Population Genetics and Genomics of Pathogenic *Leptospira*

Thesis submitted to the University of Hyderabad

for the award of

DOCTOR OF PHILOSOPHY

in Biotechnology and Bioinformatics

By

S. A. Vardhan Kishore Nalam

Reg. No. : 09LTPH08

Department of Biotechnology and Bioinformatics

School of Life Sciences

University of Hyderabad

Hyderabad-500046

INDIA

2019

University of Hyderabad

(A Central University Established In 1974 By An Act of Parliament)

Department of Biotechnology and Bioinformatics

School of Life Sciences

DECLARATION

I, **S. A. VARDHAN KISHORE NALAM**, hereby declare that the research work presented in the thesis entitled "**Population genetics and genomics of pathogenic *Leptospira***", has been carried out by me at the Pathogen Biology Laboratory, Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad, Hyderabad, under the guidance of **Prof. Dr. Niyaz Ahmed**. I also declare that this work is a bonafide research work which is free from plagiarism and has not been submitted previously in part or full for any other degree or diploma to this university or any other university or institution.

Date:

S. A. Vardhan Kishore Nalam

Reg. No.: 09LTPH08

University of Hyderabad

(A Central University Established In 1974 By An Act of Parliament)

Department of Biotechnology and Bioinformatics

School of Life Sciences

CERTIFICATE

This is to certify that the thesis entitled "**Population genetics and genomics of pathogenic *Leptospira***" submitted by **Mr. S. A. Vardhan Kishore Nalam** bearing registration number **09LTPH08** in partial fulfillment of the requirements for award of Doctor of Philosophy in the Department of Biotechnology and Bioinformatics, School of Life Sciences is a bonafide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institution for award of any degree or diploma.

Further, **the student has the following publications** before submission of the thesis for adjudication and has produced evidence for the same in the form of the reprint in the relevant area of his research

1. **Kishore Nalam,** Ahmed Ahmed, Sundru Manjulata Devi, Paolo Francalacci, Mumtaz Baig, Leonardo A. Sechi, Rudy A. Hartskeerl, Niyaz Ahmed. 2010. 'Genetic affinities within a large global collection of pathogenic *Leptospira*: Implications for strain identification and molecular epidemiology', *PLoS ONE*, 5(8), p. e12637. doi:10.1371/journal.pone.0012637 [ISSN number : 1932-6203]
    - Chapter of the dissertation where this publication appears **: Chapter 2**

2. Ahmed Ahmed, Janjira Thaipadungpanit, Siriphan Boonsilp, Vanaporn Wuthiekanun, **Kishore Nalam**, Brian G. Spratt, David M. Aanensen, Lee D. Smythe, Niyaz Ahmed, Edward J. Feil, Rudy A. Hartskeerl, Sharon J. Peacock. 2011. 'Comparison of two multilocus sequence based genotyping schemes for *Leptospira* species', *PLoS Neglected Tropical Diseases*. Edited by M. Picardeau. Public Library of Science, 5(11), p. e1374. doi:10.1371/journal.pntd.0001374 [ISSN number: 1935-2727]

- Chapter of the dissertation where this publication appears **: Chapter 2**

and

has made presentations in the following conferences

1. ZIBI summer symposium on "Global challenges of Chronic Tropical infections" held at Berlin, Germany from 18-19 June 2012. (International)
2. UoH-AS joint workshop on "Frontiers in Biological Sciences" held at University of Hyderabad, Hyderabad, India from 8-9 April, 2013. (International)
3. Annual meetings of the Society of Biological Chemists (India) on "International Conference on Genomics: Mechanism and Function" held at University of Hyderabad, Hyderabad from December 2-5, 2013. (International)

Further, the student also has passed the following courses towards the fulfilment of course work requirements for the award of Ph.D. degree

S. No.	Course code	Subject Name	Pass/ Fail
1	BT-801	Research Methodology	Pass
2	BT-802	Research Ethics and Management	Pass
3	BT-803	Lab work	Pass
4	BT-804	Biostatistics	Pass

Prof. Dr. Niyaz Ahmed	**Head**	**Dean**
Research supervisor	Dept. of Biotechnology & Bioinformatics	School of Life Sciences

Dedicated to

My Dear

Parents

ACKNOWLEDGEMENTS

In the long process of completion of this research work, it is my honor to express my deepest gratitude with great satisfaction to the people who were involved in the journey. This thesis becomes reality with the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

First and foremost, I would like to express my deepest sense of gratitude to my supervisor ***Prof.Dr.Niyaz Ahmed A. S****., for his continuous support, constant encouragement, supervision and valuable guidance throughout the tenure of this research work. His mentorship during the tenure of Ph.D. work has helped me in many ways to overcome hurdles and also to learn and understand the basic concepts of science. His encouragement and support to excel in the area of work increased optimism and helped me in my scientific maturity. Inspiration I got from him for effective planning, execution and analyzing the results of aims are always remembered and will be carried forward in my life. Words fail to express my gratitude for his patience during the time of completion of my Ph.D.*

I thank my doctoral committee members ***Prof. Manjula Sritharan*** *and* ***Prof.P.Prakash Babu****, for their constructive suggestions and critical and timely review of my projects. I am thankful to Department of Biotechnology and Bioinformatics and its present and former* ***Heads*** *for providing facilities and the infrastructure to carry out the research. Help extended by School of Life Sciences and its present and former* ***Deans*** *for providing all the facilities to school are greatly acknowledged.*

I would like to acknowledge the faculty at Department of Biochemistry and the faculty of school of life sciences, who instigated the endeavor to pursue my career in research and for making me to learn the basic concepts of life sciences during my Master's program.

I thank ***Prof. Lothar H Weiler, Dr. Torsten Semmler*** *and* ***Dr. Esther-Maria Antao*** *at Institute of Microbiology and Epizootics, Freie University, Berlin, Germany*

for their help in MLST data analysis and guidance during my visit. Help from Dr. Keith Jolley, curator of PubMLST database, for hosting our MLST scheme online is greatly acknowledged.

I would like to thank ***Dr. Ahmed Ahmed*** *and* ***Prof. Rudy A. Hartskeerl*** *for providing bacterial genomic DNA. Suggestions and help received from* ***Dr.P.Vijayachari*** *and* ***Dr. Mumtaj Baig*** *for analyzing the MLST data is highly appreciated.*

I am grateful to my senior ***Dr. Sundru Manjulata Devi*** *for sharing all MLST data. I am extremely thankful to all my lab mates at PBL for making the environment always competitive, cheerful and lively. It is my pleasure mentioning the support and friendship I have received from Nishant, Ashutosh, Shivendra, Suhail, Arif, Sandhya, Savita, Priya, Hritha and Mallika. Their involvement in designing and discussion of projects, and support during the initial days my research is remembered forever. It would be always happy to acknowledge the support of NGS team members in the lab Narender, Ramani, Sabiha, Suma, Aditya and Kshitij for their valuable discussions and helping in writing scripts for easy analysis. Lively atmosphere created by other members of the lab Vidyuallatha, Shankar, Amit, Shamshul, Arya and Sumeet is always remembered. I thank Arun for his assistance in the lab.*

I am extremely thankful for the motivation received from my seniors Dr. Usha, Dr. Roy, Dr. Sunil, Dr. Chaitanya and Dr. Bhaskar.

It is high time to acknowledge the crucial members of my journey namely Bimolata, Jyothi, Chaitanya, Debosri, Shyama, Prasuna, Sarada, Saritha, Malati, Veronica, Subha, Gouse, Venkat, Madhu Babu, Sumit, Rajesh, Anil, Surendra, Maruthi, Kamusha, Srinivas, Sunil, Ashok, Madhu Prakash, Sugith, Ravi, Chandu, Paparao, Kumar, Satish, Vijay, Naveen and all unsung friends of SLS for making my stay at University of Hyderabad a life time memorable one.

I am thankful to all my Master's batch mates from school of life sciences for intellectually inspiring my research career. I am extremely thankful for all my JNV school friends for making my life happy all the time by sharing nostalgic moments.

I would like to acknowledge the fellowship received from CSIR. The funding support received from CSIR, DBT, ICMR, UM-HIR, FU-ZOO, IRTG-GRK 1673, DBT-BIF, DBT-CREBB, DST-FIST, UGS-SAP and UPE to the lab, department and school are highly acknowledged.

I thank all the administrative and supporting non-teaching staff of school and departments for their help and assistance.

Special thanks to my brothers ***Kumar*** *and* ***Sridhar*** *for their support and encouragement. Love and support I received from my family members and extended family members is highly acknowledged and without that I would not have reached this height.*

And I could not end my acknowledgements without the mention of my life partner and best friend of my life ***Narmadha*** *for all her encouragement, understanding, patience and support all throughout the journey. Cheerfulness and lovely smile on my little hero* ***Reyansh*** *always makes my life.*

INDEX

Abbreviations vii

List of Figures xi

List of Tables xiv

General introduction 1

1. Chapter 1 – Review of Literature

1.1 Leptospira 5

1.2 Leptospirosis 7

1.3 Epidemiology of leptospirosis 7

1.4 Mode and source of transmission 11

1.5 Clinical presentation 12

1.6 Pathogenesis 15

1.7 Virulence Factors 16

1.8 Diagnosis of leptospirosis 17

1.9 Treatment and Management 22

1.10 Characterization 23

1.11 *Leptospira* Genomics 33

1.12 Rational and objectives of the study 36

2. Chapter 2 - Evaluation of MLST for epidemiology of pathogenic *Leptospira* isolates and creation of publicly accessible *Leptospira* MLST database

2.1 Introduction 39

2.2 Materials and Methods 45

2.2.1 Bacterial strains and genomic DNA samples 45

2.2.2 Gene Loci, PCR conditions and nucleotide sequences 60

2.2.3 Nucleotide sequences and allele designation 62

2.2.4 Sequence retrieval from whole genome sequences 63

2.2.5 Sequence analysis 64

2.2.6 Phylogenetic reconstruction by Bayesian MCMC approach 65

2.2.7 Network analysis for Indian sub-continent isolates 65

2.2.8 Whole genome sequence based phylogeny 66

2.3 Results 67

2.3.A Evaluation of Ahmed's MLST scheme for global applicability 67

2.3.A.1 Nucleotide diversity of genetic loci 67
2.3.A.2 Phylogeny inferred from MLST data 68
2.3.A.3 Distribution of pathogenic *Leptospira* isolates 72
2.3.A.5 Network analysis of MLST data 74
2.3.B Creation of MLST database for Ahmed *et al* (2006) scheme 75
2.3.C Comparison of Ahmed's and Thaipadungpanit's MLST schemes 80
2.3.C.1 Nucleotide diversity of genetic loci 80
2.3.C.2 Phylogenetic relationships 82
2.3.D Comparison of MLST schemes with WGS inferred phylogeny 82
2.3.D.1 Clonality among Brazil isolates 84
2.3.D.2 Phylogenetic relationships inferred by whole genomes and MLST 84
2.4 Discussion 93
2.4.1 Ahmed's MLST scheme for global applicability of *Leptospira* strain typing 93
2.4.2 MLST as the gold standard for *Leptospira* typing 95
2.4.3 Genetic affinities within large global collection 96
2.4.4 *Leptospira* MLST database 96
2.4.5 Comparison of MLST schemes 96
3. Chapter 3 - Genomic analysis of pathogenic Leptospira isolates
3.1 Introduction 101
3.1.1 Pathogenomics of *Leptospira* 101
3.1.2 Whole genome sequencing by NGS platforms 102
3.1.3 The pan-genome 103
3.1.4 Comparative genomics 104
3.2 Materials and Methods 105
3.2.1 Genomic DNA for sequencing 105
3.2.2 Whole-Genome sequencing 105
3.2.3 Quality assessment of sequencing data 107
3.2.4 N base Filtration 107
3.2.5 Genome assembly and annotation 108
3.2.6 Submission of genomes to NCBI 109
3.2.7 Whole genome sequences 109
3.2.8 Genomic Islands and mobile genetic elements detection 109
3.2.9 Pan-genome and core genome analysis 115
3.2.10 Pathogen specific genes 116

3.2.11 COG functional classification 116
3.2.12 Lipoproteins and secretary proteins detection 117
3.3 Results 118
3.3.1 Quality analysis of sequencing 118
3.3.2 Genome Assembly and annotation 120
3.3.3 Genome submission 120
3.3.4 Genomic Island prediction 121
3.3.5 Pan and core genome analysis 122
3.3.6 Common core genome of *Leptospira* pathogenic species 124
3.3.7 Pathogen specific genes 124
3.3.8 Functional classification of proteins 125
3.3.9 Analysis of hypothetical protein 125
3.4 Discussion 137
4.0 Chapter 4 – Summary and Outlook 141
References 145
Publications 164

ABBREVIATIONS

°C	Degree Celsius
α	Alpha
β	Beta
γ	Gamma
μg	Microgram(s)
μl	Microliter(s)
μm	Micrometer(s)
ANI	Average Nucleotide Identity
ATP	Adenosine Tri-phosphate
BAPS	Bayesian Analysis for Population Structure
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
CAAT	Cross-Agglutinin Absorption Test
CF	complement fixation test
COG	Cluster of Orthologous Groups
CSF	Cerebrospinal fluid
DFM	Dark-field microscopy
DNA	Deoxyribonucleic Acid
dNTP	Deoxy nucleoside Phosphate
dsDNA	Double Stranded DNA
EDTA	Ethylene diamine tetra acetic acid
ELISA	Enzyme-Linked Immunosorbent Assay
ELISA	Enzyme Linked Immuno sorbent assay
EMJH	Ellinghausen-McCullough-Johnson-Harris

FAO	The Food and Agriculture Organization of the United Nations
FlaA	Flagellar protein A
FlaB	Flagellar protein B
GC	Guanine and Cytosine
HGT	Horizontal gene transfer
IFA	Indirect Immunofluorescent Antibody
Ig	Immunoglobulins
IM	Inner membrane
IS	Insertion Sequence
Kb	Kilo base pairs
LAMP	Loop-Mediated Isothermal Amplification
Lig	*Leptospiral*-immunoglobulin-like gene
LPS	Lipopolysaccharide
L. interrogans	*Leptospira interrogans*
L. borgpetersenii	*Leptospira borgpetersenii*
L. santarosai	*Leptospira santarosai*
L. noguchii	*Leptospira noguchii*
L. weilii	*Leptospira weilii*
L. kirschneri	*Leptospira kirschneri*
L. alexanderi	*Leptospira alexanderi*
L. alstonii	*Leptospira alstonii*
L. kmetyi	*Leptospira kmetyi*
MAT	Microscopic Agglutination Test
Mb	Mega base pairs
MCAT	Microcapsule Agglutination Test
MCL Algorithm	Markov Cluster Algorithm

MCMC	Morkov Chain Monte Carlo
MEE	Multilocus Enzyme Electrophoresis
MEGA	Molecular Evolutionary Genetics Analysis
mg	Milligram(s)
MLST	Multi Locus Sequence Typing
MLVA	Multiple-Locus Variable number of tandem repeat Analysis
NASBA	Nucleic Acid Sequence-Based Amplification
NCBI	National Centre for Biotechnology Information
ng	Nano gram(s)
NGS	Next Generation Sequencing
NJ Algorithm	Neighbor Joining Algorithm
OIE	The World Organization for Animal Health
OMP	Outer Membrane Protein
ORF	Open reading frame
PAUP	Phylogenetic Analysis Using Parsimony
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFGE	Pulsed-field gel electrophoresis
pg	Pico gram(s)
QC	Quality Control
RAPD	Randomly Amplified Polymorphic DNA
REA	Restriction-endonuclease DNA analysis
RefSeq	Reference Sequence
REP	Repetitive Extra genic Palindromic Sequences
RFLP	Restriction Fragment Length Polymorphism
RFU	Relative Fluorescence Units

RNA	Ribonucleic Acid
rpm	Revolutions per minute
RPS BLAST	Reversed Position Specific BLAST
rRNA	Ribosomal RNA
SAM	Sequence Alignment/Map
SNPs	Single Nucleotide Polymorphisms
Sps	Species
S.S.	Sensu stricto
ST	Sequence Type
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UV	Ultraviolet
VFDB	Virulence Factor Database
VNTR	Variable Number Tandem Repeats
WGS	Whole Genome Sequencing
WHO	The World Health Organization

LIST OF FIGURES

Figure 1A: High resolution scanning electron micrograph of *Leptospira interrogans*

Figure 1B: Schematic representation of membrane architecture of *Leptospira*

Figure 2: Worldwide burden of leptospirosis in terms of morbidity

Figure 3: Transmission cycle of leptospirosis

Figure 4: The clinical course of leptospirosis

Figure 5: Phylogenetic tree (unrooted) generated with 16S rRNA gene sequences of *Leptospiraceae* family

Figure 6: Venn diagram showing the comparison of pathogenic, intermediate and saprophytic counterparts of *Leptospira* genus

Figure 7: Overview of salient features of Ridom SeqSphere software used for designating allelic profiles for isolates

Figure 8: Genetic clustering of global *Leptospira* isolates based on MLST analysis. Phylogenetic tree constructed using NJ algorithm in MEGA

Figure 9: Bayesian inference tree showing the phylogeny, generated from MLST data with relaxed molecular clock approach in Bayesian MCMC analysis using BEAST v1.5.2

Figure 10: Median-Joining network analysis using Network v 4.5.0.0 for the MLST data obtained from Indian sub-continent subsample of *L. interrogans*.

Figure 11: Phylogenetic tree generated with altered concatenated sequences for Ahmed's MLST scheme using NJ algorithm in MEGA 5.0 with 1000 bootstrap values

Figure 12: PubMLST database mentioning MLST scheme description and its curators

Figure 13: *Leptospira* locus/sequence definition database at PubMLST

Figure 14: Isolates database at PubMLST

Figure 15: Interface for submission of new data for curation by curators at PubMLST

Figure 16: NJ tree generated for 46 strains using Thaipadungpanit's scheme (A) and Ahmed' scheme (B) MLST data

Figure 17: Heat-plot generated based on BLASTN similarity scores obtained from fragmented alignment of whole genomes of Fiocruz isolates belonging to *L. interrogans* Serovar Copenhageni using Gegenees software

Figure 18: Heat-plot generated based on BLASTN similarity scores obtained from fragmented alignment of whole genomes of *Leptospira* isolates using Gegenees software

Figure 19: NJ phylogenetic tree generated from whole genome similarity scores for *Leptospira* isolates obtained using Gegenees software

Figure 20: NJ trees generated using whole genome based similarity scores (A) and MLST profiles of Boonsilp's scheme (B), Varni's scheme(C) and Ahmed's scheme (D)

Figure 21: NJ trees, showing *L. interrogans* (Li) species cluster, generated from whole genome based similarity scores(A) and MLST profiles of Boonsilp's scheme(B), Varni's scheme(C) and Ahmed's scheme(D)

Figure 22: NJ trees, showing *L. weilii* (Lw), *L. borgpetersenii* (Lb), *L. santarosai* (Ls), *L. kirschneri* (Lk) *and L. noguchii*(Ln) species clusters, generated from whole genome based similarity scores(A) and MLST profiles of Boonsilp's scheme(B), Varni's scheme(C) and Ahmed's scheme(D)

Figure 23 A: Schematic representation of whole genome sequencing method

Figure 23 B: Pipeline of steps involved in sequencing of whole genome using Illumina Platform

Figure 24: FASTq format of the read sequence output from sequencer

Figure 25: Read quality graphs showing average Phred quality score at each base position for NASR61 strain Read 1 (A) and Read 2 (B), and NAMuis5 strain Read 1 (C) and Read 2 (D)

Figure 26: Graphical representation of species wise pan (blue) and core (green) genomes

Figure 27: Graphical representation of number of new gene clusters being identified upon addition of a new genome to the pan-genome analysis

Figure 28: Venn diagram showing the core genome size of individual species and common core genome of pathogenic *Leptospira* species

Figure 29: COG Functional classification of core genomes of *Leptospira* species and common core genome of pathogenic species

Figure 30 A: Functional classification of common core genome of pathogenic *Leptospira* species

Figure 30 B: Functional classification of pathogen specific gene content of *Leptospira*

LIST OF TABLES

Table 1: *Leptospira* species classification using DNA-DNA hybridization method

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme for global applicability

Table 3: Details of *Leptospira* isolates employed for comparing Ahmed's and Thaipadungpanit's scheme

Table 4: Details of *Leptospira* isolates used for comparison of *Leptospira* MLST schemes with WGS based phylogeny

Table 5: Details of Fiocruz isolates belonging to serovar Copenhageni

Table 6: Details of MLST loci and primers used for amplification

Table 7: Allelic diversity of the MLST loci calculated using START2 software

Table 8: Distribution and phylogenetic affiliation of *Leptospira* isolates

Table 9: Altered locus size considered for Ahmed's scheme

Table 10: Discriminatory ability of Thaipadungpanit's and Ahmed's schemes

Table 11: Details of whole genome sequence of *Leptospira* isolates

Table 12: Species wise, host wise and geographic distribution of whole genomes

Table 13: Quality statistics of sequencing for both *Leptospira* isolates

Table 14: Statistics of genome assembly performed using Velvet with different hash lengths for both the strains.

Table 15: Genome statistics of whole-genome assembly

Table 16: Island viewer predicted genomic regions and their major proteins in *Leptospira interrogans* genome

Table 17: Pan and core genome analysis of pathogenic *Leptospira* species

Table 18: Details of pathogen specific gene content for *Leptospira*

General Introduction

Leptospira, belonging to class spirochetes of eubacteria, is the causative agent of disease leptospirosis in both humans and animals. Leptospirosis is a zoonotic disease transmitted to humans by direct contact with infected animal urine or tissue or by indirect contact via contact with contaminated water or soil and wild animals which are natural hosts and acts as carriers of *Leptospires*. Leptospirosis incidence is higher in tropical regions which receives higher rainfalls (Levett, 2001). Its burden has increased drastically in past decades worldwide and epidemic outbreaks were also reported from different parts of India (John, 2005).

Disease manifestations of leptospirosis differs in humans and animals, and their effect on live veterinary stock is of utmost importance because it results in devastating stillbirths. Leptospirosis in humans is characterized by a large array of symptoms ranging from mild febrile illness to severe form of the disease known as Weil's syndrome principally characterized by jaundice, renal failure, conjunctival suffusion and multi organ failure culminating in increased morbidity incidents (Bharti *et al.*, 2003).

Diagnosis of leptospirosis, based on the disease outcome is hindered because of wide array of symptoms that exactly overlap with the other commonly occurring disease outbreaks such as dengue fever, viral hepatitis and malaria (Levett, 2001). Microscopic agglutination test (MAT) based on circulating anti-*Leptospira* antibodies in blood is widely accepted and considered as gold standard for the detection of *Leptospira* infection. However, MAT assay requires maintenance of large panel of live reference strains and expertise to perform and interpret results (Faine, 1999). Moreover, the anti *Leptospira* antibodies in blood can be detected only after 4-7 days of infection. Therefore, early diagnosis of the infection is indispensable for proper treatment regimen and control over the disease severity. Although there are some commercially available diagnostic kits based on IgM antibodies, for the detection of causative agent in circulation, they lack sensitivity and specificity (Rajapakse *et al.*, 2015). Treatment of

leptospirosis in humans is majorly symptomatic and is generally supported by broad spectrum antibiotics like doxycycline, ampicillin and amoxicillin (Levett, 2001).

Leptospirosis can be caused by more than 7 different species with in genus *Leptospira* and proper classification of bacteria can help in clearly deciphering the cause and source of infection to effectively tackle the disease transmission and mitigation. Hence, characterization of *Leptospires* will potentially help in understanding the epidemiology and surveillance of the disease and to develop public health interventions like immunization campaigns that play important role in treatment and control of this zoonotic agent (Zuerner *et al.*, 2000).

Leptospira genus was classified into more than 20 species of pathogenic and saprophytic using different molecular biology tools. Advancement in the Molecular Biology has paved ways to design and develop methods for diagnosis and characterization of *Leptospira*. A wide array of typing methods were developed using different techniques such as polymerase chain reaction (PCR), sequencing, genome polymorphisms, repetitions, DNA hybridization, southern blot, restriction digestion and pulsed field gel electrophoresis (PFGE). Characterization of *Leptospira* using these methods have helped type the organism to species level and in some cases to sub species level (Ahmed, and P. Grobusch, 2012). Multilocus sequence typing (MLST) is one such method that has received wider acceptance for diagnosis and typing of several bacterial species (Urwin and Maiden, 2003). Reports have shown that use of MLST for characterizing *Leptospira* isolates is simple and effective, without the need of tedious and laborious typing methods (Ko, Goarant and Picardeau, 2009)

With the advancement of whole genome sequencing technologies, a large number of *Leptospira* isolates were sequenced and genomes of multiple isolates of different species were deposited in public domain. Availability of these genomes has opened up ways to visualize the genetic makeup of the organism providing novel insights for development of new research strategies, validation of molecular typing methods and identification of possible

virulent and diagnostic markers, recombination events and specific genomic islands present in pathogenic isolates (Lukjancenko, Wassenaar and Ussery, 2010). Understanding the genetic relationship with the pathogenicity and their ability to survive in diverse environmental niches is also very critical to decipher the potential pathogenicity of *Leptospira*.

Hence with different typing methods available, and humongous data in the public domain, the primary goal of the present study is to validate a molecular typing method, Multilocus sequence typing (MLST), developed by Ahmed *et al*., 2006, for typing *Leptospira* isolates. This objective includes extensive validation of MLST scheme against other prevalent typing tools and other MLST schemes available for typing *Leptospira,* and hosting the Ahmed *et al* MLST scheme in the public database for easy access by research fraternity. The secondary aim is to decipher pathogenic characteristics encoded by genomes of *Leptospira* isolates. This objective includes whole genome sequencing of two pathogenic *Leptospira* isolates and comparison of multiple pathogenic and saprophytic genomes of the genus in order to identify pathogen specific genomic entities that can be used as potential targets for efficient diagnosis and vaccine candidates for *Leptospira*.

Chapter 1

Review of Literature

1.1. *Leptospira*

Leptospira is a diverse eubacterial genus belonging to the family *Leptospira*ceae in the order Spirochetales and is divided into 20 species based on the DNA relatedness studies (Yasuda *et al.*, 1987; Brenner *et al.*, 1999; Smythe *et al.*, 2013). Genus *Leptospira* consists of saprophytic and pathogenic bacteria. Saprophytes are non-pathogenic and lives in soil. Whereas pathogenic bacteria requires maintenance host for their survival and propagation.

1.1.1 Taxonomy

Domain	:	Bacteria
Phylum	:	Spirochetes
Class	:	Spirochetes
Order	:	Spirochetales
Family	:	*Leptospira*ceae
Genus	:	*Leptospira*

1.1.2 Morphology

The name *Leptospira* is derived from two Greek words "Leptos" meaning "fine" and "spira" meaning a "coil". *Leptospires* are right handed spiral-shaped spirochetes with 0.25 μm in diameter and 6-20 μm in length with one or more distinctive hooked thick ends (**Figure 1A**).

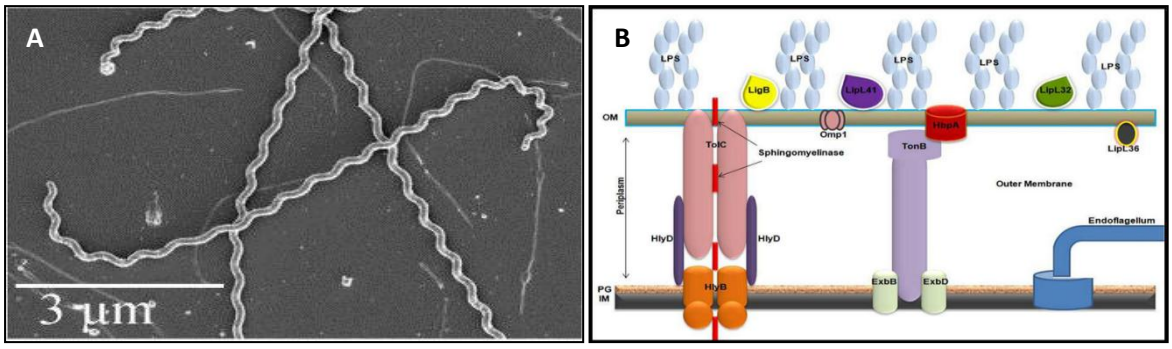

Figure 1: (A) **High resolution scanning electron micrograph of *Leptospira interrogans***. Spiral shaped bacteria with characteristic hooked ends is seen (Image adopted from (Stewart *et al.*, 2012))
(B) **Schematic representation of membrane architecture of *Leptospira*** showing the distribution of major cytoplasmic membrane and outer membrane proteins. Association of lipopolysaccharide (LPS) with outer membrane (OM) and peptidoglycan (PG) with inner membrane (IM) is shown (Image adapted from (Sritharan et al., 2012))

It has two internal flagella situated periplasmically and arising from sub terminally placed basal bodies at each end. *Leptospires* are highly motile with typical cork-screw motility and rotates along its longitudinal axis (Bharti *et al.*, 2003).

Leptospira is characterized by a double membrane structure with cytoplasmic membrane and peptidoglycan cell wall covered by outer membrane (**Figure 1B**). Although *Leptospira* is a Gram negative bacteria, it shares features of both Gram positive and Gram negative bacteria. Membranes separated by periplasmic space gives Gram negative nature whereas attachment of peptidoglycan cell wall to the inner membrane gives Gram positive nature. Staining of the organism with general aniline based dyes is of no use because of its thin structure and are best stained by silver impregnating dyes and by artificial thickening using immunoperoxidase (Faine, 1999). Lipopolysaccharide (LPS) of outer membrane constitutes the major antigenic repertoire for *Leptospira* (Levett, 2001; Cullen, Haake and Adler, 2004)

1.1.3 Culture characteristics

Leptospires are obligate aerobes and grow in microaerophilic conditions with an optimum growth temperature of 28ºC to 30ºC and optimum pH of 7.2 to 7.6. (Faine, 1999; Levett, 2001). The unique nutritional requirement of *Leptospires* makes this organism fastidious and the organism depends on the β-oxidation of long chain fatty acids as sole source of carbon (Henneberry and Cox, 1970). *Leptospires* are often cultured in Ellinghausen McCullough medium modified by Johnson and Harris (EMJH) containing 10% rabbit serum or 1% bovine serum albumin with long chain fatty acids (Ellis and Michno, 1976; Faine, 1999). Both liquid and semi solid media are used for culturing bacteria along with addition of 5-fluorouracil and other antibiotics like rifampicin, amphotericin or neomycin to inhibit contaminants. Growth of *Leptospira* is very slow in cultures and has to be retained for at least 3 months before discarding to confirm the organism presence or absence.

1.2 Leptospirosis

Leptospirosis is a neglected tropical infectious disease and is considered as an emerging zoonotic disease caused by pathogenic species of the genus *Leptospira*. This spirochetal disease is characterized by the involvement of multiple organs and is considered as a major health problem in developing countries particularly in tropical and subtropical regions with higher rainfall (Bharti *et al.*, 2003). *Leptospira* is transmitted from animals to humans directly or indirectly. Leptospirosis pose major health challenge to the impoverished populations of developing countries, especially to farmers and daily wage workers who reside in poor hygienic environments. Incidences of infection are also reported from urban areas with increasing slums and decreased sanitation (Karande *et al.*, 2005; Haake and Levett, 2015).

Leptospirosis was discovered way back in 1883 by Landouzy and its severe icteric nature of infection was described by Adolf Weil in 1886 in Heidelberg with particular type of jaundice accompanied by splenomegaly, renal dysfunction, conjunctivitis and skin rashes, naming the disease after his name as Weil's disease (Vinetz *et al.*, 1996). Causative agent of leptospirosis was first isolated from sera of a patient suffering from Weil's disease in 1917 by Inada and Noguchi in Japan (Kobayashi, 2001). Although leptospirosis was first described in 1886, it was first reported officially in India in the early 20th century from patients suffering from Andaman hemorrhagic fever (AHF), which used to be considered as mysterious disease till that time (Farr, 1995; Vijayachari, Sugunan and Shriram, 2008).

1.3 Epidemiology of leptospirosis

1.3.1 Disease burden

Estimation of global prevalence of the infection is always ambiguous and less reported because of lack of awareness, absence of global surveillance programs and improper diagnostic tools. Occurrence of leptospirosis is highly associated with wet seasons with increased incidents during floods, hurricanes and rains. Disease incidence are quite high in

people who are involved in fishing, farming, mining and sanitation who might have regular contact with contaminated water sources and also in persons involved in animal exposure like veterinarians, animal handlers, hunters and workers in slaughter houses. Increasing incidences of infection were also reported from people involved in water based sports and recreational activities (Levett, 2001; Bharti *et al.*, 2003; Karande *et al.*, 2005; Abela-Ridder, Sikkema and Hartskeerl, 2010; Adler and de la Peña Moctezuma, 2010).

Incidences of regular outbreaks were reported majorly from tropical countries with higher rainfall than the temperate regions attributing to the survival of bacteria in warm and humid conditions for longer times in the environment. Indian subcontinent, Latin America, Caribbean islands, Oceania and Southeast Asia are considered as the most significant foci for leptospirosis (**Figure 2**) (Bharti *et al.*, 2003; Costa *et al.*, 2015). Before discovering the causative agent of leptospirosis, outbreaks were reported in ancient times in the name of rice field jaundice, autumn fever, yellow fever, seven days fever, cane cutter's fever, swine herd's disease, mud fever and Andaman hemorrhagic fever, naming the disease after the possible source of infection (Kobayashi, 2001; Vijayachari, Sugunan and Shriram, 2008).

Despite lack of rapid diagnostics, more than 500,000 cases of severe leptospirosis were reported worldwide every year with about 10% fatality rate (Vinetz *et al.*, 1996; World Health Organisation, 2003). The Leptospirosis Burden Epidemiology Reference Group (LERG) established by World health organization (WHO) for reporting global prevalence of the disease, estimated that the mean global burden of endemic leptospirosis to be 5 per 100,000 people. As these numbers does not include epidemic leptospirosis and are based only on severe and scarcely notified cases, this is considered as an underestimated number (WHO report, 2011).

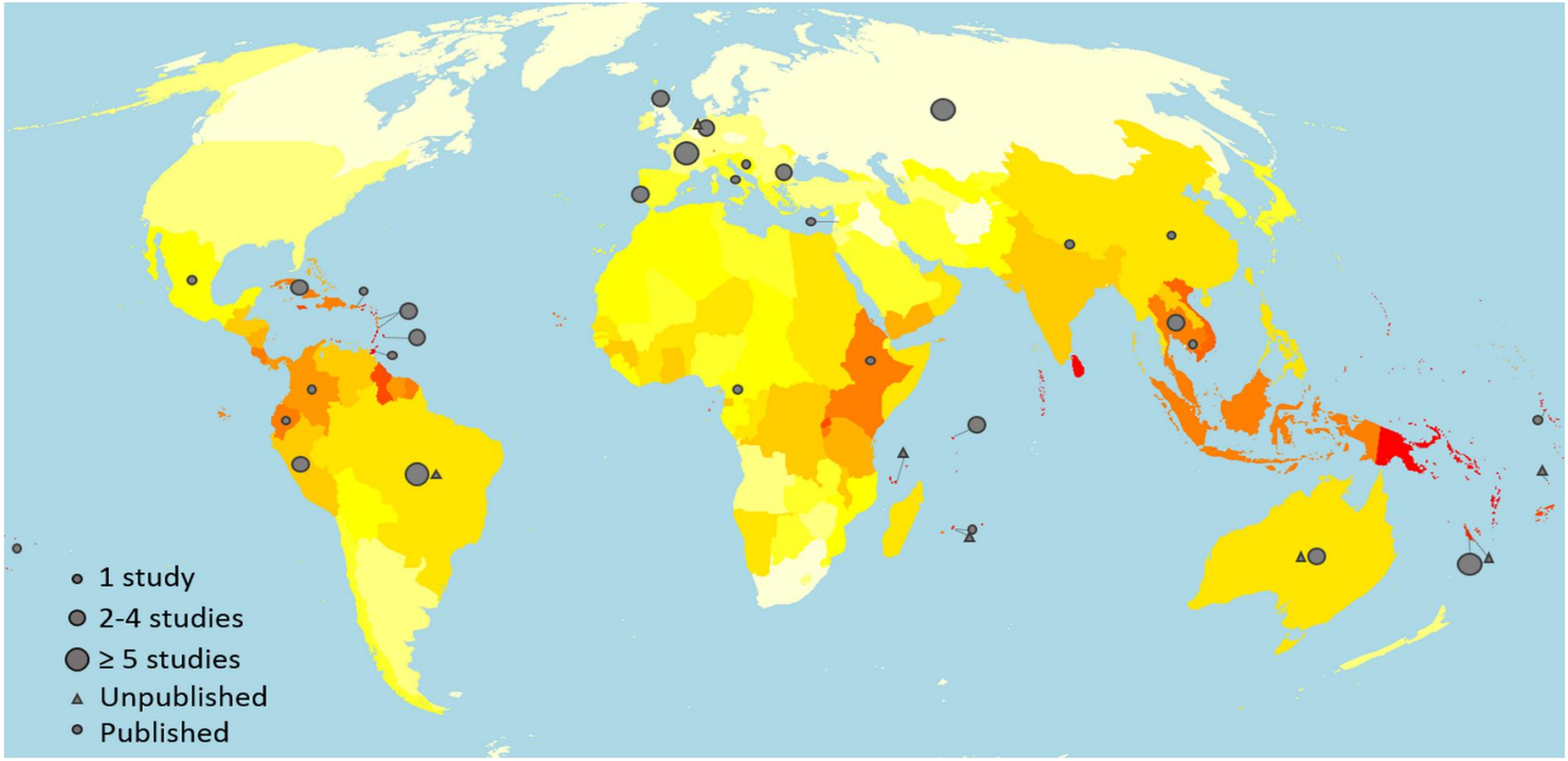

Figure 2: Worldwide burden of leptospirosis in terms of morbidity. Annual disease incidence is represented as an exponential color gradient from white (0–3), yellow (7–10), orange (20–25) to red (over 100), in cases per 100,000 population. Circles and triangles indicate the countries of origin for published and grey literature quality-assured studies, respectively. (Picture adopted from (Costa *et al.*, 2015).

1.3.2 Leptospirosis in India

After it was first reported in 1926 from Andaman and Nicobar islands, leptospirosis outbreaks were reported from different parts of the country and is recognized as increasingly infectious disease in the recent times (Sehgal, 2006). Taylor and Goyle (1931) were the first to report the etiology of *Leptospira* in patients suffering from mysterious fever with jaundice popularly known as Andaman Hemorrhagic Fever (AHF) (Barker, 1926). In 1988, an endemic leptospirosis outbreak was reported with disease symptoms related to Weil's disease (Vijayachari *et al.*, 2015). Outbreaks were repeatedly reported from Andaman and Nicobar Islands and also the coastal states of main land India especially Tamil Nadu, Karnataka, Kerala, Gujarat, Orissa and Maharashtra. Approximately 2000 cases of disease were reported with fatality rates from 0.7 to 13.9% with increasing incidents even in northern India confirming its presence pan-India (Sehgal, 2006; Shivakumar, 2008; Sethi *et al.*, 2010; Routray *et al.*, 2018).

In addition to these, a large number of cases of pyrexia with unknown origin during rainy season and floods, in many parts of India were also considered to be due to *Leptospira* infection imputing to the lack of proper diagnosis for confirmation (Ramakrishnan *et al.*, 2003). Kerala is the major state which has reported higher incidence of leptospirosis mainly with hepato-renal complications in areas surrounding Trivandrum, Kottayam, Alleppey and Kozhikode (Kuriakose *et al.*, 2008). According to the India's National Center for Disease Control reports, there were total 3326 cases of suspected leptospirosis with 1711 confirmed ones for the year 2018 and responsible for the death of 70 persons. This was reported to be the devastating effect of sudden floods in the state.

The Government of India has also initiated a programme for prevention and control of leptospirosis under its XII five year plan to implement in the states of Maharashtra, Gujarat, Kerala, Karnataka, Tamilnadu and Andaman and Nicobar Islands with an aim to strengthen

the diagnostics, inter sectoral coordination, patient management facilities, training manpower and creating awareness in general community (John, 2005; *National Guidelines Diagnosis, Case Management Prevention and Control of Leptospirosis Programme for Prevention and Control of Leptospirosis*, 2015).

There were also reports from the southern districts of Tamilnadu that an increase in the number of cases of *Leptospiral* Uveitis were reported at Arvind eye hospital, Madurai (Chu *et al.*, 1998; Rathinam, 2002; Priya *et al.*, 2003). Owing to the high population density and deteriorating sanitation conditions, leptospirosis has become a major health hazard for the Indian population. It is considered that lack of reports of leptospirosis cases in other parts of the country is merely just because of lack of proper knowledge, awareness among health care workers and robust diagnostics for detection of *Leptospira*.

1.4 Mode and source of transmission

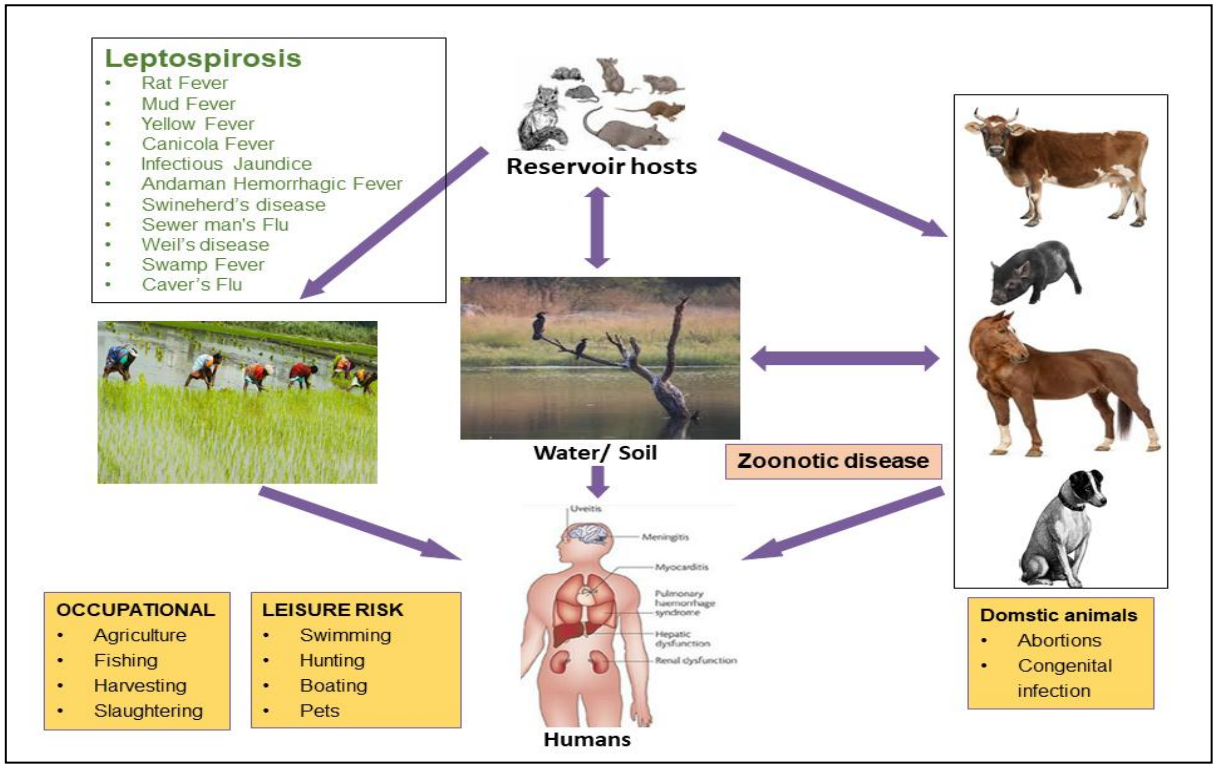

Figure 3: Transmission cycle of leptospirosis. Wild and domestic animals act as carriers and humans are accidental hosts for the *Leptospira* infection. (Leptospirosis known by different names based on geographical location is also enlisted)

In maintenance hosts, pathogenic *Leptospira* live and multiply in immune compromised organs like urogenital tract and proximal renal tubules (Leonard *et al.*, 1992). Maintenance hosts, majorly mammalian species comprising whole range of animals worldwide, shows no signs of infection and act as reservoirs. These animals can excrete the pathogenic *Leptospira* through urine for their entire lifetime and contaminate the environment, thereby serving as a source for infection (Faine, 1999). *Leptospira* transmission to humans is always accidental and occurs when pathogenic organism enters into the body through cuts and aberrations present on the skin or through mucous membrane such as oral, conjunctival or genital surfaces of the individual (**Figure 3**). Disease transmission is either by direct contact with tissue, blood, body fluids, urine of the infected or carrier animal, or indirectly when contacted with *Leptospira* contaminated sources like soil or water (Faine, 1999; Bharti *et al.*, 2003; Haake and Levett, 2015). Rarely, reports of transient *Leptospir*a shedding from urine of humans and human-human transmission during sexual intercourse and lactation were also reported (Bolin and Koellner, 1988; Harrison and Fitzgerald, 1988).

