangalore, India: Karnataka.

Ganoza, C. A., Matthias, M. A., Collins-Richards, D., Brouwer, K. C., Cunningham, C. B., Segura, E. R., . . . Vinetz, J. M. (2006). Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic Leptospira. *PLoS Medicine, 3*(8), e308.

Goldstein, S. F., & Charon, N. W. (1990). Multiple-exposure photographic analysis of a motile spirochete. *Proceedings of the National Academy of Sciences, 87*(13), 4895-4899.

Gravekamp, C., Korver, H., Montgomery, J., Everard, C. O., Carrington, D., Ellis, W., & Terpstra, W. J. (1991). Leptospires isolated from toads and frogs on the Island of Barbados. *Zentralblatt für Bakteriologie, 275*(3), 403-411.

Guerreiro, H., Croda, J., Flannery, B., Mazel, M., Matsunaga, J., Reis, M. G., . . . Haake, D. A. (2001). Leptospiral proteins recognized during the humoral immune response to leptospirosis in humans. *Infection and Immunity, 69*(8), 4958-4968.

Haake, D. A. (2000). Spirochaetal lipoproteins and pathogenesis. *Microbiology, 146*(7), 1491-1504.

Hartskeerl, R., Collares-Pereira, M., & Ellis, W. (2011). Emergence, control and re-emerging leptospirosis: Dynamics of infection in the changing world. *Clinical Microbiology and Infection, 17*(4), 494-501.

Hisham, B., Marzukhi, M., & Daud, A. (2009). Spectrum of flood-related diseases encountered during flood disaster in Johore, Malaysia. *Malaysian Journal of Community Health, 15*(S), 15-23.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution, 34*(8), 2115-2122.

Inada, R., Ido, Y., Hoki, R., Kaneko, R., & Ito, H. (1916). The etiology, mode of infection, and specific therapy of Weil's disease (spirochaetosis icterohaemorrhagica). *The Journal of Experimental Medicine, 23*(3), 377.

James, S., Sathian, B., Van Teijlingen, E., & Asim, M. (2018). Outbreak of Leptospirosis in Kerala. *Nepal Journal of Epidemiology, 8*(4), 745.

Johnson, M. A., Smith, H., Joseph, P., Gilman, R. H., Bautista, C. T., Campos, K. J., . . . Terry, H. (2004). Environmental exposure and leptospirosis, Peru. *Emerging Infectious Diseases, 10*(6), 1016.

Karlsen, O. A., Kindingstad, L., Angelskår, S. M., Bruseth, L. J., Straume, D., Puntervoll, P., . . . Jensen, H. B. (2005). Identification of a copper-repressible C-type heme protein of Methylococcus capsulatus (Bath). *The FEBS Journal, 272*(24), 6324-6335.

Kassegne, K., Hu, W., Ojcius, D. M., Sun, D., Ge, Y., Zhao, J., . . . Yan, J. (2013). Identification of collagenase as a critical virulence factor for invasiveness and transmission of pathogenic Leptospira species. *The Journal of Infectious Diseases, 209*(7), 1105-1115.

Kearns, D. B., Chu, F., Rudner, R., & Losick, R. (2004). Genes governing swarming in Bacillus subtilis and evidence for a phase variation mechanism controlling surface motility. *Molecular Microbiology, 52*(2), 357-369.

Ko, A., Reis, M. G., Dourado, C. R., Johnson, W., & Riley, L. (1999). Salvador Leptospirosis Study Group Urban epidemic of severe leptospirosis in Brazil. *Lancet, 354*(9181), 820-825.

Koutis, C. (2007). Special epidemiology. *Editions, Technological Educational Institute of Athens. Athens, Greece, 7*, 87-90.

Lapierre, P., & Gogarten, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends in Genetics, 25*(3), 107-110.

Lawrence, J. G., & Hendrickson, H. (2005). Genome evolution in bacteria: Order beneath chaos. *Current Opinion in Microbiology, 8*(5), 572-578.

Lee, S. H., Kim, K. A., Park, Y. G., Seong, I. W., Kim, M. J., & Lee, Y. J. (2000). Identification and partial characterization of a novel hemolysin from Leptospira interrogans serovar lai. *Gene, 254*(1), 19-28.