1.5 Clinical presentation

1.5.1 Human leptospirosis

Clinical manifestation of the disease varies in humans and animals by a wide array of symptoms from mild anicteric presentation to severe icteric presentation. In humans, classical presentation of the disease is an acute biphasic febrile illness with an incubation period of 5-14 days (Feigin and Anderson, 1975; Haake and Levett, 2015). Initial bacteremia phase or septicemic phase, which lasts for 4-7 days, is manifested by febrile illness with clinical symptoms such as high fever, headache, nausea, vomiting, anorexia, conjunctival suffusion, chills, severe myalgia, prostration and vomiting (Katz *et al.*, 2001). *Leptospira* during this phase can be detected in blood stream and cerebrospinal fluid. Bacteremia phase is then followed by the immune phase during which bacteria is cleared from blood but remains persistent in immune compromised sites such as cerebrospinal fluid, eyes and kidneys. This phase lasts for

days to weeks and is characterized by elicitation of antibody and excretion of *Leptospira* in urine. Based on the involvement of organs in the immune phase, leptospirosis is categorized into Anicteric leptospirosis and Icteric leptospirosis (**Figure 4**). Ocular involvement in the form of Uveitis is reported during the convalescent phase of leptospirosis, nearly after 2-6 months of onset of disease (Rathinam, 2002). Leptospirosis during pregnancy has been associated with fetal loss (Coghlan and Bain, 1969).

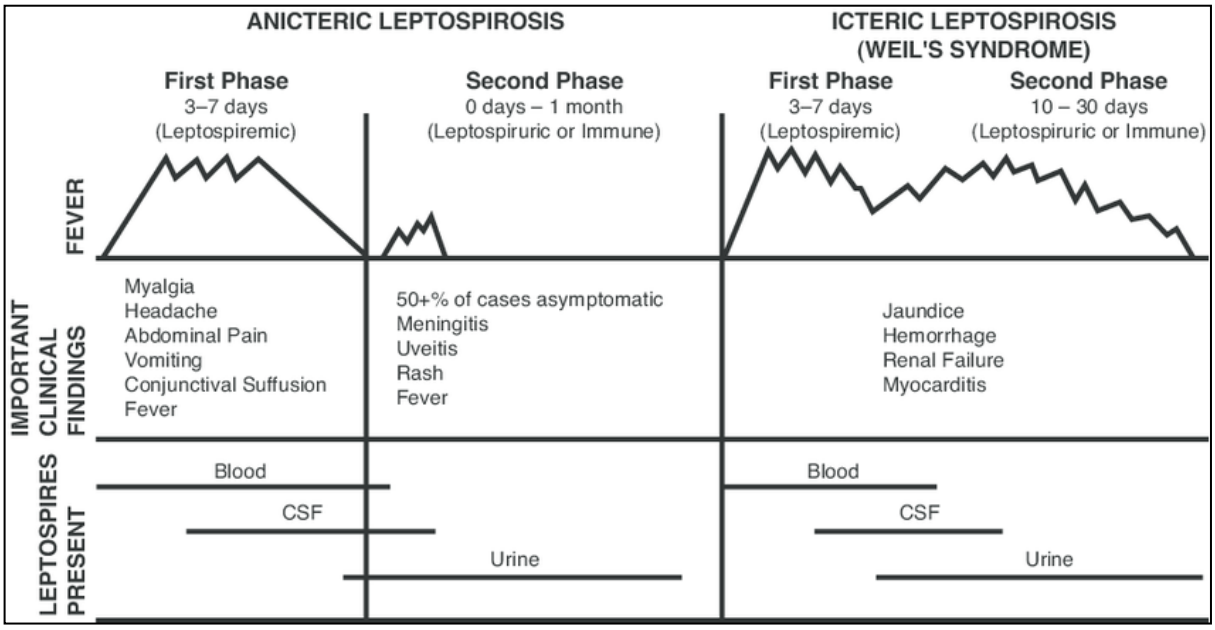

Figure 4: The clinical course of leptospirosis. Biphasic nature of leptospirosis leading to anicteric and icteric types of disease manifestation and the pathological sample used for detection of *Leptospira* during different clinical phases (CSF- cerebrospinal fluid) [Adopted from (Feigin and Anderson, 1975)**].**

<u>Anicteric leptospirosis</u>

Majorly, 85-90% of leptospirosis cases are categorized as anicteric form which is the milder form of disease. Before onset of the immune phase there will be 1-3 days long lag phase with minimal symptoms. All the symptoms of bacteremia phase re-emerge and last for 4-30 days. Abdominal pain, severe myalgia, high fever, mild proteinuria, cough, chest pain and muscle tenderness are observed in addition to the bacteremia phase symptoms (Faine, 1999; Haake and Levett, 2015).

Icteric leptospirosis

Icteric form is more severe form of the disease and is characterized with association of jaundice and also involvement of other organs. 5-10% of cases develop into icteric form and has a mortality rate of 5-15% (Haake and Levett, 2015). Major clinical manifestation of this form is high and prolonged fever, severe myalgia, headache, abdominal pain, nausea, hypotension, acute renal failure with proteinuria, oliguria or anuria and circulatory collapse. Severe form of this leptospirosis is known as Weil's disease involving liver and kidneys and is associated with hepatomegaly and tenderness of liver, pulmonary damage with intra alveolar hemorrhage, acute renal failure with interstitial nephritis or acute tubular necrosis, meningitis, hypotension, arrhythmia, and cardiac shock. Most of the cases presented with Weil's disease has high fatality rate because of multi-organ failure. (Bharti *et al.*, 2003; *National Guidelines Diagnosis, Case Management Prevention and Control of Leptospirosis Programme for Prevention and Control of Leptospirosis*, 2015; Haake and Levett, 2015). Reports of leptospirosis leading to sill births in pregnant women was also reported (Sharma *et al.*, 2011)

Because of its various symptoms and protean nature, leptospirosis is always misdiagnosed with other endemic and epidemic infectious diseases such as hepatitis, meningitis, dengue, influenza or viral hemorrhagic fever. This makes the diagnosis of leptospirosis based on the background of clinical symptoms a major challenge and thereby makes laboratory confirmation essential (Marquez *et al.*, 2017).

1.5.2 Animal leptospirosis

Leptospira infection in animals majorly goes off unnoticed with mild or no symptoms in many cases. But its infection in cattle and horses were well documented with remarkable loss to the animal husbandry. Because of recurrent infection of *Leptospira* in horses, equine leptospirosis is well documented with its characteristic autoimmune Uveitis and respiratory problems

(Verma, Stevenson and Adler, 2013). Congenital infection to the fetus in uterus of cattle leads to the abortions and still born fetus accounting for a higher economic losses (Ellis, 2015).

1.6 Pathogenesis

Leptospirosis, being a bacterial infection, unlike other pathogenic spirochetes, does not localize at the site of entry and spreads hematogenously throughout the body. First step in the pathogenesis is the entry of *Leptospira* in to the body by crossing through tissue barriers. Portals of entry are the cuts or abrasions on the skin and mucous membranes of conjunctivae or oral cavity (Corwin *et al.*, 1990; Lingappa *et al.*, 2004). Major factors contributing to its virulence are its motility and ability to swim through viscous media and presence of fibronectin binding protein, which helps in initial adhesion and invasion at the site of entry (Merien *et al.*, 2000). These characteristic features of *Leptospira* helps the organism to rapidly invade and penetrate skin or mucous membrane and spread through the blood stream to different tissues including central nervous system (CNS) and aqueous humor (Lux, Moter and Shi, 2000). Collagenase mediated injury to vascular endothelium of host by platelet activated factor acetyl hydrolase, vonwillebrand factor type A domain and paraoxonase is considered as the major factor for pathogenesis of *Leptospira* (Merien *et al.*, 2000). It is also believed that hemorrhages are because of sever vasculitis with endothelial damage caused by bacterial proteins which leads to capillary injury (Nally *et al.*, 2004). After damaging membranes of endothelial cells of the small blood vessels, the immediate effect is loosening of junctions between cells allowing *Leptospires* and fluid to enter into the extracellular space of the tissues. This follows the migration of erythrocytes from the site of damage to the extracellular space.

Liver being the major target for *Leptospira*, damage to hepatocytes and their apoptosis is documented resulting in damage of hepatocellular junctions (Merien *et al.*, 1998). Jaundice is the outcome of impaired effect of sub cellular host enzyme systems leading to hepatic lesions and secretion of bile from bile canaliculi into sinusoidal blood vessels, increasing the levels of

direct bilirubin in the icteric form. Entry of *Leptospires* into the lung tissue leads to the interstitial and intra alveolar hemorrhages (Arean, 1962). These pulmonary hemorrhages are responsible for the severe respiratory distress and may lead to death of the patient (Silva *et al.*, 2002; Nally *et al.*, 2004). Renal involvement in leptospirosis varies from mild non oliguric dysfunction of renal tubules to failure of complete renal system. Acute tubular necrosis and interstitial nephritis is because of secondary effects of host immune system after recognition of LipL32 antigen by TLR-2 receptors (Yang *et al.*, 2006). Secondary effect of ischemic changes, anoxia and increased pressure in the tissue leads to cellular functional disintegration followed by death. Long term persistence of *Leptospires* in the aqueous humor causes chronic and recurrent latent uveitis (Faine, 1999).

The diverse range of events contributing to the variable clinical manifestations of the disease makes understanding of *Leptospira* pathogenesis elusive and limited. A solid, clear and proven mechanism by which *Leptospira* causes infection in humans and livestock are not well understood. Looking at its genetic content particularly the virulence factors like LPS, hemolysins and surface exposed proteins may provide further insights into the pathogenesis of *Leptospira*.

1.7 Virulence factors

A wide array of virulence factors were identified by different groups which facilitate infection and successful colonization of *Leptospira* with in the host. Although a clear molecular mechanism by which *Leptospira* causes infection is still not clear, genes that play an important role in infection were studied. Adhesins such as *Leptospiral*-immunoglobulin like Proteins LigA, LigB and LigC, fibronectin binding protein and lamin binding proteins Lsa24/LfhA and Lsa21 which might play an important role in initial adhesion and invasion of bacteria to the host tissues and extracellular matrix were identified in the genome of pathogenic *Leptospires* (Chirathaworn and Kongpan, 2013; Faisal *et al.*, 2016). Outer membrane proteins

LipL21, LipL32, LipL41, LipL36, LigA, Qlp42 and Loa22 which might be important in signal transduction, host immune evasion, receptors for various host molecules were found to be regulated when grown in host like conditions *in-vitro* (Lo *et al.*, 2006; Matsunaga *et al.*, 2007). Sphingomyelinase and hemolysins encoded by pathogenic species were considered to be involved in tissue and cellular damage because of their phospholipase, pore forming and sphingomyelinase activities. Several members of sphingomyelinases were observed in pathogenic species of *Leptospira* which might help the bacteria to establish a niche in the host environment (Narayanavari *et al.*, 2012). In addition to these pathogenic factors, major pathogenic mechanism is driven by bacterial lipopolysaccharide (LPS) which is coded by a large 100kb fragment of genomic DNA. Unusual composition of Lipid A moiety in LPS is attributed to its role in successful evasion of complement mediated cytotoxicity by host immune system (Raja and Natarajaseenivasan, 2013).

1.8 Diagnosis of leptospirosis

Protean clinical manifestations of leptospirosis makes it resemble with other common endemic and epidemic disease such as malaria, influenza, dengue fever, viral hepatitis, scrub typhus, typhoid, viral pneumonia, tuberculosis, pyelonephritis, pneumonitis and pulmonary tuberculosis. Normal or slightly elevated levels of aspartate transaminase (AST) and alanine transaminase (ALT) and high level of Creatinine Phosphokinase (CPK) in circulation suggests possible leptospirosis in patients and differentiates with other common viral infections (*National Guidelines Diagnosis, Case Management Prevention and Control of Leptospirosis Programme for Prevention and Control of Leptospirosis*, 2015).

Conventional diagnosis is broadly divided into direct evidence by isolation of organism, examination by dark field microscopy or *Leptospira* specific fragment amplification by PCR and indirect evidence by detection of antibodies to *Leptospira*.

1.8.1 Isolation of bacteria

Isolation of *Leptospira* by culturing from tissue and clinical specimens is the strongest evidence for the confirmation of leptospirosis. *Leptospires* can be cultured successfully from body fluids like blood and cerebrospinal fluid till first 10 days and urine samples till 30 days of onset of disease. *Leptospires* are fastidious organisms and require several days to weeks to grow with prolonged incubation times from weeks to months, making the culturing method less sensitive and unsuitable for early diagnosis of infection. Additionally, infusion of antibiotics to patient further reduces the chances of successful isolation (Faine and World Health Organization, 1982).

1.8.2 Microscopic identification

Leptospires can be visualized in clinical specimen using Dark Field Microscopy. This method is a simple and rapid procedure for diagnosis of leptospirosis when circulating titers are more than 10^4 *Leptospires*/ml (Vinetz, 2001). But in real scenario this method often yields false positive and false negative results as the fibrin and other cell fragments in the blood mimics live bacterium. As the circulating *Leptospira* is present only for short time during the acute phase of infection and then localizes in various tissues, thus decreasing the circulating levels, makes the specimen preparation difficult and was shown that this method's positive confirmation diminishes with the duration of the infection (Chandrasekaran and Gomathi, 2004). In addition, this method requires strong technical expertise to understand the artifacts and lack of motility for the bacteria after sample preparation due to mechanical injury or presence of reactive antibodies further complicates the testing method (Vijayachari, 2007).

Staining of *Leptospires* is often performed with various silver impregnation techniques for detection of bacteria in tissues and body fluids (Skilbeck and Chappel, 1987). Immunofluorescence staining using anti-*Leptospiral* antibodies increases the sensitivity and specificity of microscopic determination (Wild *et al.*, 2002).

1.8.3 Serological tests

Microscopic agglutination test (MAT)

MAT stands as a conventional test and is considered as a gold standard for diagnosis of leptospirosis because of its proven sensitivity and specificity. MAT has serovar specificity and isolates belongs to serovars, representative of all major serogroups and locally prevalent serovars, are used as antigens and patient serum dilutions are used as antibody source. Agglutination of the antigen-antibody mixture is the evidence for serovar/serogroup specific antibodies in the serum (Goris and Hartskeerl, 2014). However, MAT has limitations in terms of maintenance of live cultures of panel of reference strains representing all major serogroups and locally prevalent serovars, technical expertise to interpret results, complexity, time consuming and applicability of the test only during late acute phase of the disease when anti-*Leptospira* antibodies are generated (Smythe *et al.*, 2009; Tiengrim *et al.*, 2009). This method for confirmation of *Leptospira* infection will be too late to determine an effective antibiotic treatment (Bharti *et al.*, 2003). Further the test cannot reveal if the agglutination is because of IgM indicating the current infection or IgG indicating the past infection.

Enzyme Linked Immuno Sorbent Assay (ELISA)

IgM specific ELISAs were developed using whole cell lysates or LPS as antigen to detect at the genus level and to differentiate present and past infection in clinical samples. This technique is simple and rapid method for diagnosis of leptospirosis but lacks reproducibility because of variations in preparation of antigen and poor specificity (Raja and Natarajaseenivasan, 2013).

ELISA methods based on recombinant antigenic proteins that are found to be conserved across all the pathogenic *Leptospira* with antigenic nature like LipL21, Hap1/LipL32, Lip41, LigA, LigB, Loa22, OmpL1 and Hsp58 have been shown to be effective for diagnosis with their documented reproducibility and sensitivity (Priya *et al.*, 2003; Chalayon *et al.*, 2011).

ELISA methods employing recombinant proteins alone or in combination have their advantages over the whole cell lysate based methods in terms of higher concentration of antigens, consistency in antigen preparation and lack of interfering moieties thus increasing its sensitivity and reproducibility (Chalayon *et al.*, 2011). As these methods were developed based on the locally prevalent pathogenic organisms, their applicability and usefulness outside the particular geographical area is always doubtful.

Rapid diagnostic test (RDT)

LEPTO-dipstick ELISA has been developed and used effectively as RDT at primary health centers. It has longer shelf life and does not require special equipment to perform the assay. It also has good sensitivity and specificity in comparison with regular IgM based ELISA tests (Sehgal *et al.*, 1999; Goris *et al.*, 2013). However, to obtain a better sensitivity for the diagnosis, it requires testing of two samples per patient and RDT alone cannot confirm the leptospirosis in early stages of infection.

Other serological methods

Other serology based tests employed at genus level include macroscopic slide agglutination test, DriDOT ELISA, latex agglutination test, complement fixation, indirect hemagglutination and indirect immunofluorescence assay. Most of the antigen preparation used in these tests are prepared from nonpathogenic *Leptospira* species and have varying sensitivities from population to population (Marquez *et al.*, 2017).

As the serology based tests are dependent on the presence of anti-*Leptospira* antibodies which can be detected only after 5 to 7 days of onset of disease, a timely based accurate and reliable diagnostic test that can be used during early acute phase is of prime importance for early detection of infection and to effectively tackle the disease burden.

1.8.4 Molecular Methods

PCR and RT-PCR based methods are successfully employed to detect *Leptospira* during the early acute phase infection, before the antibodies are detected and organism is cleared from the circulation. PCR was successfully employed for samples isolated from urine, blood, tissues and CSF (Musso and La Scola, 2013). A PCR reaction based on set of primers namely G1/G2 and B64-I/B64-II were successfully employed to clinical samples to diagnose *Leptospira* infection for all major seven pathogenic species (Gravekamp,' *et al.*, 1993).

Real-time PCR which has advantage over traditional PCR was also developed to rule out false positive results during diagnosis. Multiple targets were employed and most successful ones are based on 16S rRNA targets with increased sensitivity and specificity among pathogenic and saprophytic strains of *Leptospira* (Agampodi *et al.*, 2012).

Because of requirement of sophisticated and expensive equipment to perform these tests, availability of these tests in resource poor countries has become a night mare.

1.8.5 Loop- mediated Isothermal amplification (LAMP)

LAMP employs amplification of target DNA at isothermal conditions when incubated with DNA polymerase and target specific primers. Successful amplification of target increases fluorescence or turbidity in the mixture without the need for electrophoretic identification on gel. *LipL41* and *rrs* genes were used as potential targets for developing LAMP based leptospirosis diagnostic kit. With its advantage of using successfully in resource poor settings with sensitivity of detecting 2 *Leptospiral* cells per reaction, it is considered as the better alternative for PCR based techniques (Koizumi *et al.*, 2012). But its applicability in endemic zones is yet to be assessed.

In conclusion, although culturing and MAT tests stand gold standard for detection of *Leptospira*, they could not be implemented given the fastidious nature of bacteria and time

taken to perform the tests. Serological tests which employ majorly antibodies raised against OMPs have shown to have higher sensitivity and specificity but their applicability over global isolates was in question as the OMP against which antibodies were raised may not be conserved across all pathogenic strains (Rajapakse *et al.*, 2015). Thus, lack of sensitive and specific diagnostic tools and overlapping disease symptoms with other prevalent endemic diseases remains the major reasons for neglected nature of this disease and increased fatality.

1.9 Treatment and Management

1.9.1 Treatment

Symptomatic supportive therapy is given for different symptoms of the disease. Leptospirosis can be treated effectively with antibiotics like doxycycline, ampicillin, amoxicillin, tetracycline or penicillin given the condition that the disease is diagnosed within 5 days of the onset of illness (Guidugli, Castro and Atallah, 2000). Therapy with oral doxycycline (100 mg orally twice per day) was shown to be effective in treating the disease in adults if diagnosed early. However, these antibiotics are ineffective on the disease manifestations, mainly during acute leptospirosis (World Health Organisation, 2003; Haake and Levett, 2015).

1.9.2 Vaccination

Vaccination of humans with inactivated and killed *Leptospiral* cells was proved to be effective but required repeated vaccine doses to retain the immunity (Faine, 1999; Koizumi and Watanabe, 2005). Most of the available human vaccines require two booster doses after initial immunization and a repeated immunization for every two years (Laurichesse *et al.*, 2007). Several vaccines prepared from heat killed, whole cell bacteria were shown to be protective only from homologous serovars but not completely from heterologous serovars having different antigenic repertoire on their cell wall (Chapman, Faine and Adler, 1990). Further, polyvalent subunit vaccines that can be effective against all pathogenic species of *Leptospira* are being considered using conserved OMPs with proven antigenicity for inducing higher

immune response (Bashiru and Bahaman, 2018). Important conserved proteins being considered for vaccine development are Omp L1, LipL32, LipL41, LemA, LigA, LigB and LigC (Dellagostin *et al.*, 2011).

Unacceptable side effects, short and incomplete protection, varying disease manifestations, potential autoimmune disease induction and incomplete knowledge of mechanisms of infection by *Leptospira* confront the development of vaccines for human leptospirosis (Rajapakse *et al.*, 2015). Sequencing of complete genome of pathogenic *Leptospira* isolates has assisted in studying detailed genetic background of different molecular coordinates and in-depth evaluation of genome encoded properties. This has made possibility of identifying conserved membrane proteins among all pathogenic species using computer assisted programs and has laid new ways to develop a sophisticated, simple, sensitive, reproducible diagnostic method and a potential vaccine for leptospirosis (Nascimento *et al.*, 2004; Gamberini *et al.*, 2005).

1.9.3 Prevention and control

Leptospirosis can be effectively controlled by educating the people residing in endemic regions and preventing disease transmission from animals that act as reservoirs of *Leptospira.* Maintenance of proper hygiene and protective clothing by occupational risk groups will keep the disease at bay. Avoiding contact with potentially contaminated water sources during adventures and vacations, and acquaintance with the leptospirosis to the health workers will help to prevent and contain the disease outbreaks.

1.10 Characterization

Current classification system to characterize *Leptospira* is based on two different parameters, one based on serology, defining serovar as basic taxon and other based on DNA composition, defining species as basic taxon (Ramadass *et al.*, 1992; Cerqueira and Picardeau, 2009).

Genus *Leptospira* was historically classified into two groups *L. interrogans sensu lato* and *L. biflexa sensu lato* containing pathogenic and non-pathogenic strains respectively (Faine, 1999). *Leptospira* isolates were classified into pathogenic and non-pathogenic strains based on their ability to grow at low temperature (13°C) and resistance to 8-azaguanine (225µg/ml). Pathogenic strains cannot grow at low temperature and are sensitive to 8-azaguanine (Johnson and Rogers, 1964). Further classification of *Leptospira* isolates based on their serological characteristics has resulted in serovars and serogroups.

DNA hybridization studies has shown that more amount of genetic heterogeneity was observed among strains of genus suggesting for species classification of pathogenic strains. Application of molecular taxonomic methods, majorly DNA based homology studies, has led to the description of total 21 species in the genus *Leptospiraceae* (Yasuda *et al.*, 1987; Ramadass *et al.*, 1992; Perolat *et al.*, 1993; Brenner *et al.*, 1999; Vijayachari *et al.*, 2004; Cerqueira and Picardeau, 2009; Smythe *et al.*, 2013). There are nine major species which are designated as pathogenic and are *L. interrogans, L. borgpetersenii, L. santarosai, L. noguchii, L. weilii, L. kirschneri, L. alexanderi, L. alstonii and L. kmetyi* (Kmety and Dikken, 1993; Smythe *et al.*, 2013; Varni *et al.*, 2014).

1.10.1 Serological characterization

Conventionally *Leptospira* isolates were differentiated into serovars using rabbit antisera, on the basis of serological characteristics which depend on the cell wall composition. According to the taxonomical society classification for *Leptospira*, 1986 "two strains are said to belongs to different serovars if after cross absorption with adequate amounts of heterologous antigen, more than 10% of the homologous titer regularly remains in at least one of the two antisera in repeated tests" (Kmety and Dikken, 1993). Although serovar classification does not have any taxonomical standing, this classification over the time has proved useful for serological diagnosis and understanding of disease epidemiology. Serological classification has divided

pathogenic *Leptospires* into more than 300 serovars and grouping serovars with close serological affinities into around 25 higher order serogroups (Kmety and Dikken, 1993). Serological classification is performed using following methods.

Microscopic agglutination test (MAT)

MAT is performed using rabbit antisera to the known *Leptospira* isolates. Group sera, which is rabbit antiserum that reacts with all serovars of the serogroup, is used to determine the serogroup status of the unknown strain and reference antisera for all known serovars of the serogroup are used to determine the serovar status of isolate.

Cross agglutinin absorption test (CAAT)

Serovar being the basic taxon of the *Leptospira* taxonomy, characterization of unknown strain to the level of serovar is important. CAAT is the standard assay for designating serovar status (Dikken and Kmety, 1978).

Briefly the method involves, testing of antigen or unknown strain against all relevant reference antisera which has a MAT titer of 1:5120 and testing of relevant reference strains against the antisera produced against the antigen or unknown strain. After absorption of the antigen – antibody complex for overnight, antisera left in the supernatant is used for homologous titration with live and killed antigen using MAT (Dikken and Kmety, 1978; Vijayachari, 2007).

MAT titers are expressed according to the following formula

$$\boldsymbol{Tui = \frac{Tgs}{Ths} \times 100}$$

Where in,

Tui = agglutination titre of unknown strain

Tgs = Reciprocal titre of antisera from positive groups with unknown strain

Ths = Reciprocal titre of reference antiserum with homologous strain

And, $$\boldsymbol{Tus = \frac{Tgi}{Thi} \times 100}$$

Where in,

Tus = agglutination titre of unknown sera

Tgi = Reciprocal titre of unknown sera with reference strains from positive groups

Thi = Reciprocal titre of unknown serum with homologous strain

However, serological based classification has limitations in terms of their feasibility to perform only in reference laboratories, ambiguous identification of isolates, results archiving, growing number of controversial typing results and maintenance of live reference strains (Cerqueira and Picardeau, 2009). To improve the sensitivity of these methods, mouse monoclonal antibodies were developed for few serovars for typing but their universal applicability is still at large (Terpstra *et al.*, 1985; Masuzawa *et al.*, 1988).

1.10.2 Molecular characterization

With the advancement of molecular tools for diagnosis and classification of microorganisms, several molecular methods were also employed for *Leptospira* sps. Characterization of *Leptospira* isolates requires both identification of serovar and species status. Molecular methods which are DNA-based techniques were exploited as supplementary or alternative methods to replace the tedious serological classification methods. DNA-based characterization is based on genetic makeup and can easily be translated into genetic relationship and affinities across isolates.

1.10.2.1 DNA-hybridization

This is performed by using labelled whole genomic DNA or specific probes, harnessing dot and in situ hybridization techniques (Yasuda *et al.*, 1987; Ramadass *et al.*, 1992; Ahmed, and P. Grobusch, 2012). This has been successfully implemented for detecting pathogenic *Leptospira* from plasma sediment, liver smears, tissue samples, body fluids and urine samples dotted

onto nitrocellulose membrane. As the technique uses radioactive labels and enzymatic staining, requirement of special safety and well equipped laboratories are mandatory.

1.10.2.2 DNA-DNA hybridization

Haapala and co-workers were the first to employ this method for species identification of *Leptospira*. DNA homology is determined by DNA-DNA hybridization using thermal elution technique on the basis of duplexes trapped in Agar (Yasuda *et al.*, 1987; Busse *et al.*, 2010). As genomic characterization using DNA-DNA hybridization has shown little correlation with the existing serological classification for serovar designation, this method was used to separate *Leptospira* into species. Its application is limited because of requirement of higher amounts of isotope labelled pure DNA. Based on the DNA-DNA hybridization studies, *Leptospira* has been divided in to different species (**Table 1**).

Table 1: *Leptospira* species classification using DNA-DNA hybridization method

(Adopted from Ahmed A, and Grobusch MP *et al*, 2012, J. Bacteriol Parasitol)

Pathogenic	Intermediate	Non-pathogenic
L. interrogans	*L. inadai*	*L. biflexa*
L. santarosai	*L. fainei*	*L. wolbachii*
L. weilii	*L. broomii*	*L. vanthielii*
L. borgpetersenii	*L. wolffii*	*L. terpstrae*
L. noguchii.	*L. licerasiae*	*L. yanagawae*
L. kirschneri		*L. meyeri*
L. alexanderi		*L. idonii*
L. alstonii		
L. kmetyi		

1.10.2.3 Bacterial restriction –endonuclease DNA analysis (BRENDA)

Non-sequence based classification method employed for *Leptospira* is BRENDA which relays on the agarose gel fingerprints generated by treating the DNA with restriction endonucleases (Marshall, Wilton and Robinson, 1981; Venkatesha and Ramadass, 2001). Southern blot hybridization of BRENDA generated profiles simplified the process. Employing labeled probes corresponding to pathogenic elements or repetitive elements, *Leptospira* classification

was simplified and has shown its efficiency in classification in accordance with the DNA homology methods (Van Eys *et al.*, 1991). However, its difficulty in terms of data interpretation because of presence of large number of bands, profile matching and exchange of results between laboratories, made this method less universally applicable.

1.10.2.4 Ribotyping

This is based on employing probes designed against the conserved homologous rRNA coding sequences detected by southern blotting. Restriction fragment length polymorphism (RFLP) profiles will be generated by digesting chromosomal DNA after hybridizing with *rrs* and *rrl* gene probes. This method has been used for universal phylogenetic typing of bacteria for taxonomy purposes and sub group classification of microorganisms (Grimont and Grimont, 1986). Ribotyping for *Leptospira* classification has put isolates into separate species clusters in accordance with DNA homology studies (Perolat *et al.*, 1993). Because of its limitation in distinguishing only few serovars and presence of small number of rRNA genes, this typing method is proved to be not highly discriminative (Kositanont *et al.*, 2007).

1.10.2.5 Pulsed field gel electrophoresis (PFGE)

PFGE is the technique in which larger chromosomal DNA molecules are cut with a rare cutting endonuclease like NotI and the fragments generated are segregated on agarose gel by applying perpendicularly oriented electric field with alternate pulses for better resolution. *Leptospira* genome consists of two genomic DNA copies, one is large chromosome ranging from 3.1kb to 5kb and another smaller one with 0.35kb size. PFGE generated DNA fragment profile of panel of serovars will be compared with the PFGE fragment profile of an unknown isolate to assign serovar status to it. Modifications in terms of computational analysis for the fragment profile generated and creation of database for easy sharing of data between laboratories without the need of shipping strains has made this technique a powerful one and an alternate to serotyping (Galloway and Levett, 2010). Macro restriction profiles generated

using PFGE were reported to have good concordance with serotyping results and this technique is considered as gold standard for molecular typing tools (Herrmann *et al.*, 1992). However, discrepancies observed during differentiation of few serovars and distinction of few isolates altogether as new genomo species, restricted this method applicability. In addition, this method is labor-intensive and is not available in all laboratories.

1.10.2.6 Nucleic acid amplification

Polymerase chain reaction (PCR) based nucleic acid amplification methods were employed successfully for detecting and characterization of *Leptospira* specific nucleic acid in clinical samples. PCR was introduced as early as in 1989 for detection of *Leptospira* in clinical samples and was shown to be useful in both diagnostic and epidemiological studies. Several targets were identified and proved to be good for diagnosis of leptospirosis in humans and animals. Methods suggested by Gravekamp *et al.*, 1993 and Merien *et al.*, 1992 utilizing two sets of primers (G1/G2 and B64-I/B64-II) and a set of primers for *rrs* gene respectively have been validated for use in the early diagnosis and confirmation of leptospirosis in clinical samples (Merien *et al.*, 2005). Various studies have been performed to show the useful ness of PCR in *Leptospira* epidemiology using primers for genes such as *rrs, rrl, ompL1, gyrB, flaB, hbpA, hap1, rpoB, lipL32, lipL21, lipL44* and *ligB* (Branger *et al.*, 2005; Sridhar *et al.*, 2008). As the clinical evaluation of this conventional PCR was performed at limited scale and false positive results because of contamination of DNA were reported, this conventional PCR applicability was limited (Ahmed *et al.*, 2009).

Real time PCR (qPCR) which uses conventional PCR methodology and labelled primers or dyes for monitoring and detection of amplification process, was shown to be highly sensitive and specific. This has made qPCR based techniques superior to conventional PCR and this method has been employed for *Leptospira* detection targeting rRNA genes, housekeeping genes and genes specific to pathogenic *Leptospira* like *rrs, rrl, 23S rRNA, LipL32, gyrB, ligA,*

ligB and *secY* genes (Ahmed *et al.*, 2009; Bourhy *et al.*, 2011). Further this method's accuracy was proven during diagnosis in blood samples, even in the initial stages of early acute phase (Ahmed *et al.*, 2009; Bourhy *et al.*, 2011; Agampodi *et al.*, 2012).

PCR based techniques are considered advantageous due to their easy standardization and possibility of sharing and archival of results. This in combination with sequencing of the amplified fragment or in combination of other typing method has made this a potential method to use for categorizing *Leptospira* species and its diagnosis. Further sensitivity of the PCR was also increased by using more specific nested primers, targeting repetitive elements, insertion elements and fragments obtained from restriction digestion (Zuerner and Bolin, 1997; Barocchi *et al.*, 2001; Romero and Yasuda, 2006), and by using arbitrary primers to randomly amplify polymorphic regions of DNA (Ralph *et al.*, 1993; Tulsiani *et al.*, 2010).

Randomly amplified polymorphic DNA (RAPD) and **arbitrarily primed PCR (AP-PCR)** are the two techniques which amplify arbitrary genomic regions using low stringent primers for generation of strain specific fingerprints for characterization of *Leptospira* (Ralph *et al.*, 1993; Tulsiani *et al.*, 2010). These methods have shown to produce results in consistent with the traditional 16S rRNA sequencing and DNA-DNA hybridization studies (Ciceroni *et al.*, 2002; Natarajaseenivasan *et al.*, 2004). However, because of poor reproducibility and difficulty in comparison of results across the laboratory, these methods were found to be not suitable for large scale studies.

1.10.2.7 16S rRNA sequencing

With the advancement of gene sequencing technologies, the methods based on sequencing of conserved regions of the genomes were applied successfully for elucidating the evolution, taxonomy and molecular epidemiology of micro-organisms (Fox *et al.*, 1980). 16S rRNA sequences were used to deduce phylogeny of *Leptospira* and has revealed that spirochetes represented an ancient branch of eubacteria (Woese, 1987). *Leptospira* phylogeny using 16S

rRNA sequencing has divided the species into pathogenic, non-pathogenic and saprophytic clades (**Figure 5**) (Morey *et al.*, 2006). In recent times many other variable genes were targeted for constructing phylogeny but till date sequencing of *rrs* gene is the regular method employed for phylogeny construction.

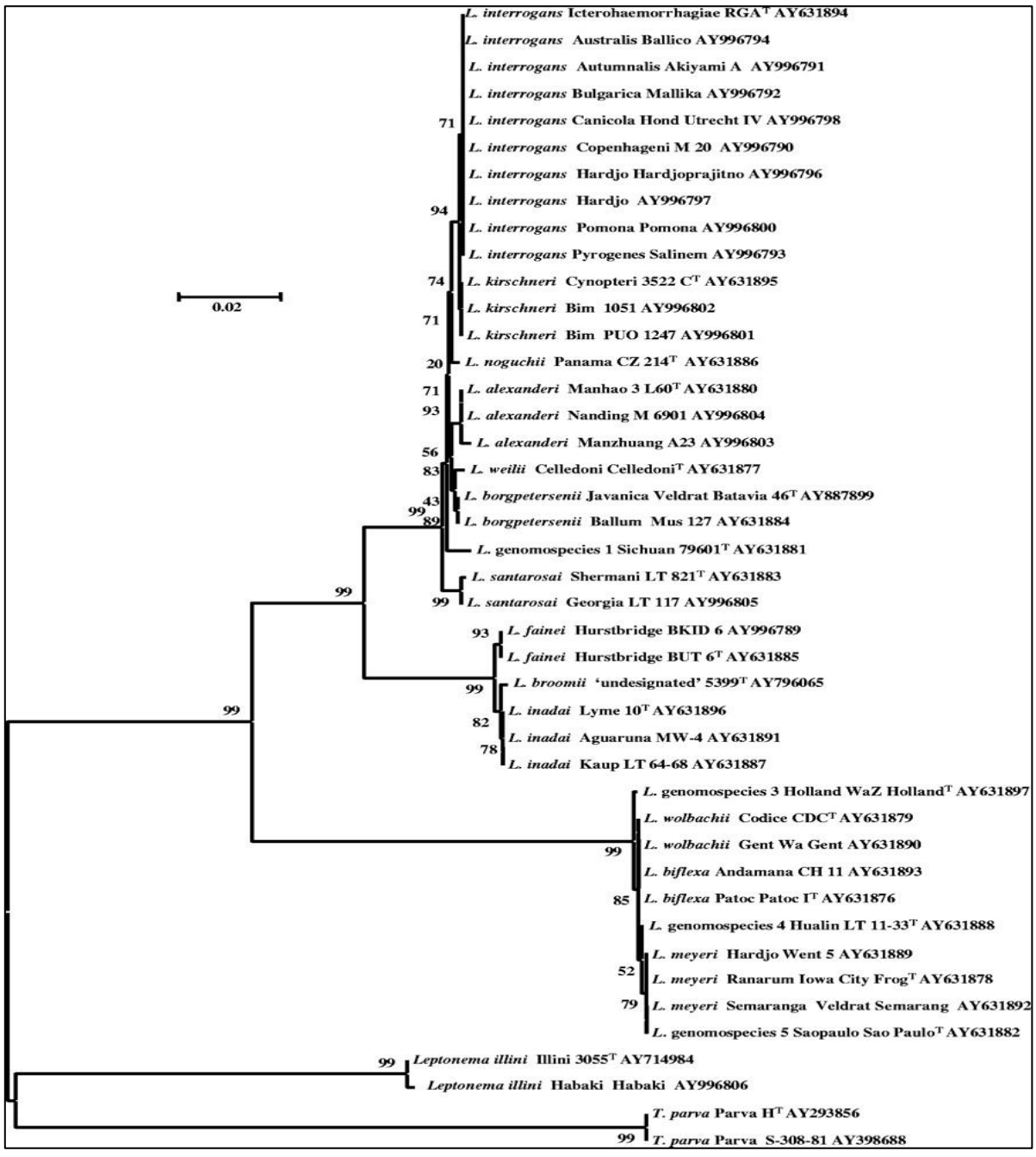

Figure 5: Phylogenetic tree (unrooted) generated with 16S rRNA gene sequences of *Leptospiraceae* family. 15 species of *Leptospira* genus were resolved into pathogenic, intermediate and non-pathogenic clades (Morey *et al.*, 2006) (Figure adopted from Roger E. Morey et al. J. Clin. Microbiol. 2006).

1.10.2.8 Multiple locus variable number of tandem repeats analysis (MLVA)

MLVA which is based on PCR amplification of variable number of tandem repeats (VNTR) present on genomes of microorganisms is used for categorization of strains. This method was employed for characterization of *Leptospira* and could discriminate all serovars and grouped them as per their serogroups (Majed *et al.*, 2005; Salaun *et al.*, 2006; Nalam *et al.*, 2010). But use of this method directly to clinical samples and environmental samples need further improvement.

1.10.2.9 Amplified fragment length polymorphism (AFLP)

AFLP combined with fluorescent labelled primers (FAFLP) employs digestion of genomic DNA with specific restriction enzymes, ligation of adapters to digested fragments and their amplification by PCR using adapter specific primers. Electrophorograms of amplified fragments are used to generate fingerprints for each isolate and are then analyzed by comparison to deduce the clonality in outbreaks and epidemiological studies of *Leptospira* (Vijayachari *et al.*, 2004). However, requirement of large quantity of purified DNA makes this method less applicable for epidemiological studies with high number of isolates.

1.10.2.10 Multilocus sequence typing (MLST)

MLST is a robust and efficient genotyping method for identifying ancestral relationships and segregating outbreak associated strains according to their species. This method is based on sequencing of multiple genes which are not under positive selection (Maiden *et al.*, 1998). MLST has been applied successfully for *Leptospira* classification in several epidemiological studies (Ahmed *et al.*, 2006; Thaipadungpanit *et al.*, 2007). MLST makes use of sequence deduced from PCR amplified DNA segments and thus depends on the success of amplification, which in turn depends on the annealing efficiency of the PCR primers. This method has its own merits in terms of simplicity to perform, requirement of small quantity of purified DNA, reproducibility, data management and exchange of results across laboratories without the need of physical exchange of live cultures or materials.

1.11 *Leptospira* Genomics

1.11.1 Whole genome sequencing and genetic analysis

With the advancement of next generation sequencing technologies in the last one decade, complete genome sequences of *Leptospira* isolates were made available in the public domain. *Leptospira* genome is of size ranging from 3.9 Mb to 4.6 Mb and is typically distributed into two circular chromosomes, one larger and another smaller one. A 74 kb replicon designated as p74 is also reported from saprophytic strain *L. biflexa* (Picardeau *et al.*, 2008).

Availability of complete genomes of pathogenic, intermediate and non-pathogenic strains of *Leptospira* has facilitated to gain insights into genetic potential to understand the molecular basis of pathogenesis, course of evolution of pathogenic strains, mechanism behind lateral gene transfers and to develop various sequence based molecular typing methods to study pathogen evolution patterns and distribution in detail (Zuerner *et al.*, 2000; Haake *et al.*, 2004).

Genome analysis of *Leptospira* species revealed high level genome plasticity in terms of duplications, rearrangements and changes in the conserved regions of genome indicating a rapid adaptation to new environments by pathogenic species (Ren *et al.*, 2003; Nascimento *et al.*, 2004; Bulach *et al.*, 2006). It also has been proposed based on genome resemblance that pathogenic *Leptospira* might had a common progenitor like that of *L. biflexa* (Picardeau *et al.*, 2008).

Whole genome sequences of *Leptospira biflexa* serovar Patoc strain Patoc1 and strain Ames has revealed that its genome encodes for 3590 protein coding genes whereas *L. interr*ogans serovar Copenhageni strain Fiocruz encodes 3379 and *L. borgpetersenii* serovar Hardjo strain L550 encodes 2842 genes (Nascimento *et al.*, 2004; Bulach *et al.*, 2006; Picardeau *et al.*, 2008). Higher coding density, which is four times the number predicted for other members of spirochetes is considered as responsible for highly variable clinical manifestations and differential expression of the organism (Ren *et al.*, 2003). It has also revealed that very high

number (approx. 30%) of coding genes are hypothetical in nature and does not have any predicted function. Presence of higher number of transposases, pseudogenes and gene fragments in *L. borgpetersenii* genome suggests their role in IS-mediated genome reduction and possible speciation in the lineage (Bulach *et al.*, 2006).

1.11.2 Comparative genomics

For addressing a basic question "what makes bacteria pathogenic", one has to understand the basic functional differences between pathogenic and nonpathogenic isolates of a species or genus. As the evolutions of pathogenic and intermediate pathogenic species are believed to be from a common progenitor like that of *L. biflexa*, a close comparison of their whole genomes might reveal the underlying evolutionary mechanisms of *Leptospira* pathogenesis. Attempts have been made to compare the pathogenic, intermediate and saprophytic genomes at genetic level to get insights into the virulence mechanism and factors responsible for adaptation of bacteria in different host environments (Picardeau *et al.*, 2008; Ricaldi *et al.*, 2012).

It is believed that *L. borgpetersenii* evolution underwent a process of genome erosion and loss of gene function limited its viability outside the host environment. Thus disease transmission by *L. borgpetersenii* is believed to be possible only through host – host contact. Whereas the evolution of *L. interrogans* is believed to be by a process of gene gain aiding its viability in the environment for longer periods and transmission of disease by it through contaminated environment sources like soil and water too (Nascimento *et al.*, 2004; Bulach *et al.*, 2006).

Pathogenic mechanism of *Leptospira* remains poorly understood and varying pathogenic potentials across different species is also not explored (Ko, Goarant and Picardeau, 2009). Lateral transfer of genes has been documented in *Leptospira* but underlying mechanism has not been deciphered yet (Haake *et al.*, 2004). Getting insights into the evolution of pathogenicity, *Leptospiral* tropism towards different mammalian reservoirs and horizontal gene transfer mechanisms is very important.

Genetic comparison of *Leptospira* genomes revealed that there are 1431 genes present only in pathogenic isolates which does not have any orthologues in their saprophytic counterparts (Picardeau *et al.*, 2008). Out of the pathogenic genes a majority of 893 genes do not have any assigned function and fall in to hypothetical proteins. Further, comparison with intermediate pathogen has revealed that there are approximately 452 pathogen and intermediate pathogen conserved proteins (**Figure 6**) (Ricaldi *et al.*, 2012).

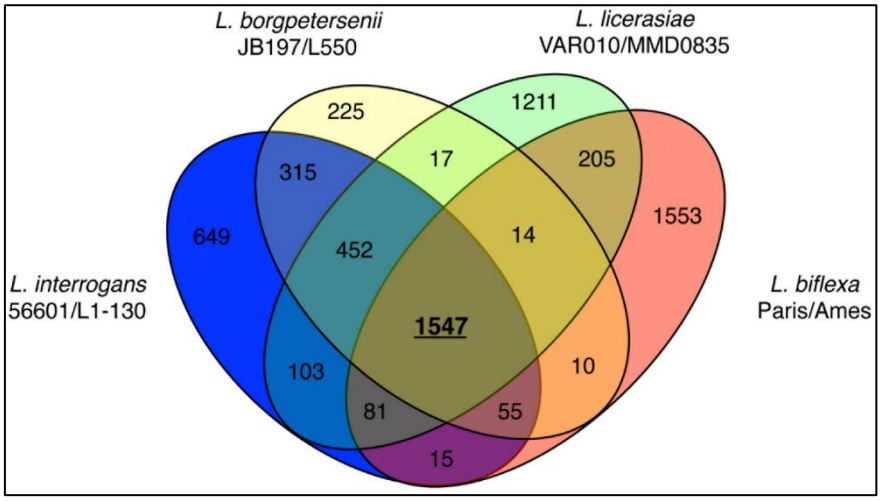

Figure 6: Venn diagram showing the comparison of pathogenic, intermediate and saprophytic counterparts of *Leptospira* genus (Figure adopted from (Ricaldi *et al.*, 2012))

Comparative genomics has opened up areas to focus on critical mechanisms underlying the transmission modes, metabolite requirement, pseudogenes, virulent determinants, genomic islands and events of lateral gene transfers. Comparison of this data along with valid transcriptomic data will aid in the better understanding of the pathogenicity of the organism.

1.12 Rational and objectives of the study

Characterization of pathogenic bacteria stands a basic requirement for better understanding of epidemiology of a disease. Epidemiology, which analyses the distribution and determinants of disease and health conditions, is very important for designing and shaping effective policies for implementing ttherapeutic interventions to contain the disease progression and transmission. Epidemiology of leptospirosis is complex and dynamic as one *Leptospira* serovar can be maintained in several maintenance hosts and one host can maintain several *Leptospira* serovars. As identification of maintenance host and infecting serovar is very important and essential for designing control and prevention strategies, effective and reliable tools for characterization of *Leptospira* are very important (Hartskeerl *et al.*, 2004). The enormous repertoire of serovars of *Leptospira* which are classified based on the ever changing surface antigens represents an unreliable scenario for strain identification, as the molecular identity keeps changing depending on host and environment niches they inhabit and cross through.

Conventional techniques used for leptospirosis diagnosis and characterization are mainly based on serology with major drawbacks such as lack of simplicity, laborious procedures, long turnaround time for confirmation and difficulty in exchanging results among research communities. Sensitive and specific detection of *Leptospira* during early stages of infection is also stands as a major challenge. Molecular techniques have shown to be advantageous over the conventional methods in terms of their simplicity, repeatability, and applicability in early stages of infection, robustness and possibility to exchange data across research community for diagnosis and characterization. Most of the molecular techniques were able to differentiate *Leptospira* to species level and deciphered the status of many difficult to distinguish serovars of same serogroup belonging to different species (Ko, Goarant and Picardeau, 2009).

MLST is one such molecular method developed based on the genome sequence and has established promise in unravelling phylogeny and diversity. MLST has an advantage in terms

of generating electronic portable data with a possibility of establishing a central database to share the data among research community. It is also considered that advantages of MLST technique can only be undermined by phylogeny established using whole genome sequences. MLST for characterization of *Leptospira* was developed and validated by different research groups using different set of genes. This has created confusion in the research community to follow a consensus scheme for characterization of *Leptospira* at global level.

The pathogenicity of *Leptospira* species remains inconclusive because of varied reasons. It is also found that pathogenic *Leptospira* forms a different branch in the evolution, distinct from saprophytes and intermediate group of *Leptospira.* It was also reported that pathogenic potential can be attributed to the acquisition of genes responsible for survival in host determined environments by lateral gene transfer mechanisms (Picardeau *et al.*, 2008).

Therefore deciphering genetic conservation at pathogenic population level will provide insights into the infection potential, host adaptation and evolution of pathogenic *Leptospira.* The availability of humongous sequence data in public domains facilitated the possibility to evaluate the genetic repertoire of *Leptospira* isolates and comparison of genomes of isolates to put forward the genome structure of pathogenic species.

Having said that, we framed our objectives to evaluate the MLST methods currently available for typing *Leptospira* and to harness whole genome sequence information to identify conserved regions in pathogenic *Leptospira* genomes than can be better targets for diagnosis and deciphering pathogenic potential of the organism. This study is thus broadly divided into following objectives:

1. Evaluation of MLST for epidemiology of pathogenic *Leptospira* isolates and creation of publicly accessible *Leptospira* MLST database
2. Genomic analysis of pathogenic *Leptospira* isolates

Chapter 2

Evaluation of MLST for Epidemiology of Pathogenic *Leptospira* isolates

And

Creation of Publicly Accessible *Leptospira* MLST Database

2.1 Introduction

Identification and typing of bacterial species plays an important role in understanding epidemiology of the disease in terms of distribution, progression and outbreaks. Clinical microbiologists and epidemiologists require accurate and reproducible isolate characterization data for proper understanding of the disease, investigation of local and global outbreaks, and for development of measures to counter the disease.

Leptospirosis is one of the most prevalent zoonotic disease worldwide yet underestimated because of the lack of awareness and lack of accurate and rapid diagnostic approaches (Bharti *et al.*, 2003; Ko, Goarant and Picardeau, 2009). Causative agents of the disease are helically coiled Gram negative bacteria of the genus *Leptospira* (Faine, 1999).

Historically, *Leptospira* classification was based on the pathogenicity and pathogenic isolates were classified into *interrogans* species and free living saprophytic isolates into *biflexa* species. Later, serological classification using Cross Agglutinin Absorption Test (CAAT) was used to determine the serovar status of isolates based on their serology expressed by membrane proteins which in turn depends on the heterogeneity of lipopolysaccharide (Kmety and Dikken, 1993). Closely related serovars based on the antigenicity are grouped into serogroups using Microscopic agglutination test (MAT) (Dikken and Kmety, 1978). Till date more than 300 serovars are identified with in the species *L. interrogans* which are grouped into almost 25 serogroups (Levett, 2001). Serological classification which is majorly based on the ever-changing surface repertoire of bacteria has proven its importance in clinical and epidemiological investigations to point out the host reservoirs involved in transmission. However, it was considered as an ambiguous, arbitrary and confusing approach, throwing challenges for studying epidemiology of leptospirosis and tracking of strains back to the point of origin during endemics and epidemics (Tiengrim *et al.*, 2009).

With the advancement of molecular typing methods and availability of whole genome sequences, development of more reliable, portable, simple and sensitive methods for classification and diagnosis of *Leptospira* has taken its pace (Cerqueira and Picardeau, 2009). Myriad of molecular techniques for typing *Leptospira* have often lead to inappropriate elucidation of epidemiology (Cerqueira and Picardeau, 2009). Molecular methods like DNA-DNA hybridization, restriction endonuclease assay (REA), pulsed field gel electrophoresis (PFGE), restriction fragment length polymorphism (RFLP), arbitrary primed PCR, variable number of tandem repeats (VNTR), fluorescent amplified fragment length polymorphism (FAFLP) and multi locus sequence typing (MLST) are applied successfully for the classification of *Leptospira* genus (Ko, Goarant and Picardeau, 2009). However, these techniques are having their own limitations in terms of requirement of a large quantity of pure and high-quality DNA, ambiguous interpretation of results, complex procedures and lack of result archival and data transfer mechanism across laboratories.

MLST, proposed way back in 1998 (Maiden *et al.*, 1998), in general has overcome most of the disadvantages of other molecular typing tools and delivered accurate and portable data that can be harnessed for successful epidemiological investigations (Urwin and Maiden, 2003). It utilizes simple PCR technique to amplify specific gene fragments and determination of their nucleotide sequences for further analysis. Thus, success of MLST technique depends only on the success of PCR reaction which in turn depends on the primer annealing. Choice and selection of loci is the key factor for success of MLST and its applicability for molecular typing of organism (Enright and Spratt, 1999; Maiden, 2006).

MLST, which targets the variations present in the multiple loci, is built on the success of multi locus enzyme electrophoresis (MLEE). Variations in the multiple loci are analysed from nucleotide sequences of amplicons. In population with more events of recombination, selection of loci is important and should be done from multiple chromosomal locations that

are unlikely to be co-inherited in a single genetic event and by avoiding parts of chromosome that are evolving rapidly under selection pressures (Urwin and Maiden, 2003).

All unique sequences for a given locus will be assigned an allele number numerically and allele number profile of all MLST loci for a given isolate is considered as the sequence type (ST) which is also numbered numerically. Each and every ST thus represents a unique nucleotide sequence obtained by the combination of allelic sequences of MLST loci. The relationship between isolates is determined by comparing their STs and two isolates are considered to be closely related if both the isolates have same ST or STs that differ at few of the MLST loci (Maiden *et al.*, 1998).

As the outcome of MLST is dependent on the determination of nucleotide sequences, the results can be easily validated, stored electronically and shared across different laboratories with an ease (Enright and Spratt, 1999). With inherent advantages of sharing protocols along with the primer details, transportation of non-infective clinical material i.e. purified DNA, easy automation and scalability of technique from single isolate to hundreds and thousands of isolates, MLST stands out among all other molecular typing tools to characterize isolates (Maiden, 2006).

Design and applicability of MLST scheme depend on three factors, one is choice of isolates to be included during evaluation, second is the choice of genetic loci and third is primer designing for amplification and sequencing of loci (Maiden *et al.*, 1998). Selection of isolates for initial evaluation of the method should include diverse population based on the current typing method, instead of a clonal population taken from a single place or disease outbreak. Choice of genetic loci for MLST should be preferably the house keeping genes which do not change or alter under selection pressure or during genetic events. Primer designing for a genetic locus should amplify and yield a fragment of size around 450bp (Maiden *et al.*, 1998; Enright and Spratt, 1999).

Availability of whole genome sequences has facilitated the better selection of genetic loci for establishing good diversity among isolates using MLST (Parkhill, 2002; Maiden, 2006). Levels of diversity established at each locus will also be examined for success of MLST. The number of loci selected for a scheme depends on the level of resolution achieved by the combination of chosen allele, as lower number of loci will increase chances of association of alleles for different isolates and higher number of alleles makes the scheme expensive.

MLST has played a very important role in epidemiology in investigating the extent of genetic structure in bacterial populations and is considered as a cornerstone technique for molecular typing of bacteria (Pérez-Losada *et al.*, 2006, 2013). Also, its applicability to identify clusters of closely related isolates in outbreaks and epidemics was well documented (Romero, Blanco and Galloway, 2011). Since from the development of first MLST method for *Neisseria meningitides* in 1998, the method has been developed for more than 100 species and genera, mainly bacteria and few eukaryotes (Pérez-Losada *et al.*, 2013). MLST databases containing locus detail, allelic profiles and identifiers, sequence types and isolate information about provenance and pathogenicity for each scheme are hosted majorly on www.pubmlst.org maintained at University of Oxford in United Kingdom, to facilitate the sharing of information over the internet for all researchers (Jolley and Maiden, 2013; Jolley, Bray and Maiden, 2018). The major disadvantage of MLST is the requirement of availability of whole-genome sequence for designing the scheme and variability in the selected candidate genes (Urwin and Maiden, 2003).

MLST method for *Leptospira* characterization was developed and reported by several groups employing different set of gene loci. All these schemes differ at the choice of loci where in Ahmed's scheme used *adk, icdA, secY, rrs2, lipL41, lipL32* genes (Ahmed *et al.*, 2006), Thaipadungpanit's scheme used *pntA, sucA, pfkB, tpiA, mreA, glmU, fadD* genes (Thaipadungpanit *et al.*, 2007), Leon's scheme used *accA2, ccmF, czcA, gcvP, groEL, polA, recF*

genes (Leon *et al.*, 2010), Boonsilp's scheme used *glmU, pntA, sucA, tpiA, pfkB, mreA, caiB* genes (Boonsilp *et al.*, 2013) and Varni's scheme used *adk, glmU, icdA, lipL32*, *lipL41, mreA, pntA* genes (Varni *et al.*, 2014).

Initially Ahmed's scheme is the only scheme that allowed typing of all major pathogenic *Leptospira* species but was evaluated with limited number of strains (n=120) or strains with restricted geographic prevalence. Although, allelic sequences of loci for all the isolates are freely accessible to download for offline analysis, a dedicated online database was not available for the scheme.

Thaipadungpanit's and Leon's schemes have their inherent disadvantage in terms of their limited applicability to only two pathogenic species *L. interrogans* and *L. kirschneri*. Whereas Boonsilp's scheme is the modified scheme of Thaipadungpanit's scheme by excluding *fadD* locus and including *caiB* locus, with modified primer sequences in order to amplify loci from all seven major pathogenic species of *Leptospira* (Boonsilp *et al.*, 2013). Varni's scheme was developed based on the re-assessment of available MLST schemes for *Leptospira* typing, majorly Ahmed's and Boonsilp's schemes, to come up with a consensus scheme that was proposed to have higher level of intra-species discrimination among global strain collection (Varni *et al.*, 2014).

With advent of next-generation sequencing (NGS) platforms, whole-genome sequencing became affordable and more number of bacterial genomes are being sequenced. As a part of *Leptospira* genomics and Human health project of National institute of allergy and infectious diseases (NIAID), a large number of genomes of pathogenic and intermediate *Leptospira* isolates were sequenced and deposited at NCBI, in addition to other genomes available (Lehmann *et al.*, 2014). Various studies have shown that phylogenies constructed using the whole-genome sequences rather than using a small portion of genome yielded better evolutionary relationships and better understanding of the epidemiology, making the WGS

based phylogeny a benchmark for genome-based phylogenetic methods (Foster *et al.*, 2009; Ahrenfeldt *et al.*, 2017). A web-based tool, hosted at www.cbs.dtu.dk/services/MLST, is also designed to extract the MLST typing results from short sequence reads and partially or completely assembled genomes without the need of PCR amplification and sequencing (Larsen *et al.*, 2012). This has opened a window to compare the capability of individual MLST schemes in deciphering relationships among isolates in comparison to whole genome-based relationships.

Availability of different MLST schemes has led to the uncertainty in adaptation of a single uniform scheme that can be applied globally. In this regard the major focus of this part of the study is divided into following **sub-objectives**:

A. Evaluation of Ahmed's MLST scheme for global applicability for *Leptospira* typing
B. Creation of publicly accessible database for Ahmed's MLST scheme
C. Comparison of Ahmed's and Thaipadungpanit's MLST schemes
D. Comparison of *Leptospira* MLST schemes with WGS based phylogeny

2.2 Materials and Methods

2.2.1 Bacterial strains and genomic DNA samples

2.2.1.A **Evaluation of Ahmed's MLST scheme for global applicability for *Leptospira* typing**

A total of 271 *Leptospira* pathogenic isolates **(Table 2),** representing the global collection and wide array of hosts were included in the study. All the strains were cultured to mid logarithmic phase and DNA was isolated at WHO Reference Laboratory for Leptospirosis at the KIT Biomedical Research Centre at The Royal Tropical Institute (currently part of Department of Medical Microbiology at Academic Medical Center) Amsterdam, The Netherlands, the Veterinary Sciences Division (VSD), The Queen's University of Belfast, United Kingdom and Regional Medical Research Centre, Port Blair, India. All these isolates were collected and cultured over a period of time as a part of routine diagnostic/ epidemiological investigation and do not include any cohorts or recruited patients. DNA was isolated at source from late log phase cultures using QIAmp DNA min kit (Qiagen, Germany) following manufacturer's instructions.

2.2.1.C Comparison of Ahmed's and Thaipadungpanit's MLST schemes

A total of 48 isolates **(Table 3)** belonging to only *L. interrogans* (40 no.) and *L. kirschneri* (8 no.) species were included in the study as the Thaipadungpanit's scheme has restriction for its applicability for these two species. 17 of these isolates were from reference collection and 31 were clinical isolates collected as a part of routine diagnostic and epidemiological investigations in Thailand. Reference isolates were cultured at WHO/FAO/OIE Collaborating Centers for Reference and Research on Leptospirosis located at Centre for Public Health Sciences, Queensland Health Scientific Services, Brisbane, Australia and Department of Biomedical Research, Royal Tropical Institute (KIT), Amsterdam, Netherlands. Clinical isolates were cultured at Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University MORU, Bangkok, Thailand. DNA

was isolated at source from late log phase cultures using QIAmp DNA min kit (Qiagen, Germany) following manufacturer's instructions.

2.2.1.D Comparison of *Leptospira* MLST schemes with WGS based phylogeny

Leptospira MLST schemes by Boonsilp's, Varni's and Ahmed's were considered for comparison with WGS based phylogeny. A total of 149 isolates **(Table 4)** for which whole genome sequences were available at public databases, either in complete or in partial status, were included in the study. Selection of genomes was random and a search term containing "*Leptospira*" was used to retrieve information from NCBI genome database. This search has yielded a total of 276 different *Leptospira* isolates with whole genome sequences deposited in complete or partial status during the initiation of the study in 2014. Genomes with more than 5000 contigs and of non-pathogenic *Leptospira* isolates were excluded from the study. Genomes representing all major pathogenic species, possible hosts and geographic locations were considered for the analysis. This criteria has ended with a total of 224 isolates belongs to *L. interrogans* (n=154), *L. borgpetersenii* (n=14), *L. kirschneri* (n=16), *L. santarosai* (n=20), *L.noguchii* (n=8), *L. weilii* (n=7), *L. alexanderi* (n=1) and an isolate of unknown *Leptospira* species.

Further, a large set consisting of 78 isolates **(Table 5)** belonging to serovar Copenhageni of *L. interrogans* species were found to be isolated at Fiocruz, Brazil and were analyzed for clonality among one another using Varni's MLST scheme and whole genome based phylogeny. Whole genomes of these 76 isolates were used to deduce sequence types using web based *in-silico* MLST as per the methodology described in section 2.2.4 and whole-genome based phylogeny was constructed using neighbor joining algorithm using Gegenees version 2.2.1 program as per the methodology described in section 2.2.8. A total of 6 representative isolates of the set were chosen to include in the final study making the *L.interrogans* isolates count to a total of 82 and total genomes to 149.

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme for global applicability

S.No	Genome species	Serogroup	Serovar	Strain	Country	Host
1	*L. interrogans*	Ictero	Copenhageni	Fiocruz L1-130	Brazil	Human
2	*L. interrogans*	Ictero	Lai	56601	China	Human
3	*L. interrogans*	Sehgali	Portblari	DS2	Andaman	Human
4	*L. interrogans*	Australis	Australis	Ballico	Australia	Human
5	*L. interrogans*	Ictero	Ictero	RGA	Belgium	Human
6	*L. interrogans*	Canicola	Sumneri	Sumner	Malaysia	Human
7	*L. interrogans*	Canicola	Portlandere	MY1039	Jamaica	Human
8	*L. interrogans*	Pomona	Pomona	Pomona	Australia	Unknown
9	*L. interrogans*	Pomona	Proechimys	1161 U	Panama	Rat
10	*L. interrogans*	Pomona	Kenniwicki	LT1026	USA	Bovine
11	*L. interrogans*	Grippotyphosa	Grippotyphosa	Moskva Y	Russia	Unknown
12	*L. interrogans*	Grippotyphosa	Muelleri	RM2	Malaysia	Unknown
13	*L. interrogans*	Hebdomadis	Goiano	Unknown	Unknown	Unknown
14	*L. interrogans*	Sejroe	Roumanica	LM 294	Romania	Mouse
15	*L. interrogans*	Sejroe	Saxkoebing	Mus24	Denmark	Mouse
16	*L. interrogans*	Sejroe	Hardjo type prajitno	Hardjoprajitno	Indonesia	Human
17	*L. interrogans*	Ictero	Lai	Lai	China	Human
18	*L. interrogans*	Ictero	Copenhageni	M20	Denmark	Rat
19	*L. interrogans*	Grippotyphosa	Valbuzzi	Valbuzzi	Australia	Human
20	*L. interrogans*	Pyrogenes	Manilae	LT398	Philippines	Rat
21	*L. interrogans*	Australis	Fugis	Fudge	Malaysia	Human
22	*L. interrogans*	Australis	Hawain	LT 62-68	New Guinea	Bandicoot
23	*L. interrogans*	Australis	Lora	Lora	Italy	Human
24	*L. interrogans*	Australis	Muenchen	C 90	Germany	Human
25	*L. interrogans*	Autumnalis	Bangkinang	Bangkinang I	Indonesia	Human
26	*L. interrogans*	Autumnalis	Carlos	C 3	Philippines	Toad
27	*L. interrogans*	Autumnalis	Mooris	Moores	Malaysia	Human
28	*L. interrogans*	Autumnalis	Nanla	A 6	China	Human
29	*L. interrogans*	Autumnalis	Weerasinghe	Weerasinghe	Sri Lanka	Human
30	*L. interrogans*	Bataviae	Bataviae	Swart	Indonesia	Human
31	*L. interrogans*	Bataviae	Losbanos	LT 101-69	Philippines	Rat
32	*L. interrogans*	Bataviae	Paidjan	Paidjan	Indonesia	Human
33	*L. interrogans*	Canicola	Benjamini	Benjamin	Indonesia	Human
34	*L. interrogans*	Canicola	Bindjei	Bindjei	Indonesia	Human
35	*L. interrogans*	Canicola	Broomii	Patane	Australia	Human
36	*L. interrogans*	Canicola	Jonsis	Jones	Malaysia	Human
37	*L. interrogans*	Canicola	Malaya	H 6	Malaysia	Human
38	*L. interrogans*	Djasiman	Djasiman	Djasiman	Indonesia	Human
39	*L. interrogans*	Djasiman	Gurungi	Gurung	Malaysia	Human
40	*L. interrogans*	Djasiman	Huallaga	M 7	Peru	Opossum
41	*L. interrogans*	Djasiman	Sentot	Sentot	Indonesia	Human
42	*L. interrogans*	Grippotyphosa	Muelleri	RM 2	Malaysia	Rat
43	*L. interrogans*	Ictero	Gem	Simon	Sri Lanka	Human
44	*L. interrogans*	Ictero	Hongchon	18R	Korea	Mouse
45	*L. interrogans*	Ictero	Smithi	Smith	Malaysia	Human
46	*L. interrogans*	Ictero	Yeonchon	HM 3	Korea	Human
47	*L. interrogans*	Javanica	Kalimantani	Amos	Indonesia	Unknown
48	*L. interrogans*	Louisiana	Lanka	R 740	Sri Lanka	Human

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme *(Continued)*

S.No	Genome species	Serogroup	Serovar	Strain	Country	Host
49	*L. interrogans*	Pyrogenes	Abramis	Abraham	Malaysia	Human
50	*L. interrogans*	Pyrogenes	Biggis	Biggs	Malaysia	Human
51	*L. interrogans*	Pyrogenes	Camlo	LT 64-67	Vietnam	Human
52	*L. interrogans*	Pyrogenes	Guaratuba	An 7705	Brazil	Opossum
53	*L. interrogans*	Pyrogenes	Pyrogenes	Salinem	Indonesia	Human
54	*L. interrogans*	Pyrogenes	Robinsoni	Robinson	Australia	Human
55	*L. interrogans*	Pyrogenes	Zanoni	Zanoni	Australia	Human
56	*L. interrogans*	Sejroe	Geyaweera	Geyaweera	Sri Lanka	Human
57	*L. interrogans*	Sejroe	Haemolytica	Marsh	Malaysia	Human
58	*L. interrogans*	Sejroe	Ricardi	Richardson	Malaysia	Human
59	*L. interrogans*	Sejroe	Saxkoebing	Mus 24	Denmark	Mouse
60	*L. interrogans*	Sejroe	Wolffi	3705	Indonesia	Human
61	*L. interrogans*	Canicola	Canicola	M12/90	Brazil	Dog
62	*L. interrogans*	Ictero	Copenhageni	M9/99	Brazil	Rat
63	*L. interrogans*	Australis	Rushan	L01	Brazil	Dog
64	*L. interrogans*	Canicola	Canicola	L02	Brazil	Dog
65	*L. interrogans*	Canicola	Canicola	L03	Brazil	Swine
66	*L. interrogans*	Canicola	Canicola	L09	Brazil	Cow
67	*L. interrogans*	Ictero	Copenhageni	L10	Brazil	Cow
68	*L. interrogans*	Canicola	Canicola	L14	Brazil	Cow
69	*L. interrogans*	Lyme	Lyme	K30B	UK	Mouse
70	*L. interrogans*	Australis	Australis	K9H	UK	Mouse
71	*L. interrogans*	Ictero	Copenhageni	K13A	UK	Unknown
72	*L. interrogans*	Unknown	Unknown	Isolate 7	Costa Rica	Human
73	*L. interrogans*	Shermani	Unknown	Isolate 8	Costa Rica	Human
74	*L. interrogans*	Ictero	Copenhageni	Isolate 9	Costa Rica	Human
75	*L. interrogans*	Unknown	Unknown	Isolate 10	Costa Rica	Human
76	*L. interrogans*	Australis	Lora	1992	Tanzania	Mastomys
77	*L. interrogans*	Australis	Lora	2324	Tanzania	Crocedura
78	*L. interrogans*	Australis	Lora	2364	Tanzania	Mastomys
79	*L. interrogans*	Australis	Lora	2366	Tanzania	Mastomys
80	*L. interrogans*	Ballum	Kenya	4885	Tanzania	Crocedura
81	*L. interrogans*	Ballum	Kenya	4883	Tanzania	Crocedura
82	*L. interrogans*	Hebdomadis	Hebdomadis	Hebdomadis	Unknown	Unknown
83	*L. interrogans*	Grippotyphosa	Grippotyphosa	CH31	India	Human
84	*L. interrogans*	Ballum	Ballum	Mus127	Denmark	Mouse
85	*L. interrogans*	Grippotyphosa	Valbuzzi	DS15	India	Unknown
86	*L. interrogans*	Grippotyphosa	Valbuzzi	DS18	India	Human
87	*L. interrogans*	Grippotyphosa	Valbuzzi	D22	India	Unknown
88	*L. interrogans*	Grippotyphosa	Valbuzzi	DCHCF-30	India	Human
89	*L. interrogans*	Grippotyphosa	Valbuzzi	Duyster-H2	Unknown	Unknown
90	*L. interrogans*	Grippotyphosa	Unknown	ICI Pod 179	Unknown	Unknown
91	*L. interrogans*	Grippotyphosa	Ratnapura	GC-1	Andaman	Human
92	*L. interrogans*	Ictero	Copenhageni	GC-3	Andaman	Unknown
93	*L. interrogans*	Grippotyphosa	Ratnapura	TB-6	Andaman	Unknown
94	*L. interrogans*	Grippotyphosa	Ratnapura	TB-19	Andaman	Unknown
95	*L. interrogans*	Grippotyphosa	Valbuzzi	JAMES	Andaman	Human
96	*L. interrogans*	Ictero	Copenhageni	Yasuodamma	Andaman	Human
97	*L. interrogans*	Grippotyphosa	Valbuzzi	DS-18	Andaman	Unknown

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme *(Continued)*

S.No	Genome species	Serogroup	Serovar	Strain	Country	Host
98	*L. interrogans*	Grippotyphosa	Valbuzzi	DCHCF-3	Andaman	Human
99	*L. interrogans*	Grippotyphosa	Ratnapura	MG-11	Andaman	Unknown
100	*L. interrogans*	Grippotyphosa	Ratnapura	MG-17	Andaman	Unknown
101	*L. interrogans*	Grippotyphosa	Ratnapura	MG-23	Andaman	Unknown
102	*L. interrogans*	Hebdomadis	Hebdomadis	MG-37	Andaman	Unknown
103	*L. interrogans*	Grippotyphosa	Unknown	MG-47	Andaman	Unknown
104	*L. interrogans*	Sejroe	Saxkoebing	MG-73	Andaman	Unknown
105	*L. interrogans*	Pomona	Unknown	MG-90	Andaman	Human
106	*L. interrogans*	Grippotyphosa	Ratnapura	MG-100	Andaman	Human
107	*L. interrogans*	Australis	Ramisi	MG-347	Andaman	Unknown
108	*L. interrogans*	Grippotyphosa	Unknown	MG-79	Andaman	Unknown
109	*L. interrogans*	Grippotyphosa	Valbuzzi	MG-342	Andaman	Human
110	*L. interrogans*	Grippotyphosa	Valbuzzi	MG-373	Andaman	Unknown
111	*L. interrogans*	Australis	Australis	MG-375	Andaman	Unknown
112	*L. interrogans*	Australis	Australis	MG-392	Andaman	Unknown
113	*L. interrogans*	Grippotyphosa	Valbuzzi	MG-472	Andaman	Unknown
114	*L. interrogans*	Canicola	Canicola	H-12	South India	Unknown
115	*L. interrogans*	Autumnalis	Unknown	AUT(N)	South India	Unknown
116	*L. interrogans*	Canicola	Unknown	PAI	South India	Unknown
117	*L. interrogans*	Ictero	Unknown	Thahkchan	South India	Unknown
118	*L. interrogans*	Canicola	Unknown	G-1	Central India	Human
119	*L. interrogans*	Canicola	Unknown	G-2	Central India	Human
120	*L. interrogans*	Canicola	Unknown	G-3	Central India	Human
121	*L. interrogans*	Djasmin	Unknown	G-4	Central India	Human
122	*L. interrogans*	Bataviae	Unknown	G-5	Central India	Unknown
123	*L. interrogans*	Canicola	Unknown	G-6	Central India	Human
124	*L. interrogans*	Canicola	Unknown	G-7	Central India	Human
125	*L. interrogans*	Canicola	Unknown	G-8	Central India	Human
126	*L. interrogans*	Canicola	Unknown	G-10	Central India	Human
127	*L. interrogans*	Grippotyphosa	Unknown	ALC-10	South India	Human
128	*L. interrogans*	Hebdomadis	Unknown	ALC-11	South India	Unknown
129	*L. interrogans*	Pomona	Unknown	H-3	South India	Human
130	*L. interrogans*	Pomona	Unknown	H-41	South India	Unknown
131	*L. interrogans*	Pomona	Unknown	H-61	South India	Unknown
132	*L. interrogans*	Pomona	Unknown	H-518	South India	Renal
133	*L. interrogans*	Pomona	Unknown	H-578	South India	Unknown
134	*L. interrogans*	Pomona	Unknown	289-M.C.Calicut	South India	Unknown
135	*L. kirschneri*	Grippotyphosa	Grippotyphosa	Moskva V	Russia	Human
136	*L. kirschneri*	Grippotyphosa	Ratnapura	Wumalasena	Sri Lanka	Human
137	*L. kirschneri*	Cynopteri	Cynopteri	3522_C	Indonesia	Bat
138	*L. kirschneri*	Canicola	Kuwait	136/2/2	Kuwait	Rat
139	*L. kirschneri*	Canicola	Schueffneri	Vleermuis 90C	Indonesia	Bat
140	*L. kirschneri*	Pomona	Mozdok	5621	Russia	Vole
141	*L. kirschneri*	Grippotyphosa	Vanderhoedeni	Kipod 179	Israel	Hedgehog
142	*L. kirschneri*	Pomona	Tsaratsovo	B81/7	Bulgaria	Mouse
143	*L. kirschneri*	Autumnalis	Bulgarica	Nikolaevo	Bulgaria	Human
144	*L. kirschneri*	Autumnalis	Butembo	Butembo	Zaire	Human

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme *(Continued)*

S.No	Genome species	Serogroup	Serovar	Strain	Country	Host
145	*L. kirschneri*	Autumnalis	Erinaceiauriti	Erinaceus Auritus 670	Russia	Hedgehog
146	*L. kirschneri*	Autumnalis	Lambwe	Lambwe	Kenya	Grass Rat
147	*L. kirschneri*	Autumnalis	Mujunkumi	Yezsh 237	Kazakhstan	Hedgehog
148	*L. kirschneri*	Bataviae	Djatzi	HS 26	Puerto Rico	Human
149	*L. kirschneri*	Canicola	Bafani	Bafani	Zaire	Human
150	*L. kirschneri*	Djasmin	Agogo	Agogo	Ghana	Human
151	*L. kirschneri*	Ictero	Bogvere	LT 60-69	Jamaica	Rat
152	*L. kirschneri*	Ictero	Zimbabwe	SBF 23	Zimbabwe	Cattle
153	*L. kirschneri*	Pomona	Kunming	K5	China	Mouse
154	*L. kirschneri*	Grippotyphosa	Valbuzzi	Duyster	Netherlands	Bovine
155	*L. kirschneri*	Ictero	Sokoine	745	Tanzania	GP Rat
156	*L. kirschneri*	Ictero	Sokoine	771	Tanzania	GP Rat
157	*L. kirschneri*	Ictero	Mwogolo	826	Tanzania	GP Rat
158	*L. kirschneri*	Ictero	Mwogolo	845	Tanzania	GP Rat
159	*L. kirschneri*	Canicola	Qunjian	2980	Tanzania	GP Rat
160	*L. kirschneri*	Ictero	Sokoine	4602	Tanzania	GP Rat
161	*L. kirschneri*	Sejroe	Ricardi /Saxkoebing	1499	Ireland	Bank Vole
162	*L. kirschneri*	Sejroe	Ricardi /Saxkoebing	1501	Ireland	Wood Mouse
163	*L. kirschneri*	Unknown	Kenya	Nijenga	Kenya	Rat
164	*L. santarosai*	Grippotyphosa	Canalzonae	CZ188	Panama	Rat
165	*L. santarosai*	Shermani	Shermani	1342K	Panama	Rat
166	*L. santarosai*	Mini	Georgia	LT117	USA	Unknown
167	*L. santarosai*	Sejroe	Caribe	Unknown	Unknown	Unknown
168	*L. santarosai*	Pyrogenes	Guaratuva	An7705	Brazil	Opossum
169	*L. santarosai*	Sejroe	Recreo	380	Nicaragua	Unknown
170	*L. santarosai*	Pyrogenes	Varella	1019	Nicaragua	Unknown
171	*L. santarosai*	Autumnalis	Alice	Alice	Sri Lanka	Human
172	*L. santarosai*	Ballum	Peru	MW 10	Peru	Opossum
173	*L. santarosai*	Bataviae	Balboa	735 U	Panama	Spiny Rat
174	*L. santarosai*	Bataviae	Brasiliensis	An 776	Brazil	Opossum
175	*L. santarosai*	Bataviae	Kobbe	CZ 320	Panama	Spiny Rat
176	*L. santarosai*	Bataviae	Rioja	MR 12	Peru	Opossum
177	*L. santarosai*	Cynopteri	Tingomaria	M 13	Peru	Opossum
178	*L. santarosai*	Grippotyphosa	Huanuco	M 4	Peru	Opossum
179	*L. santarosai*	Javanica	Fluminense	Aa 3	Brazil	Field Mouse
180	*L. santarosai*	Mini	Tabaquite	TRVL 3214	Trinidad	Human
181	*L. santarosai*	Pyrogenes	Alexi	HS 616	Puerto Rico	Human
182	*L. santarosai*	Pyrogenes	Princestown	TRVL 112499	Trinidad	Human
183	*L. santarosai*	Sarmin	Machiguenga	MMD 3	Peru	Opossum
184	*L. santarosai*	Sarmin	Rio	Rr 5	Brazil	Rat
185	*L. santarosai*	Sarmin	Weaveri	CZ390	Panama	Human
186	*L. santarosai*	Sejroe	Gorgas	1413 U	Panama	Spiny Rat
187	*L. santarosai*	Shermani	Babudieri	CI 40	Peru	Pig
188	*L. santarosai*	Shermani	Luis	M 6	Peru	Opossum
189	*L. santarosai*	Tarassovi	Atchafalaya	LSU 1013	USA	Opossum
190	*L. santarosai*	Tarassovi	Atlantae	LT 81	USA	Opossum
191	*L. santarosai*	Tarassovi	Bakeri	LT 79	USA	Opossum

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme *(Continued)*

S.No	Genome species	Serogroup	Serovar	Strain	Country	Host
192	*L. santarosai*	Tarassovi	Chagres	1913 K	Panama	Spiny Rat
193	*L. santarosai*	Tarassovi	Darien	637 K	Panama	Opossum
194	*L. santarosai*	Tarassovi	Gatuni	1473 K	Panama	Opossum
195	*L. santarosai*	Tarassovi	Rama	316	Nicaragua	Opossum
196	*L. santarosai*	Javanica	Vargonicas	24	Peru	Rodent
197	*L. santarosai*	Bataviae	Brasiliensis	An 776	Brazil	Opossum
198	*L. santarosai*	Sejroe	Guaricura	Bov.G	Brazil	Cow
199	*L. santarosai*	Sejroe	Guaricura	M4/98	Brazil	Buffalo
200	*L. santarosai*	Grippotyphosa	Bananal	2ACAP	Brazil	Capybara
201	*L. santarosai*	Grippotyphosa	Bananal	16CAP	Brazil	Capybara
202	*L. santarosai*	Pyrogenes	Unknown	Isolate 1	Costa Rica	Human
203	*L. santarosai*	Sarmin	Weaveri/Rio	Isolate 2	Costa Rica	Human
204	*L. santarosai*	Tarassovi	Rama	Isolate 3	Costa Rica	Human
205	*L. santarosai*	Tarassovi	Rama	Isolate 5	Costa Rica	Human
206	*L. santarosai*	Bataviae	Claytoni	Isolate 6	Costa Rica	Human
207	*L. borgpetersenii*	Javanica	Poi	Poi	Italy	Human
208	*L. borgpetersenii*	Mini	Mini	Sari	Italy	Human
209	*L. borgpetersenii*	Sejroe	Istrica	Bratislava M84	Slovakia	Unknown
210	*L. borgpetersenii*	Sejroe	Sejroe	M84	Denmark	Unknown
211	*L. borgpetersenii*	Javanica	Dehong	De 10	China	Unknown
212	*L. borgpetersenii*	Javanica	Javanica	Veldrat Batavia	Indonesia	Unknown
213	*L. borgpetersenii*	Javanica	Zhenkang	L 82	China	Rat
214	*L. borgpetersenii*	Australis	Pina	LT 932	Panama	Opossum
215	*L. borgpetersenii*	Cellodoni	Whitcombi	Whitcomb	Malaysia	Human
216	*L. borgpetersenii*	Ictero	Tonkini	LT96-68	Vietnam	Human
217	*L. borgpetersenii*	Javanica	Ceylonica	Piyasena	Sri Lanka	Human
218	*L. borgpetersenii*	Javanica	Javanica	VeldratBatavia46	Indonesia	Rat
219	*L. borgpetersenii*	Javanica	Menoni	Kerala	India	Bandicoot
220	*L. borgpetersenii*	Javanica	Sorexjalna	Sorex Jalna	Czechoslovakia	Shrew
221	*L. borgpetersenii*	Pyrogenes	Kwale	Julu	Kenya	Human
222	*L. borgpetersenii*	Sejroe	Balcanica	1627 Burgas	Bulgaria	Human
223	*L. borgpetersenii*	Sejroe	Nyanza	Kibos	Kenya	Human
224	*L. borgpetersenii*	Tarassovi	Gengma	M 48	China	Pig
225	*L. borgpetersenii*	Tarassovi	Kisuba	Kisuba	Zaire	Human
226	*L. borgpetersenii*	Tarassovi	Tarassovi	Perepelitsin	Russia	Human
227	*L. borgpetersenii*	Tarassovi	Tunis	P 2/65	Tunisia	Pig
228	*L. borgpetersenii*	Tarassovi	Yunxian	L 100	China	Pig
229	*L. borgpetersenii*	Canicola	Canicola	HondUtrecht IV	Netherlands	Dog
230	*L. borgpetersenii*	Ballum	Kenya	153	Tanzania	Mastomys
231	*L. borgpetersenii*	Ballum	Kenya	159	Tanzania	MuskShrews
232	*L. borgpetersenii*	Ballum	Kenya	723	Tanzania	MuskShrews
233	*L. borgpetersenii*	Ballum	Kenya	766	Tanzania	GPRat
234	*L. borgpetersenii*	Ballum	Kenya	1605	Tanzania	MuskShrews
235	*L. borgpetersenii*	Ballum	Kenya	1610	Tanzania	MuskShrews
236	*L. borgpetersenii*	Ballum	Kenya	2062	Tanzania	MuskShrews
237	*L. borgpetersenii*	Ballum	Kenya	2348	Tanzania	MuskShrews
238	*L. borgpetersenii*	Ballum	Kenya	2447	Tanzania	MuskShrews
239	*L. borgpetersenii*	Ballum	Kenya	4880	Tanzania	Mouse
240	*L. borgpetersenii*	Ballum	Kenya	4787	Tanzania	MuskShrews

Table 2: Details of *Leptospira* isolates used in evaluation of Ahmed's MLST scheme *(Continued)*

S.No	Genome species	Serogroup	Serovar	Strain	Country	Host
241	*L. borgpetersenii*	Hebdomadis	Kremastos/ Hebdomadis	873	Ireland	Dog
242	*L. borgpetersenii*	Hebdomadis	Kremastos/ Hebdomadis	871	Ireland	Pygmy Shrew
243	*L. borgpetersenii*	Sejroe	Saxkoebing	1498	Ireland	GemsBuck
244	*L. borgpetersenii*	Sejroe	Ricardi / Saxkoebing	1522	Ireland	Wood Mouse
245	*L. borgpetersenii*	Sejroe	Ricardi/ Saxkoebing	1525	Ireland	Dog
246	*L. borgpetersenii*	Pomona	Kunming	RIM 139	Portugal	Mouse
247	*L. borgpetersenii*	Pomona	Kunming	RIM 201	Portugal	Mouse
248	*L. borgpetersenii*	Sejroe	Ricardi / Saxkoebing	RIM 156	Portugal	Mouse
249	*L. borgpetersenii*	Unknown	Sokoine	RM1	Laos	Unknown
250	*L. noguchii*	Louisiana	Louisiana	LSU_1945	USA	Armadillo
251	*L. noguchii*	Panama	Panama	Cz214k	Panama	Opossum
252	*L. noguchii*	Pyrogenes	Myocastoris	LSU 1551	USA	Unknown
253	*L. noguchii*	Autumnalis	Fortbragg	Fort Bragg	USA	Human
254	*L. noguchii*	Bataviae	Argentiniensis	Peludo	Argentina	Armadillo
255	*L. noguchii*	Bataviae	Claytoni	1348 U	Panama	Spiny Rat
256	*L. noguchii*	Louisiana	Orleans	LSU 2580	USA	Coypu
257	*L. noguchii*	Panama	Cristobali	1996 K	Panama	Opossum
258	*L. noguchii*	Pyrogenes	Guaratuba	Isolate 4	Costa Rica	Human
259	*L. weilii*	Cellodoni	Cellodoni	Cellodoni	Unknown	Unknown
260	*L. weilii*	Cellodoni	Hainan-Whitcombi	6712	China	Human
261	*L. weilii*	Javanica	Coxi	Cox	Malaysia	Human
262	*L. weilii*	Javanica	Mengma	S 590	China	Human
263	*L. weilii*	Javanica	Menrun	A 102	China	Human
264	*L. weilii*	Manhao	Qingshui (Manhao2)	L105	China	Human
265	*L. weilii*	Mini	Hekou	H 27	China	Human
266	*L. weilii*	Tarassovi	Ngavi	SBF 16	Zimbabwe	Cattle
267	*L. inadai*	Manhao	Lincang	L 14	China	Human
268	*L. inadai*	Panama	Mangus	TRVL/CAREC 137774	Trinidad	Mongoose
269	*L. inadai*	Shermani	Aguaruma	MW 4	Peru	Opossum
270	*L. meyeri*	Mini	Perameles	Bandicoot 343	Australia	Perameles
271	*L. alexanderi*	Manhao	Manhao3	L60	China	Human