Lee, S. H., Kim, S., Park, S. C., & Kim, M. J. (2002). Cytotoxic activities of Leptospira interrogans hemolysin SphH as a pore-forming protein on mammalian cells. *Infection and Immunity, 70*(1), 315-322.

Lehmann, J. S., Matthias, M. A., Vinetz, J. M., & Fouts, D. E. (2014). Leptospiral pathogenomics. *Pathogens, 3*(2), 280-308.

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics, 22*(13), 1658-1659.

Liao, S., Sun, A., Ojcius, D. M., Wu, S., Zhao, J., & Yan, J. (2009). Inactivation of the fliY gene encoding a flagellar motor switch protein attenuates mobility and virulence of Leptospira interrogans strain Lai. *BMC Microbiology, 9*(1), 253.

Lim, J. K., Murugaiyah, V. A., Ramli, A. S., Rahman, H. A., Mohamed, N. S. F., Shamsudin, N. N., & Tan, J. C. (2011). *A case study: leptospirosis in Malaysia*. Retrieved on 1st April 2018 from https://www.webmedcentral.com/article_view/2764

Louvel, H., Bommezzadri, S., Zidane, N., Boursaux-Eude, C., Creno, S., Magnier, A., . . . Bouchier, C. (2006). Comparative and functional genomic analyses of iron transport and regulation in Leptospira spp. *Journal of Bacteriology, 188*(22), 7893-7904.

Matsunaga, J., Barocchi, M. A., Croda, J., Young, T. A., Sanchez, Y., Siqueira, I., . . . Haake, D. A. (2003). Pathogenic Leptospira species express surface-exposed proteins belonging to the bacterial immunoglobulin superfamily. *Molecular Microbiology, 49*(4), 929-946.

Matthias, M., & Levett, P. N. (2002). Leptospiral carriage by mice and mongooses on the island of Barbados. *The West Indian Medical Journal, 51*(1), 10-13.

McNally, D. J., Schoenhofen, I. C., Houliston, R. S., Khieu, N. H., Whitfield, D. M., Logan, S. M., . . . Brisson, J. R. (2008). CMP pseudaminic acid is a natural potent inhibitor of PseB, the first enzyme of the pseudaminic acid pathway in Campylobacter jejuni and Helicobacter pylori. *ChemMedChem: Chemistry Enabling Drug Discovery, 3*(1), 55-59.

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development, 15*(6), 589-594.

Mendoza, R. L. (2010). Leptospirosis in the tropics: When prevention doesn't easily sell as a ton of cure. *American Journal of Economics and Business Administration, 2*(3), 307.

Merhej, V., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biology Direct, 4*(1), 13.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research, 15*(12), 1767-1776.

Mira, A., Martín-Cuadrado, A. B., D'Auria, G., & Rodríguez-Valera, F. (2010). The bacterial pan-genome: A new paradigm in microbiology. *International Microbiology, 13*(2), 45-57.

Mohammed, H., Nozha, C., Hakim, K., Abdelaziz, F., & Rekia, B. (2011). Leptospira: Morphology, classification and pathogenesis. *Journal of Bacteriol Parasitology, 2*(06).

Munoz, F., & Jarquin, C. (1995). Outbreak of acute febrile illness and pulmonary hemorrhage--Nicaragua, 1995. *MMWR: Morbidity & Mortality Weekly Report, 44*(44), 841-843.

Muzzi, A., Masignani, V., & Rappuoli, R. (2007). The pan-genome: Towards a knowledge-based discovery of novel targets for vaccines and antibacterials. *Drug Discovery Today, 12*(11-12), 429-439.

Nascimento, A., Ko, A. I., Martins, E., Monteiro-Vitorello, C., Ho, P., Haake, D., . . . Oliveira, M. (2004). Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. *Journal of Bacteriology, 186*(7), 2164-2172.

Ogier, J.-C., Calteau, A., Forst, S., Goodrich-Blair, H., Roche, D., Rouy, Z., . . . Tailliez, P. (2010). Units of plasticity in bacterial genomes: New insight from the comparative genomics of two bacteria interacting with invertebrates, Photorhabdus and Xenorhabdus. *BMC Genomics, 11*(1), 568.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., . . . Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691-3693.