UK – United Kingdom, USA – United States of America, GP Rat - Giant Pouched rat, Ictero - Icterohaemorrhagiae

Table 3: Details of *Leptospira* isolates employed for comparing Ahmed's and Thaipadungpanit's scheme

S.No	Genomospecies	Serovar	Strain	Country	Host	Source
1.	*L. interrogans*	Copenhageni	M20	Denmark	Human	Reference
2.	*L. interrogans*	Guaratuba	An7705	Brazil	Opossum	Reference
3.	*L. interrogans*	Hardjo	Hardjoprajitno	Indonesia	Human	Reference
4.	*L. interrogans*	Ictero	RGA	Belgium	Human	Reference
5.	*L. interrogans*	Kenniwicki	LT1026	USA	Unknown	Reference
6.	*L. interrogans*	Kuwait	136/2/2	Kuwait	Rat	Reference
7.	*L. interrogans*	Lai	Lai	China	Human	Reference
8.	*L. interrogans*	Pomona	Pomona	Australia	Human	Reference
9.	*L. interrogans*	Portlandvere	MY1039	Jamaica	Human	Reference
10.	*L. interrogans*	Scheuffneri	Vleermuis 90C	Indonesia	Bat	Reference
11.	*L. interrogans*	Sumneri	Sumner	Malaysia	Human	Reference
12.	*L. interrogans*	Valbuzzi	Valbuzzi	Australia	Human	Reference
13.	*L. interrogans*	Autumnalis	3	Thailand	Human	Clinical
14.	*L. interrogans*	Autumnalis	86	Thailand	Human	Clinical
15.	*L. interrogans*	Autumnalis	L0020	Thailand	Human	Clinical
16.	*L. interrogans*	Autumnalis	L0661	Thailand	Human	Clinical
17.	*L. interrogans*	Autumnalis	L1151	Thailand	Human	Clinical
18.	*L. interrogans*	Autumnalis	UT227	Thailand	Human	Clinical
19.	*L. interrogans*	Autumnalis	548	Thailand	Human	Clinical
20.	*L. interrogans*	Autumnalis	729	Thailand	Human	Clinical
21.	*L. interrogans*	Autumnalis	LP101	Thailand	Human	Clinical
22.	*L. interrogans*	Bataviae	L1111	Thailand	Human	Clinical
23.	*L. interrogans*	Bataviae	UT229	Thailand	Human	Clinical
24.	*L. interrogans*	Bataviae	UT234	Thailand	Human	Clinical
25.	*L. interrogans*	Medanensis	L0448	Thailand	Human	Clinical
26.	*L. interrogans*	Medanensis	L0887	Thailand	Human	Clinical
27.	*L. interrogans*	Medanensis	L0941	Thailand	Human	Clinical
28.	*L. interrogans*	Unknown	UT364	Thailand	Human	Clinical
29.	*L. interrogans*	Pyrogenes	UD009	Thailand	Human	Clinical
30.	*L. interrogans*	Pyrogenes	L0443	Thailand	Human	Clinical
31.	*L. interrogans*	Pyrogenes	L0374	Thailand	Human	Clinical
32.	*L. interrogans*	Unknown	654	Thailand	Human	Clinical
33.	*L. interrogans*	Unknown	M04	Thailand	Human	Clinical
34.	*L. interrogans*	Unknown	M08	Thailand	Human	Clinical
35.	*L. interrogans*	Unknown	UT126	Thailand	Human	Clinical
36.	*L. interrogans*	Unknown	L1085	Thailand	Human	Clinical
37.	*L. interrogans*	Unknown	L0996	Thailand	Human	Clinical
38.	*L. interrogans*	Unknown	UT053	Thailand	Human	Clinical
39.	*L. interrogans*	Unknown	M10	Thailand	Human	Clinical
40.	*L. interrogans*	Unknown	L1207	Thailand	Human	Clinical
41.	*L. kirschneri*	Grippotyphosa	Moskva V	Russia	Human	Reference
42.	*L. kirschneri*	Mozdok	5621	Russia	Vole	Reference
43.	*L. kirschneri*	Ratnapura	Wumalasena	Sri Lanka	Human	Reference
44.	*L. kirschneri*	Tsaratsovo	B 81/7	Bulgaria	Mouse	Reference
45.	*L. kirschneri*	Vanderhoedeni	Kipod 179	Israel	Hedgehog	Reference
46.	*L. kirschneri*	Grippotyphosa	UT130	Thailand	Human	Clinical
47.	*L. kirschneri*	Unknown	M06	Thailand	Human	Clinical
48.	*L. kirschneri*	Unknown	M07	Thailand	Human	Clinical

Ictero - Icterohaemorrhagiae

Table 4: Details of *Leptospira* isolates used for comparison of *Leptospira* MLST schemes with WGS based phylogeny

S. No	Genomospecies	Serogroup	Serovar	Strain	Country	Host
1	*L. alexanderi* [2,3]	Manhao	Manhao 3	L60	China	Human
2	*L. borgpetersenii*	Ballum	Ballum	Muis5	Netherlands	Mouse
3	*L. borgpetersenii*	Unknown	Castellonis	200801910	Guadeloupe	Human
4	*L. borgpetersenii*	Sejroe	Hardjo-Bovis	JB197	USA	Bovine
5	*L. borgpetersenii*	Sejroe	Hardjo-Bovis	L550	Australia	Human
6	*L. borgpetersenii*	Sejroe	Hardjo-Bovis	Lely 607	Netherlands	Cow
7	*L. borgpetersenii*	Javanica	Javanica	L0864	Thailand	Human
8	*L. borgpetersenii*	Javanica	Javanica	MK146	Thailand	Human
9	*L. borgpetersenii*	Javanica	Javanica	UI09931	Laos	Human
10	*L. borgpetersenii*	Javanica	Javanica	UI09149	Laos	Human
11	*L. borgpetersenii*	Ballum	Kenya	TE 0159	Tanzania	MuskShrews
12	*L. borgpetersenii*	Javanica	Poi	Brem 171	Unknown	Unknown
13	*L. borgpetersenii*	Mini	Unknown	200801926	Mayotte	Human
14	*L. borgpetersenii*	Unknown	Unknown	Brem 328	Germany	Horse
15	*L. borgpetersenii*	Unknown	Unknown	Noumea 25	NewCaledonia	Unknown
16	*L. interrogans* [1]	Australis	Australis	200703203	French Polynesia	Human
17	*L. interrogans*	Autumnalis	Autumnalis	LP101	Thailand	Human
18	*L. interrogans*	Bataviae	Bataviae	Kariadi-Satu	Unknown	Human
19	*L. interrogans*	Bataviae	Bataviae	L1111	Thailand	Human
20	*L. interrogans*	Bataviae	Bataviae	Swart	Indonesia	Human
21	*L. interrogans*	Bataviae	Bataviae	UI08561	Laos	Human
22	*L. interrogans*	Bataviae	Bataviae	UT234	Thailand	Human
23	*L. interrogans*	Bataviae	Bataviae	HAI135	Peru: Iquitos	Unknown
24	*L. interrogans*	Unknown	Bim	P2529	Unknown	Unknown
25	*L. interrogans*	Unknown	Bratislava	Brem 137	Unknown	Unknown
26	*L. interrogans*	Autumnalis	Bulgarica	Mallika	India	Human
27	*L. interrogans*	Canicola	Canicola	FiocruzLV133	Brazil	Human
28	*L. interrogans*	Canicola	Canicola	HAI0024	Peru: Iquitos	Human
29	*L. interrogans*	Canicola	Canicola	LT1962	Taiwan	Human
30	*L. interrogans*	Canicola	Canicola	P2655	Portugal	Mouse
31	*L. interrogans* [1]	Ictero	Copenhageni	2006006972	Unknown	Unknown
32	*L. interrogans*	Ictero	Copenhageni	2007005490	Unknown	Unknown
33	*L. interrogans*	Ictero	Copenhageni	FiocruzLV130	Brazil	Human
34	*L. interrogans*	Ictero	Copenhageni	FiocruzLV2953	Brazil	Human
35	*L. interrogans*	Ictero	Copenhageni	FiocruzLV3094	Brazil	Human
36	*L. interrogans* [1,2,3]	Ictero	Copenhageni	FiocruzLV3726	Brazil	Human
37	*L. interrogans*	Ictero	Copenhageni	FiocruzLV4034	Unknown	Unknown
38	*L. interrogans*	Ictero	Copenhageni	Fiocruz R83	Brazil	Rat
39	*L. interrogans*	Ictero	Copenhageni	HAI0156	Peru: Iquitos	Human
40	*L. interrogans*	Ictero	Copenhageni	LT2050	South America	Human

Table 4: Details of *Leptospira* isolates used for comparison of *Leptospira* MLST schemes *(Continued)*

S. No	Genomospecies	Serogroup	Serovar	Strain	Country	Host
41	*L. interrogans* [1,2,3]	Ictero	Copenhageni	MMD1562	Peru: Iquitos	Bat
42	*L. interrogans*	Ictero	Copenhageni	P2431	Unknown	Unknown
43	*L. interrogans*	Ictero	Copenhageni	R066	Colombia	Human
44	*L. interrogans*	Djasiman	Djasiman	LT1649	Thailand	Human
45	*L. interrogans*	Grippotyphosa	Grippotyphosa	2006006986	Egypt	Human
46	*L. interrogans*	Grippotyphosa	Grippotyphosa	Andaman	Unknown	Unknown
47	*L. interrogans*	Grippotyphosa	Grippotyphosa	LT2186	Thailand	Human
48	*L. interrogans*	Grippotyphosa	Grippotyphosa	UI08434	Laos	Human
49	*L. interrogans*	Hebdomadis	Hebdomadis	L0996	Thailand	Human
50	*L. interrogans*	Hebdomadis	Hebdomadis	R499	Sri Lanka	Human
51	*L. interrogans*	Ictero	Ictero	201100516	Unknown	Unknown
52	*L. interrogans*	Ictero	Ictero	Kantorowic	Unknown	Unknown
53	*L. interrogans*	Ictero	Ictero	P2422	Unknown	Unknown
54	*L. interrogans*	Ictero	Ictero	RGA	Belgium	Human
55	*L. interrogans*	Ictero	Ictero	Verdun LP	France	Human
56	*L. interrogans*	Unknown	Jalna	2008720116	Unknown	Unknown
57	*L. interrogans*	Unknown	Jalna	2008720117	Unknown	Unknown
58	*L. interrogans*	Ictero	Lai	56601	China	Human
59	*L. interrogans*	Ictero	Lai type Langkawi	Langkawi	Malaysia	Human
60	*L. interrogans*	Ictero	Lai type Langkawi	SR61	Sri Lanka	Human
61	*L. interrogans*	Australis	Lora	1992	Tanzania	Mouse
62	*L. interrogans*	Pyrogenes	Manilae	K56	Unknown	Unknown
63	*L. interrogans*	Sejroe	Medanensis	L0448	Thailand	Human
64	*L. interrogans*	Sejroe	Medanensis	L0887	Thailand	Human
65	*L. interrogans*	Australis	Muenchen	Brem 129	Germany	Horse
66	*L. interrogans*	Ictero	Naam	Naam	Indonesia	Human
67	*L. interrogans*	Pomona	Pomona	2006006962	Unknown	Unknown
68	*L. interrogans*	Pomona	Pomona	CSL10083	USA	Sea Lion
69	*L. interrogans*	Pomona	Pomona	Fox 32256	USA	Fox
70	*L. interrogans*	Pomona	Pomona	Kennewicki LC82-25	USA	Human
71	*L. interrogans*	Pomona	Pomona	Pomona	Australia	Human
72	*L. interrogans*	Pomona	Pomona	UT364	Thailand	Human
73	*L. interrogans*	Pyrogenes	Pyrogenes	C10069	Thailand	Rat
74	*L. interrogans*	Pyrogenes	Pyrogenes	2006006956	Unknown	Unknown
75	*L. interrogans*	Pyrogenes	Pyrogenes	2006006960	Egypt	Human
76	*L. interrogans*	Pyrogenes	Pyrogenes	200701872	Mayotte	Human
77	*L. interrogans*	Pyrogenes	Pyrogenes	L0374	Thailand	Human
78	*L. interrogans*	Pyrogenes	Pyrogenes	R168	Sri Lanka	Human
79	*L. interrogans*	Pyrogenes	Pyrogenes	Sri Lanka 14	Unknown	Unknown
80	*L. interrogans*	Pyrogenes	Pyrogenes	Sri Lanka 30	Unknown	Human

Table 4: Details of *Leptospira* isolates used for comparison of *Leptospira* MLST schemes *(Continued)*

S. No	Genomospecies	Serogroup	Serovar	Strain	Country	Host
81	*L. interrogans*	Pyrogenes	Pyrogenes	Sri Lanka 46	Unknown	Human
82	*L. interrogans*	Pyrogenes	Pyrogenes	Srilanka1	Unknown	Unknown
83	*L. interrogans*	Unknown	Szwajizak	Szwajizak	Unknown	Human
84	*L. interrogans* [2, 3]	Unknown	Unknown	L1207	Thailand	Human
85	*L. interrogans*	Unknown	Unknown	MMD3731	Peru: Iquitos	Rat
86	*L. interrogans*	Unknown	Unknown	2002000621	USA	Human
87	*L. interrogans*	Unknown	Unknown	2006001854	Thailand	Human
88	*L. interrogans*	Unknown	Unknown	Brem 329	German	Horse
89	*L. interrogans*	Unknown	Unknown	FPW1039	Thailand	Human
90	*L. interrogans*	Unknown	Unknown	FPW2026	Thailand	Human
91	*L. interrogans*	Unknown	Unknown	UI 08452	Laos	Human
92	*L. interrogans*	Unknown	Unknown	UI 09600	Laos	Human
93	*L. interrogans*	Unknown	Unknown	UI 12758	Laos	Human
94	*L. interrogans*	Unknown	Unknown	UI 13372	Laos	Human
95	*L. interrogans*	Unknown	Unknown	HAI1536	Peru: Iquitos	Human
96	*L. interrogans*	Grippotyphosa	Valbuzzi	Valbuzzi	Australia	Human
97	*L. interrogans*	Pyrogenes	Zanoni	Zanoni	Australia	Human
98	*L. kirschneri*	Autumnalis	Bim	1051	Barbados	Canine
99	*L. kirschneri*	Cynopteri	Cynopteri	3522_C	Indonesia	Bat
100	*L. kirschneri*	Grippotyphosa	Grippotyphosa	UT130	Thailand	Human
101	*L. kirschneri*	Grippotyphosa	Grippotyphosa	Duyster-Boelhouwer	Unknown	Unknown
102	*L. kirschneri*	Grippotyphosa	Grippotyphosa	Moskva V	Russia	Human
103	*L. kirschneri*	Grippotyphosa	Grippotyphosa	RM52	USA	Pig
104	*L. kirschneri*	Grippotyphosa	Honghe	H2	Thailand	Human
105	*L. kirschneri*	Pomona	Mozdok	B 81/7 Type 3/ Tsaratsovo	Bulgaria	Mouse
106	*L. kirschneri*	Pomona	Mozdok	Brem 166	Unknown	Unknown
107	*L. kirschneri*	Pomona	Mozdok 1	Vehlefans 2	Netherlands	Cow
108	*L. kirschneri*	Ictero	Sokoine	RM1	Tanzania	Cow
109	*L. kirschneri*	Grippotyphosa	Unknown	H1	Thailand	Human
110	*L. kirschneri*	Unknown	Unknown	200801774	Mayotte	Human
111	*L. kirschneri*	Unknown	Unknown	2008720114	Croatia	Rodent
112	*L. kirschneri*	Grippotyphosa	Valbuzzi	200702274	France	Human
113	*L. kirschneri*	Grippotyphosa	Valbuzzi	Brem 179	Unknown	Unknown
114	*L. noguchii*	Autumnalis	Autumnalis	ZUN142	Peru: Iquitos	Human
115	*L. noguchii*	Panama	Panama	CZ214	Panama	Opossum
116	*L. noguchii* [1, 2]	Unknown	Unknown	1993005606	USA	Human
117	*L. noguchii*	Unknown	Unknown	2006001870	USA	Human
118	*L. noguchii*	Unknown	Unknown	2007001578	USA	Human
119	*L. noguchii*	Autumnalis	Unknown	Bonito	Brazil	Human
120	*L. noguchii*	Bataviae	Unknown	Cascata	Brazil	Human
121	*L. noguchii*	Australis	Unknown	Hook	Brazil	Dog

Table 4: Details of *Leptospira* isolates used for comparison of *Leptospira* MLST schemes *(Continued)*

S. No	Genomospecies	Serogroup	Serovar	Strain	Country	Host
122	*L. santarosai*	Unknown	Arenal	11	Unknown	Unknown
123	*L. santarosai*	Unknown	Arenal	7	Unknown	Human
124	*L. santarosai*	Unknown	Arenal	MAVJ 401	Costa Rica	Human
125	*L. santarosai*	Shermani	Shermani	LT 821	Panama	Rat
126	*L. santarosai*	Unknown	Szwajizak	Oregon	USA	Cow
127	*L. santarosai*	Unknown	Unknown	2000027870	USA	Human
128	*L. santarosai*	Unknown	Unknown	2000030832	USA	Human
129	*L. santarosai*	Unknown	Unknown	200702252	Guadeloupe	Human
130	*L. santarosai*	Unknown	Unknown	AIM	Colombia	Human
131	*L. santarosai*	Unknown	Unknown	CBC1531	Peru: Iquitos	Buffalo
132	*L. santarosai*	Unknown	Unknown	CBC379	Peru: Iquitos	Pig
133	*L. santarosai*	Unknown	Unknown	CBC523	Peru: Iquitos	Cattle
134	*L. santarosai*	Unknown	Unknown	HAI134	Peru: Iquitos	Human
135	*L. santarosai*	Unknown	Unknown	HAI1380	Peru: Iquitos	Human
136	*L. santarosai*	Unknown	Unknown	HAI821	Peru: Iquitos	Human
137	*L. santarosai*	Unknown	Unknown	JET	Colombia	Human
138	*L. santarosai*	Unknown	Unknown	MOR084	Peru: Iquitos	Human
139	*L. santarosai*	Unknown	Unknown	ST188	Trinidad	Dog
140	*L. santarosai*	Unknown	Unknown	ZUN179	Peru: Iquitos	Human
141	*L. santarosai*	Unknown	Unknown	HAI1594	Peru: Iquitos	Human
142	*L. weilii*	Celledoni	Mengdeng	LNT1194	Laos	Human
143	*L. weilii*	Unknown	Topaz	LT2116	Australia	Human
144	*L. weilii*	Unknown	Unknown	2006001853	Thailand	Human
145	*L. weilii* [3]	Unknown	Unknown	2006001855	Thailand	Human
146	*L. weilii*	Hebdomadis	Unknown	LNT1234	Laos	Human
147	*L. weilii*	Unknown	Unknown	UI 13098	Laos	Human
148	*L. weilii*	Unknown	Unknown	UI 14631	Unknown	Human
149	*Leptospira sps*	Unknown	Unknown	FiocruzLV4135	Brazil	Human

[1]– ST could not be generated using Boonsilp's scheme

[2]– ST could not be generated using Varni's scheme

[3] – ST could not be generated using Ahmed's scheme

Ictero - Icterohaemorrhagiae

Table 5: Details of Fiocruz isolates belonging to serovar Copenhageni

S.No	Genomospecies	Serovar	Strain	Host	Country
1.	*L. interrogans* *	Copenhageni	Fiocruz LV130	human	Brazil
2.	*L. interrogans*	Copenhageni	Fiocruz LV192	human	Brazil
3.	*L. interrogans*	Copenhageni	Fiocruz LV199	human	Brazil
4.	*L. interrogans*	Copenhageni	Fiocruz LV204	human	Brazil
5.	*L. interrogans*	Copenhageni	Fiocruz LV212	human	Brazil
6.	*L. interrogans*	Copenhageni	Fiocruz LV224	human	Brazil
7.	*L. interrogans*	Copenhageni	Fiocruz LV237	human	Brazil
8.	*L. interrogans*	Copenhageni	Fiocruz LV239	human	Brazil
9.	*L. interrogans*	Copenhageni	Fiocruz LV251	human	Brazil
10.	*L. interrogans*	Copenhageni	Fiocruz LV256	human	Brazil
11.	*L. interrogans*	Copenhageni	Fiocruz LV2750	human	Brazil
12.	*L. interrogans*	Copenhageni	Fiocruz LV2752	human	Brazil
13.	*L. interrogans*	Copenhageni	Fiocruz LV2756C6	human	Brazil
14.	*L. interrogans*	Copenhageni	Fiocruz LV2759	human	Brazil
15.	*L. interrogans*	Copenhageni	Fiocruz LV2763	human	Brazil
16.	*L. interrogans*	Copenhageni	Fiocruz LV2766	human	Brazil
17.	*L. interrogans*	Copenhageni	Fiocruz LV2767	human	Brazil
18.	*L. interrogans*	Copenhageni	Fiocruz LV2769	human	Brazil
19.	*L. interrogans*	Copenhageni	Fiocruz LV2772	human	Brazil
20.	*L. interrogans*	Copenhageni	Fiocruz LV2776	human	Brazil
21.	*L. interrogans*	Copenhageni	Fiocruz LV2787	human	Brazil
22.	*L. interrogans*	Copenhageni	Fiocruz LV2790	human	Brazil
23.	*L. interrogans*	Copenhageni	Fiocruz LV2791	human	Brazil
24.	*L. interrogans*	Copenhageni	Fiocruz LV2799	human	Brazil
25.	*L. interrogans*	Copenhageni	Fiocruz LV2804	human	Brazil
26.	*L. interrogans*	Copenhageni	Fiocruz LV2805	human	Brazil
27.	*L. interrogans*	Copenhageni	Fiocruz LV2806	human	Brazil
28.	*L. interrogans*	Copenhageni	Fiocruz LV2807	human	Brazil
29.	*L. interrogans*	Copenhageni	Fiocruz LV2811	human	Brazil
30.	*L. interrogans*	Copenhageni	Fiocruz LV2812	human	Brazil
31.	*L. interrogans*	Copenhageni	Fiocruz LV2816	human	Brazil
32.	*L. interrogans*	Copenhageni	Fiocruz LV2825	human	Brazil
33.	*L. interrogans*	Copenhageni	Fiocruz LV2832	human	Brazil
34.	*L. interrogans*	Copenhageni	Fiocruz LV2840	human	Brazil
35.	*L. interrogans*	Copenhageni	Fiocruz LV2897	human	Brazil
36.	*L. interrogans*	Copenhageni	Fiocruz LV2908	human	Brazil
37.	*L. interrogans*	Copenhageni	Fiocruz LV2919	human	Brazil
38.	*L. interrogans*	Copenhageni	Fiocruz LV2933	human	Brazil
39.	*L. interrogans*	Copenhageni	Fiocruz LV2948	human	Brazil
40.	*L. interrogans* *	Copenhageni	Fiocruz LV2953	human	Brazil
41.	*L. interrogans*	Copenhageni	Fiocruz LV2958	human	Brazil
42.	*L. interrogans*	Copenhageni	Fiocruz LV2959	human	Brazil
43.	*L. interrogans*	Copenhageni	Fiocruz LV2973	human	Brazil
44.	*L. interrogans*	Copenhageni	Fiocruz LV3076	human	Brazil

Table 5: Details of Fiocruz isolates belonging to serovar Copenhageni *(Continued)*

S.No	Genomospecies	Serovar	Strain	Host	Country
45.	*L. interrogans*	Copenhageni	Fiocruz LV3086	human	Brazil
46.	*L. interrogans* *	Copenhageni	Fiocruz LV3094	human	Brazil
47.	*L. interrogans*	Copenhageni	Fiocruz LV3096	human	Brazil
48.	*L. interrogans*	Copenhageni	Fiocruz LV3213	human	Brazil
49.	*L. interrogans*	Copenhageni	Fiocruz LV3244	human	Brazil
50.	*L. interrogans*	Copenhageni	Fiocruz LV3323	human	Brazil
51.	*L. interrogans*	Copenhageni	Fiocruz LV3373	human	Brazil
52.	*L. interrogans*	Copenhageni	Fiocruz LV3409	human	Brazil
53.	*L. interrogans* *	Copenhageni	Fiocruz LV3726	human	Brazil
54.	*L. interrogans*	Copenhageni	Fiocruz LV3737	human	Brazil
55.	*L. interrogans*	Copenhageni	Fiocruz LV3738	human	Brazil
56.	*L. interrogans*	Copenhageni	Fiocruz LV3834	human	Brazil
57.	*L. interrogans*	Copenhageni	Fiocruz LV3879	human	Brazil
58.	*L. interrogans* *	Copenhageni	Fiocruz LV4034	Unknown	Unknown
59.	*L. interrogans*	Copenhageni	Fiocruz LV4108	human	Brazil
60.	*L. interrogans*	Copenhageni	Fiocruz LV4113	human	Brazil
61.	*L. interrogans*	Copenhageni	Fiocruz LV4118	human	Brazil
62.	*L. interrogans*	Copenhageni	Fiocruz LV4152	human	Brazil
63.	*L. interrogans*	Copenhageni	Fiocruz LV4160	human	Brazil
64.	*L. interrogans*	Copenhageni	Fiocruz LV4173	human	Brazil
65.	*L. interrogans*	Copenhageni	Fiocruz LV4174	human	Brazil
66.	*L. interrogans*	Copenhageni	Fiocruz LV4188	human	Brazil
67.	*L. interrogans*	Copenhageni	Fiocruz LV4211	human	Brazil
68.	*L. interrogans*	Copenhageni	Fiocruz LV4212	human	Brazil
69.	*L. interrogans*	Copenhageni	Fiocruz LV4217	human	Brazil
70.	*L. interrogans*	Copenhageni	Fiocruz LV4234	human	Brazil
71.	*L. interrogans*	Copenhageni	Fiocruz LV999	human	Brazil
72.	*L. interrogans*	Copenhageni	Fiocruz R154	human	Brazil
73.	*L. interrogans* *	Copenhageni	Fiocruz R83	rat	Brazil
74.	*L. interrogans*	Copenhageni	Fiocruz_LV3992	human	Brazil
75.	*L. interrogans*	Copenhageni	Fiocruz_LV4231	human	Brazil
76.	*L. interrogans*	Copenhageni	Fiocruz LV4114	human	Brazil
77.	*L. interrogans*	Copenhageni	Fiocruz LV4225	human	Brazil
78.	*L. interrogans*	Copenhageni	Fiocruz L1-130	human	Brazil

* Isolates included in the final study

2.2.2 Gene Loci and PCR conditions

Gene loci and PCR conditions were selected as per the MLST schemes proposed elsewhere (Ahmed *et al.*, 2006; Thaipadungpanit *et al.*, 2007; Boonsilp *et al.*, 2013). Details of all gene loci and primer sequences for amplification of each locus are listed in **Table 6** along with the locus size considered for all MLST schemes included in the study.

Table 6: Details of MLST loci and primers used for amplification

Gene	Function	Primer sequences	Amplified fragment size (bp)	Size of MLST locus (bp)
Ahmed *et al* (2006) scheme				
adk	Adenylate Kinase	F-gggctggaaaaggtacacaa R-acgcaagctccttttgaatc	531	430
icdA	Isocitrate Dehydrogenase	F-gggacgagatgaccaggat R-ttttttgagatccgcagcttt	674	557
LipL32	Outer membrane Lipoprotein LipL32	F-atctccgttgcactctttgc R-accatcatcatcatcgtcca	474	474
LipL41	Outer membrane Lipoprotein LipL41	F-taggaaattgcgcagctaca R-gcatcgagaggaattaacatca	520	518
rrs2	16S ribosomal RNA	F-catgcaagtcaagcggagta R-agttgagcccgcagttttc	542	452
secY	Pre- protein translocase secY	F-atgccgatcatttttgcttc R-ccgtcccttaattttagacttcttc	549	549
	Concatenated sequence length			**2980**
Thaipadungpanit *et al* (2007) scheme				
glmU	UDP-N-acetyl glucosamine pyro phosphorylase	F -ggaagggcacccgtatgaa R -tccctgagcgttttgattt	557	444
pntA	NAD(P) transhydrogenase subunit alpha	F -tgccgatcctacaacatta R -aagaagcaagatccacaactac	638	525
sucA	2-oxoglutarate dehydrogenase decarboxylase component	F -agaagaggccggttatcatcag R -cttccgggtcgtctccattta	560	447
fadD	Probable long chain fatty acid CoA ligase	F -agtatggcgtatcttcctcctt R -ttcccactgtaatttctcctaa	577	456
tpiA	Triose phosphate isomerase	F -aagccgttttcctagcacattc R -aggcgcctacaaaaagaccaga	555	426
pfkB	Ribokinase	F -ccgaagataaggggcatacc R -caagctaaaaccgtgagtgatt	560	432
mreA	Rod shape determining protein rodA	F -gtaaaagcggccaacctaacac R -acgatcccagacgcaagtaa	602	435
	Concatenated sequence length			**3165**

Table 6: Details of MLST loci and primers used for amplification (Continued)

Gene	Function	Primer sequences	Amplified fragment size (bp)	Size of MLST locus (bp)
Boonsilp *et al* (2007) scheme				
glmU	UDP-N-acetyl glucosamine pyro phosphorylase	F -aggataaggtcgctgtggta R -agtttttttccggagtttct	650	444
pntA	NAD(P) transhydrogenase subunit alpha	F –taggaaaratgaaaccrggaac R -aagaagcaagatccacaaytac	621	525
sucA	2-oxoglutarate dehydrogenase decarboxylase component	F -tcattccacttytagatacgat R -tctttttgaatttttgacg	640	447
tpiA	Triose phosphate isomerase	F -ttgcaggaaactggaaaatgaat R -gttttacrgaacchccgtagagaat	639	426
pfkB	Ribokinase	F -cggagagttttataaraaggacat R -agaacacccgccgcaaaacaat	588	432
mreA	Rod shape determining protein rodA	F -ggctcgctctygacggaaa R -tccrtaactcataaamgacaaagg	719	435
caiB	carnitine dehydratase	F -caacttgcggayataggaggag R -attatgttccccgtgaytcg	650	403
	Concatenated sequence length			**3112**
Varni *et al* (2013) scheme				
adk	Adenylate Kinase	F1 -gggctggaaaaggtacacaa R1 -acgcaagctccttttgaatc	564	430
glmU	UDP-N-acetyl glucosamine pyro phosphorylase	F1 -aggataaggtcgctgtggta R1 -agtttttttccggagtttct	650	444
icdA	Isocitrate Dehydrogenase	F1 -gggacgagatgaccaggat R1 -ttttttgagatccgcagcttt	1197	557
LipL32	Outer membrane Lipoprotein LipL32	F1-atctccgttgcactctttgc R1-accatcatcatcatcgtcca	819	450
LipL41	Outer membrane Lipoprotein LipL41	F1-taggaaattgcgcagctaca R1-gcatcgagaggaattaacatca	1068	493
mreA	Rod shape determining protein rodA	F1-ggctcgctctygacggaaa R1-tccrtaactcataaamgacaaagg	719	435
pntA	NAD(P) transhydrogenase subunit alpha	F1-taggaaaratgaaaccrggaac R1-aagaagcaagatccacaaytac	621	525
	Concatenated sequence length			**3334**

The PCR amplification of different MLST target genes of Ahmed's scheme was performed using 200 µM of each dNTP (Fermentas), 50–100 ng template DNA, 10 pmol of each primer, 1.5 mM $MgCl_2$ and 1.0U of DreamTaq DNA polymerase (Fermentas) for 20 µl reaction volume using Mastercycler Pro (Eppendorf AG, Germany) PCR system. Cycling conditions were as described elsewhere (Ahmed *et al.*, 2006) and briefly include initial denaturation at 95°C for 5 min followed by 35 cycles of 30 sec denaturation at 95°C, 30 sec

annealing at 58 °C and 1 min extension at 72°C. The final extension was carried out for 7 min at 72°C.

2.2.3 Nucleotide sequences and allele designation

Sequence chromatogram files for PCR amplified fragments, in both forward and reverse directions, were obtained from Dr. Manjulatha Devi, CDFD, Hyderabad. Sequence chromatograms were viewed using ChromasLite version 2.01 (Technelysium Pty Ltd, Australia) and sequences were aligned using forward and reverse read sequences in Seqscape software (Applied Biosystems, Foster City, USA) to generate consensus sequence at each locus. Chromatograms were also analyzed using BioEdit software available at http://www.mbio.ncsu.edu/BioEdit/bioedit.html (Hall, 1999), Genedoc (version 2.6.002) (Nicholas *et al.*, 1997) and SeqSphere software version 0.9 beta (Ridom Bioinformatics GmbH; Munster, Germany) to generate consensus sequence for each sample.

Figure 7: Overview of salient features of Ridom SeqSphere software used for designating allelic profiles for isolates

SeqSphere software was used to curate raw sequencer files to designate new alleles at each locus. Salient features, highlighting the functions, of SeqSphere software are detailed in **Figure 7**. Each variant call observed in the consensus sequence against the reference sequence was confirmed using the forward and reverse sequence chromatograms and allele designation at each locus was done numerically in serial order. Alleles with its designation were stored in the local database and allelic profile of all loci for each isolate was used to designate sequence type (ST).

2.2.4 Sequence retrieval from whole genome sequences

Nucleotide sequences of alleles for all loci were retrieved from whole-genome sequences of isolates for Ahmed's scheme, Boonsilp's scheme and Varni's scheme using web-based MLST tool version 1.7 hosted at https://cge.cbs.dtu.dk/services/MLST/ (Larsen *et al.*, 2012). This tool can handle sequence raw reads, whole-genome alignments and draft genomes as input source and retrieves respective loci sequences as per the MLST scheme definition based on BLAST based ranking method. Individual contig files of a draft genome of an isolate were combined together into single file and was used as input for the MLST tool. Sequences were also confirmed in case of ambiguity using *in-silico* PCR simulation tool hosted at http://insilico.ehu.es/PCR/ (Bikandi *et al.*, 2004).

As most of the whole genome sequences used in the study were draft sequences with gaps in their assemblies, allele sequences at one or more loci could not be retrieved owing to the incidents of truncation of the sequence or missing of the sequence in the gaps.

L. interrogans serovar Copenhageni Strain Fiocruz LV3726 was excluded from all three MLST schemes as sequences for *glmU, pntA, sucA, tpiA, adk, icdA* loci could not be retrieved. *L.interrogans* serovar Copenhageni strain MMD1562 was also excluded from all three MLST schemes as sequence for *mreA* and *LipL32* loci could not be retrieved.

Sequences for *tpiA* locus from genomes of *L. interrogans* Serovar Australis strain 2006001855 and *L. interrogans* Serovar Copenhageni strain 200703203, and for *glmU* locus from genome of *L. noguchii* strain 193005606 could not be retrieved and so were excluded from both Boonsilp's scheme and Varni's scheme.

L. interrogans strain L1207 was found to contain a 3 nucleotide gap in *LipL32* sequence and sequence for *adk* locus from genome of *L. alexanderi* Serovar Manhao 3 Strain L 60 could not be retrieved. So, both these isolates were excluded from Varni's scheme and Ahmed's scheme. *L. weilii* strain 2006001855 was excluded from the Ahmed's scheme as the complete sequence for *rrs*2 locus could not be retrieved.

Thus in total 145, 144 and 146 isolates were typed using *in-silico* MLST method for Boonsilp's scheme, Varni's scheme and Ahmed's scheme respectively.

2.2.5 Sequence analysis

Individual allelic sequences of all MLST loci for each isolate were concatenated end to end in the order specified in **Table 6** for Ahmed's scheme, Thaipadungpanit's scheme, Boonsilp's scheme, and for Varni's scheme (Ahmed *et al.*, 2006; Boonsilp *et al.*, 2013; Varni *et al.*, 2014).

Multiple sequence alignment was carried out for concatenated sequences of a MLST scheme of all isolates using Clustal X software version 2.0 (Larkin *et al.*, 2007). Molecular Evolutionary Genetics Analysis (MEGA) software version 3.0 and version 5.0 were used for phylogenetic analysis (Kumar *et al.*, 2008; Tamura *et al.*, 2011) employing neighbor-joining (NJ) algorithm using Kimura 2 parameter at 1000 bootstrap values and maximum likelihood (ML) with general time reversible parameter at 1000 bootstrap replications (Ahmed *et al.*, 2006, 2011; Varni *et al.*, 2014).

To evaluate the evolutionary pressure on protein coding MLST loci, synonymous (d_S) and nonsynonymous (d_N) substitutions were determined using program START version 2 (Jolley

et al., 2001). The ratio of number of d_N to number of d_S substitutions (d_N/d_S) was used to determine the evolutionary selective pressure on each locus (Kryazhimskiy and Plotkin, 2008).

Simpson's index of diversity (D) was calculated with 95% confidence interval ($CI_{95\%}$) using LIAN version 3.0 web tool hosted at www.pubmlst.org to determine the discriminatory ability of individual MLST method (Hunter and Gaston, 1988; Haubold and Hudson, 2000). Its value close to 0 indicates a less or little diversity and value close to 1 reflects higher diversity.

2.2.6 Phylogenetic reconstruction by Bayesian Markov Chain Monte Carlo (MCMC) approach

BEAST (Bayesian Evolutionary Analysis Sampling trees) software version 1.5.2 was used for Bayesian analysis of MLST data for molecular sequence variation and relationship in the phylogenetic tree (Drummond and Rambaut, 2007; Drummond *et al.*, 2012). Relaxed molecular clock approach was used to generate phylogenetic tree by employing coalescent constant population size and Yule speciation tree prior model of evolution. Two independent runs for each model were achieved for 30,000,000 steps and sampled every 1000 steps. The first 1,00,000 steps of each run were discarded as burn-in. The tree was annotated using the program TreeAnnotator v1.5.2 (http://tree.bio.ed.ac.uk/) and viewed using the program FigTree v1.2.2 (http://tree.bio.ed.ac.uk/).

2.2.7 Network analysis for Indian sub-continent isolates

Network version 4.5.0.0 program (www.fluxus-engineering.com) was used to reconstruct phylogenetic network using Median-joining algorithm to a set of *L. interrogans* isolates of Indian sub-continent (**Table 2**) (Bandelt *et al.*, 1995). Concatenated MLST sequences of those isolates were used to generate networks to infer ancestral relationships among isolates (Bandelt, Forster and Rohl, 1999).

2.2.8 Whole genome sequence (WGS) based phylogeny

Whole genome sequence based phylogeny was constructed using Gegenees version 2.2.1 software (Agren *et al.*, 2012) for all 152 genomes considered for the study **(Table 4)**. Gegenees employs fragmented all-against-all comparison using BLASTN method. Genomes were subjected to the analysis with fragmentation size of 200 and step size of 200. After the alignment, a 5% threshold was used to generate heat plot for phylogenomic data showing average normalized BLAST score values of all fragments. Phylogenomic data was exported to nexus file and is imported to SplitsTree4 version 4.1.3.1 software (Huson and Bryant, 2006). Phylogram was constructed using NJ distance based method and the tree was exported into Newick format. Phylogenetic tree in Newick format was processed further using FigTree version 1.4.0 available at http://tree.bio.ed.ac.uk/software/figtree/ for better visualization.

2.3 **Results:**

2.3.A Evaluation of Ahmed's MLST scheme for global applicability of *Leptospira* strain typing

Ahmed's MLST scheme was applied successfully to a set of 271 isolates representing global dispersal of *Leptospira* species, isolated from wide array of hosts and geographical locations. Consensus sequences using forward and reverse sequences of respective locus were generated for all isolates. Loci sequences of individual isolate were concatenated end to end to generate sequence types.

2.3.A.1 Nucleotide diversity of genetic loci

Multiple sequence alignment of concatenated sequences has confirmed that there were no large insertions or deletions at any locus. d_N/d_S ratios of all the protein-coding genes were found to be less than 1 (**Table 7**) indicating that none of the genes were under positive selection as ascertained earlier (Ahmed *et al.*, 2006). Diversity among genes varied from 64 alleles at *rrs2* locus to 121 alleles at *LipL41* locus among these isolates. Percentage of G+C content varied from 38.85 % at *secY* locus to 51.87% at *rrs2* locus. Discriminatory power with 95% confidence interval was found to be 91.8% resulting in 249 unique concatenated sequences for the total 271 isolates. Presence of these many unique sequences has ruled out the possibility of convenient sampling of isolates for the study.

Table 7: Allelic diversity of the MLST loci calculated using START2 software

Parameters ↓ **loci→**	***adk***	***icdA***	***LipL32***	***LipL41***	***rrs2***	***secY***	**Concat.**
Length in bases	430	557	474	518	452	549	2980
GC content (%)	41.61	41.08	46.43	42.85	51.87	38.85	43.78
No. of alleles	114	116	74	121	64	111	249
Polymorphic sites	158	188	62	137	40	185	770
Synonymous sites (S)	100	126	112	122	NA*	126	
Non-synonymous sites (N)	328	428	362	393	NA*	423	
d_N/d_S	0.042	0.023	0.086	0.056	NA*	0.020	
Discriminatory power (CI 95%)							91.8

* d_N/d_S ration for rrs2 was not calculated as the gene does not code for a protein,

Concat.: concatenated sequence

2.3.A.2 Phylogeny inferred from MLST data

This MLST data has revealed a highly organized phylogenetic tree and did not affect the embranchment pattern of the tree and branch composition upon using various phylogenetic methods like ML and NJ by MEGA 3.0, Median Joining by Network and Bayesian MCMC analysis confirming the robustness of the method.

Phylogenetic tree constructed employing NJ algorithm using concatenated nucleotide sequences generated six different major clusters representing *L. interrogans, L. kirschneri, L.noguchii, L. weilii, L. borgpetersenii* and *L. santarosai* species. A single isolate of *L. meyeri* used in the study was found to be clustered with *L. interrogans* on a separate branch, and another single isolate of *L. alexanderi* is found to be clustered with *L. santarosai.* Three isolates of *L. inadai* species used in the study were found to be present on a separate branch and among *L. santarosai* cluster representing their intermediate nature of pathogenicity (**Figure 8**).

Bayesian MCMC analysis which yields genetic relationships among various branches and generates a rooted phylogeny is considered as the state of the art method for the phylogenetic reconstruction. To know the extent of genetic affinities among various branches of phylogeny and to reproduce and validate the associations among species, MCMC analysis was performed with coalescent constant population size and Yule speciation tree prior and found that both trees showed similar tree topology. Bayesian inference (BI) tree (**Figure 9**) has split the global collection of *Leptospira* isolates into 12 branches. These 12 branches showed 5 major clades A1, D, G1, J1 and K and 7 minor clades A2, A3, A4, E, G2, J2 and J3. All these major and minor clades were well supported by 95% highest posterior density (HPD) intervals. BI tree was found to be rooted with *L. inadai* (clade E) and its intermediate nature in genus *Leptospira* is well supported by previous studies (Brenner *et al.*, 1999; Victoria *et al.*, 2008). Major clades A1, D, J1 and K represented *L. interrogans, L. noguchii, L. santarosai and L. kirschneri* respectively.

The *L. interrogans* clade A1 comprises the largest collection of isolates and emerged as a tight cluster comprising of 123 isolates. Further, four visible sub-branching within the *L. interrogans* clade could perhaps suggest strain specific or host/environment specific genetic alterations. The minor clades A2 and A4 include four and three *L. interrogans* isolates respectively and behave as outliers. Taking into account the genetic distance, the minor clade A4 shows affinities towards *L. kirschneri* and not towards *L. interrogans*. In terms of genetic distance, the position of a sole *L. interrogans* isolate depicted as branch A3 in the tree lies intermediate between the *L. interrogans* and *L. kirschneri*. Thus the minor clade A2 which harbours two isolates from rodents (one each from a bandicoot and a necked field mouse) seems to be the most recent plausible ancestor of the present day *L. interrogans* found in varied hosts.

Major and minor clades G1 and G2 respectively, containing *L. borgpetersenii* isolates shows clustering with *L. weilii* isolates. The clade D containing entirely of *L. noguchii* strains has inclusion of an *L. inadai* isolate. Further, *L. noguchii* shows closer genetic affinity towards *L.kirschneri* than *L. interrogans*. The basal position of *L. noguchii* cluster in the vicinity of *L.kirschneri* and *L. interrogans* clusters suggests the ancestral nature of *L. noguchii* or *L. noguchii* like ancestor for both the species. In contrast to the less distinct evolution of *L. kirschneri* and *L. interrogans*, the evolution of *L. santarosai*, *L. borgpetersenii* and *L. weilii* was marked by more distinct speciation events along the intermediate *L. inadai* or *L. inadai* like ancestor. The minor clade G2, comprises of two *L. weilii* and one *L. borgpetersenii* and *L. interrogans* isolate each. The basal position of *L. weilii* in the minor clade G2 and in the major clade G1, suggests that *L.borgpetersenii* evolution was from *L. we*ilii or *L. weilii* like ancestor.

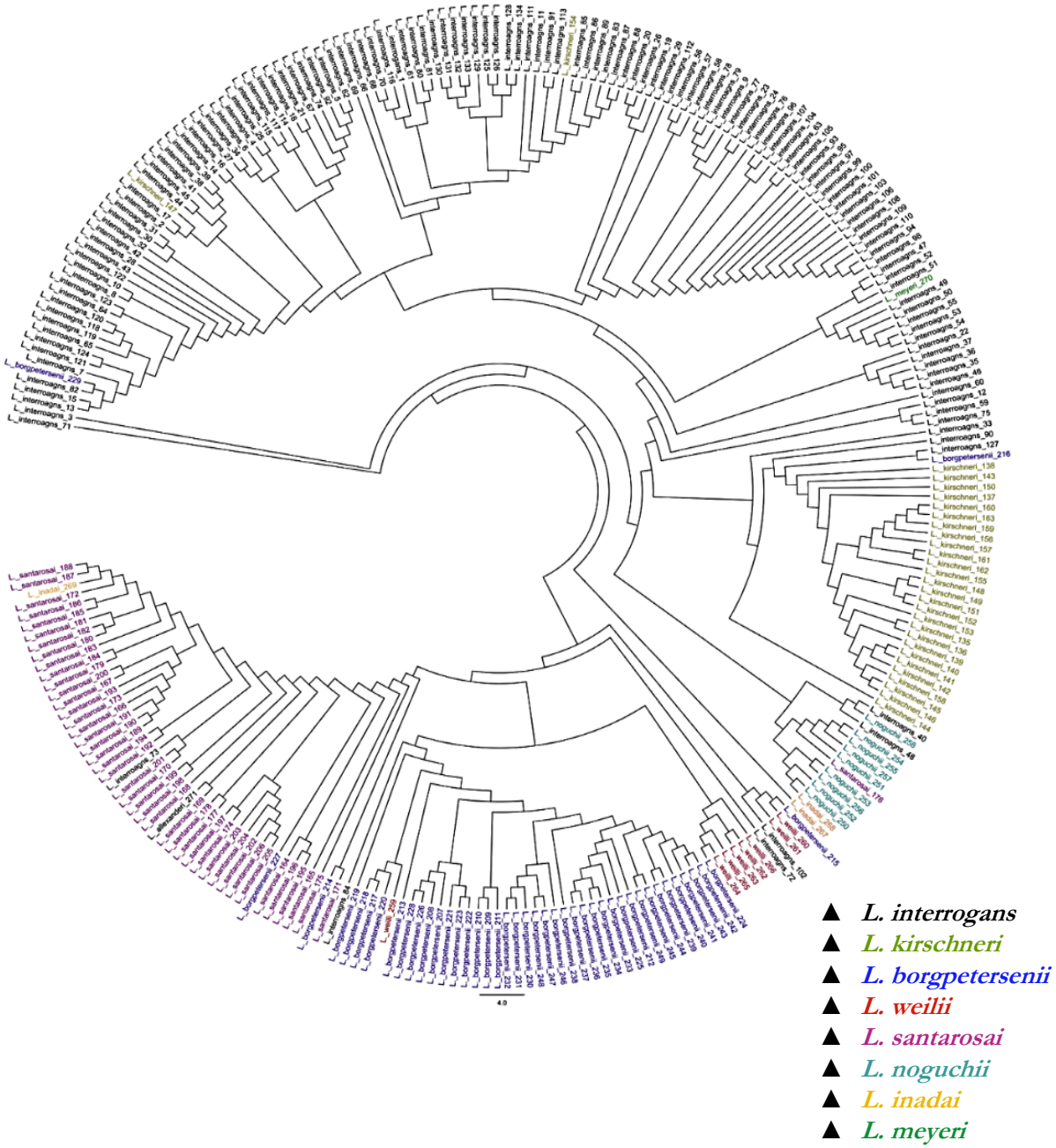

Figure 8: Genetic clustering of global *Leptospira* isolates based on MLST analysis. Phylogenetic tree constructed using NJ algorithm in MEGA. (Different colours were used to each species for easy identification) (Detailed information of isolates are mentioned in Table 2)

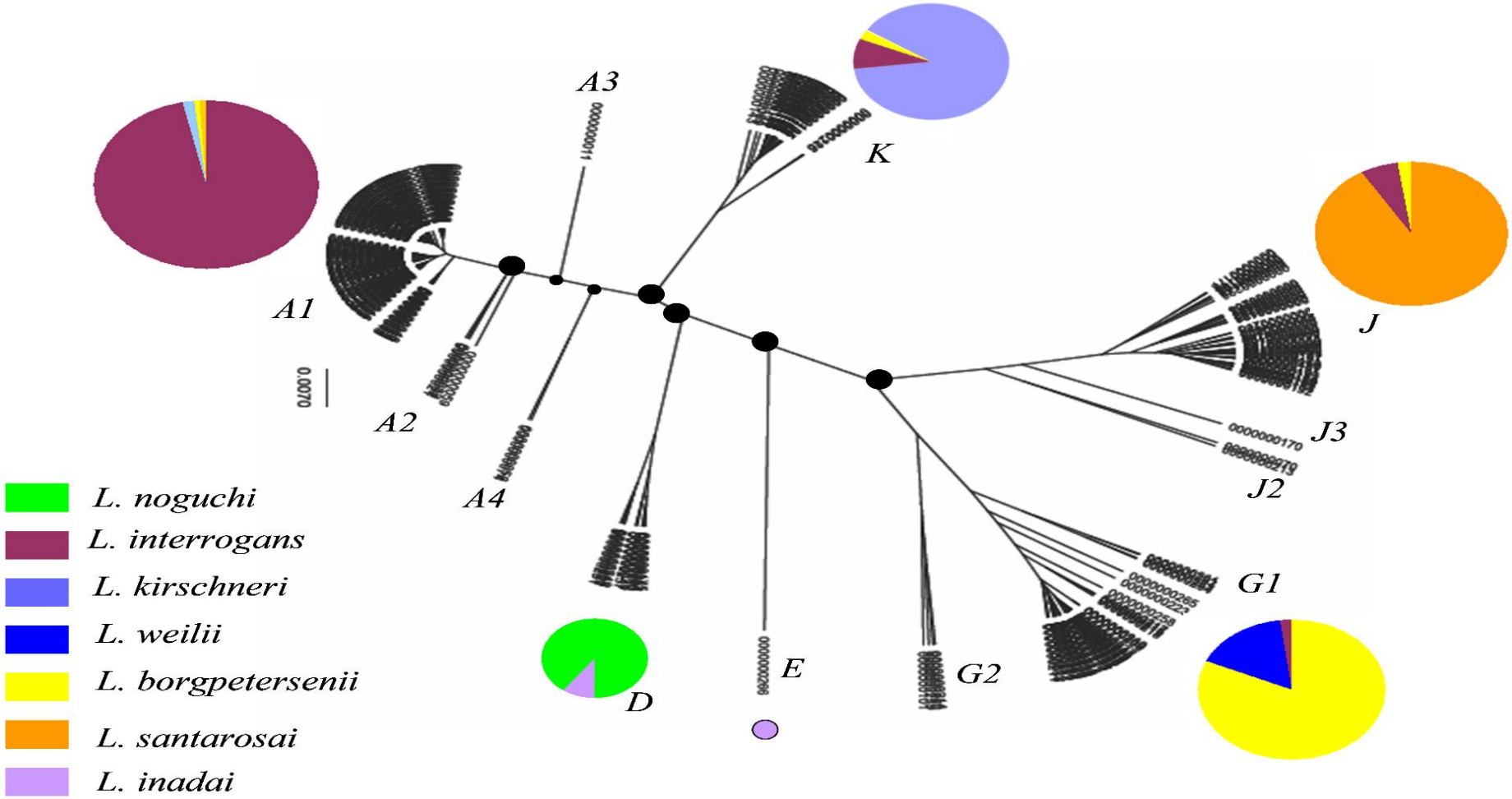

Figure 9: Bayesian inference tree showing the phylogeny, generated from MLST data with relaxed molecular clock approach in Bayesian MCMC analysis using BEAST v1.5.2

In addition, few isolates were found clustering unusually with isolates of other species in both phylogenetic trees which was not reported earlier by Ahmed *et al.* One isolate previously classified serologically as *L. interrogans* was clustering with *L. kirschneri* clade and three isolates classified previously as *L. interrogans* was clustered with *L. santarosai* clade. Two isolates each of *L. kirschneri* and *L. borgpetersenii* and an isolate of *L. meyeri* clustered with *L. interrogans* clusters. Taking into consideration of the genetic distances between the major clades and evolutionary pattern, possibility of clade switching between *L. kirschneri* and *L. interrogans* can be anticipated but not between *L. interrogans* with *L. borgpetersenii* or *L. santarosai*. Such unusual clustering could only be explained by incorrect serological classification or clade switching due to horizontal gene transfer events.

2.3.A.3 Distribution of pathogenic *Leptospira* isolates

It is evident from the global isolate collection and molecular typing by MLST that *L.interrogans, L. kirschneri* and probably *L. borgpetersenii* are ubiquitous species (**Table 8**). Among these isolates *L. interrogans* seems to be the more frequently isolated species from humans and was largely reported from South Asia. This trend perhaps reflects endemicity and the maintenance of this species in that region. Whereas, *L. weilii* was largely confined to Asia and *L. santarosai* and *L. noguchii* were found to be adapted to America.

Table 8: Distribution and phylogenetic affiliation of *Leptospira* isolates

Region	Country	No. of Isolates	% of total collection	Predominant species	Major serogroups	Major Source/Host	Major MLST clade(s)*
South Asia		94	34.7	*L. interrogans* *L. borgpetersenii*	Grippotyphosa, Icterohaemhoragie, Australis, Canicola,	Human (74)	A
	Indian sub-continent	60					
	South East Asia	34					
Rest of Asia		24	9	*L. weilii*	Javanica	Human (9)	G2
	China	16					
	West Asia	8					
Europe		26	9.5	*L. borgpetersenii* *L. interrogans*	Sejroe	Mainly rodents/ animals (12)	A, G
Africa		34	12.5	*L. borgpetersenii* *L. kirschneri* *L. interrogans*	Ballum	Rodents (25)	G, K
North/ Central America		35	13	*L. santarosai* *L. interrogans* *L. noguchii*	Tarassovi	Opossum (10) Human (10)	A, D, J
South America and Caribbean		35	13	*L. santarosai* *L. interrogans*	diverse	Opossum (11) Livestock (10)	A, J
Australia		7	2.5	*L. Interrogans*	Pyrogenes	Human (3)	A

* Clade designation is in accordance with the Bayesian inference tree drawn from MLST data as depicted in Figure 8

2.3.A.4 Network analysis of MLST data

A total of 44 *L. interrogans* isolates of these 271 isolates were collected from the Indian sub-continent as a part of routine outbreak investigations majorly from humans, over a period of time. Indian isolates, based on MLST data, when analysed using Network (**Figure 10A**), revealed that few sequences from the region were divergent, possibly because of recombination. The enlargement of the core sequence network (**Figure 10B**) revealed four sequences from the Andaman in central position, posing as a possible source of the Indian *L.interrogans* variability with distinct, derived clusters corresponding to South India, Central India and Andaman Islands. Given this, it is possible to think of an early spread to and from Andaman to mainland India and to other adjoining countries, possibly through rodents that travelled in vessels and ships to India from Andaman and vice-versa.

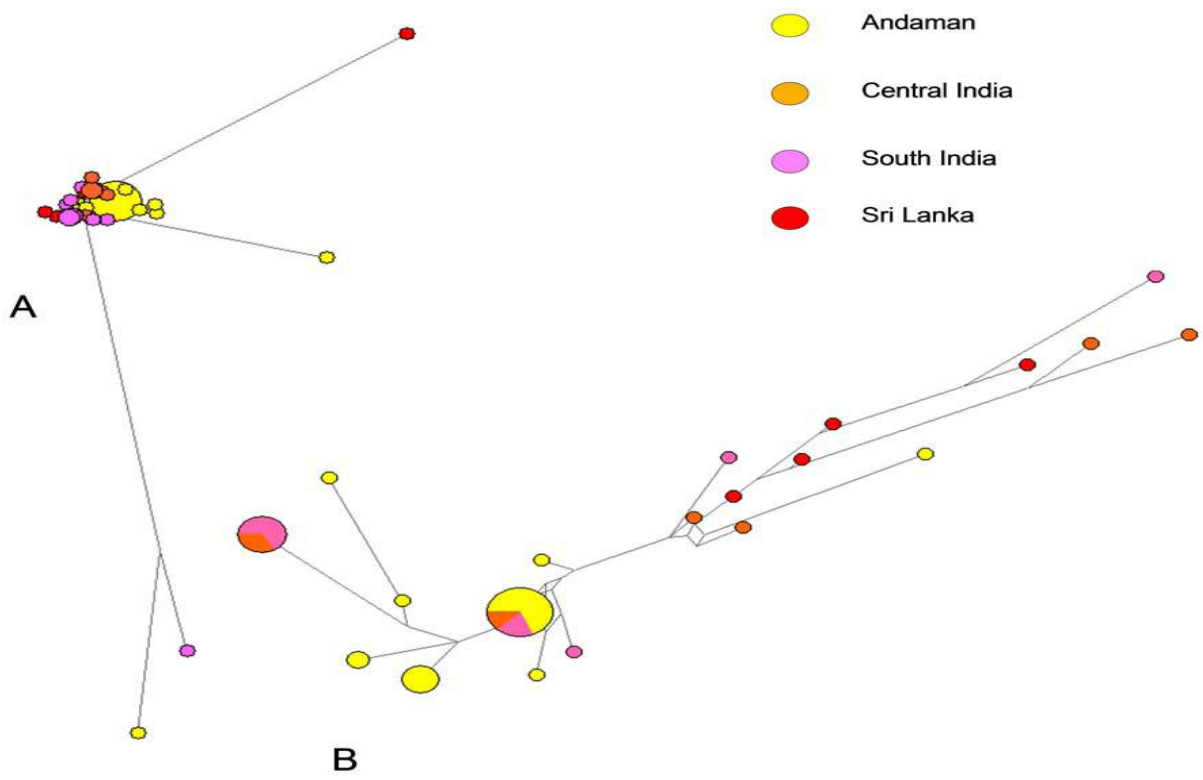

Figure 10: Median-Joining network analysis using Network v 4.5.0.0 for the MLST data obtained from Indian sub-continent subsample of *L. interrogans.* Tree B is the enlargement of core sequences of Tree A.

2.3.B Creation of MLST database for Ahmed *et al* (2006) scheme

2.3.B.1 Congruence with MLST locus designation

SeqSphere analysis of the raw sequencer files for allele designation has shown some chromatogram disturbances at the start and end of locus. So, the original locus size defined for MLST by Ahmed *et al* (2006) was altered to keep the protein-coding loci in translational frame and to get high quality sequence tracers every time the PCR fragment is sequenced in future and were detailed in **Table 9**. With the altered locus size, concatenated sequence length changed to 2877 bp from original 2980 bp. Phylogenetic tree was constructed employing NJ algorithm using the new concatenated sequences generated using altered loci size and found that the changes did not alter the clustering of isolates and did not change phylogenetic relationships across one another (**Figure 11**).

Table 9: Altered locus size considered for Ahmed's scheme

Gene	Function	Gene size (bp)	Original MLST locus size (bp)	Altered MLST locus size (bp)
adk	Adenylate Kinase	564	430	429
icdA	Isocitrate Dehydrogenase	1197	557	555
LipL32	Outer membrane Lipoprotein LipL32	819	474	450
LipL41	Outer membrane Lipoprotein LipL41	1068	518	492
rrs2	16S ribosomal RNA	1512	452	450
secY	Pre- protein translocase secY	1383	549	501
	Concatenated sequence length		2980	**2877**

2.3.B.2 Web hosting of Ahmed's MLST scheme

PubMLST, a publicly accessible database intended for molecular typing and microbial genome diversity, hosts majority of published MLST schemes of prokaryotes and eukaryotes (Jolley and Maiden, 2013). A separate database with the name MLST scheme #3 under *Leptospira* was created at PubMLST domain (www.pubmlst.org) for Ahmed's MLST scheme which can be accessed publicly (**Figure 12**). Ridom SeqSphere software was used to designate

and number allele types at each locus as per the altered size and to generate the allelic profile of each isolate. Allele sequences and allelic profiles were uploaded and each unique allelic profile was assigned a sequence type (ST).

Sequence/profile definitions database in PubMLST includes allele sequences of each locus and allelic profiles for sequence types. This section gives public access to view, search database against query sequence and download sequence details of one or more loci and sequence types (**Figure 13**). *Leptospira* isolates metadata, containing information about the organism, geographic isolation and host from where it was isolated, was also uploaded along with their allelic profiles. Isolate database gives access to view, search and download the isolate information for public (**Figure 14**). Information section of database contains details of genes and experimental conditions including PCR conditions, primers and MLST locus size considered (**Figure 12**).

PubMLST database allows users to submit their own data regarding new alleles, new allelic profile and new isolates to the database (**Figure 15**). These submissions will be uploaded to the MLST database after physical curation. Till date *Leptospira* MLST scheme #3 has in total 416 *Leptospira* isolates belonging to total 143 sequence types in the database.

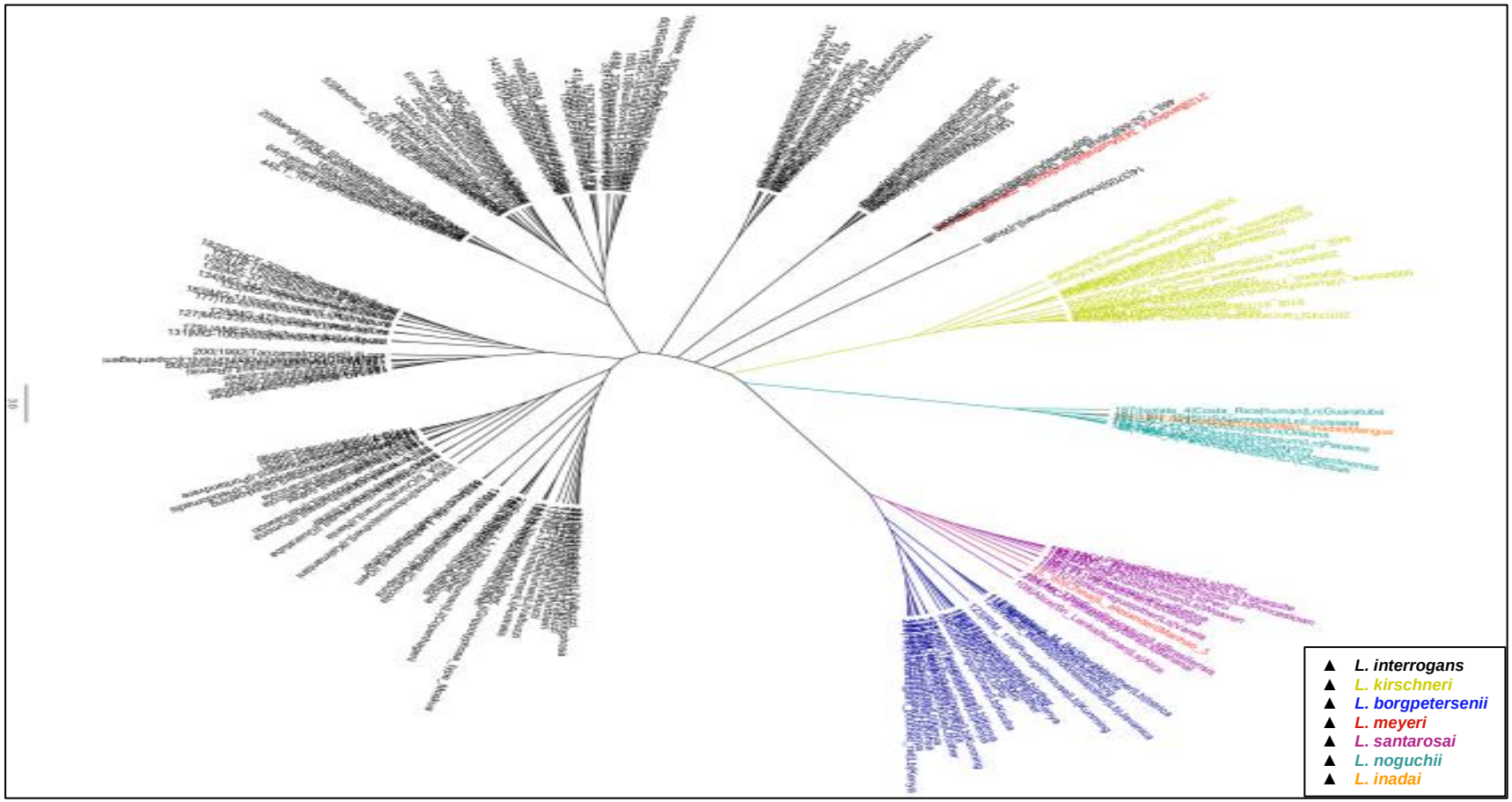

Figure 11: Phylogenetic tree generated with altered concatenated sequences for Ahmed's MLST scheme using NJ algorithm in MEGA 5.0 with 1000 bootstrap values (isolates were given color as per their species)

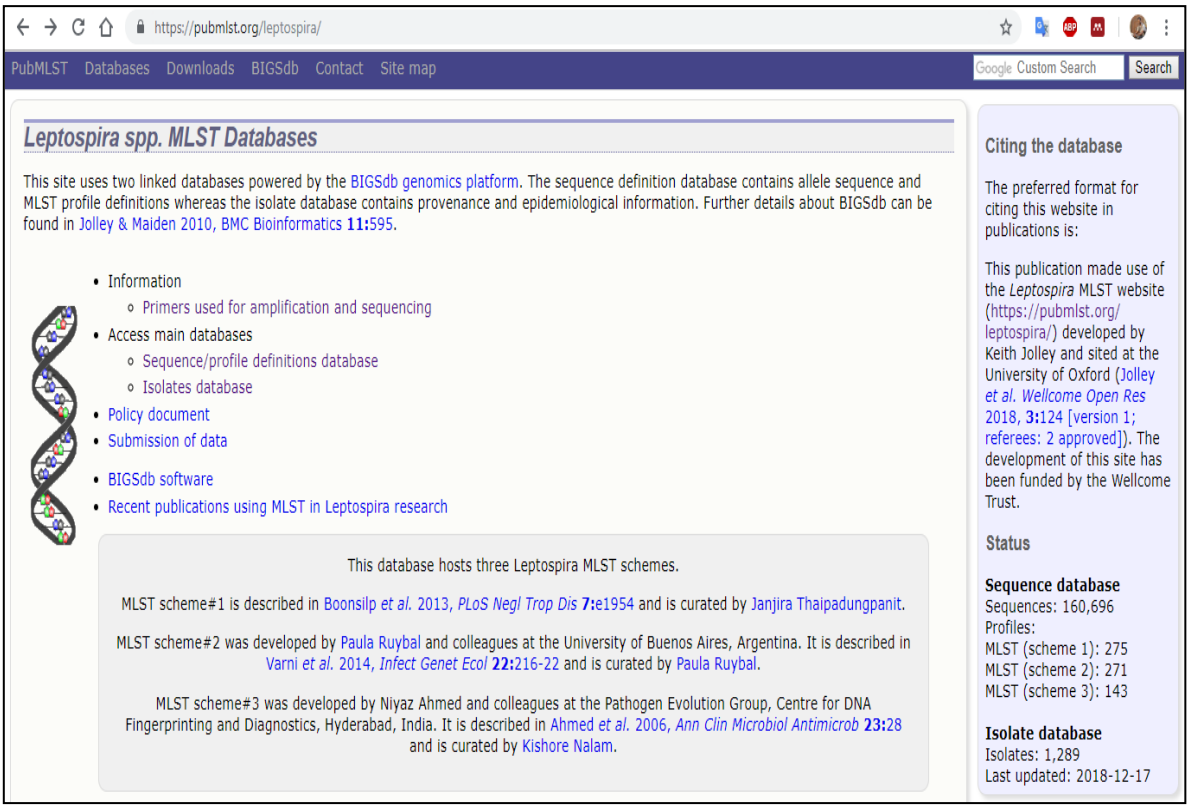

Figure 12: PubMLST database mentioning MLST scheme description and its curators

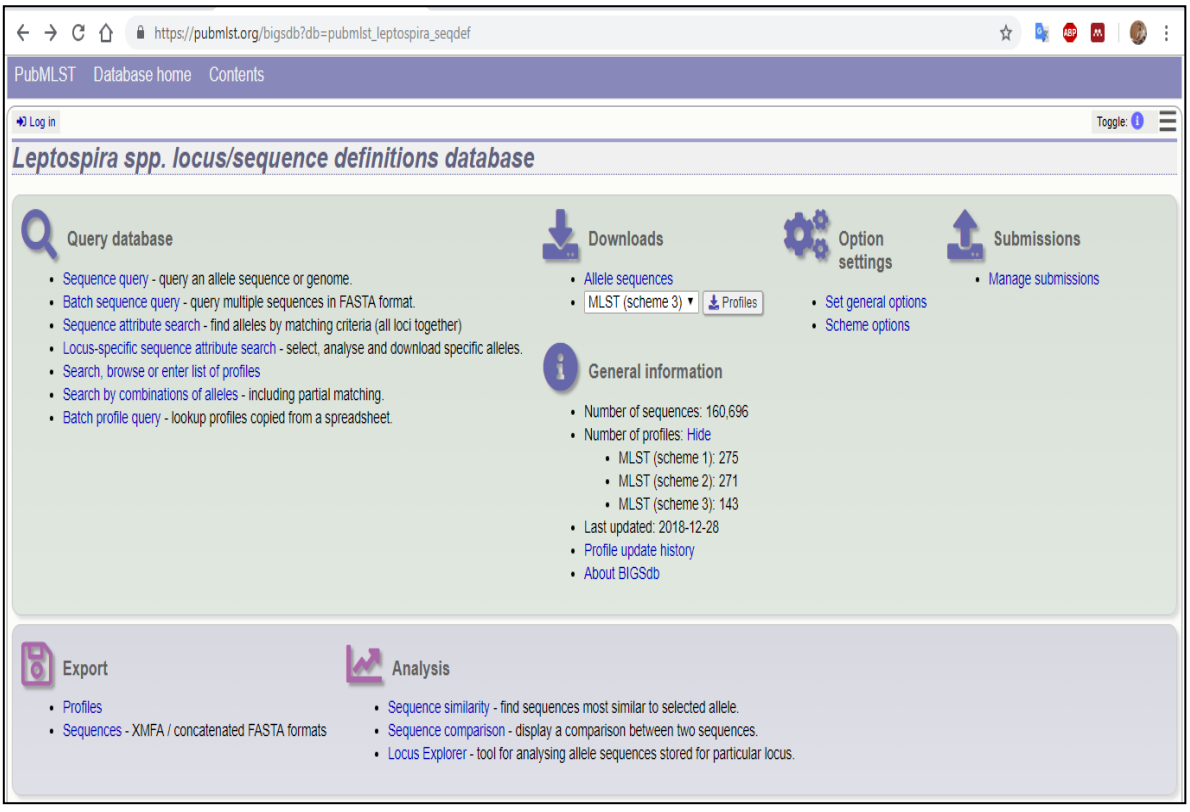

Figure 13: *Leptospira* locus/sequence definition database at PubMLST

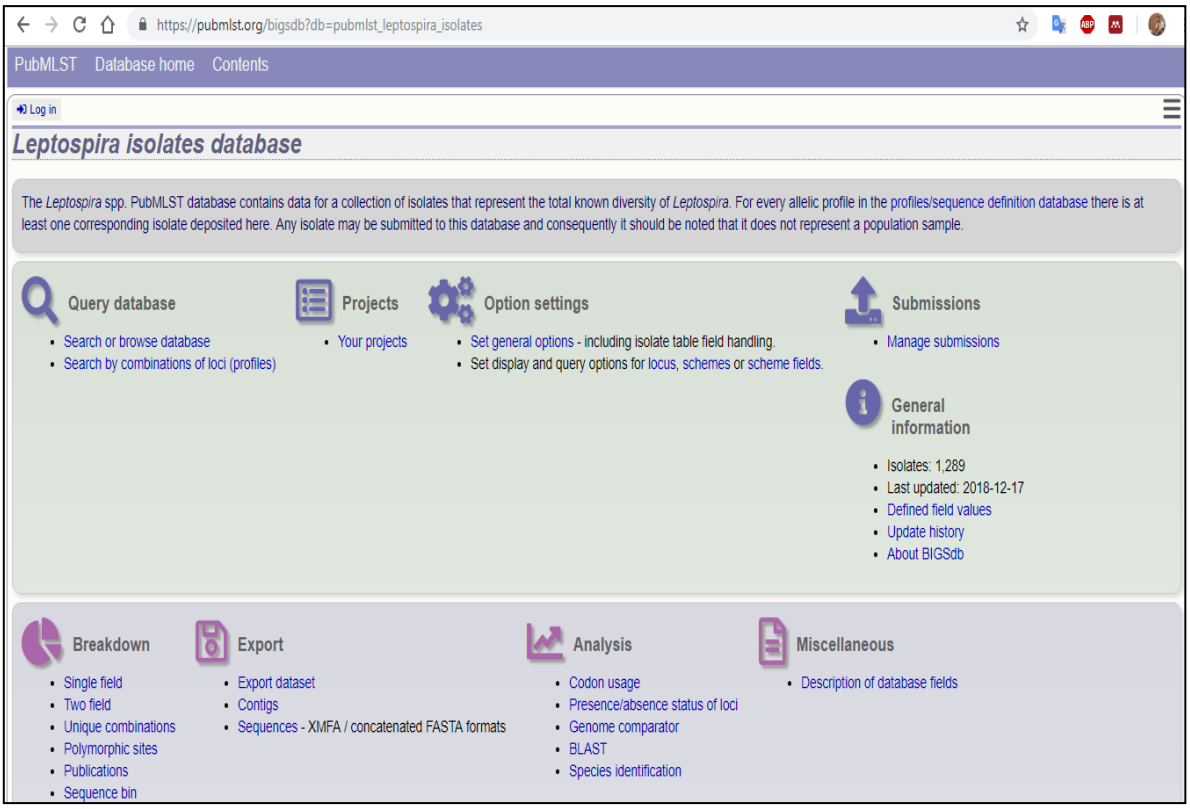

Figure 14: Isolates database at PubMLST

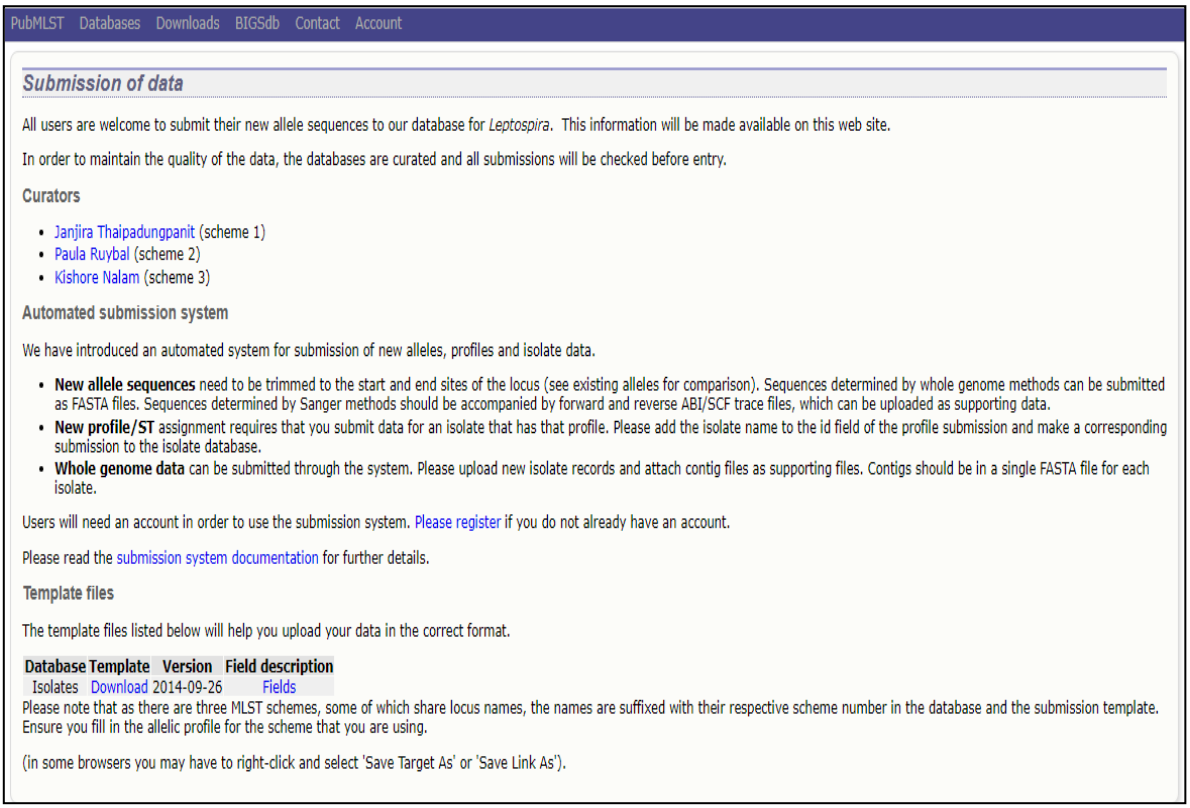

Figure 15: Interface for submission of new data for curation by curators at PubMLST

2.3.C Comparison of Ahmed's and Thaipadungpanit's MLST schemes

A total of 48 strains belonging to only *L. interrogans* (n=40) and *L. kirschneri* (n=8) were used in this comparative analysis as the Thaipadungpanit's scheme was restricted for these two species. All the clinical isolates and the reference strains included in the study were evaluated for Thaipadungpanit's scheme in their previous study (Thaipadungpanit *et al.*, 2007) and only 9 reference strains were evaluated previously using Ahmed's scheme (Nalam *et al.*, 2010). All other isolates which were not evaluated using Ahmed's scheme were evaluated at KIT biomedical research center, Amsterdam. Typing results of two *L. interrogans* strains, a reference strain 136/2/2 and a clinical strain L1207 for Ahmed's scheme were not considered because of presence of a three-nucleotide deletion at their *lipL32* locus which was not observed in previous studies, restricting total isolates to 46.

2.3.C.1 Nucleotide diversity of genetic loci

All 46 isolates were resolved into 30 STs by Ahmed's scheme and into 21 STs by Thaipadungpanit's scheme confirming the higher discriminatory ability of the former. Most of the alleles were also found to be species specific, either for *L. interrogans* or *L. kirschneri* with very few exceptions. The overall level of diversity at 95% confidence interval (CI) was found to be 93.5% and 92% for Ahmed's and Thaipadungpanit's schemes respectively (**Table 10**). d_N/d_S ratios of loci suggested that none of them were under positive selection.

Table 10: Discriminatory ability of Thaipadungpanit's and Ahmed's schemes

Thaipadungpanit's scheme				Ahmed's scheme			
Gene	**Number of alleles**	**dN/dS**	**Discriminatory ability (%) (95% CI)**	**Gene**	**Number of alleles**	**dN/dS**	**Discriminatory ability (%) (95% CI)**
glmU	11	0.073	86.9	***adk***	10	0.057	70.2
pntA	11	0.012	64.3	***icdA***	12	0.022	74.8
sucA	7	0.007	59.3	***lipL32***	7	0.154	71.9
fadD	7	0.066	76.3	***lipL41***	7	0.01	81.9
tpiA	10	0.093	84.7	***rrs2***	6	NA*	66.3
pfkB	14	0.048	83.4	***secY***	20	0.019	91.8
mreA	12	0.007	86.9				
ST	21		92.0	**ST**	30		93.5

* dN/dS ratio was not applicable (NA) to rrs2 locus as it does not encode a protein

Figure 16: NJ tree generated for 46 strains using Thaipadungpanit's scheme (A) and Ahmed' scheme (B) MLST data

2.3.C.2 Phylogenetic relationships

Phylogenetic trees were generated employing NJ algorithm using concatenated sequences of both the schemes (**Figure 16**). Isolates were numbered with the ST numbers of Thaipadungpanit's scheme for easy comparison. Both the schemes were able to discriminate isolates and placed all the isolates into two distinct groups corresponding to their species.

Clonal structure of many isolates, like in the case of ST34 and ST46, was retained in both the schemes. As it is found that discriminatory power of Ahmed's scheme was more than Thaipadungpanit's scheme, former scheme could able to split isolates of a later scheme ST into closely related STs, like in the case of ST17, ST37, ST42, ST49 and ST68. In addition, *L.kirschneri* strains, Kipod179 and MoskvaV, which are designated as ST110 in Thaipadungpanit's scheme were placed distantly apart in Ahmed's scheme and was supported by their different serovar classification. Another discrepancy observed was for two *L.interrogans* strains, 654 and Hardjoprajitno, where Ahmed's scheme showed them to be closely related and Thaipadungpanit's scheme separated them distinctly apart. The relationship among STs and isolates was found to be not conserved among many isolates across the schemes which can only be further confirmed by any other genotyping method.

In addition to the higher discrimination power, major advantage of Ahmed's scheme is that it is applicable to all major pathogenic species of the *Leptospira* genus where as Thaipadungpanit's scheme, although supported by a web based database, limited only to two species of *Leptospira*.

2.3.D Comparison of MLST schemes with WGS inferred phylogeny

It was evident from the previous studies that the whole genome-based phylogenetic studies are on par with any other established methods and maintains a bench mark for comparing and validating other molecular methods for drawing phylogeny (Agren *et al.*, 2012; Ahrenfeldt *et al.*, 2017).

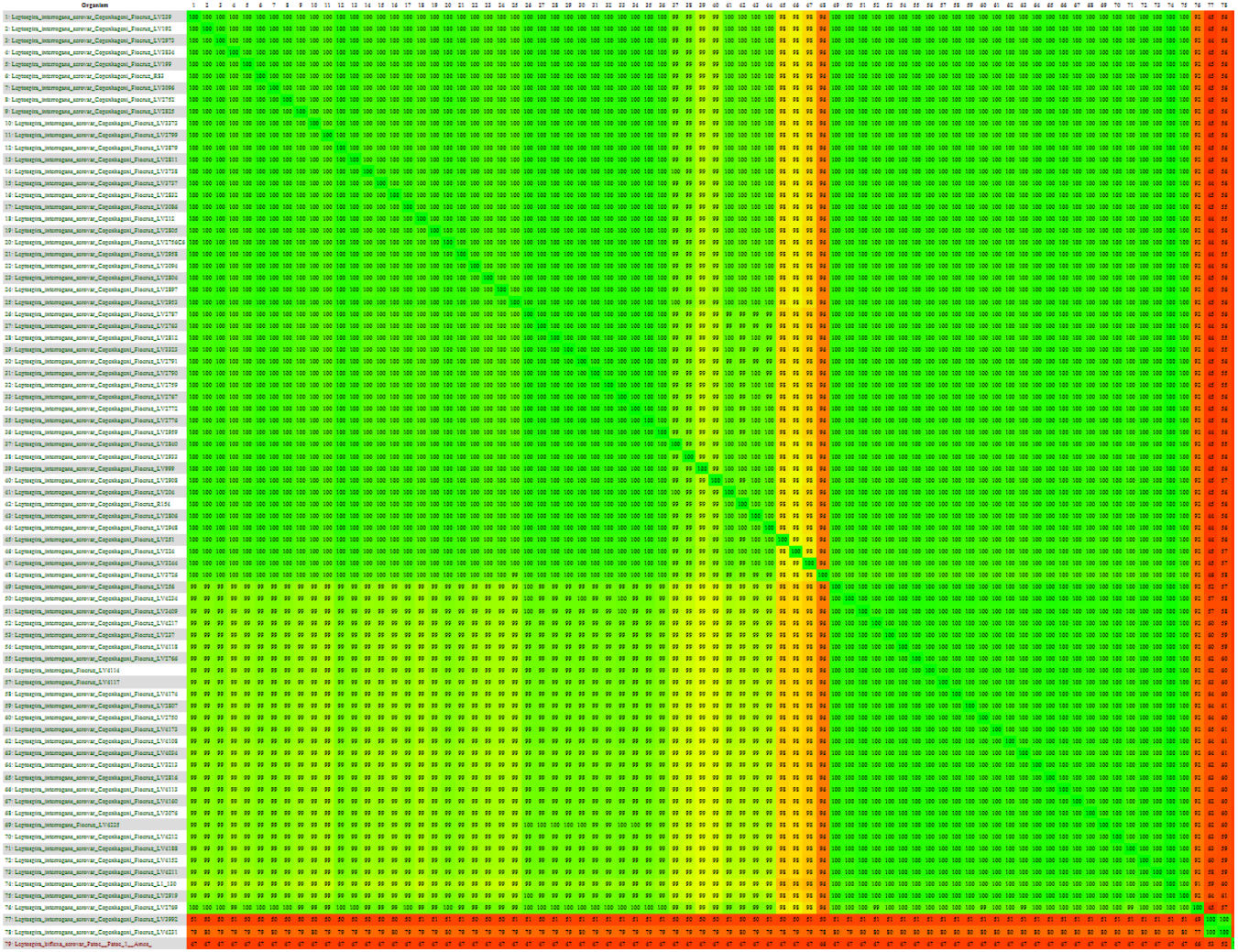

Figure 17: Heat-plot generated based on BLASTN similarity scores obtained from fragmented alignment of whole genomes of Fiocruz isolates belonging to *L. interrogans* Serovar Copenhageni using Gegenees software

2.3.D.1 Clonality among Brazil isolates

As MLST evaluation has to be performed using isolates from diverse population, 78 isolates isolated at Fiocruz, Brazil were analyzed for the possibility of clonality or dominance of a single clone in the data set. An initial whole-genome comparison was made using Gegenees version 2.2.1 and heat-plot generated from BLASTN similarity scores of fragmented alignment approach showed that most of the genomes were having 100% identical average nucleotide identity (ANI) (**Figure 17**). In addition, sequence types retrieved for Varni's MLST scheme using *in-silico* MLST method has shown that most of these isolates belongs to ST47 and ST17. Based on these two methods, a total of six Fiocruz isolates belonging to serovar Copenhageni were included in the final study as representative isolates (**Table 5**).

2.3.D.2 Phylogenetic relationships inferred by whole genomes and MLST methods

A total of 149 whole genome sequences of pathogenic isolates along with two biflexa isolates were used to generate whole genome-based phylogeny. Normalized mean values of fragmented genome alignment BLSATN scores obtained using Gegenees program were used to generate heat-plot (**Figure 18**) and to construct phylogenetic tree (**Figure 19**) using NJ algorithm. Heat plot of whole genomes, showing one to one similarity scores based on average nucleotide identity, has confirmed the genetic similarity at species level and remarkable difference across the species of *Leptospira*.