Pappas, G., Papadimitriou, P., Siozopoulou, V., Christou, L., & Akritidis, N. (2008). The globalization of leptospirosis: worldwide incidence trends. *International Journal of Infectious Diseases, 12*(4), 351-357.

Patti, J. M., Allen, B. L., McGavin, M. J., & Hook, M. (1994). MSCRAMM-mediated adherence of microorganisms to host tissues. *Annual Reviews in Microbiology, 48*(1), 585-617.

Paul, K., Nieto, V., Carlquist, W. C., Blair, D. F., & Harshey, R. M. (2010). The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. *Molecular Cell, 38*(1), 128-139.

Picardeau, M. (2017). Virulence of the zoonotic agent of leptospirosis: Still terra incognita? *Nature Reviews Microbiology, 15*(5), 297.

Picardeau, M., Brenot, A., & Saint Girons, I. (2001). First evidence for gene replacement in Leptospira spp. Inactivation of L. biflexa flaB results in non-motile mutants deficient in endoflagella. *Molecular Microbiology, 40*(1), 189-199.

Pinne, M., & Haake, D. A. (2009). A comprehensive approach to identification of surface-exposed, outer membrane-spanning proteins of Leptospira interrogans. *PLoS One, 4*(6), e6071.

Ptak, C. P., Hsieh, C.-L., Lin, Y.-P., Maltsev, A. S., Raman, R., Sharma, Y., . . . Chang, Y.-F. (2014). NMR solution structure of the terminal immunoglobulin-like domain from the Leptospira host-interacting outer membrane protein, LigB. *Biochemistry, 53*(32), 5249-5260.

Rao, R. S., Gupta, N., Bhalla, P., & Agarwal, S. (2003). Leptospirosis in India and the rest of the world. *Brazilian Journal of Infectious Diseases, 7*(3), 178-193.

Ratnam, S., Everard, C., Alex, J., Suresh, B., & Thangaraju, P. (1993). Prevalence of leptospiral agglutinins among conservancy workers in Madras City, India. *The Journal of Tropical Medicine and Hygiene, 96*(1), 41-45.

Read, T. D., & Ussery, D. W. (2006). Opening the pan-genomics box. *Current Opinion in Microbiology, 5*(9), 496-498.

Ren, S.-X., Fu, G., Jiang, X.-G., Zeng, R., Miao, Y.-G., Xu, H., . . . Lu, L.-F. (2003). Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. *Nature, 422*(6934), 888.

Rust, A. G., Mongin, E., & Birney, E. (2002). Genome annotation techniques: New approaches and challenges. *Drug Discovery Today, 7*(11), S70-S76.

Saier Jr, M., & García-Lara, J. (2001). The spirochetes: Molecular and cellular biology: *Horizon Scientific Press*.

Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J.-R., & Logan, S. M. (2009). The CMP-legionaminic acid pathway in Campylobacter: Biosynthesis involving novel GDP-linked precursors. *Glycobiology, 19*(7), 715-725.

Schwarz-Linek, U., Höök, M., & Potts, J. R. (2004). The molecular basis of fibronectin-mediated bacterial adherence to host cells. *Molecular Microbiology, 52*(3), 631-641.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069.

Segers, R., Van Gestel, J., Van Eys, G., Van der Zeijst, B., & Gaastra, W. (1992). Presence of putative sphingomyelinase genes among members of the family Leptospiraceae. *Infection and Immunity, 60*(4), 1707-1710.

Sharma, M., & Yadav, A. (2008). Leptospirosis: epidemiology, diagnosis, and control. *Journal of Infectious Diseases and Antimicrobial Agents, 25*(2), 93-103.

Sivashankari, S., & Shanmughavel, P. (2007). Comparative genomics-a perspective. *Bioinformation, 1*(9), 376.

Slack, A. T., Khairani-Bejo, S., Symonds, M. L., Dohnt, M. F., Galloway, R. L., Steigerwalt, A. G., . . . Smythe, L. D. (2009). Leptospira kmetyi sp. nov., isolated from an environmental source in Malaysia. *International Journal of Systematic and Evolutionary Microbiology, 59*(4), 705-708.