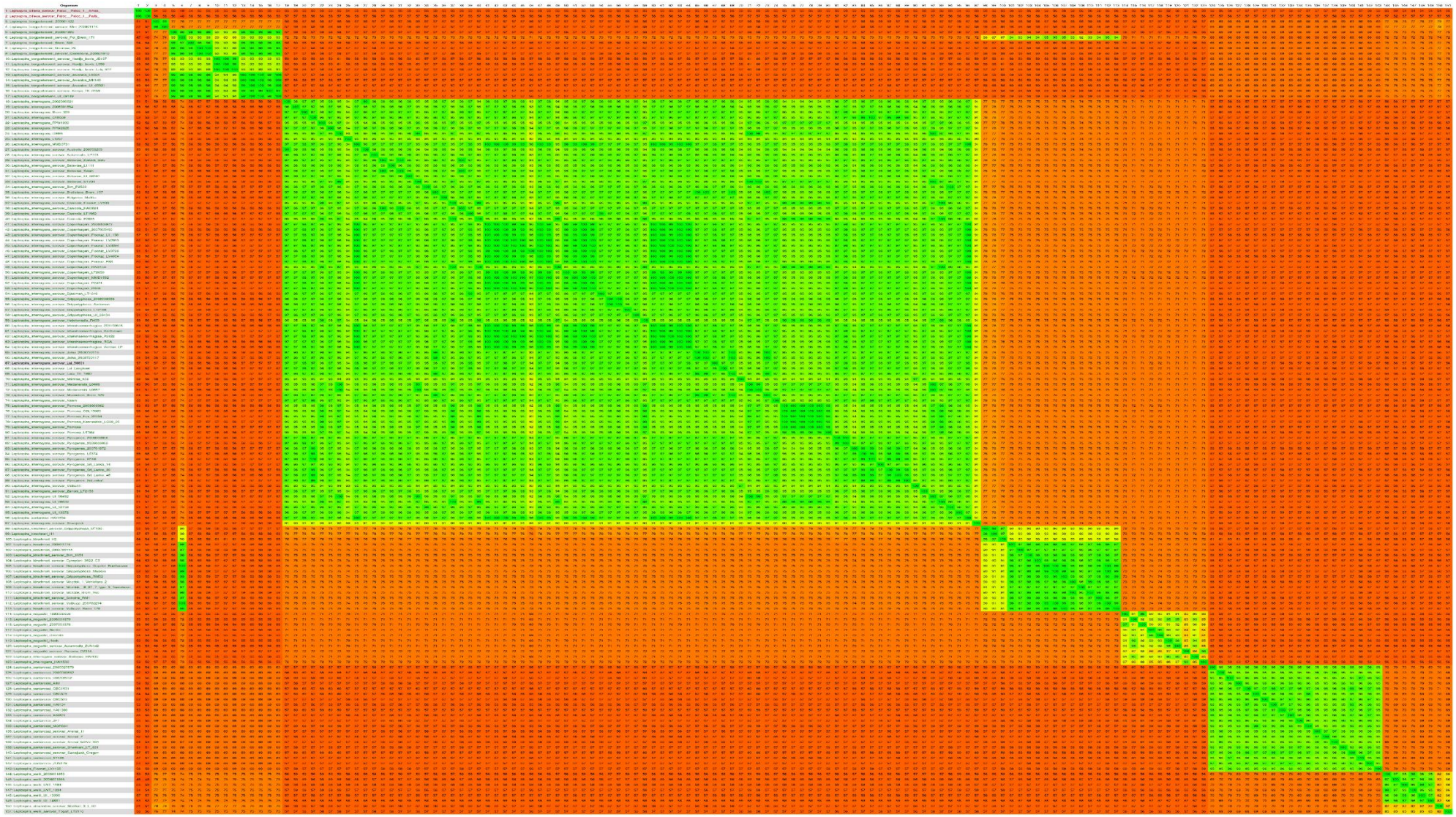

Figure 18: Heat-plot generated based on BLASTN similarity scores obtained from fragmented alignment of whole genomes of *Leptospira* isolates using Gegenees software

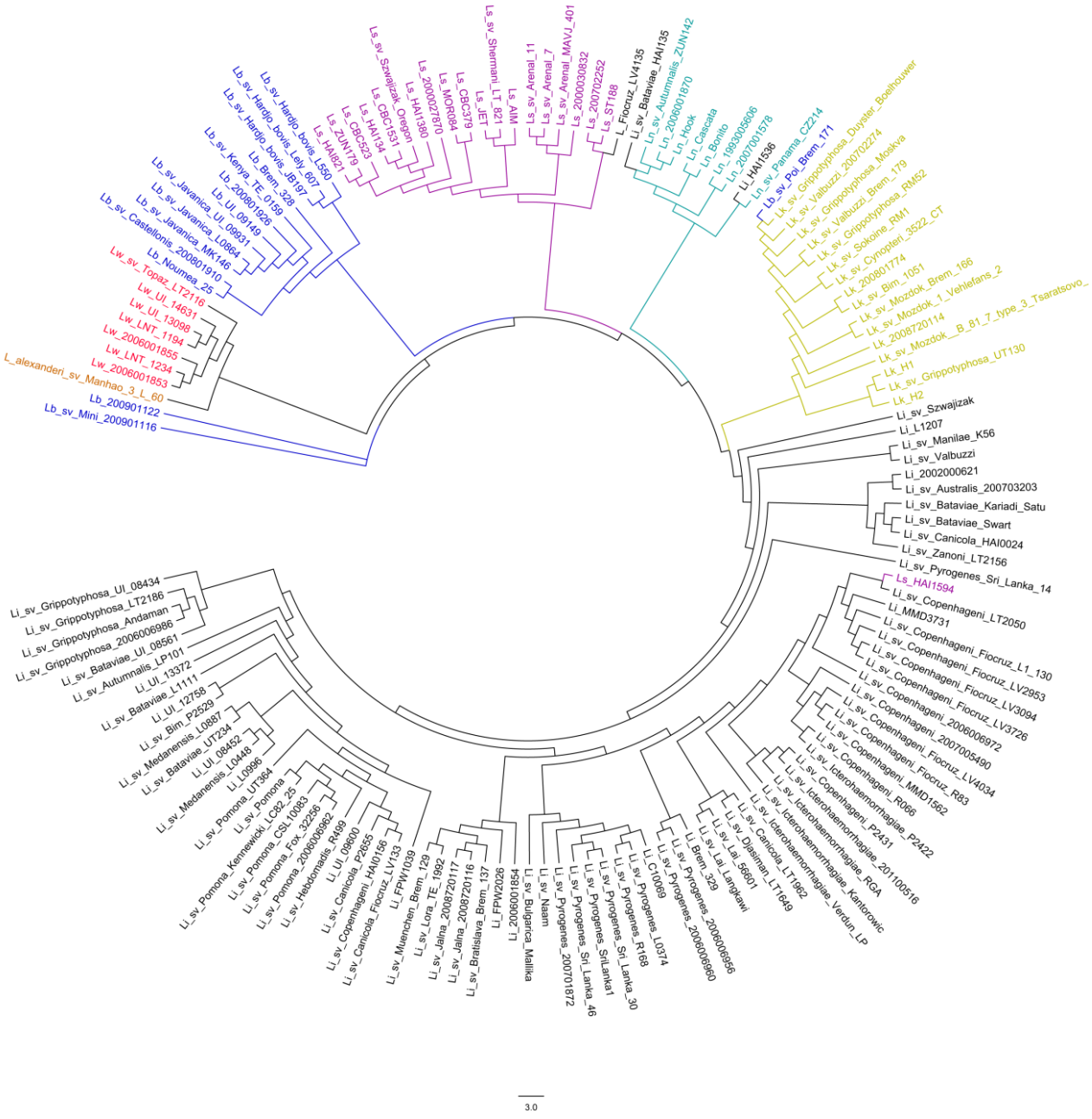

Figure 19: NJ phylogenetic tree generated from whole genome similarity scores for *Leptospira* isolates obtained using Gegenees software (Different colors were used for different species).

Allele sequences retrieved from whole genomes sequences were used to generate sequence types for all three MLST schemes and phylogenetic trees were constructed using NJ algorithm in MEGA software (**Figure 20**). Phylogenetic trees constructed using MLST schemes and whole genome similarity showed similar clustering of isolates and species with very few exceptions.

In all phylogenetic trees, species wise discrete clusters were observed. Monophyletic origin of clusters for *L. interrogans, L. kirschneri, L. noguchii, L. santarosai* species were observed in all schemes except in Boonsilp's scheme where in *L. kirschneri* and *L. noguchii* clustered individually in the same branch. *L. borgpetersenii* and *L. weilii* clusters were found to be on the same branch in all phylogenies.

When analyzed for clustering and grouping of isolates, a total of 28 clusters comprising 80 isolates in Boonsilp's scheme tree were found to be in accordance with whole genome inferred phylogeny whereas a total of 30 clusters comprising 94 isolates and 92 isolates in Varni's scheme and Ahmed's scheme, respectively, were found to be in accordance with whole genome inferred phylogeny (**Figure 21 & 22**).

Clade switching was observed for few isolates in whole genome based phylogeny where in isolate characterized as one species by serological method was clustered with isolates of other species. As WGS based phylogeny is dependent on the total genome similarity, an isolate's taxonomical position in the WGS phylogeny is considered as best justified than any other typing method. Clustering of *L. interrogans* strain HAI1536 and *L. interrogans* strain HAI135 with *L. noguchii* species, *L. santarosai* strain HAI1594 with *L. interrogans* species and *L.borgpetersenii* strain Brem171 with *L. kirschneri* species was observed in all phylogenies. *Leptospira* strain Fiocruz LV4135 was found to cluster with *L. santarosai* species suggesting its species status which was not confirmed earlier. All these clade switching events were also found to be conserved in phylogeny inferred from all three MLST schemes (**Figure 21 & 22)**.

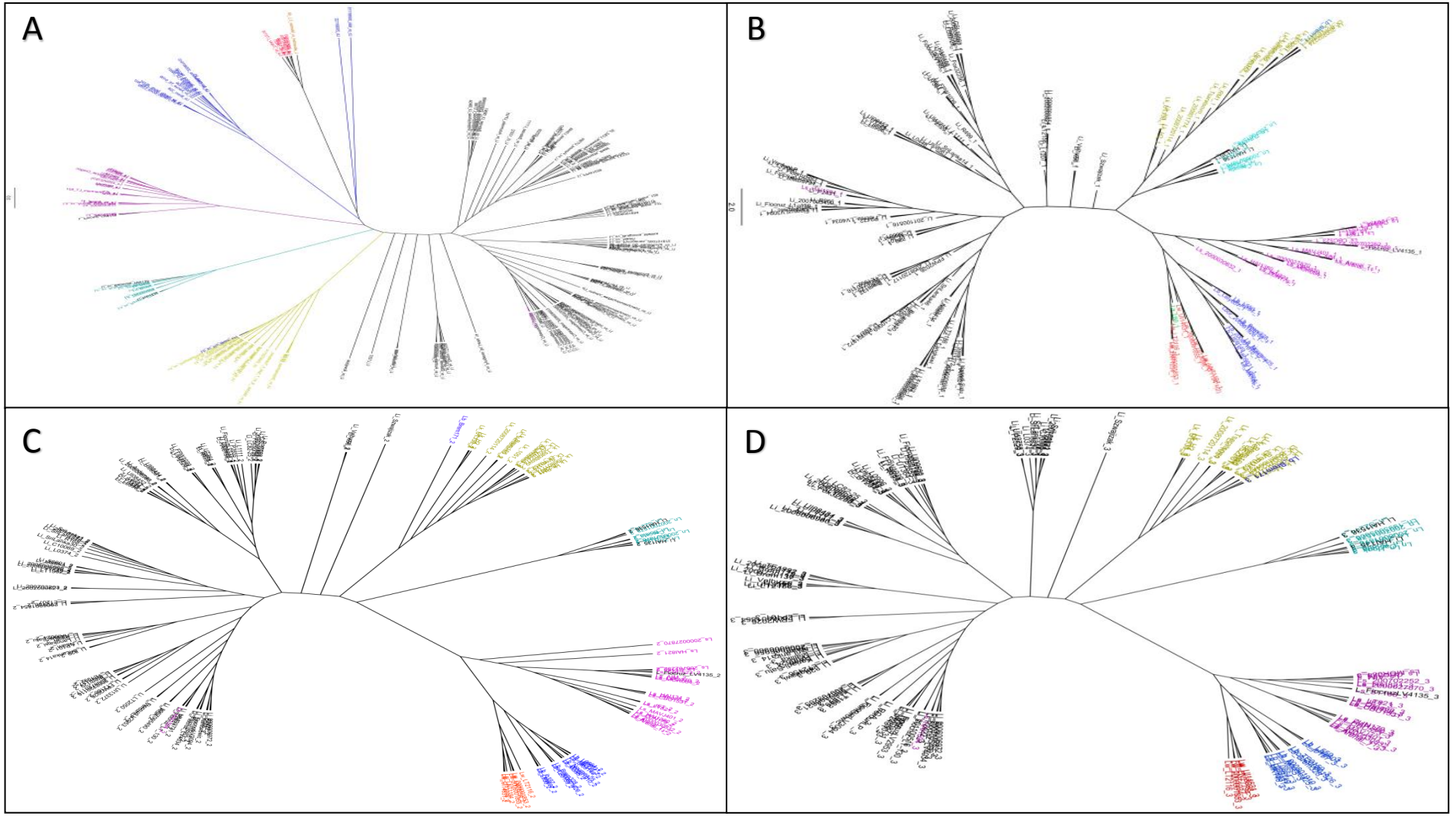

Figure 20: NJ trees generated using whole genome based similarity scores (A) and MLST profiles of Boonsilp's scheme (B), Varni's scheme(C) and Ahmed's scheme (D) (Different colors were used for different species).

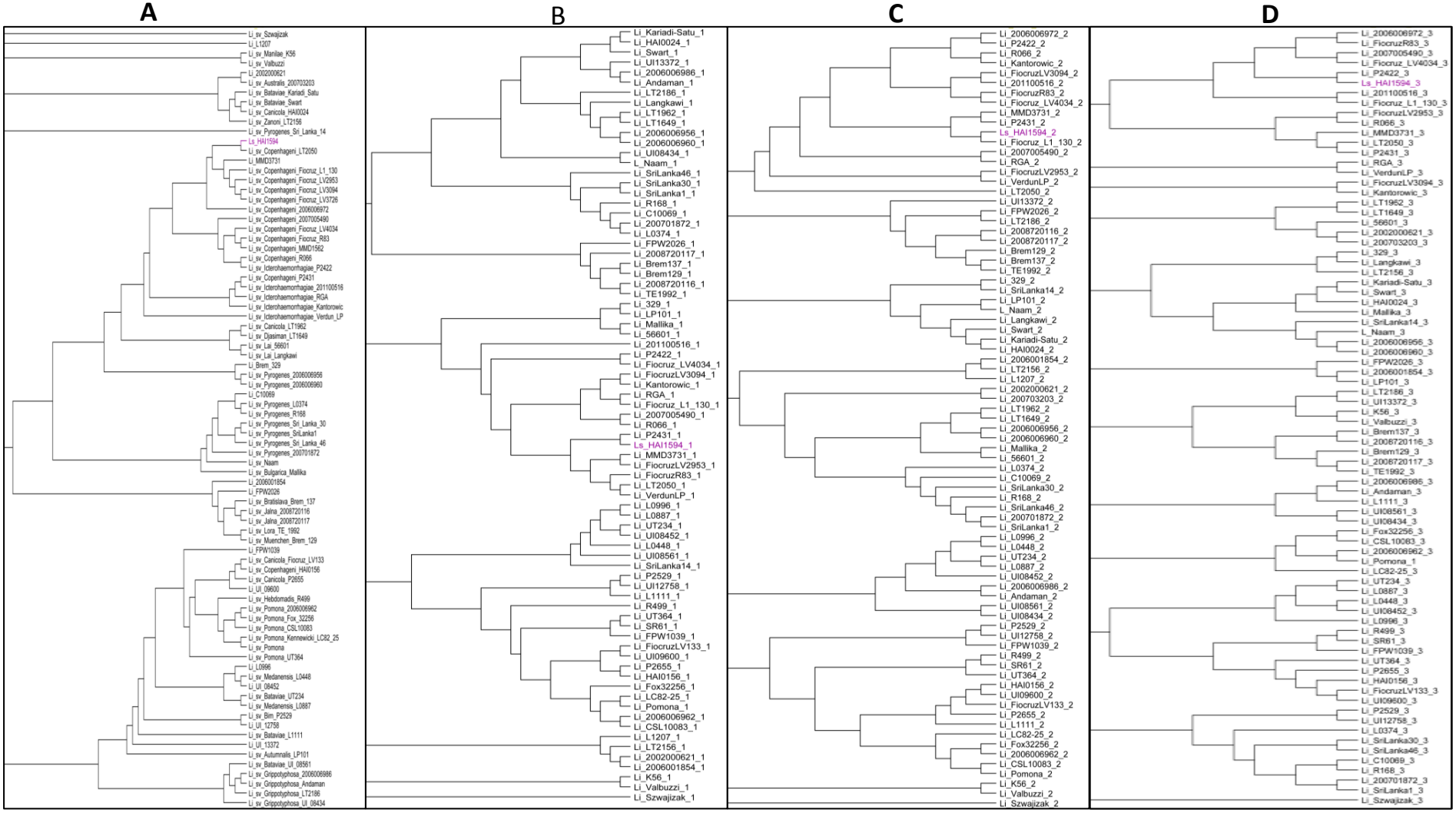

Figure 21: NJ trees, showing *L. interrogans* (Li) species cluster, generated from whole genome based similarity scores(A) and MLST profiles of Boonsilp's scheme(B), Varni's scheme(C) and Ahmed's scheme(D). (Sv- Serovar) (Species wise coloring was given for isolates)

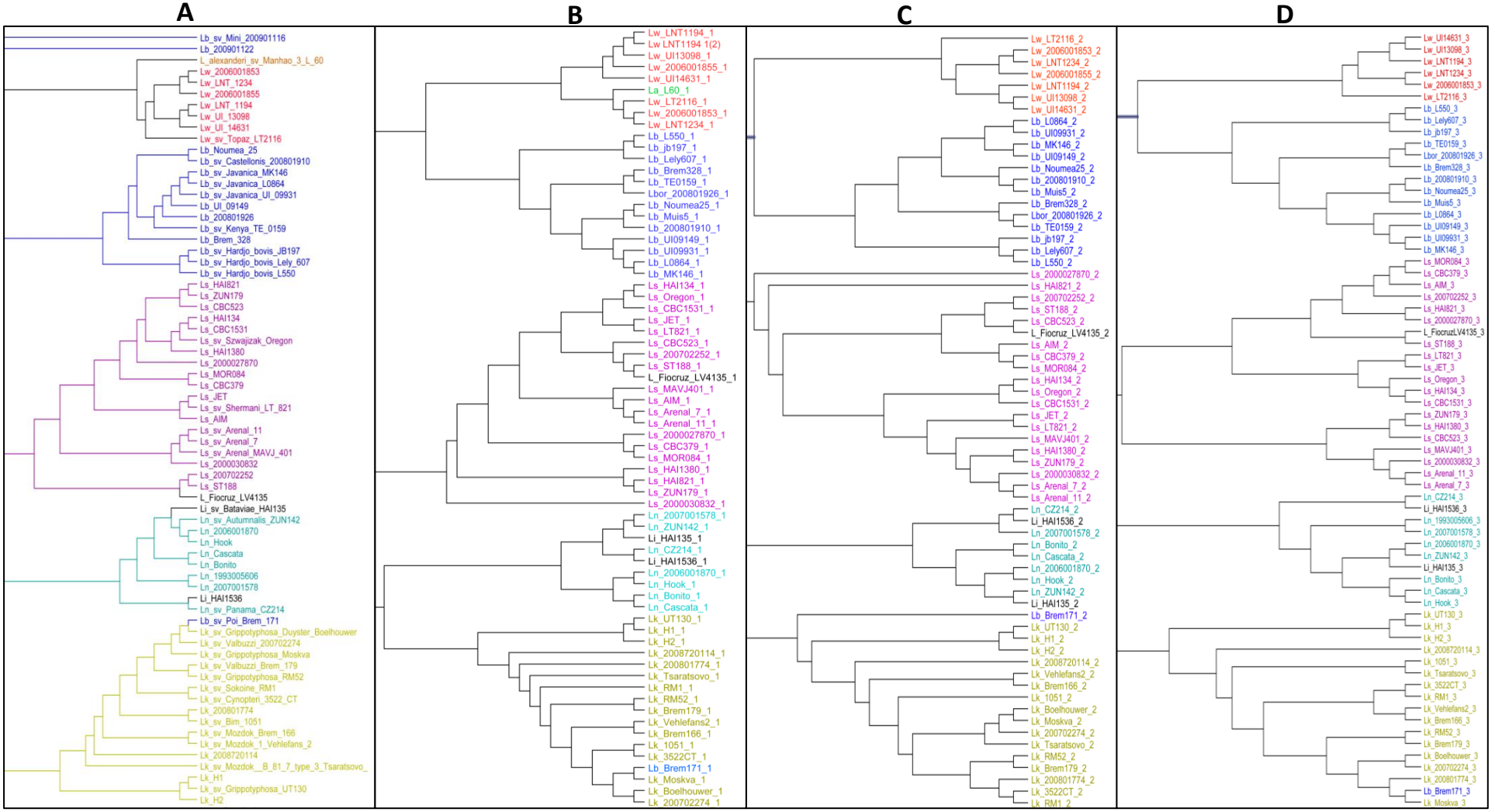

Figure 22: Neighbor Joining trees, showing *L. weilii* (Lw), *L. borgpetersenii* (Lb), *L. santarosai* (Ls), *L. kirschneri* (Lk) *and L. noguchii*(Ln) species clusters, generated from whole genome based similarity scores(A) and MLST profiles of Boonsilp's scheme(B), Varni's scheme(C) and Ahmed's scheme(D). (Sv – Serovar) (Species wise coloring was given for isolates)

Discrete clustering was observed for many isolates in *L. interrogans* species as per their serovar nomenclature, especially for Copenhageni, Icterohaemorrhagiae, Pomona, Grippotyphosa and Pyrogenes serovars in whole genome based and MLST based phylogenies. Similar clustering was also observed for isolates of *L. borgpetersenii* serovars Javanica and Hardjo, and *L. santarosai* serovar Arenal in all the methods employed (**Figure 21 & 22**).

Heat-plot of genome similarity shows that *L. alexanderi* Serovar Manhao 3 Strain L 60 was found to be genetically similar to *L. weilii* species and its presence as a singleton in the phylogenetic tree suggest its subspecies level or a closely allied species of *L. weilii* which can further be confirmed by including more number of *L. alexanderi* isolates in the study.

Many isolates belonging to one serovar were observed to be in clustering with other serovar isolates which can be best explained by the possible ever-changing antigenic repertoire of the organisms or the incorrect serological typing method performed earlier.

2.4 Discussion

2.4.1 Ahmed's MLST scheme for global applicability of *Leptospira* strain typing

Although earlier Ahmed's scheme was tested on a limited set of *Leptospira* isolates (n=120), it proved that the six loci employed was able to individualize isolates up to the strain level (Ahmed *et al.*, 2006). This analysis using 271 pathogenic *Leptospira* isolates sourced from different *Leptospira* Reference Laboratories distributed across the world, representing global collection of *Leptospira* isolates has confirmed that the scheme can be applied at global level across all pathogenic species of the genus. This analysis has also proven that all six MLST loci included in the study have exhibited a high degree of sequence diversity and resolution.

It was observed that a greater level of sequence diversity was present among MLST loci and sequence types across the species. Thus the dendrogram generated could group isolates belonging to *L. interrogans ss, L. kirschneri, L. noguchii, L. weilii, L. borgpetersenii* and *L. santarosai* species explicitly according to their species classification. Analysis of the *Leptospira* strains using Ahmed's MLST scheme confirms earlier findings that the serovars and serogroups, that do not have any taxonomic status, are clustered according to their species based on DNA relatedness (Brenner *et al.*, 1999). In addition, analysis by Bayesian MCMC confirmed that the method is capable of individualizing *Leptospira* isolates up to species level with flexibility to type isolates with many different taxonomic identities.

Having said this, we should also consider limitations of MLST in terms of its failure to resolve the horizontal variome, which might play an important role in speciation. Horizontal variome depends on the extent and impact of HGT in different bacterial species. This was evident in our study where in few isolates switched their clades in the phylogenetic tree, also might be because of incorrect and ambiguous typing methods used before for characterization. An explanation for this requires a detailed evaluation at the whole genome sequence of the isolate. Inclusion of target genes other than the housekeeping genes, namely,

the envelope proteins, LipL32 and LipL41 in Ahmed's MLST scheme has allowed sampling of genome variation beyond the core genome and might have relevance in epidemiologic/taxonomic resolution of the strains (Ahmed *et al.*, 2006; Victoria *et al.*, 2008).

In an extended study by our group, these isolates were also typed using another genome based molecular typing methods, FAFLP and MLVA. In the study all 271 isolates were typed using FAFLP and phylogenetic tree was constructed. A close comparison between the phylogenetic trees generated from FAFLP and MLST method was performed to draw ancestral origin, genetic affinities and linkages among one another to shed light on to the typing of *Leptospira* and its epidemiology. FAFLP method, which depends on whole-genome micro restriction patterns to construct phylogeny, had revealed quite complex and confusing genetic affinities among various *Leptospira* species. This was not surprising given the resolution power of the FAFLP technique. Further analysis of the bifurcating NJ tree revealed broadly species-specific clusters although clade-switching by few strains, in almost each cluster, was clearly evident with respect to their projected genomic DNA-based species status. Thus, more than one cluster was observed for *L. interrogans, L. kirschneri,* and *L.borgpetersenii*. Further, it was also noted that the splitting of these clusters was not in agreement with geographical descent or the host species types and believed that discrete genetic associations arising due to recombination events could lead to such sub clustering. Nevertheless, plausible ancestral associations were found in the tree in terms of co-clustering of *L. kirschneri* with *L. interrogans*; *L. borgpetersenii* with *L. santarosai,* and *L. noguchii* with *L.kirschneri*. (Nalam *et al.*, 2010).

Further, MLVA method proposed by Majed *et al* for *Leptospira* classification, when applied to *L. interrogans* isolates (n=134) of the collection, revealed embranchment broadly confirming to the serotypical positions and corresponding to serogroups Icterohaemorrhagiae, Djasiman, Autumnalis, Australis, Canicola, Sejroe, Pyrogenes, Hebdomadis, Pomona, and

Grippotyphosa. Although MLVA technique may not be applicable across all species of *Leptospira*, its phylogeny revealed that these isolates belong to different serogroups representing different serotypes thus ruling out the convenient sampling of *L. interrogans* isolates which was the largest set of isolates for this study (Nalam *et al.*, 2010).

2.4.2 MLST as the gold standard for *Leptospira* typing

Given the advantages of MLST in terms of its robustness, simplicity and efficiency in identifying ancestral relationships and segregating strains according to the genome species status, we believe that MLST is sufficient to replace tedious serotyping procedures currently in practice and implementation of MLST as the typing method for *Leptospira* isolates for accurate identification and classification. We also strongly propose that issues related to strain diversity as well as the taxonomic organization and accuracy of the reference collections can be set to rest in the best possible way. Allelic profiles and sequence types generated by MLST could also be successfully used to gain insights in to the evolution of pathogen based on their phylogeographic affinities. This genotypic characterization method can also be applied successfully to establish whole genome based metagenome to investigate diagnostic markers, vaccine candidates and strain-specific coordinates. Strain –specific coordinates can be used in reconstructing the evolutionary history of the organism to understand its emerging or re-emerging status in a particular epidemiological catchment area. These investigations also will have promising impact on strengthening the cause of "Functional molecular infection Epidemiology (FMIE)" of *Leptospira*.

Given the capability of MLST method in distinguishing the outcomes of genetic recombination and erroneous serotyping responsible for clade switching, MLST appears to be inevitable and can be used as an alternative typing method to the traditional tedious serotyping method.

2.4.3 Genetic affinities within large global collection

In summation of genetic affinities with the large global collection of *Leptospira*, it is evident that *L. interrogans, L. kirschneri and L. borgpetersenii* are considered as ubiquitous species with *L.interrogans* being the major species responsible for human infections particularly from South Asia. Whereas *L. weilii* is considered to be largely confined to Asia and *L. santarosai* and *L.noguchii* species were confined to America. The trend observed with the isolates collected over the period of time at different reference laboratories will shed light on understanding transmission dynamics and *Leptospira* evolution.

2.4.4 *Leptospira* MLST database

A sophisticated database developed and hosted at www.pubmlst.org for Ahmed's MLST scheme will help the research community and public health across the world to access and query sequenced isolates and sequence types. PubMLST being the major MLST database for all organisms was aimed to cater the needs of scientific community to easily mine and analyse the data. This database lets researchers to view, download and analyse sequences already deposited and to query or submit their own sequences. This database will be permanently available and new submissions will be curated by us regularly to update isolates and sequence databases.

2.4.5 Comparison of MLST schemes

With the availability of more than one MLST method for *Leptospira*, it was inevitable to compare the available methods to clear uncertainty among the research community for adopting a single scheme. During the initiation of this study in 2007 there were only two schemes reported and are by Ahmed *et al* (2006) and Thaipadungpanit *et al* (2007). Initial comparison of these two MLST methods collectively as a metacentric study by involving representatives of the scientific groups who originally developed these methods was carried out to detail the advantages and disadvantages in an unbiased way. Major difference with the

MLST scheme proposed by Thaipadungpanit *et al* is that it was applicable to only two pathogenic species *L. interrogans* and *L. kirschneri,* whereas Ahmed's scheme was applicable to all major pathogenic species of the genus. Although, Ahmed's scheme does not completely developed on conventional strategy of MLST by employing only housekeeping genes, none of the employed genes were found to be under positive selection, which is the basis for selecting housekeeping genes. More number of genotypes were resolved using Ahmed *et al* scheme than Thaipadungpanit's scheme, indicating its higher level of discrimination within species. Whereas, on sliding window analysis it was shown that Thaipadungpanit's scheme has better resolution between species. In addition Ahmed's scheme during the time of this comparison was not having any online accessible database compelling researchers to download all sequence types, available in the public domain and to analyze them offline. This handicap was resolved in the following years by developing a dedicated website for Ahmed's MLST scheme.

Further, shortcomings outlined in the above comparison were overcome in the following years through modifications in the Thaipadungpanit's scheme by Boonsilp *et al* by changing *fadD* locus with *CaiB* locus and changing primers sequences to make it applicable to all seven pathogenic species of *Leptospira* (Boonsilp *et al*., 2013).

In addition to Boonsilp's and Ahmed's MLST schemes, a consensus MLST scheme was developed by Varni *et al* using three loci from Boonsilp's scheme and four loci from Ahmed's scheme to make it applicable to all pathogenic species of the *Leptospira* genus with maximum power of discrimination by adhering to all the conventional strategy of MLST development (Varni *et al*., 2014). This scheme was developed citing the facts that the Boonsilp's scheme was tested with isolates primarily isolated from a single human outbreak in Thailand questioning the methods global applicability and inclusion of non-housekeeping genes in the Ahmed's scheme (Goarant, 2014).

Whole genome sequences hold great potential in developing methods for routine characterization of infectious agents and microbial organisms that are difficult to grow under laboratory conditions. Whole genome sequence datasets were proven successful in constructing benchmark phylogeny for an organism. So, our approach to compare phylogenies generated by all three publicly available MLST schemes with WGS based phylogeny has revealed clear insights into the genetic relationships of organism and accuracy of methods. Although discrete clusters were observed for all species, Boonsilp's scheme has limited in terms of keeping *L. kirschneri* and *L. noguchii* on the same clade with different branches.

The only *L. alexanderi* isolate L60 included in the study was grouped together with *L. weilii* isolates in both WGS and Boonsilp's scheme phylogenies. Although, in the present study this isolate could not be typed using Ahmed's and Varni's schemes because of sequence termination for *adk* locus, previous studies has shown that *L. alexanderi* isolate was shown to cluster with *L. borgpetersenii* and *L. santarosai* isolates respectfully (Ahmed *et al.*, 2006; Varni *et al.*, 2014). Inclusion of more number of *L. alexanderi* isolates in the future study will help in resolving its proper position in the phylogenetic tree. Clade switching events in MLST phylogenies were also found to be consistent in WGS based phylogeny ruling out the possibility of genetic variome and alluding a possible incorrect serological classification or mis-nomenclature during transportation or storage of isolates. Given the genetic distance across the species clusters and the average nucleotide identity evident by the genome similarity heat plot generated, the discrepancies in the isolate typing were sorted out without any ambiguity.

WGS phylogeny has reflected well accepted serovar classification of *L. interrogans* for the serovars Copenhageni, Icterohaemorrhagiae, Pyrogenes, Pomona and Grippotyphosa in clustering the isolates of the serovars together. It is also evident that serovars Copenhageni

and Icterohaemorrhagiae are showing very close relationship in all the four schemes employed demonstrating their evolution from a common ancestor. This analysis has also shown dissimilarities with the serovar classification where in few isolates belonging to a different serovar are sitting tightly in the clusters of other serovars implying discrepancies in serotyping results or ever-changing outer membrane protein repertoire, which is the basis for serological classification.

Further, it was evident that all MLST schemes employed for typing *Leptospira* are more or less similar in typing bacteria and conserving the relationships among the strains in accordance with WGS inferred phylogeny. In-depth analysis of dendrogram has shown that Varni's and Ahmed's scheme has conserved the relationships of almost 94 strains (65%) in accordance with WGS phylogeny whereas only 80 strains (55%) were conserved in Boonsilp's scheme. Although performance of individual alleles was not tested in the present study, discrepancies in terms of few loci was observed in previous studies. It was reported that *caiB* locus of Boonsilp's scheme has a 78 bp deletion in few isolates questioning this locus validity at global collection. A deletion event of triplet codon was also observed in an allelic sequence for *LipL32* loci of Varni's and Ahmed's scheme and was not included in the study for further confirmation.

With all these analyses it is evident that Varni's and Ahmed's schemes are having their superiority in clear speciation of isolates of all major pathogenic species in to different branches and conserving the relationships among strains and species as per the WGS inferred phylogeny. In addition to the above advantages, Ahmed's scheme employs only 6 loci for typing, making the method economical and more preferable than the other MLST schemes published online.

Chapter 3

Genomic analysis of pathogenic *Leptospira* isolates

3.1 Introduction

Leptospirosis, a globally prevalent tropical disease, is caused by more than 250 serovars belongs to pathogenic species of *Leptospira*. Disease burden is predominant in developing countries and impoverished population living under inadequate sanitary conditions (Bharti *et al.*, 2003). Yet, a comprehensive understanding of mechanism of pathogenesis of pathogenic *Leptospira* isolates remains poorly understood owing to different factors such as the slow and fastidious growth of the organism in the laboratory, poor transportability of infectious strains across diagnostic centres and laboratories (Ko, Goarant and Picardeau, 2009). Containment measures for disease propagation are also not very successful because of unavailability of efficient and globally applicable diagnostics (Musso and La Scola, 2013).

3.1.1 Pathogenomics of *Leptospira*

Pathogenomics utilizes the genomic and metagenomics data of pathogens to understand the underlying mechanisms of pathogenicity, diversity and host-pathogen interactions. Pathogenomics thus addresses the basic question of "What makes an organism pathogenic?" in context of its genome dictated data. This can be best achieved by comparing the functional differences between pathogenic and non-pathogenic strains or species of the genus (Tettelin *et al.*, 2008; Kumar *et al.*, 2015).

DNA relatedness studies and phylogenetic analysis has divided the *Leptospira* genus into three clusters comprising nine pathogenic species, five intermediates and six saprophytic species revealing its extensive genetic diversity (Ricaldi *et al.*, 2012). Genomic comparison of pathogenic, intermediate and saprophytic isolates of *Leptospira* using representative genomes has revealed that saprophytic *biflexa* genomes are relatively stable whereas pathogenic species genomes undergoes considerable insertion sequence (IS) mediated rearrangements (Gamberini *et al.*, 2005; Bulach *et al.*, 2006; Picardeau *et al.*, 2008). It is also believed that evolution of pathogenic isolates of *Leptospira* has taken a path from *L. biflexa* like ancestors via

a transitional group of intermediate pathogens probably mediated by horizontal DNA transfers (Haake *et al.*, 2004; Ricaldi *et al.*, 2012). Thus genomic comparison of complete set of pathogenic isolates with non-pathogenic isolates at large scale will identify features that are unique to pathogenic and saprophytic species there by providing insights into the factors responsible for pathogenicity, evolution, virulence of the genus and providing new experimental directions.

3.1.2 Whole genome sequencing by NGS platforms

With the advancement of Next-Generation Sequencing (NGS) technologies, sequencing of complete genome of organisms became more economical. Automation of the assembly and annotation of genomes from millions of short sequence reads, generated from NGS platforms, has further decreased the time required to generate whole genome sequence of an organism. Affordability of these techniques and generation of huge amount of data in very less time, made researchers to sequence complete genomes of strains of interest to understand the organisms underlying evolution at genetic level (Metzker, 2010). Analysis of genomic data of an organism in the context of its genetic background made this a better approach to understand the population structure, epidemiology and adaptation to new environmental conditions (Ahmed *et al.*, 2008; Koboldt *et al.*, 2013; Kao *et al.*, 2014). The availability of the whole genome sequence of organisms isolated from outbreaks has led to the characterization and deciphering of the virulence potential of the organism, helping to plan strategies to contain the disease-causing organisms more effectively (Karlin, 2001).

To understand the genetic potential of Indian sub-continent *Leptospira* organisms, an isolate belonging to *L. interrogans* species collected from a patient in Sri Lanka was sequenced. As *L. interrogans* species isolates are most frequently isolated causative organism from patients of sub-continent leptospirosis, analysing its genetic background will aid in understanding the transmission dynamics and its evolution.

Another isolate isolated from a mouse belonging to *L. borgpetersenii* was also sequenced to understand the pathogenic potential of the species which has host specificity. Understanding its genetic background will facilitate in identifying factors responsible for its host specificity and events confining the organism to only host-host transmission in its lifecycle.

As a part of *Leptospira* genome project with the support of NIAID and international Leptospirosis society, a large number of pathogenic and saprophytic *Leptospira* strains isolated from an array of hosts, geographical locations and pathological conditions were sequenced generating a huge amount of data into the public domain. This has made possible to implement the functional genomic analysis of pathogenic *Leptospira* (Lehmann *et al.*, 2014). The availability of whole-genome sequences of pathogenic *Leptospira* organisms has also facilitated the community to effectively design and validate diagnostics, detect organism without the need of bacterial isolation and to avoid laborious and cumbersome serological tests.

3.1.3 The pan-genome

Pan-genome of a species or genus reflects the total genes of the dataset and consists of three parts, first core genome, representing genes present in all genomes, second accessory or dispensable genome, representing genes absent in some genomes and third strain-specific or species-specific genes those are present specifically only in single genome (Medini *et al.*, 2005; Carlos Guimaraes *et al.*, 2015). Analysing the pan-genome is important to understand selective advantages in terms of host adaptation, antibiotic resistance, pathogenicity, bacterial evolution, niche adaptation, lateral and horizontal gene transfers and identification of virulence genes (Medini *et al.*, 2005). In addition, studies on pan-genome also brings inferences in vaccine development and drug design.

Nature of pan-genome describes the genome plasticity at the species/ genus level and is considered as open if its genome size increases on addition of new genome to the analysis

and will be considered closed if the size of genome seizes and will not add any new genes upon addition of a genome to the analysis (H Tettelin *et al.*, 2005). Nature of pan-genome determines the number of genomes to be sequenced to obtain complete gene repertoire of a species/genus. This also give the approximate number of genes that are added to pan-genome upon addition of a newly sequenced genome (Hervé Tettelin *et al.*, 2005).

3.1.4 Comparative genomics

Comparative analysis of genomic features of multiple organisms will provide insights into genome conferred characteristics responsible for the fitness of an organism. With the availability of more number of whole genomes of a species, comparative genomics approach has helped to address several unanswered questions pertaining to the species adaptation or pathogenesis. As the *Leptospira* genus contain pathogenic, intermediate pathogenic and saprophytic species, it is quite predictable that the genetic differences at the level of species may give insights into leptospiral virulence and pathogenicity (Picardeau *et al.*, 2008; Ricaldi *et al.*, 2012). A close comparison of core genomes of pathogenic species and saprophytic species may highlight the possible genome encoded properties and help in understanding the prospective of leptospiral evolution, environmental persistence and causation of the disease (Lehmann *et al.*, 2014).

In this study an attempt has been made to get insights of genome encoded properties that may probably confer pathogenicity to *Leptospira* by effectively utilizing the whole genome sequences available at the NCBI database. To pursue the objective following **sub-objectives** were framed

A. Whole-Genome sequencing, assembly of the genomes of two *Leptospira* isolates

B. Pan-genome analysis of pathogenic *Leptospira* species

C. Identification of pathogen specific genes in *Leptospira* genus

3.2 Materials and Methods

3.2.1 Genomic DNA for sequencing

Genomic DNA of two *Leptospira* isolates was sourced from Dr. Rudy Hartskeerl, KIT Biomedical Research, The Royal Tropical Institute, Amsterdam, The Netherlands as a part of collaborative study. Both isolates were collected as a part of a routine diagnostic and surveillance. Single colony was isolated, cultured to mid logarithmic phase and genomic DNA was isolated. These isolates were initially characterized using MAT and MLST to determine the strains. Details of two strains are as follows

1. *L. interrogans* Serovar Lai type Langkawi strain SR61 (NASR61)
2. *L. borgpetersenii* Serogroup Ballum strain Muis 5 (NAMuis5)

Strain SR61 is an infectious isolate, isolated from a patient suffering from icteric fever in Sri Lanka. Strain Muis 5 is a carrier strain, isolated from a mouse in the Netherlands during routine epidemiological investigations.

Quality of the DNA integrity was analysed using 0.8% Agarose gel electrophoresis and quantity was estimated by Nano-drop spectrophotometer. Approximately 10 μg of high quality genomic DNA with very less smearing visible on gel was sent for whole genome sequencing.

3.2.2 Whole-Genome sequencing

Paired-end sequencing of genomes was performed using Illumina Genome Analyzer IIx system with a read length of 73 bp. The sequencing is based on the principle of sequencing by synthesis (Quail *et al.*, 2008). The major steps involved in the sequencing of complete genomic DNA are shown in **Figure 23**.

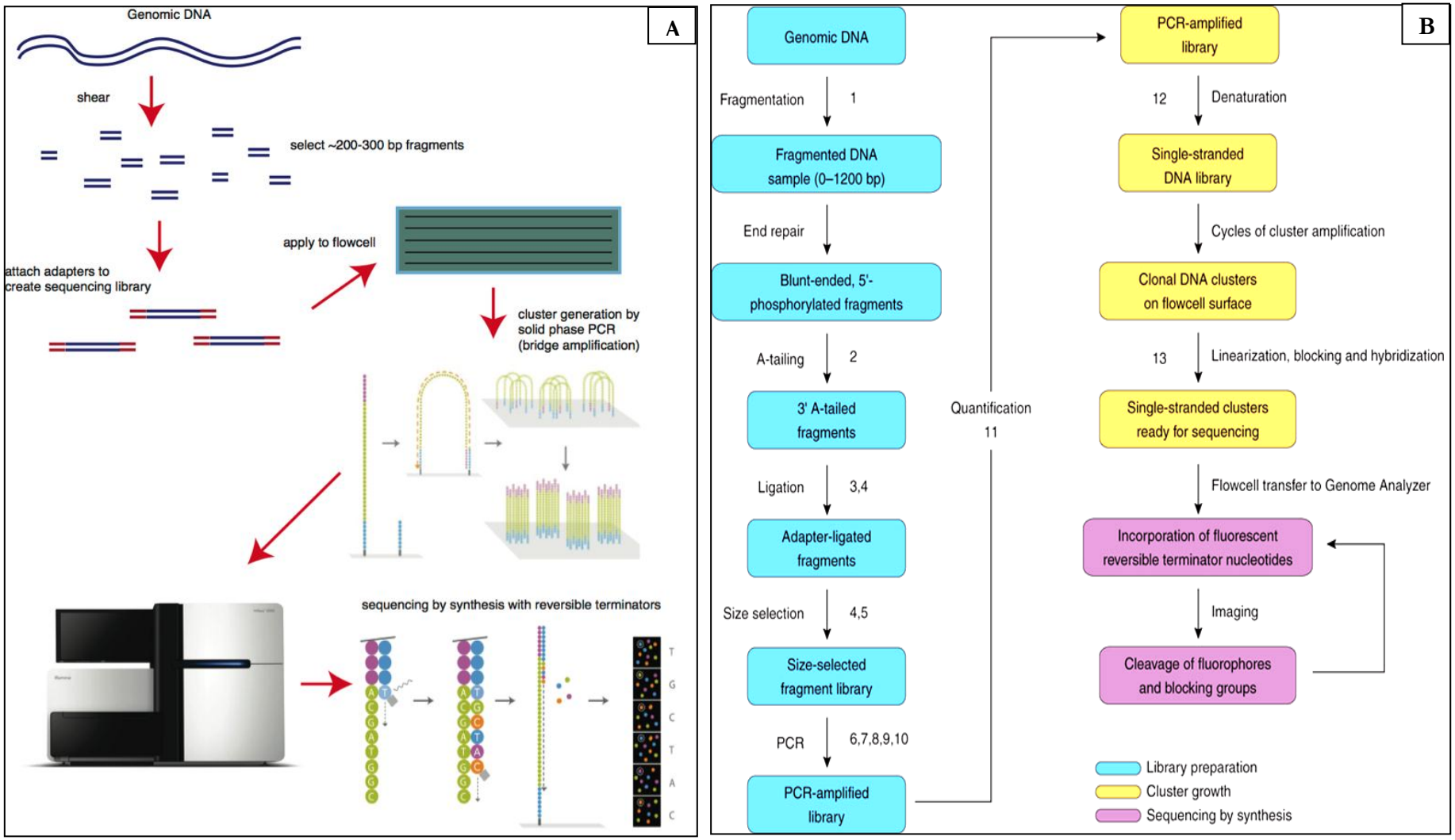

Figure 23: (A). Schematic representation of whole genome sequencing method and (B). pipeline of steps involved in sequencing of whole genome using Illumina Platform (Figure adopted from (Stuart M. Brown, no date; Quail *et al.*, 2008))

3.2.3 Quality assessment of sequencing data

Sequencing data was generated in .bcl format and was converted into a readable .fastq format. Sequencing of a genome in paired end fashion generates two read files, R1 and R2. A typical FASTq file represents all the information captured from the sequencer in four lines for each read as depicted in **Figure 24.**

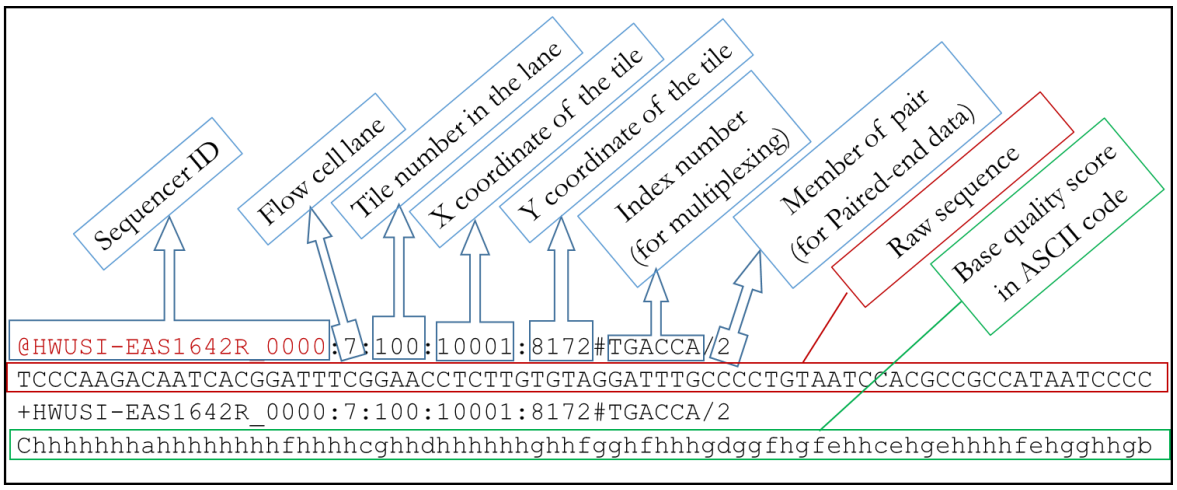

Figure 24: FASTq format of the read sequence output from sequencer

FASTX tool kit, available at http://hannonlab.cshl.edu/fastx_toolkit/, was used to determine quality of sequencing for each .fastq file by calculating total number of reads, mean read length, total number of bases, reads with ATGC characters, average base quality score at each base positon of read and nucleotide composition.

Reads were trimmed from ends to remove nucleotides which have Phred quality score of less than 20 and filtered to remove reads which have more than 30% of bases with less than 20 Phred score to obtain the read data with high mean base quality using in-house developed scripts.

3.2.4 N base Filtration

In addition to ATGC bases, there will be N bases in the read sequences because of ambiguity in base calling or because of very low quality of the base call. Reads were trimmed from ends to remove nucleotides with N bases using "awk" language based command as follows.

awk'BEGIN{i=0}{i++;if(i=2){rem=0;rem=gsub(/N*$/,"",$0);if(rem>0){len=length($0)
;trimmed=1}else{trimmed=0};print$0}else{if(i==4){if(trimmed=1){sseq=substr($o,1,
len);print sseq}else{print$0};i=0}else {print $0}}}'filtered.fastq>terminalNfiltered.fastq

After trimming terminal N bases, reads containing more than 5% of N bases in their sequences are removed using following "awk" language based commands

awk'{if($0~/@HWUSI/){if(i>0){printf"\n"$0}else{printf$0}}else{printf"\t"$0};i++}E
ND {printf"\n"}'filtered.fastq>filtered_tab .txt

awk-F"\t"'BEGIN{i=0}{i++;Ncount=gsub(/N/,"N",$2);len=length($2);percN=
Ncount/len*100;if(percN<=5){print$1"\n"$2"\n"$3"\n"$4}}' filtered_tab.txt> file_final
_filtered.fastq

3.2.5 Genome assembly and annotation

Sequence read data after filtering out low quality reads were used for assembly of genome using multiple open source freely accessible programs and in-house developed scripts. Matched reads of R1 and R2 files of a genome were isolated and shuffled to generate a single file for each genome containing sequences of both read files using in-house developed Perl script. Shuffled paired end data was used to assemble genome by *de novo*.

The reads were assembled *de novo* into contigs using *De Bruijn* graph based tool Velvet version 1.1.05 (Zerbino and Birney, 2008). Velvet has two programs Velveth and Velvetg and runs assembly in two stages. Velveth performs hashing of sequences using specified hash length, known as *k*-mer, for making base pairs and outputs two files, sequences and roadmaps, which were then utilized by Velvetg to assemble reads into contigs. Multiple runs were performed with different k-mer values for determining the appropriate hash length. Hash length that generated least number of contigs with a maximum N50 value, utilizing the maximum number of reads and generates approximate genome length was selected for contigs

generation. N50 represents the size of the smallest contig which when added to a set of larger contigs yields 50% of genome size. Contigs generated were viewed against paired end reads in NGS assembly visualization tool, Tablet, for manual screening of possible errors (Milne *et al.*, 2010). Contigs containing less than 200 bp length, if any, were filtered out.

Gene prediction and annotation of contigs was carried out using a fully automated NCBI prokaryotic genome annotation pipeline (Tatusova *et al.*, 2016).

3.2.6 Submission of genomes to NCBI

Assembled whole genomes were submitted along with their metadata to the NCBI Bioprojects as per their instructions and respective bioproject IDs were obtained. The .fsa file containing all contigs having more than 200bp length was submitted to NCBI for obtaining GenBank accession number.

3.2.7 Whole-genome sequences

A total of 172 whole-genome sequences of *Leptospira* species (**Table 11**) along with their annotations were downloaded from the NCBI database and their species wise, host wise and geography wise distribution is detailed in **Table 12**.

3.2.8 Genomic Islands and mobile genetic elements detection

IsalandViewer available at http://www.pathogenomics.sfu.ca/islandviewer/ was used to predict probable genomic inserts and insertion sequences in both chromosomes of *Leptospira interrogans* serovar Lai 56601 by integrating sequence composition based prediction methods and comparative prediction based methods (Langille and Brinkman, 2009). Artemis genome browser was used to visualize and highlight the regions of probable genomic inserts (Carver *et al.*, 2012)

Table 11: Details of whole genome sequence of *Leptospira* isolates

S.No	Species	Strain	Source	Country	Size (Mb)	GC (%)	Gene	Protein	WGS
1	*L. interrogans*	200703203	Human	French Polynesia	5.01	35.4	5772	5725	AHNY02
2	*L. interrogans*	LP101	Human	Thailand	5.03	35.1	4831	4782	AHNF02
3	*L. interrogans*	2006006959	Unknown	Unknown	4.70	35	3932	3832	AHPY01
4	*L. interrogans*	2006006976	Unknown	Unknown	4.70	35	3927	3825	AHPX01
5	*L. interrogans*	HAI135	Human	Peru	4.45	35.8	6084	6039	AHOI02
6	*L. interrogans*	Kariadisatu	Unknown	Unknown	4.68	35	3891	3801	AHQF01
7	*L. interrogans*	L1111	Human	Thailand	4.85	35.1	4630	4585	AHND02
8	*L. interrogans*	Swart	Unknown	Unknown	4.72	35	3942	3846	AHQE01
9	*L. interrogans*	UI 08561	Human	Laos	4.84	35.1	4892	4843	AHNM02
10	*L. interrogans*	UT234	Unknown	Unknown	4.83	35.1	4084	3983	AHQV01
11	*L. interrogans*	Brem 137	Unknown	Unknown	4.62	35	3792	3675	AHQH01
12	*L. interrogans*	Mallika	Human	India	4.66	35	4314	4272	AFLS02
13	*L. interrogans*	Fiocruz LV133	Human	Brazil	4.71	35	4386	4343	AKWU02
14	*L. interrogans*	HAI0024	Human	Peru	4.71	35	3914	3822	AFLQ01
15	*L. interrogans*	LT1962	Human	Taiwan	4.70	35	4671	4629	AFMC02
16	*L. interrogans*	2001025091	Human	Hawaii	4.52	35	3716	3645	AFMH01
17	*L. interrogans*	2006007831	Human	French Guiana	4.54	35	3737	3668	AFMG01
18	*L. interrogans*	Fiocruz L1-130	Human	Brazil	4.62	35	3762	3667	NC_005823.1 NC_005824.1
19	*L. interrogans*	Fiocruz LV192	Human	Brazil	4.54	35	3741	3672	AFJP01
20	*L. interrogans*	Fiocruz LV199	Human	Brazil	4.55	35	3735	3665	AFJQ01
21	*L. interrogans*	Fiocruz LV204	Human	Brazil	4.54	35	3771	3694	AFJR01
22	*L. interrogans*	Fiocruz LV212	Human	Brazil	4.54	35	3774	3695	AFJS01
23	*L. interrogans*	Fiocruz LV239	Human	Brazil	4.54	35	3732	3663	AFJU01
24	*L. interrogans*	Fiocruz LV2756C6	Human	Brazil	4.55	35	3717	3647	AFMI01
25	*L. interrogans*	Fiocruz LV2772	Human	Brazil	4.55	35	3710	3640	AFKF01
26	*L. interrogans*	Fiocruz LV2787	Human	Brazil	4.55	35	3728	3657	AFKH01
27	*L. interrogans*	Fiocruz LV2799	Human	Brazil	4.55	35	3773	3700	AFKK01
28	*L. interrogans*	Fiocruz LV2804	Human	Brazil	4.55	35	3776	3702	AFKL01
29	*L. interrogans*	Fiocruz LV2805	Human	Brazil	4.55	35	3749	3676	AFKM01
30	*L. interrogans*	Fiocruz LV2806	Human	Brazil	4.54	35	3779	3699	AFKN01

Table 11: Details of whole genome sequence of *Leptospira* isolates (continued)

S.No	Species	Strain	Source	Country	Size (Mb)	GC (%)	Gene	Protein	WGS
31	*L. interrogans*	Fiocruz LV2812	Human	Brazil	4.55	35	3712	3643	AFKP01
32	*L. interrogans*	Fiocruz LV2825	Human	Brazil	4.55	35	3746	3676	AFKQ01
33	*L. interrogans*	Fiocruz LV2832	Human	Brazil	4.54	35	3761	3687	AFKR01
34	*L. interrogans*	Fiocruz LV2897	Human	Brazil	4.54	35	3740	3667	AFKT01
35	*L. interrogans*	Fiocruz LV2948	Human	Brazil	4.54	35	3757	3684	AFKW01
36	*L. interrogans*	Fiocruz LV2953	Human	Brazil	4.54	35	3777	3704	AFKX01
37	*L. interrogans*	Fiocruz LV2958	Human	Brazil	4.53	35	3716	3649	AFKY01
38	*L. interrogans*	Fiocruz LV2959	Human	Brazil	4.55	35	3717	3647	AFKZ01
39	*L. interrogans*	Fiocruz LV2973	Human	Brazil	4.54	35	3720	3650	AFLA01
40	*L. interrogans*	Fiocruz LV3373	Human	Brazil	4.54	35	3735	3662	AFLG01
41	*L. interrogans*	Fiocruz LV3737	Human	Brazil	4.54	35	3747	3674	AFLI01
42	*L. interrogans*	Fiocruz LV3738	Human	Brazil	4.54	35	3768	3694	AFLJ01
43	*L. interrogans*	Fiocruz LV3834	Human	Brazil	4.54	35	3746	3672	AFLK01
44	*L. interrogans*	FiocruzR154	Rat	Brazil	4.54	35	3812	3741	AFMJ01
45	*L. interrogans*	HAI0156	Human	Peru	4.62	34.9	3848	3740	AFLP01
46	*L. interrogans*	HAI0188	Human	Peru	4.61	35	4278	4236	AHOG02
47	*L. interrogans*	LT2050	Human	S. America	4.46	35.2	5440	5396	AFMD02
48	*L. interrogans*	P2518	Unknown	Unknown	4.54	35	3753	3681	AHQO01
49	*L. interrogans*	R066	Rat	Colombia	4.54	35	3715	3645	AFLN01
50	*L. interrogans*	R103	Unknown	Unknown	4.55	35	3721	3650	AHQU01
51	*L. interrogans*	LT1649	Human	Thailand	4.67	34.9	4364	4321	AFMB02
52	*L. interrogans*	2006006986	Human	Egypt	4.93	35.1	4652	4610	AKXC02
53	*L. interrogans*	Andaman	Unknown	Andaman	4.93	35.1	4646	4604	AKXG02
54	*L. interrogans*	LT2186	Human	Thailand	4.84	35.1	5671	5626	AFME02
55	*L. interrogans*	UI 08368	Human	Laos	4.87	35	4599	4555	AHNJ02
56	*L. interrogans*	UI 08434	Human	Laos	4.74	34.9	4440	4394	AHNK02
57	*L. interrogans*	UI 12764	Human	Laos	5.07	35.1	4860	4809	AHNS02
58	*L. interrogans*	UI 12769	Human	Laos	4.90	35	4609	4567	AHNT02
59	*L. interrogans*	R499	Human	Sri Lanka	4.70	35.1	4455	4410	AHNI02
60	*L. interrogans*	P2422	Unknown	Unknown	4.53	35	3730	3658	AHQP01
61	*L. interrogans*	P2547	Unknown	Unknown	4.54	35	3729	3658	AHQR01
62	*L. interrogans*	P2554	Unknown	Unknown	4.54	35	3763	3687	AHQQ01

Table 11: Details of whole genome sequence of *Leptospira* isolates (continued)

S.No	Species	Strain	Source	Country	Size	GC%	Gene	Protein	WGS
63	*L. interrogans*	Verdun LP	Human	France	4.60	35	4256	4214	AKWP02
64	*L. interrogans*	56601	Human	China	4.69	35.01	3741	3683	NC_004342.2 NC_004343.2
65	*L. interrogans*	IPAV	Human	China	4.71	35.01	3759	3711	NC_017551.1 NC_017552.1
66	*L. interrogans*	Lai	Unknown	Unknown	4.60	34.9	3786	3672	AHQB01
67	*L. interrogans*	Langkawi	Unknown	Unknown	4.77	35.1	4016	3912	AHQD01
68	*L. interrogans*	SR61	Human	Sri Lanka	4.89	35.1	4228	4118	JPUB01
69	*L. interrogans*	TE 1992	Rat	Tanzania	4.78	35.1	4895	4850	AKWW02
70	*L. interrogans*	L0448	Human	Thailand	4.95	35.2	4736	4690	AHNA02
71	*L. interrogans*	L0887	Human	Thailand	4.76	35.1	4004	3905	AHQW01
72	*L. interrogans*	UT053	Human	Thailand	5.15	35.2	4989	4944	AHNW02
73	*L. interrogans*	Brem 129	Horse	Germany	4.63	35	4288	4245	AHMQ02
74	*L. interrogans*	CSL10083	Sea lion	USA	4.57	35	4276	4232	AOHJ01
75	*L. interrogans*	Fox 32256	Fox	USA	4.58	35	4308	4263	AOHG01
76	*L. interrogans*	Kennewicki LC82-25	Human	USA	4.61	35	4315	4269	AHMK02
77	*L. interrogans*	Pomona	Human	Australia	4.58	35	4300	4255	AFLT02
78	*L. interrogans*	UT364	Human	Thailand	5.11	35.2	4938	4891	AHNX02
79	*L. interrogans*	2006006956	Unknown	Unknown	4.77	34.9	3918	3803	AHPW01
80	*L. interrogans*	2006006960	Human	Egypt	4.87	35	4547	4505	AHME02
81	*L. interrogans*	200701872	Human	Mayotte	4.35	35.5	5321	5279	AKWN02
82	*L. interrogans*	L0374	Human	Thailand	4.85	35.2	5310	5266	AHMZ02
83	*L. interrogans*	R168	Human	Sri Lanka	4.87	35.2	4606	4561	AHNH02
84	*L. interrogans*	Valbuzzi	Human	Australia	4.54	34.9	4278	4236	AKXF02
85	*L. interrogans*	LT2156	Human	Australia	4.89	35.1	5316	5269	AFMF02
86	*L. interrogans*	2002000621	Human	Hawaii	4.92	35.2	4682	4639	AFLU02
87	*L. interrogans*	2002000623	Human	Hawaii	4.93	35.2	4721	4676	AHMG02
88	*L. interrogans*	2002000624	Human	Hawaii	4.92	35.2	4676	4633	AFJK02
89	*L. interrogans*	2002000626	Human	Hawaii	4.85	35.3	5156	5113	AFJL02
90	*L. interrogans*	2006001854	Human	Thailand	4.95	35.1	5698	5652	AFLW02
91	*L. interrogans*	Brem 329	Horse	Germany	4.75	35	4473	4428	AKXA02
92	*L. interrogans*	C10069	Human	Thailand	4.74	35.1	4461	4418	AFLZ02
93	*L. interrogans*	FPW1039	Human	Thailand	5.11	35	5035	4990	AKWR02
94	*L. interrogans*	FPW2026	Human	Thailand	4.88	35.1	4641	4597	AHMX02
95	*L. interrogans*	HAI1536	Human	Peru	4.69	35.3	4336	4293	AKWD02
96	*L. interrogans*	L0996	Human	Thailand	4.84	35.2	4685	4639	AHNB02
97	*L. interrogans*	L1207	Human	Thailand	4.39	35	4692	4650	AHNE02
98	*L. interrogans*	MMD3731	Rat	Peru	4.59	35	4215	4173	AHOL02
99	*L. interrogans*	UI 08452	Human	Laos	4.86	35.2	4659	4614	AHNL02

Table 11: Details of whole genome sequence of *Leptospira* isolates (continued)

S.No	Species	Strain	Source	Country	Size	GC%	Gene	Protein	WGS
100	*L. interrogans*	UI 09600	Human	Laos	4.72	35	4383	4341	AHNO02
101	*L. interrogans*	UI 12758	Human	Laos	5.10	35.1	4856	4812	AHNR02
102	*L. interrogans*	UI 13372	Human	Laos	5.10	35.1	4965	4921	AHNV02
103	*L. interrogans*	FiocruzR83		Brazil	4.55	35	3979	3727	AFLM01
104	*L. interrogans*	LaiLPS mutant	-	-	4.60	34.9	4074	3801	AHQC01
1	*L.borgpetersenii*	200801910	Human	USA	3.97	40		4083	AHOB02
2	*L.borgpetersenii*	JB197	Human	USA	3.88	40.22	3242	2880	NC_008510.1 NC_008511.1
3	*L.borgpetersenii*	L550	Human	Australia	3.93	40.2	3273	2945	NC_008508.1 NC_008509.1
4	*L.borgpetersenii*	Lely 607	Cattle	Netherlands	3.77	40.2	3433	3186	AOWM01
5	*L.borgpetersenii*	L0066	Human	Thailand	3.88	40.1	3455	3352	AOUW01
6	*L.borgpetersenii*	MK146	Human	Thailand	3.89	40.1	4073	4030	AHNG02
7	*L.borgpetersenii*	UI 09931	Human	Laos	3.89	40.1	4053	4010	AHNP02
8	*L.borgpetersenii*	200901116	Human	Mayotte	4.13	39.5	4241	4199	AKWB02
9	*L.borgpetersenii*	200901868	Human	Mayotte	4.27	40.2	5157	5111	AKWF02
10	*L.borgpetersenii*	200701203	Human	Mayotte	3.97	40.2	4822	4776	AKWO02
11	*L.borgpetersenii*	200801926	Human	Mayotte	3.97	40.2	4193	4149	AKWJ02
12	*L.borgpetersenii*	200901122	Human	Mayotte	4.16	39.4	4292	4250	AKWM02
13	*L.borgpetersenii*	Brem 307	Horse	Germany	3.83	40.1	4192	4147	AHMR02
14	*L.borgpetersenii*	Brem 328	Horse	Germany	3.83	40.2	4178	4133	AHMS02
15	*L.borgpetersenii*	Noumea 25	Human	New Caledonia	3.96	40	4528	4484	AHOD02
16	*L.borgpetersenii*	UI 09149	Human	Laos	3.89	40.1	4033	3991	AHNN02
17	*L.borgpetersenii*	Muis5	Mouse	Netherlands	3.89	40	3478	3341	JPUC01
1	*L. kirschneri*	1051	Dog	Barbados	4.40	35.9	3992	3948	AHML02
2	*L. kirschneri*	3522 CT	Bat	Indonesia	4.41	35.9	4029	3986	AHMN02
3	*L. kirschneri*	Moskva	Human	Russia	4.34	35.9	4011	3968	AHMV02
4	*L. kirschneri*	RM52	Pig	USA (Iowa)	4.36	35.9	3967	3924	AHMJ02
5	*L. kirschneri*	Vehlefans 2	Cattle	Netherlands	4.41	35.9	3678	3576	AOWL01
6	*L. kirschneri*	RM1	Cattle	Tanzania	4.43	35.9	4048	4004	AHMW02
7	*L. kirschneri*	200702274	Human	France	4.32	35.9	3977	3933	AHOC02
8	*L. kirschneri*	200801774	Human	Mayotte	4.64	36.1	4239	4195	AKWL02
9	*L. kirschneri*	200801925	Human	Mayotte	4.51	36.2	4980	4937	AKWK02
10	*L. kirschneri*	200802841	Human	Mayotte	4.69	36.2	4355	4310	AKWH02
11	*L. kirschneri*	200803703	Human	Mayotte	4.70	36	4385	4343	AKWG02
12	*L. kirschneri*	2008720114	Rat	Croatia	4.39	35.9	4022	3976	AKXD02
13	*L. kirschneri*	H1	Human	Thailand	4.59	36.1	4340	4295	AHMY02
14	*L. kirschneri*	H2	Human	Thailand	4.68	36.3	4540	4498	AKWQ02

Table 11: Details of whole genome sequence of *Leptospira* isolates (continued)

S.No	Species	Strain	Source	Country	Size	GC	Gene	Protein	WGS
1	*L. santarosai*	MAVJ 401	Human	Costa Rica	4.24	41.6	4412	4367	AHMU02
2	*L. santarosai*	1342KT	Human	Panama	3.99	41.8	4049	4002	AOHB02
3	*L. santarosai*	LT 821	Spiny rat	Panama	3.88	41.8	4073	4033	ADOR01
4	*L. santarosai*	Oregon	Bovine	USA	3.91	41.9	3657	3546	AOXB01
5	*L. santarosai*	2000027870	Human	USA	3.91	41.7	3948	3903	AFLX02
6	*L. santarosai*	2000030832	Human	USA	3.99	41.7	4045	4003	AFJN02
7	*L. santarosai*	200403458	Human	USA	4.00	41.9	4005	3967	AKWI02
8	*L. santarosai*	200702252	Human	USA	4.03	41.9	4005	3963	AHOA02
9	*L. santarosai*	AIM	Human	Colombia	4.07	41.7	4158	4114	AKWT02
10	*L. santarosai*	CBC1416	cattle	Peru	3.97	41.9	4263	4221	AKWE02
11	*L. santarosai*	CBC1531	Cattle	Peru	3.83	41.8	3857	3815	APGN01
12	*L. santarosai*	CBC379	Pig	Peru	4.17	41.7	4315	4273	AHOE02
13	*L. santarosai*	CBC523	cattle	Peru	4.04	41.9	4079	4035	AHOF02
14	*L. santarosai*	HAI134	Human	Peru	4.06	41.6	4235	4193	AHOH02
15	*L. santarosai*	HAI1380	Human	Peru	3.99	41.8	4011	3967	AHOJ02
16	*L. santarosai*	HAI821	Human	Peru	3.99	41.7	4022	3977	AHOK02
17	*L. santarosai*	JET	Human	Colombia	4.13	41.6	4184	4138	AKWS02
18	*L. santarosai*	MOR084	Human	Peru	4.14	41.6	4225	4180	AHON02
19	*L. santarosai*	ST188	Dog	Trinidad	4.06	41.8	4196	4153	AOHA02
20	*L. santarosai*	ZUN179	Human	Peru	4.07	41.7	4191	4146	AHOQ02
21	*L. santarosai*	HAI1594	Human	Peru	4.61	35	4111	3772	AKWC02
1	*L. noguchii*	ZUN142	Human	Peru	4.84	35.5	4634	4590	AHOP02
2	*L. noguchii*	1993005606	Human	USA	4.97	35.8	5078	5026	AHMF02
3	*L. noguchii*	2001034031	Human	Hawaii	4.94	35.7	4823	4776	AKXB02
4	*L. noguchii*	2006001870	Human	USA	4.81	35.5	4414	4368	AFLY02
5	*L. noguchii*	2007001578	Human	Hawaii	4.96	35.7	4833	4780	AHMH02
6	*L. noguchii*	Bonito	Human	Brazil	4.50	35.7	4133	4091	AOHH01
7	*L. noguchii*	Cascata	Human	Brazil	4.55	35.7	4245	4193	AOUB01
8	*L. noguchii*	Hook	Dog	Brazil	4.54	35.6	4136	4092	AOUC01
1	*L. weilii*	LT2116	Human	Australia	4.32	40.5	5023	4981	AHOR02
2	*L. weilii*	2006001853	Human	Thailand	4.37	40.8	4696	4655	AFLV02
3	*L. weilii*	2006001855	Human	Thailand	4.28	40.7	4711	4671	AFJM02
4	*L. weilii*	LNT 1234	Human	Laos	4.26	40.8	4478	4436	AHNC02
5	*L. weilii*	UI 13098	Human	Laos	4.55	40.7	4796	4755	AHNU02
6	*L. weilii*	UI 14631	Human	Unknown	4.31	40.7	3981	3882	AHQY01
1	*L. licerasiae*	MMD0835	Opossum	Peru	4.20	41.1	3868	3821	AFLO01
2	*L. licerasiae*	MMD4847	Bat	Peru	4.20	41.1	3998	3954	AHOM02

Table 12: Species wise, host wise and geographic distribution of whole genomes

Species	Total genomes	Host	Total genomes	Region	Total genomes
L. interrogans	104	Human	127	N. America	21
L. borgpetersenii	17	Rodent	07	S. America	55
L. kirschneri	14	Dog	04	Caribbean islands	3
L. santarosai	21	Horse	04	Europe	11
L. noguchii	08	Bovine	07	Africa	15
L. weilii	06	Pig	02	Asia	7
L. licerasiae	02	Bat	02	South east Asia	35
		Others	19	Australia	7
				Unknown	18

3.2.9 Pan-genome and core genome analysis

Amino acid sequences of proteins of each isolate were used for species wise pan-genome analysis. Proteins with less than 50 amino acid in length were omitted from the analysis. All vs all similarity scores were generated using the BLASTp program for all proteins. A similarity threshold of 70% score for bi directional best hits was used to identify orthologs. Orthologous genes across isolates were grouped into orthologous clusters using the Markov Cluster Algorithm (MCL) of OrthoMCL program (Li, Stoeckert and Roos, 2003).

As most of the genomes used in the study are draft genomes with possible gaps, orthologous clusters having genes from at least 90% of genomes of a species were considered as a part of core genome. Clusters having orthologs from less than 90% of genomes of a species were considered as a part of accessory genome. Core genome and accessory genome of a species were used to build the pan-genome of a species (Medini *et al.*, 2005; Liang *et al.*, 2012).

Pan-genome nature is explained by using mathematical modelling and it represents the diversity of species (Tettelin *et al.*, 2008; Liang *et al.*, 2012).

The nature of pan-genome for each species was determined by curve fitting using Heap's law and core genome by least square fit of exponential regression decay as described previously (Tettelin *et al.*, 2008).

Heap's law $\mathbf{n=k\,N^{-\alpha}}$

Where "n" denotes the pan genome size, "N" denotes the number of genomes, "*k*" and "γ" are curve constants and exponential "α = 1 - γ". Pan genome is considered as closed if $\alpha \geq 1$ and considered as open if $\alpha<1$. Orthologous clusters were used to analyse the pan genome profiles of species using Distance guide algorithm (DG) in Pan Genome Profile Analyzer tool (PanGP) (Zhao *et al.*, 2014).

3.2.10 Pathogen specific genes

Core genomes of individual pathogenic species were used to generate common core genome of pathogenic *Leptospira* isolates. Orthologous clusters were generated using core genomes of individual pathogenic species using BLASTp and OrthoMCL programs as described above, and clusters having orthologs from at least 90% of core genomes of species were considered as a part of common core genome of pathogenic *Leptospira* isolates.

Common core genome of pathogenic *Leptospira* isolates and genomes of saprophytic *L. biflexa* serovar Patoc strain Patoc 1 (Ames) and strain Patoc 1 (Paris) were used to generate orthologous clusters. Orthologous clusters having genes from all three gene sets are considered as *Leptospira* genus core genome. Genes specific to common core genome of pathogenic *Leptospira* were isolated as pathogen specific genes.

3.2.11 COG functional classification

Conserved Domain (CD) Database search tool (Web CD Search) available with NCBI at https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi was used to elucidate the functional classification of proteins (Marchler-Bauer *et al.*, 2014). The best hit for each protein

along with its COG (Cluster of orthologous group) class was extracted. Proteins which yielded two or more classes are classified as multiple class and which did not yield any best hit with conserved domains are considered as hypothetical proteins.

3.2.12 Lipoproteins and secretary proteins detection

Online available prediction servers were used to analyse hypothetical proteins for their predicting their localization based on the presence of signal peptide. LipoP 1.0 server was used to predict probable lipoproteins by detecting lipoprotein signal peptide (Rahman *et al.*, 2008). PrediSi (Prediction of Signal peptide) available at http://www.predisi.de/ and SignalP 4.0 server were used to predict the presence of signal peptidase cleavage site required for classical secretion of proteins (Hiller *et al.*, 2004; Emanuelsson *et al.*, 2007; Petersen *et al.*, 2011). TatP 1.0 server was used to predict the presence of Twin arginine signal peptide required for secretory dependent translocation of proteins to periplasmic space and extracellular environment (Jannick Dyrløv Bendtsen *et al.*, 2005). SecretomeP 2.0 server was used to predict the proteins involved in non-classical pathway of secretion, independent of signal peptide (Jannick D Bendtsen *et al.*, 2005). All prediction servers were accessed from http://www.cbs.dtu.dk/services/.

3.3 RESULTS

3.3.1 Quality analysis of sequencing

Paired-end sequencing data generated at a read length of 73 bp for each genome was analysed for its quality using FASTX and generated statistics by it were detailed in **Table 13**. It was observed that high quality (HQ) reads in all the read files were more than 95% with negligible primer-adapter contamination. The percentage of unambiguous base call was observed to be less than 1.7%. Base composition analysis of reads confirmed that GC content was in accordance with the already sequenced *Leptospira* isolates. Average Phred quality score of each base of the read for all four read files were shown in **Figure 25** and it confirms that the sequencing till end of the read was performed with good quality of approximately above 30 Phred score.

Table 13: Quality statistics of sequencing for both *Leptospira* isolates

Description	NASR61		NAMuis5	
	Read 1	**Read 2**	**Read 1**	**Read 2**
Maximum Read Length	**73**	**73**	**73**	**73**
Minimum Read Length	**73**	**73**	**73**	**73**
Median Read Length	**73**	**73**	**73**	**73**
Total no. of reads ($x10^6$)	**2.23**	**2.23**	**2.94**	**2.94**
Total no. of HQ reads * ($x10^6$)	**2.21**	**2.20**	**2.91**	**2.90**
Percentage of HQ reads	**99.05**	**98.62**	**98.82**	**98.40**
Total no. of Bases ($x10^6$)	**163.13**	**163.13**	**214.95**	**214.95**
Total no. of HQ bases # ($x10^6$)	**161.55**	**160.93**	**212.42**	**211.44**
% of HQ bases	**99.03**	**98.65**	**98.82**	**98.36**
Total no. of Non-ATGC bases ($x10^6$)	**1.58**	**2.20**	**2.52**	**3.51**
% of Non-ATGC bases	**0.97**	**1.35**	**1.18**	**1.63**
No. of reads with Non-ATGC bases	**9259**	**6019**	**12228**	**7778**
% of reads with Non-ATGC bases	**0.41**	**0.27**	**0.41**	**0.26**
% of A	**33.934**	**29.515**	**31.909**	**27.698**
% of T	**29.322**	**33.695**	**27.405**	**31.565**
% of G	**18.995**	**17.694**	**20.941**	**19.762**
% of C	**17.742**	**18.946**	**19.738**	**20.827**
% of Non-ATGC	**0.006**	**0.150**	**0.006**	**0.148**

* >70% of bases in a read with >20 Phred score, # bases with >20 Phred score

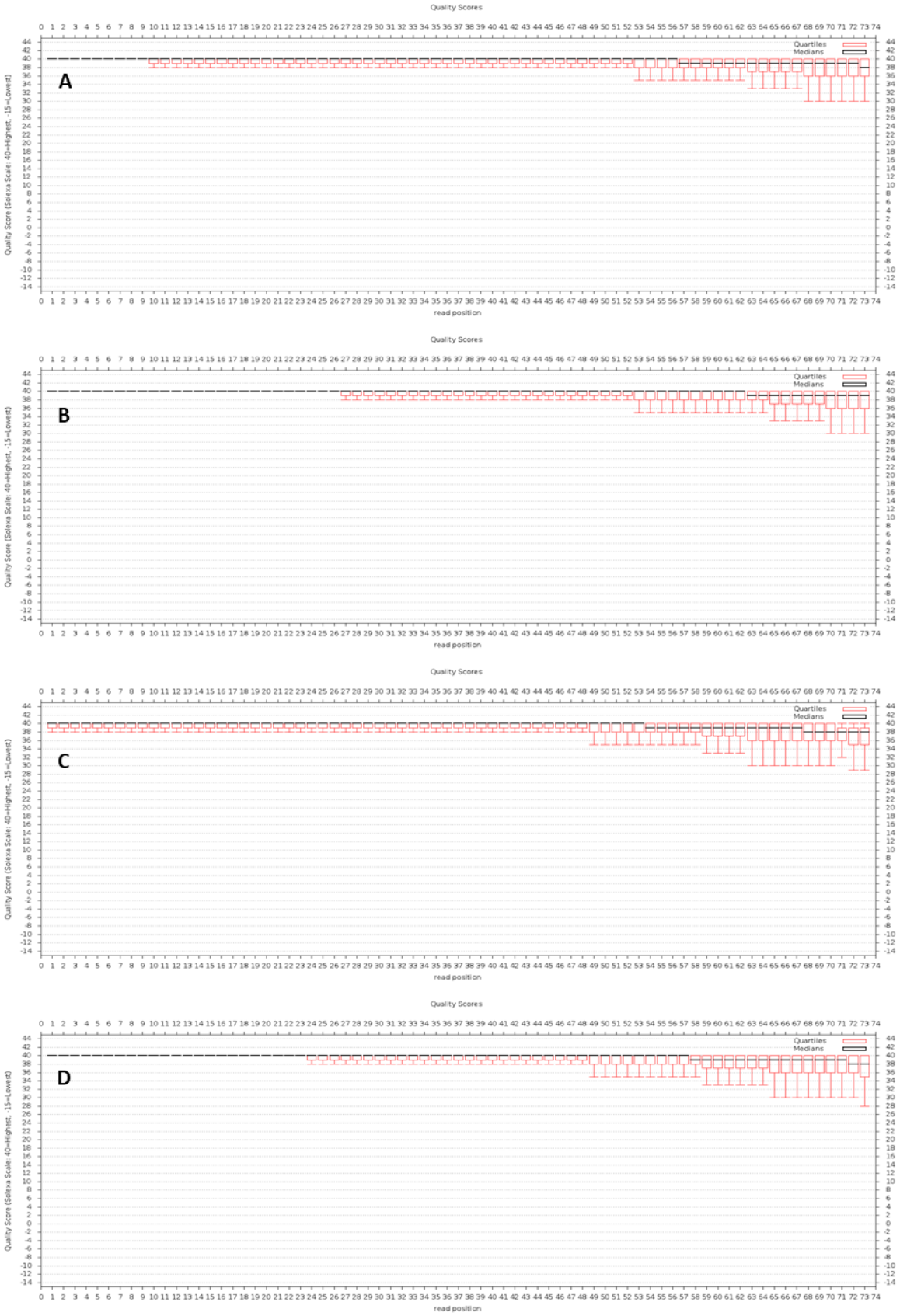

Figure 25: Read quality graphs showing average Phred quality score at each base position for NASR61 strain Read 1 (A) and Read 2 (B), and NAMuis5 strain Read 1 (C) and Read 2 (D).

3.3.2 Genome assembly and annotation

Genome assembly was carried out with different *k*-mer values and output assembly was analyzed for statistics and are detailed in **Table 14**. Optimum *k-mer* value producing assembly by utilizing the maximum number of reads with higher N50 value and lower contigs was selected as hash length. Hash length of 47 was selected for assembling genomes of both NASR61 and NAMuis5 isolates. Contigs generated with optimum hash length were then aligned to the reference genome and sorted using in-house written Perl script.

Table 14: Statistics of genome assembly performed using Velvet with different hash lengths for both the strains. Optimum hash length used for final assembly is in bold.

Hash length	No. of Contigs	N50	Max Contig Size	Genome Size (approx.)	No. of reads utilized	Total matched reads
NASR61						
41	438	31875	100392	4923060	4126383	4371248
43	431	33432	100393	4918887	4127294	4371248
45	424	34540	122562	4923613	4124653	4371248
47	**416**	**35243**	**122564**	**4928893**	**4120936**	**4371248**
49	417	34544	122566	4931380	4114193	4371248
51	429	33349	122568	4927879	4110925	4371248
NAMuis5						
41	221	48572	115042	3895722	5614525	4371248
43	223	45798	115044	3897127	5607017	4371248
45	217	48571	146534	3897546	5595984	4371248
47	**212**	**52633**	**146536**	**3898369**	**5589888**	**4371248**
49	215	48215	181775	3898838	5575404	4371248
51	220	44749	181777	3898314	5559131	4371248

Ordered contigs were then submitted to NCBI prokaryotic genome annotation pipeline for prediction of genome characteristics and annotation of genes.

3.3.3 Genome submission

Bioproject and biosample records were created for both the strains at NCBI database and completely assembled and annotated genomes were submitted to NCBI GenBank's whole-genome shotgun project as per guidelines. Complete genome characteristics of the two strains along with their NCBI accession numbers for submission were detailed in **Table 15**.

Table 15: Genome statistics of whole-genome assembly

Description	NASR61	NAMuis5
Biosample	SAMN02928167	SAMN02928168
Bioproject	PRJNA255705	PRJNA255706
Genbank accession ID	JPUB00000000	JPUC00000000
Genome coverage	34x	53x
Total contigs	415	212
Genome size in bp	4891582	3887643
% of G+C	35.1	40.0
Genes	4228	3478
Proteins	4118	3341
Pseudogenes	69	96
rRNAs	3	3
tRNAs	37	37
ncRNAs	1	1
Frame shifted genes	55	73

3.3.4 Genomic Island prediction

Genomic island prediction using Island viewer identified the regions of insertion elements in *Leptospira interrogans* genome and important proteins in those region are detailed in **Table 16**.

Table 16: Island viewer predicted genomic regions and their major proteins in *Leptospira interrogans* genome.

Islands / operons	Major proteins	Islands / operons	Major proteins
LA_0734 – LA_0766	Ribosome Assembly, adk, secY,	LA_1547 – LA_1569	Methylase, β-Lactamase, Hypo
LA_0898 – LA_0905	Multiple antibiotic resistance	LA_2300 – LA_2302	Hypothetical Proteins
LA_0281 – LA_0302	Hypothetical proteins	LA_3075 – LA_3077	Lig B, Hypo
LA_0898 – LA_0905	Hypothetical Proteins, Methylase	LA_3097 – LA_3115	Transposase, K+ transport
LA_0923 – LA_0925	Transposase	LA_3469 – LA_3491	Hypothetical
LA_1027 – LA_1029	Sphingomyelinase	LA_3544 – LA_3547	Hypothetical
LA_1396 – LA_1398	Hypothetical protein	LA_3777 – LA_3781	Secretory Proteins
LA_1420 – LA1430	Fatty Acid Synthesis	LB_0264 – LB_0272	Membrane Proteins

3.3.5 Pan and core genome analysis

Orthologous clusters were generated using OrthoMCL for each pathogenic species of *Leptospira* genus and core gene clusters were identified using in-house developed scripts. Pan and core genome sizes along with the number of orthologous clusters generated for each species are shown in **Table 17**.

PanGP analysis was performed using species wise orthologous clusters for all pathogenic species and was found to have open pan-genome nature suggesting the genome plasticity of the organism and chances of finding new genes upon addition of a genome to the analysis **(Figure 26)**. It was estimated that approximately 30 to 60 new orthologous gene clusters will be identified upon addition of a single genome to the analysis for any pathogenic species (**Figure27**). New gene cluster estimation was not performed for *L. weilii* species as the number of genomes were less.