Slamti, L., de Pedro, M. A., Guichet, E., & Picardeau, M. (2011). Deciphering morphological determinants of the helix-shaped Leptospira. *Journal of Bacteriology*, *193*(22), 6266-6275.

Snipen, L., Almøy, T., & Ussery, D. W. (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics, 10*(1), 385.

Socolovschi, C., Angelakis, E., Renvoisé, A., Fournier, P.-E., Marie, J. L., Davoust, B., . . . Raoult, D. (2011). Strikes, flooding, rats, and leptospirosis in Marseille, France. *International Journal of Infectious Diseases, 15*(10), e710-e715.

Stamm, L., & Charon, N. (1988). Sensitivity of pathogenic and free-living Leptospira spp. to UV radiation and mitomycin C. *Applied and Environmental Microbiology, 54*(3), 728-733.

Stimson, A. (1907). Note on an organism found in yellow-fever tissue. *Public Health Reports (1896-1970)*, 541-541.

ter Huurne, A. A. H., Muir, S., van Houten, M., van der Zeijst, B. A., Gaastra, W., & Kusters, J. G. (1994). Characterization of three putative Serpulina hyodysenteriae hemolysins. *Microbial Pathogenesis, 16*(4), 269-282.

Terpstra, W. (2003). Human leptospirosis: Guidance for diagnosis, surveillance and control: *World Health Organization*.

Thayaparan, S., Robertson, I., Fairuz, A., Suut, L., & Abdullah, M. (2013). Leptospirosis, an emerging zoonotic disease in Malaysia. *Malaysian Journal of Pathology, 35*(2), 123-132.

Thompson, J. C., & Manktelow, B. (1986). Pathogenesis and red blood cell destruction in haemoglobinaemic leptospirosis. *Journal of Comparative Pathology, 96*(5), 529-540.

Victoriano, A. F. B., Smythe, L. D., Gloriani-Barzaga, N., Cavinta, L. L., Kasai, T., Limpakarnjanarat, K., . . . Coulombe, C. A. (2009). Leptospirosis in the Asia Pacific region. *BMC Infectious Diseases, 9*(1), 147.

Vke, L. (2011). Leptospirosis: a re-emerging infection. *The Malaysian Journal of Pathology, 33*(1), 1-5.

Wahab, Z. A. (2015). Epidemiology and current situation of Leptospirosis in Malaysia. *Local Authority Conference on Environmental Health, Ministry of Health Malaysia* (pp. 32-49). Wilayah Persekutuan Labuan, Malaysia: Sabah.

Wangroongsarb, P., Petkanchanapong, W., Yasaeng, S., Imvithaya, A., & Naigowit, P. (2002). Survey of leptospirosis among rodents in epidemic areas of Thailand. *J Tropical Medicine Parasitology, 25*(2), 56-58.

Wolbach, S. B., & Binger, C. A. L. (1914). Notes on a filterable spirochete from fresh water. Spirocheta biflexa (new species). *The Journal of Medical Research, 30*(1), 23.

Xu, Y., Zhu, Y., Wang, Y., Chang, Y.-F., Zhang, Y., Jiang, X., . . . Zeng, L. (2016). Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic Leptospira. *Scientific Reports, 6*, 20020.

Xue, F., Yan, J., & Picardeau, M. (2009). Evolution and pathogenesis of Leptospira spp.: lessons learned from the genomes. *Microbes and Infection, 11*(3), 328-333.

Yaakob, Y., Rodrigues, K. F., & John, D. V. (2015). Leptospirosis: recent incidents and available diagnostics–a review. *Medical Journal of Malaysia, 70*(6), 351.

Yuri, K., Takamoto, Y., Okada, M., Hiramune, T., Kikuchi, N., & Yanagawa, R. (1993). Chemotaxis of leptospires to hemoglobin in relation to virulence. *Infection and Immunity, 61*(5), 2270-2272.

Zuerner, R., Haake, D., Adler, B., & Segers, R. (2000). Technological advances in the molecular biology of Leptospira. *Journal of Molecular Microbiology and Biotechnology, 2*(4), 455-462.