Table 17: Pan and core genome analysis of pathogenic *Leptospira* species

Species	No. of strains	Total no. of proteins	Orthologous clusters	Core genome size	α value	Pan genome size
L. interrogans	104	446433	6484	3386	0.71	8388
L. borgpetersenii	17	67046	4553	2750	0.84	5391
L. kirschneri	14	57866	4032	3468	0.75	4759
L. santarosai	21	85351	4729	3238	0.63	6525
L. noguchii	8	35916	4083	3795	0.86	4858
L. weilii	6	27380	3967	3913	0.87	4830
L. licerasiae	2	7839	3120	3788	ND	ND

ND- Not determined

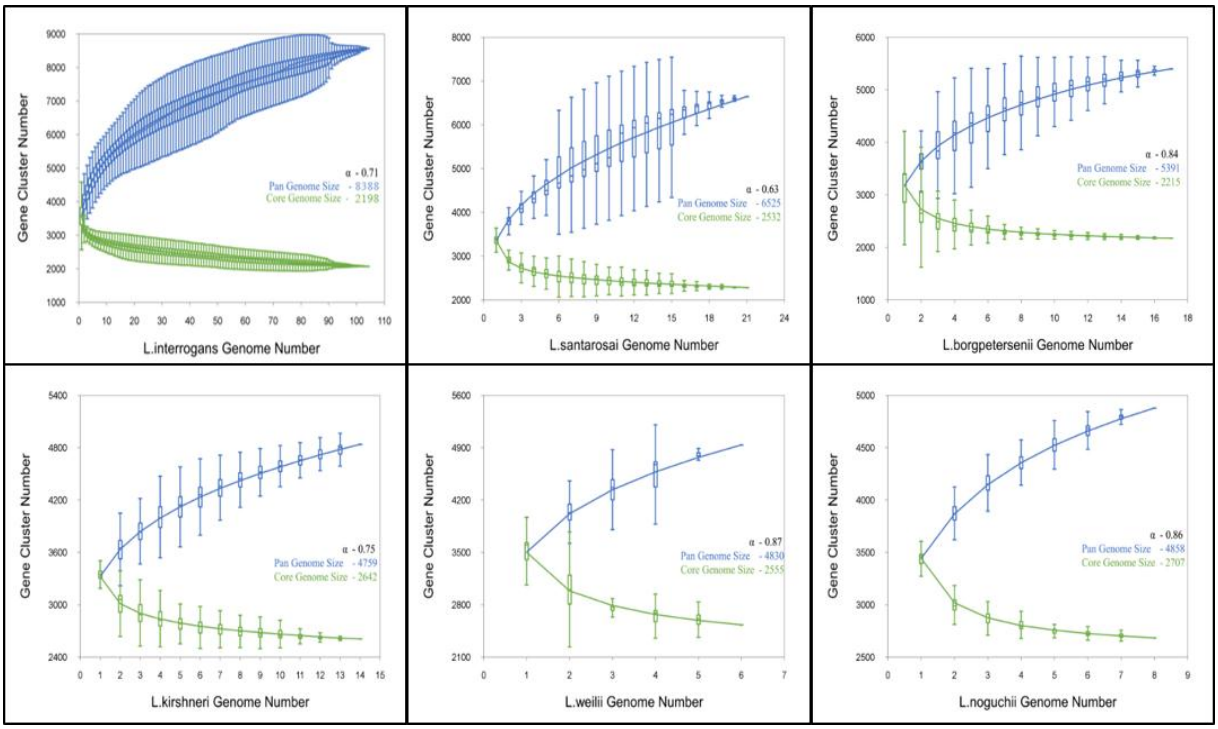

Figure 26: Graphical representation of species wise pan (blue) and core (green) genomes. ("α" value for pan-genome curve and genes constituting pan and core genomes of each curve are also depicted)

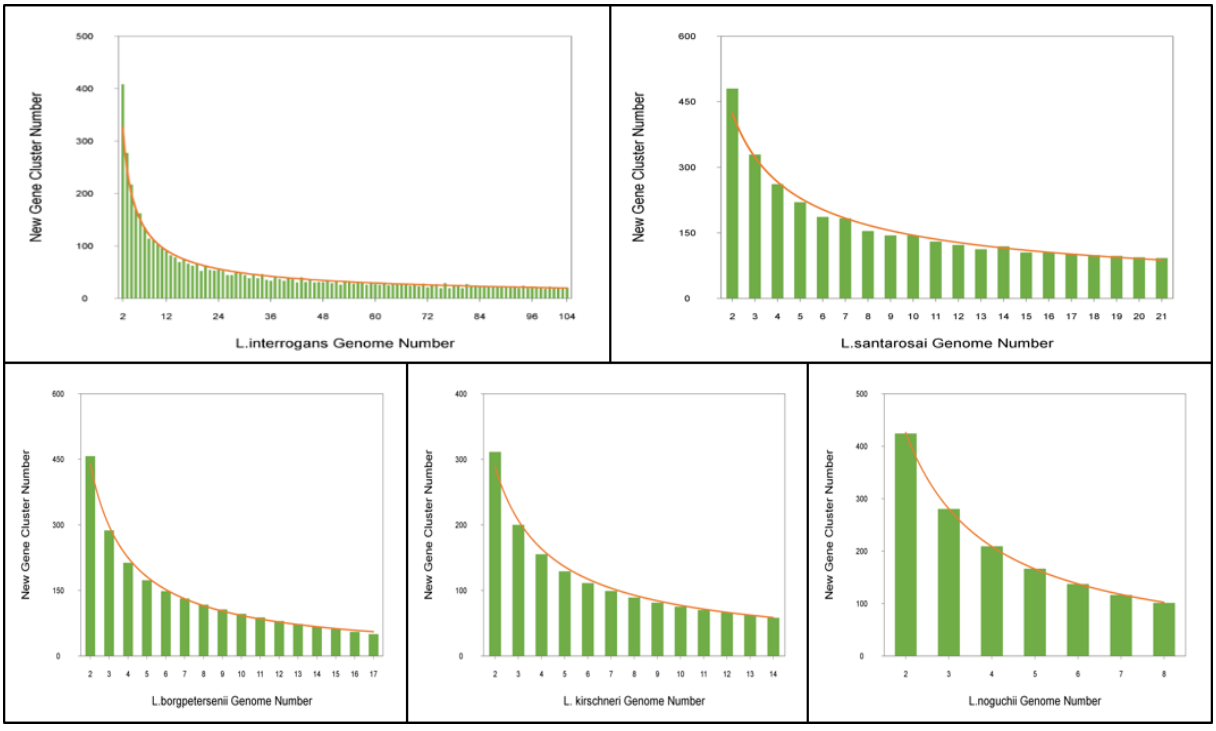

Figure 27: Graphical representation of number of new gene clusters being identified upon addition of a new genome to the pan-genome analysis.

3.3.6 Common core genome of *Leptospira* pathogenic species

When individual core genomes of all seven pathogenic species were analyzed for the presence of orthologs, a total of 2941 orthologous gene clusters were identified across species and out of them 2708 orthologous gene clusters were found to have orthologs in at least six species analyzed. Those 2708 orthologous clusters were isolated as common core genome for all pathogenic species of genus *Leptospira* (**Figure 28**).

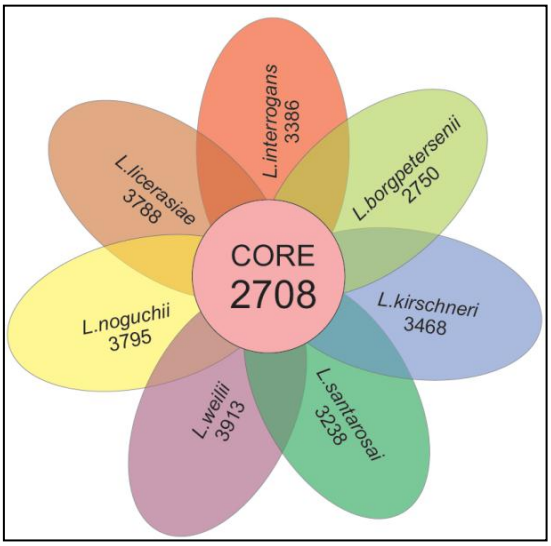

Figure 28: Venn diagram showing the core genome size of individual species and common core genome of pathogenic *Leptospira* species

3.3.7 Pathogen specific genes

Gene content specifically present in pathogenic *Leptospira* isolates was isolated. Orthologous clusters for common core genome of pathogenic *Leptospira* species and two genomes of *L. biflexa* isolates were generated. *Leptospira* genus core genome was found to contain 2281 orthologs gene clusters and were believed to consist genes responsible for basic metabolic and housekeeping functions of bacteria. In addition, 1210 genes of *L. biflexa* strains were found to have no orthologs in the common core genome and was identified as saprophytic organism specific genome. And 427 genes of common core genome of pathogenic species were found to be specific only to pathogenic species and were not having any orthologs in the saprophytic organisms of the genus, forming pathogen-specific gene content of *Leptospira*.

3.3.8 Functional classification of proteins

Functional classification of protein sequences belonging to individual core genomes of pathogenic species based on COG database identified a large number of conserved hypothetical proteins with unknown function, ranging from 28% in *L. borgpetersenii* sps to 38% in *L. interrogans* sps. COG classification of their common core genome has also shown 31% of the conserved proteins across pathogenic species are hypothetical in nature with unknown function (**Figure 29**). Further it was found that genes conserved across pathogenic species belongs to cellular process and signaling (25%), metabolism (22%), information storage and processing (15%), general function (6%), multiple class (4%) and few genes belonging to lipoproteins and mobile elements (**Figure 30 A**)

Functional characterization of pathogen specific proteins has identified a total of 285 (67 %) proteins as hypothetical in nature with unknown function, 37 proteins for cellular process and signaling, 33 proteins for metabolism, 27 proteins for information storage and processing, 25 proteins for general function prediction, 8 proteins as mobile elements, 5 proteins belonging to multiple classes and 7 proteins as lipoproteins (**Figure 30 B**). Details of 427 pathogen specific proteins with their corresponding counterpart from to *L. interrogans* Serovar Lai strain 56601 were tabulated in **Table 18**.

3.3.9 Analysis of hypothetical proteins

Pathogen specific proteins with unknown function, when analyzed for predictable signal peptides, 24 proteins were found to have predictable secretory signal peptide, 10 to have lipoprotein signal peptide and 12 to have Tat signal peptides (**Table 18**). In addition, a total of 153 hypothetical proteins were predicted to be involved in secretory pathway through non-classical pathway.

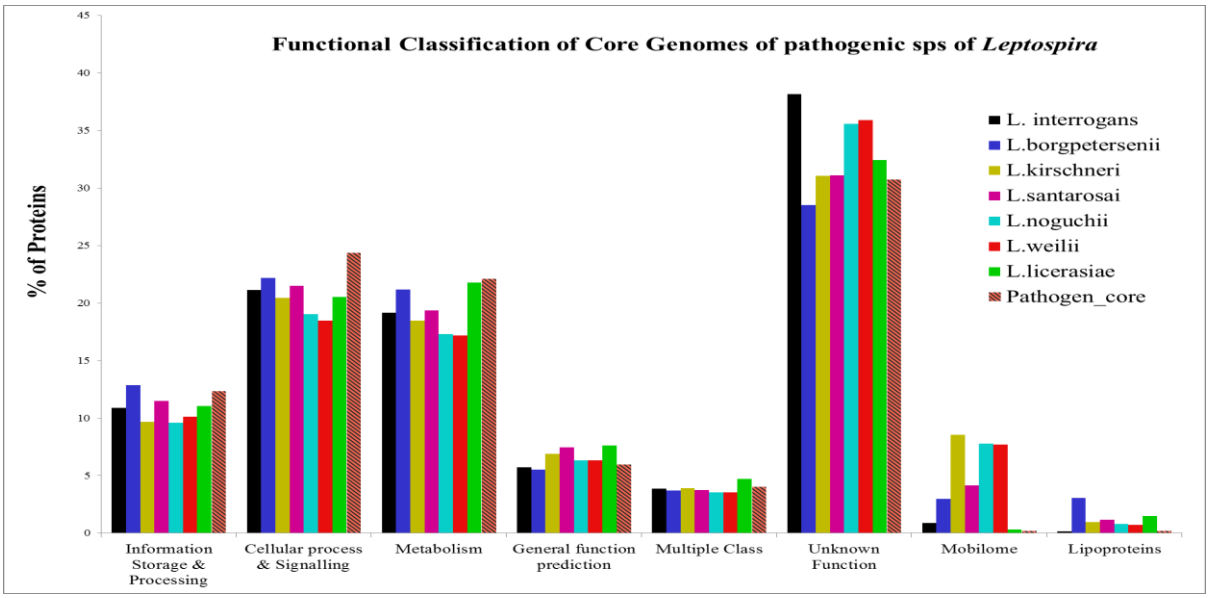

Figure 29: COG Functional classification of core genomes of *Leptospira* species and common core genome of pathogenic species.

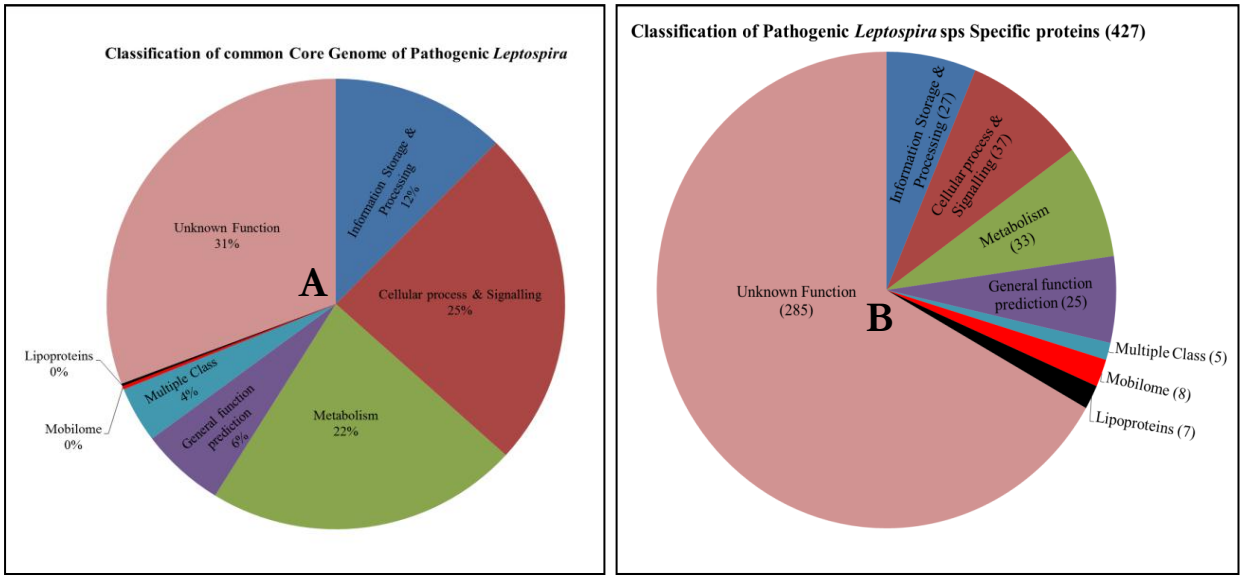

Figure 30: Functional classification of common core genome of pathogenic *Leptospira* species (A) and pathogen specific gene content of *Leptospira* (B)

Table 18: Details of pathogen specific gene content for *Leptospira*

S.No	Locus Tag	Annotation	Length
1.	LA_0016	hypothetical protein LA_0016	71
2.	LA_0018	hypothetical protein LA_0018	298
3.	LA_0019	hypothetical protein LA_0019	167
4.	LA_0031	hypothetical protein LA_0031	234
5.	LA_0032	hypothetical protein LA_0032	119
6.	LA_0065	drug: Na+ antiporter of the multi antimicrobial extrusion family	403
7.	LA_0068	hypothetical protein LA_0068	296
8.	LA_0075	hypothetical protein LA_0075	69
9.	LA_0077	hypothetical protein LA_0077	362
10.	LA_0078	hypothetical protein LA_0078	546
11.	LA_0097	hypothetical protein LA_0097	148
12.	LA_0103[s]	hypothetical protein LA_0103	347
13.	LA_0109	2'-5' RNA ligase	182
14.	LA_0121	hypothetical protein LA_0121	70
15.	LA_0128	hypothetical protein LA_0128	129
16.	LA_0129	hypothetical protein LA_0129	607
17.	LA_0142	hypothetical protein LA_0142	70
18.	LA_0149	SET domain-containing protein	132
19.	LA_0175	50S ribosomal protein L34	53
20.	LA_0184	hypothetical protein LA_0184	692
21.	LA_0189	Mu-like prophage host-nuclease inhibitor protein Gam	198
22.	LA_0195	transcriptional regulator	132

S.No	Locus Tag	Annotation	Length
23.	LA_0227	hypothetical protein LA_0227	291
24.	LA_0228	hypothetical protein LA_0228	104
25.	LA_0254	hypothetical protein LA_0254	264
26.	LA_0281	3-methyl-adenine DNA glycosylase	228
27.	LA_0283	hypothetical protein LA_0283	477
28.	LA_0284	hypothetical protein LA_0284	231
29.	LA_0286	hypothetical protein LA_0286	252
30.	LA_0299	carbon starvation protein A	701
31.	LA_0300	ADP-ribose pyrophosphatase	182
32.	LA_0301	OmpA family protein	589
33.	LA_0302	hypothetical protein LA_0302	243
34.	LA_0322	fibronectin binding protein	120
35.	LA_0348	anti-sigma factor antagonist	96
36.	LA_0350	hypothetical protein LA_0350	113
37.	LA_0353	hypothetical protein LA_0353	138
38.	LA_0365	hypothetical protein LA_0365	553
39.	LA_0368[s]	hypothetical protein LA_0368	199
40.	LA_0371	hypothetical protein LA_0371	138
41.	LA_0374	hypothetical protein LA_0374	696
42.	LA_0375	hypothetical protein LA_0375	310
43.	LA_0403	ABC transporter permease	274
44.	LA_0406	hypothetical protein LA_0406	769

Table 18: Details of pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
45.	LA_0416[S]	hypothetical protein LA_0416	359
46.	LA_0423	hypothetical protein LA_0423	480
47.	LA_0426	hypothetical protein LA_0426	516
48.	LA_0430	hypothetical protein LA_0430	244
49.	LA_0460	hypothetical protein LA_0460	193
50.	LA_0462	hypothetical protein LA_0462	268
51.	LA_0466	FOG:HEAT repeat protein	438
52.	LA_0471[S]	hypothetical protein LA_0471	157
53.	LA_0500	hypothetical protein LA_0500	225
54.	LA_0511	transcriptional regulator	98
55.	LA_0518	acetyltransferase	92
56.	LA_0525	hypothetical protein LA_0525	127
57.	LA_0532	hypothetical protein LA_0532	299
58.	LA_0573	hypothetical protein LA_0573	110
59.	LA_0574	cytoplasmic membrane protein	214
60.	LA_0575	hypothetical protein LA_0575	335
61.	LA_0578	hypothetical protein LA_0578	122
62.	LA_0580	hypothetical protein LA_0580	112
63.	LA_0589[S]	hypothetical protein LA_0589	632
64.	LA_0591	hypothetical protein LA_0591	313
65.	LA_0599	signal transduction protein	241
66.	LA_0605	SET family protein	176
67.	LA_0606	hypothetical protein LA_0606	199
68.	LA_0615	hypothetical protein LA_0615	110
69.	LA_0616[L]	hypothetical protein LA_0616	355
70.	LA_0620	hypothetical protein LA_0620	637
71.	LA_0631	hypothetical protein LA_0631	294
72.	LA_0703	Molybdate metabolism regulator	689
73.	LA_0707	transposase	304
74.	LA_0728	hypothetical protein LA_0728	175
75.	LA_0730[S]	hypothetical protein LA_0730	281
76.	LA_0734	peroxiredoxin-like protein	183
77.	LA_0735	hypothetical protein LA_0735	408
78.	LA_0747	50S ribosomal protein L29	94
79.	LA_0769	hypothetical protein LA_0769	602
80.	LA_0773	hypothetical protein LA_0773	381
81.	LA_0779	fatty acid desaturase	314
82.	LA_0792	hypothetical protein LA_0792	437
83.	LA_0793	hypothetical protein LA_0793	177
84.	LA_0817	hypothetical protein LA_0817	140
85.	LA_0835	hypothetical protein LA_0835	631
86.	LA_0872	microbial collagenase	888
87.	LA_0875	hypothetical protein LA_0875	298
88.	LA_0878	hypothetical protein LA_0878	350
89.	LA_0879	hypothetical protein LA_0879	156
90.	LA_0898	hypothetical protein LA_0898	217

Table 18: Details of pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
91.	LA_0903	methylase	543
92.	LA_0905[L]	hypothetical protein LA_0905	244
93.	LA_0913a	hypothetical protein LA_0913a	67
94.	LA_0923	IS1533 transposase	123
95.	LA_0925	IS1533 transposase	86
96.	LA_0934[T]	hypothetical protein LA_0934	638
97.	LA_0990	hypothetical protein LA_0990	232
98.	LA_0992	mechanosensitive ion channel	336
99.	LA_1013	hypothetical protein LA_1013	332
100.	LA_1027	sphingomyelinase C precursor	567
101.	LA_1029	sphingomyelinase C precursor	607
102.	LA_1074	fumarylacetoacetate hydrolase family protein	316
103.	LA_1092	hypothetical protein LA_1092	160
104.	LA_1103	hypothetical protein LA_1103	214
105.	LA_1121	hypothetical protein LA_1121	173
106.	LA_1122[S]	hypothetical protein LA_1122	317
107.	LA_1131	protein-L-isoaspartate O-methyltransferase	221
108.	LA_1141a[S]	hypothetical protein LA_1141a	248
109.	LA_1162	hypothetical protein LA_1162	372
110.	LA_1163	hypothetical protein LA_1163	221
111.	LA_1172	hypothetical protein LA_1172	134
112.	LA_1184	guanylate cyclase	530
113.	LA_1188[S]	hypothetical protein LA_1188	149
114.	LA_1210	peroxiredoxin-like protein	187
115.	LA_1213	hypothetical protein LA_1213	276
116.	LA_1263	50S ribosomal protein L32	66
117.	LA_1272	transcriptional regulator	148
118.	LA_1306	hypothetical protein LA_1306	70
119.	LA_1310	hypothetical protein LA_1310	176
120.	LA_1375	hypothetical protein LA_1375	440
121.	LA_1377	hypothetical protein LA_1377	137
122.	LA_1384	hypothetical protein LA_1384	306
123.	LA_1396	hypothetical protein LA_1396	177
124.	LA_1397	export protein	984
125.	LA_1398[S]	hypothetical protein LA_1398	218
126.	LA_1400	hypothetical protein LA_1400	573
127.	LA_1402[S]	hypothetical protein LA_1402	641
128.	LA_1413	hypothetical protein LA_1413	360
129.	LA_1420	hypothetical protein LA_1420	124
130.	LA_1421	hypothetical protein LA_1421	184
131.	LA_1423	3-oxoacyl-ACP synthase	322
132.	LA_1424	3-oxoacid CoA-transferase	593
133.	LA_1428	serine/threonine phosphatase	269
134.	LA_1429	hypothetical protein LA_1429	175
135.	LA_1430	beta-ketoacyl synthase	429
136.	LA_1453	hypothetical protein LA_1453	167

Table 18: Details of pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
137.	LA_1456	DNA repair protein RadC	232
138.	LA_1473	hypothetical protein LA_1473	178
139.	LA_1476	dehalogenase-like hydrolase	205
140.	LA_1479	hypothetical protein LA_1479	242
141.	LA_1486	hypothetical protein LA_1486	315
142.	LA_1515[S]	hypothetical protein LA_1515	158
143.	LA_1517a	hypothetical protein LA_1517a	291
144.	LA_1524	hypothetical protein LA_1524	64
145.	LA_1533	thymidylate synthase	524
146.	LA_1547	DNA methyltransferase	364
147.	LA_1549	beta-lactamase regulatory protein 1	283
148.	LA_1550	hypothetical protein LA_1550	173
149.	LA_1567	hypothetical protein LA_1567	249
150.	LA_1568[S]	hypothetical protein LA_1568	226
151.	LA_1569	hypothetical protein LA_1569	474
152.	LA_1649	membrane protein involved in the export of O-antigen and teichoic acid	447
153.	LA_1663	glycosyltransferase	301
154.	LA_1664	dTDP-rhamnosyl transferase	303
155.	LA_1666	glycosyltransferase	282
156.	LA_1686	hypothetical protein LA_1686	81
157.	LA_1691	hypothetical protein LA_1691	478
158.	LA_1715	hypothetical protein LA_1715	357
159.	LA_1734	anti-sigma factor antagonist-like protein	310
160.	LA_1741	acyltransferase	249
161.	LA_1773	hypothetical protein LA_1773	299
162.	LA_1793	transposase	296
163.	LA_1808	transposase	300
164.	LA_1830	transposase	301
165.	LA_1859	catalase	481
166.	LA_1873	hypothetical protein LA_1873	176
167.	LA_1885	hypothetical protein LA_1885	207
168.	LA_1899	aldo/keto reductase	508
169.	LA_1900	hypothetical protein LA_1900	154
170.	LA_1910	hypothetical protein LA_1910	101
171.	LA_1919	DNA-binding transcriptional activator	531
172.	LA_1937	CopG-like transcriptional regulator	178
173.	LA_1945[S]	hypothetical protein LA_1945	188
174.	LA_1954	hypothetical protein LA_1954	118
175.	LA_1957	hypothetical protein LA_1957	191
176.	LA_1962	hypothetical protein LA_1962	125
177.	LA_1971[T]	hypothetical protein LA_1971	133
178.	LA_1973	hypothetical protein LA_1973	105
179.	LA_1974	arginyl-tRNA-protein transferase	257
180.	LA_1982	O-antigen polymerase-like protein	661
181.	LA_1998	polysaccharide deacetylase	300
182.	LA_2031	hypothetical protein LA_2031	217

Table 18: Details of pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
183.	LA_2032	CopG-like transcriptional regulator	178
184.	LA_2034	hydrolase	399
185.	LA_2046	hypothetical protein LA_2046	120
186.	LA_2052	hypothetical protein LA_2052	399
187.	LA_2053	hypothetical protein LA_2053	182
188.	LA_2054	3-hydroxyisobutyrate dehydrogenase	278
189.	LA_2065	hypothetical protein LA_2065	128
190.	LA_2088	hypothetical protein LA_2088	147
191.	LA_2100^{T}	hypothetical protein LA_2100	130
192.	LA_2137	hypothetical protein LA_2137	101
193.	LA_2155	hypothetical protein LA_2155	74
194.	LA_2160	hypothetical protein LA_2160	347
195.	LA_2165	peptidyl-tRNA hydrolase	155
196.	LA_2169	lipoprotein	58
197.	LA_2170	lipoprotein with phospholipase D domain	547
198.	LA_2192	hypothetical protein LA_2192	249
199.	LA_2195	hypothetical protein LA_2195	117
200.	LA_2200	amidase	186
201.	LA_2248^{S}	hypothetical protein LA_2248	94
202.	LA_2257	hypothetical protein LA_2257	275
203.	LA_2259	hypothetical protein LA_2259	163
204.	LA_2264	ankyrin repeat-containing protein	85
205.	LA_2276	HNH family endonuclease	184

S.No	Locus Tag	Annotation	Length
206.	LA_2284	hypothetical protein LA_2284	208
207.	LA_2300	hypothetical protein LA_2300	226
208.	LA_2301	hypothetical protein LA_2301	224
209.	LA_2302	hypothetical protein LA_2302	221
210.	LA_2311	hypothetical protein LA_2311	174
211.	LA_2330	hypothetical protein LA_2330	417
212.	LA_2385	hypothetical protein LA_2385	247
213.	LA_2419	hypothetical protein LA_2419	204
214.	LA_2444	hypothetical protein LA_2444	323
215.	LA_2454	hypothetical protein LA_2454	144
216.	LA_2456	hypothetical protein LA_2456	305
217.	LA_2460	methyltransferase	259
218.	LA_2464	gliding motility ABC transporter	568
219.	LA_2465	hypothetical protein LA_2465	336
220.	LA_2484	hypothetical protein LA_2484	291
221.	LA_2488	hypothetical protein LA_2488	179
222.	LA_2499	thiol oxidoreductase	482
223.	LA_2519	hypothetical protein LA_2519	148
224.	LA_2529^{L}	hypothetical protein LA_2529	128
225.	LA_2554	phosphate sodium symporter	763
226.	LA_2578	FeoA-like protein	81
227.	LA_2582	M23 family metallo endopeptidase	380
228.	LA_2584	hypothetical protein LA_2584	353

Table 18: Details of Pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
229.	LA_2595[S]	hypothetical protein LA_2595	130
230.	LA_2601	hypothetical protein LA_2601	117
231.	LA_2626	hypothetical protein LA_2626	203
232.	LA_2628	hypothetical protein LA_2628	638
233.	LA_2637[L]	hypothetical protein LA_2637	272
234.	LA_2656	hypothetical protein LA_2656	223
235.	LA_2718	cytoplasmic membrane protein	381
236.	LA_2719	hypothetical protein LA_2719	768
237.	LA_2722	hypothetical protein LA_2722	140
238.	LA_2728	transcriptional regulator	139
239.	LA_2764[S]	hypothetical protein LA_2764	412
240.	LA_2773	hypothetical protein LA_2773	1150
241.	LA_2783a	hypothetical protein LA_2783a	109
242.	LA_2798	hypothetical protein LA_2798	146
243.	LA_2800	hypothetical protein LA_2800	368
244.	LA_2820	hypothetical protein LA_2820	669
245.	LA_2823	hypothetical protein LA_2823	349
246.	LA_2831	histidine kinase sensor protein	471
247.	LA_2845	transcriptional regulator	98
248.	LA_2850	flagellar protein	521
249.	LA_2859	hypothetical protein LA_2859	173
250.	LA_2873	hypothetical protein LA_2873	377
251.	LA_2877	hypothetical protein LA_2877	384
252.	LA_2910	hypothetical protein LA_2910	129
253.	LA_2919[L]	hypothetical protein LA_2919	161
254.	LA_2920	hypothetical protein LA_2920	496
255.	LA_2952	hypothetical protein LA_2952	570
256.	LA_2958	hypothetical protein LA_2958	377
257.	LA_2967	hypothetical protein LA_2967	104
258.	LA_2971	hypothetical protein LA_2971	686
259.	LA_2970	hypothetical protein LA_2970	100
260.	LA_2986	hypothetical protein LA_2986	230
261.	LA_3016	hypothetical protein LA_3016	90
262.	LA_3018	hypothetical protein LA_3018	455
263.	LA_3039	hypothetical protein LA_3039	390
264.	LA_3050	haemolysin	239
265.	LA_3064	hypothetical protein LA_3064	344
266.	LA_3075	LigB-like protein	1954
267.	LA_3076	hypothetical protein LA_3076	128
268.	LA_3077	hypothetical protein LA_3077	174
269.	LA_3079	hypothetical protein LA_3079	260
270.	LA_3091[S]	hypothetical protein LA_3091	245
271.	LA_3097a	IS1533 transposase	98
272.	LA_3099	IS1533 transposase	123
273.	LA_3110	potassium-transporting ATPase subunit C	190
274.	LA_3112	potassium-transporting ATPase subunitA	557

Table 18: Details of Pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
275.	LA_3114	cytoplasmic membrane protein	687
276.	LA_3115	hypothetical protein LA_3115	307
277.	LA_3118	hypothetical protein LA_3118	353
278.	LA_3120	zinc-binding carboxypeptidase	506
279.	LA_3126^{T}	hypothetical protein LA_3126	384
280.	LA_3149	TonB-dependent outer membrane hemin receptor	777
281.	LA_3150	hypothetical protein LA_3150	124
282.	LA_3152	CopG-like transcriptional regulator	178
283.	LA_3200b	hypothetical protein LA_3200b	497
284.	LA_3230	hypothetical protein LA_3230	260
285.	LA_3233	hypothetical protein LA_3233	372
286.	LA_3241	hypothetical protein LA_3241	247
287.	LA_3243	hypothetical protein LA_3243	178
288.	LA_3271^{S}	hypothetical protein LA_3271	636
289.	LA_3276^{T}	hypothetical protein LA_3276	422
290.	LA_3287	hypothetical protein LA_3287	186
291.	LA_3338^{L}	hypothetical protein LA_3338	125
292.	LA_3340^{T}	hypothetical protein LA_3340	260
293.	LA_3342	hypothetical protein LA_3342	216
294.	LA_3353	hypothetical protein LA_3353	237
295.	LA_3358	CopG-like transcriptional regulator	180
296.	LA_3370	surface antigen OrfC lipoprotein	456
297.	LA_3388	hypothetical protein LA_3388	631
298.	LA_3390	hypothetical protein LA_3390	276
299.	LA_3394^{L}	hypothetical protein LA_3394	355
300.	LA_3414	hypothetical protein LA_3414	60
301.	LA_3415	hypothetical protein LA_3415	74
302.	LA_3432	hypothetical protein LA_3432	525
303.	LA_3446^{S}	hypothetical protein LA_3446	159
304.	LA_3462	carbon-nitrogen hydrolase	527
305.	LA_3469	hypothetical protein LA_3469	440
306.	LA_3470	thiol oxidoreductase	503
307.	LA_3471^{T}	hypothetical protein LA_3471	404
308.	LA_3490	hypothetical protein LA_3490	639
309.	LA_3491^{T}	hypothetical protein LA_3491	135
310.	LA_3497	hypothetical protein LA_3497	740
311.	LA_3522	hypothetical protein LA_3522	229
312.	LA_3540	sphingomyelinase C precursor	525
313.	LA_3544	hypothetical protein LA_3544	128
314.	LA_3545^{T}	hypothetical protein LA_3545	62
315.	LA_3547	hypothetical protein LA_3547	325
316.	LA_3552^{T}	hypothetical protein LA_3552	338
317.	LA_3562	ABC transporter permease	407
318.	LA_3564	protein required for attachment to host cells	146
319.	LA_3582	gamma-glutamyl carboxylase-like protein	503
320.	LA_3586	hypothetical protein LA_3586	115

Table 18: Details of Pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
321.	LA_3623	glutamine amidotransferase	258
322.	LA_3637	hypothetical protein LA_3637	203
323.	LA_3640	hypothetical protein LA_3640	179
324.	LA_3651	hypothetical protein LA_3651	87
325.	LA_3664	hypothetical protein LA_3664	368
326.	LA_3668	uracil-DNA glycosylase	218
327.	LA_3669	hypothetical protein LA_3669	347
328.	LA_3672	hydrolase/acyltransferase	357
329.	LA_3680	hypothetical protein LA_3680	106
330.	LA_3740	acriflavine resistance protein	649
331.	LA_3762	hypothetical protein LA_3762	142
332.	LA_3770	alpha/beta hydrolase	257
333.	LA_3777	CrcB-like protein	105
334.	LA_3778	LigB-like protein	1889
335.	LA_3779	hypothetical protein LA_3779	145
336.	LA_3780	hypothetical protein LA_3780	112
337.	LA_3781	hypothetical protein LA_3781	112
338.	LA_3809	hypothetical protein LA_3809	677
339.	LA_2825a	hypothetical protein LA_2825a	481
340.	LA_3827	hypothetical protein LA_3827	184
341.	LA_3834	hypothetical protein LA_3834	475
342.	LA_3839	phospholipid binding protein	181
343.	LA_3849^{L}	hypothetical protein LA_3849	244
344.	LA_3854	hypothetical protein LA_3854	139
345.	LA_3856	hypothetical protein LA_3856	157
346.	LA_3870	hypothetical protein LA_3870	554
347.	LA_3883	uracil-DNA glycosylase	205
348.	LA_3916	hypothetical protein LA_3916	268
349.	LA_3927	TolC family protein	623
350.	LA_3957	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	373
351.	LA_3966	hypothetical protein LA_3966	175
352.	LA_3994	Fe-S oxidoreductase	565
353.	LA_4004	sphingomyelinase C precursor	510
354.	LA_4010	hypothetical protein LA_4010	148
355.	LA_4011	hypothetical protein LA_4011	337
356.	LA_4021	Bacterio ferritin-associated ferredoxin	61
357.	LA_4053	hypothetical protein LA_4053	163
358.	LA_4064	hypothetical protein LA_4064	126
359.	LA_4099	hypothetical protein LA_4099	123
360.	LA_4108	hypothetical protein LA_4108	228
361.	LA_4113	hypothetical protein LA_4113	100
362.	LA_4121	hypothetical protein LA_4121	119
363.	LA_4123	hypothetical protein LA_4123	181
364.	LA_4143^{L}	hypothetical protein LA_4143	291
365.	LA_4170	hypothetical protein LA_4170	60
366.	LA_4179	hypothetical protein LA_4179	79

Table 18: Details of Pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
367.	LA_4187	hypothetical protein LA_4187	175
368.	LA_4191	hypothetical protein LA_4191	268
369.	LA_4202[L]	hypothetical protein LA_4202	189
370.	LA_4208	hypothetical protein LA_4208	611
371.	LA_4209	hypothetical protein LA_4209	695
372.	LA_4226	hypothetical protein LA_4226	303
373.	LA_4233[s]	hypothetical protein LA_4233	140
374.	LA_4235	hypothetical protein LA_4235	176
375.	LA_4247	hydrolase	269
376.	LA_4259	transcriptional regulator	37
377.	LA_4282	hypothetical protein LA_4282	471
378.	LA_4286	hypothetical protein LA_4286	296
379.	LA_4289	multidrug ABC transporter permease	360
380.	LA_4292	hypothetical protein LA_4292	190
381.	LA_4293[s]	hypothetical protein LA_4293	224
382.	LA_4305	hypothetical protein LA_4305	270
383.	LA_4319	hypothetical protein LA_4319	195
384.	LB_008	hypothetical protein LB_008	286
385.	LB_050	hypothetical protein LB_050	204
386.	LB_070	hypothetical protein LB_070	216
387.	LB_072	hypothetical protein LB_072	178
388.	LB_080	hypothetical protein LB_080	608
389.	LB_098	xylanase/chitin deacetylase	437

S.No	Locus Tag	Annotation	Length
390.	LB_099	hypothetical protein LB_099	171
391.	LB_102	phospholysine phosphohistidine inorganic pyrophosphate phosphatase	269
392.	LB_120	hypothetical protein LB_120	365
393.	LB_127	hypothetical protein LB_127	187
394.	LB_148	hypothetical protein LB_148	113
395.	LB_158	cobalamin biosynthesis protein	385
396.	LB_159	precorrin-2 methyltransferase	252
397.	LB_190	hypothetical protein LB_190	176
398.	LB_194	hypothetical protein LB_194	192
399.	LB_196	lipoprotein	241
400.	LB_197	hypothetical protein LB_197	257
401.	LB_199	outer membrane protein	451
402.	LB_201	ATP-dependent protease LA	540
403.	LB_238	Cache family protein	323
404.	LB_240	diguanylate cyclase	507
405.	LB_250[s]	hypothetical protein LB_250	93
406.	LB_257[s]	hypothetical protein LB_257	152
407.	LB_264	hypothetical protein LB_264	305
408.	LB_265	hypothetical protein LB_265	323
409.	LB_271	permease	120
410.	LB_272	hypothetical protein LB_272	212
411.	LB_280	hypothetical protein LB_280	171
412.	LB_285	hypothetical protein LB_285	264

Table 18: Details of Pathogen specific gene content (continued)

S.No	Locus Tag	Annotation	Length
413.	LB_287	hypothetical protein LB_287	242
414.	LB_307[T]	hypothetical protein LB_307	152
415.	LB_312[T]	hypothetical protein LB_312	288
416.	LB_319	hypothetical protein LB_319	303
417.	LB_321	hypothetical protein LB_321	199
418.	LB_340	hypothetical protein LB_340	137
419.	LB_341	hypothetical protein LB_341	127
420.	LB_348	hypothetical protein LB_348	146
421.	LB_358	hypothetical protein LB_358	521
422.	-	hypothetical protein LEP1GSC117_1635	320
423.	-	hypothetical protein LEP1GSC117_3906	59
424.	LEP1GSC117_4165	PF07119 family protein	514
425.	-	hypothetical protein LEP1GSC117_2485	236
426.	-	hypothetical protein LEP1GSC117_2548	74
427.	LBL_0516	hypothetical protein LBL_0516	64

[S] -Proteins predicted with secretor signal peptide

[L] – Proteins predicted with Lipoprotein signal peptide

[T] – Proteins predicted with Tat signal peptide

3.4 Discussion

In this study two pathogenic *Leptospira* strains belonging to different species, isolated from different geographic locations and hosts were sequenced and assembled *de novo* to draft genome.

Analysis of *L. interrogans* genome revealed the presence of probable genomic islands which might be the elements of lateral gene transfer and sites of virulence-associated genes. These regions were found to harbor transposase, sphingomyelinase, β-lactamase, LigB and hypothetical genes and genes responsible for multiple antibiotic resistance. Bacterial sphingomyelinases involved in the hydrolysis of sphingomyelin and glycerophospholipids are well documented for their role in various physiological processes related to migration, growth, death and cell signaling, and also for organism's virulence by aiding it in escaping phagosomes, evading the immune system and tissue colonization (Flores-Díaz *et al.*, 2016). Conserved nature of LigB or *Leptospira* immunoglobulin like gene among all the pathogenic isolates was well documented and its role in the virulence mechanism of *Leptospira* during colonization and dissemination processes was also studied (Choy *et al.*, 2007; McBride *et al.*, 2009).

Availability of large number of genome sequences in public databases have underpinned comparative studies providing insights into the population genetic structure and evolution of *Leptospira* species. Investigations into the core and accessory gene pool of pathogenic *Leptospira* species using large number of sequenced bacteria have shown that the *Leptospira* species pan-genomes are open in nature with the potential for acquisition of new genes possibly by lateral gene transfer mechanisms.

As leptospirosis is caused by different species of *Leptospira*, we believed that analysis of core gene pool of all pathogenic species will help in identifying the possible candidate genes responsible for organism's virulence during infection, colonization and persistence in the host

tissue. Further, comparison of common core genome of pathogenic species with *L. biflexa* isolates genomes, identified genes responsible for basic housekeeping functions of the bacteria at the genus level and genes present specifically in pathogenic isolates. A total of 427 proteins were found to be conserved specifically across all pathogenic isolates included in the study. We believe that understanding the role of these pathogen specific proteins might help in revealing the pathogenicity of the organism in more focused way.

In general, bacterial pathogenesis is a collective effect of a wide range of virulence factors that help in initial binding to the host tissue, invasion and successful colonization in tissue niche, nutrient uptake and successful suppression and evasion from host immune response (Wilson, 2002). But in *Leptospira* pathogenesis, the mechanism by which *Leptospira* causes disease is largely unknown because of its unknown genetic manipulations, slow growth and difficulty to carryout gene knockout studies (Haake and Levett, 2015). This analysis thus will help in narrowing down studies onto specific candidate genes that might play an important role in pathogenicity of the organism.

A total of nine pathogen specific proteins including five hypothetical proteins (LA_1122, LA_2444, LA_3287, LA_3353 and LA_3413), LigB like protein (LA_3075), catalase (LA_1859) and two CopG like transcriptional regulators (LA_1937, LA_2032) were found to be expressed and upregulated during exposure to host innate immune system (Xue *et al.*, 2010). Possible role of catalase (KatE) in protecting bacteria during phagocytosis by host macrophage induced reactive oxygen species was well documented (Zamocky, Furtmüller and Obinger, 2008; Eshghi *et al.*, 2012). Fibronectin-binding protein, LA_0322, is an another important conserved protein, which might be playing a very important role in initial adhesion and entry of the organism in to the host (Henderson *et al.*, 2011).

Outer membrane proteins (OMPs) and lipoproteins which are part of the bacterial cytoplasmic membrane and LPS are functionally and structurally important in terms of

bacterial nutritional uptake, cell stabilization, signal transduction, host tissue attachment and immunogenicity of bacteria and can be better targets for vaccine development and diagnostic marker development (Raja and Natarajaseenivasan, 2013; Mohan and Harikrishna, 2015). A total of 18 such proteins including 11 hypothetical proteins, LipL32 (LA_2637) and LipL41 (LA_0616), two LigB like proteins (LA_3075, LA_3778), two lipoproteins (LA_2169, LB_196), two OMPs (LA_0301(OmpA), LB_199) and a surface antigen ORFc (LA_3370) were found to be conserved in all pathogen species. Conserved nature across pathogenic *Leptospira* isolates can be harnessed for designing globally applicable diagnostics and vaccines for leptospirosis.

Other major conserved genes across pathogenic species were collagenase (LA_0872), sphingomyelinase C precursor proteins (LA_1027, LA_1029, LA_3540, LA_4004) and hemolysin (LA_3050). Collagenase is a major virulence factor and it confers the potential to invade and transmit *Leptospira* in the host tissue by degrading collagen barrier (Kassegne *et al.*, 2014). Sphingomyelinase and hemolysin protect the bacteria by escaping it from phagocytosis and immune response evasion(Narayanavari *et al.*, 2012) .

Iron acquisition and regulation are the major survival strategy employed by pathogenic bacteria for establishing infection in mammalian cells and the presence of a TonB dependent hemin receptor (LA_3149) conserved only in pathogenic *Leptospira* poses a better target to tackle the disease progression during infection (Louvel *et al.*, 2006). Flagellar protein, LA_2850 and another protein, LA_3564, characterized as proteins required for host cell attachment, were found to be conserved specifically in pathogenic strains. Mu-like prophage host-nuclease inhibitor protein Gam, LA_0189, which protects linear double-stranded DNA from exonuclease degradation, was also found to be conserved suggesting its possible role in lateral gene transfers in pathogenic strains (Qin *et al.*, 2008) .

This study has also revealed that more than 50% of proteins (285) conserved among pathogenic isolates are of unknown function and are designated as hypothetical proteins. Understanding and deciphering the role of these hypothetical proteins may reveal the *Leptospira* specific pathogenic mechanism. Some of the conserved hypothetical proteins were found to be members of the PF07598 family, which includes proteins with virulence modulating properties and are LA_0589, LA_0591, LA_620, LA_0769, LA_0835, LA_0934, LA_1400, LA_1402, LA_2628, LA_3271, LA_3388 and LA_3490. Others such as LA_0031, LA_0406, LA_2259, LA_3854 and LB_272 were predicted to be membrane-associated proteins emphasizing their possible role in the antigenicity of the organism. In addition 24 hypothetical proteins were predicted to have signal peptidase 1 cleavage site and they might be secreted out of the bacterial cell. Twelve hypothetical proteins were also predicted to have tat signal peptide in their motif, suggesting their secretion to the membrane and becoming a membrane-bound protein. Apart from these proteins, a total of 147 hypothetical proteins were predicted to be involved in non-classical i.e. not signal peptide triggered secretion pathway, giving clues about their possible role in organism's virulence mechanism. Secretory proteins which play an important role in enhancing attachment of eukaryotic cells, forming niche by scavenging environment, killing target cells by disrupting the functions and playing an important role in virulence mechanism of organism were always considered as important targets (Green and Mecsas, 2016).

This study thus clearly provided a proper understanding of the population genetic structure of pathogenic *Leptospira* isolates and provided insights into genome encoded properties for the possible pathogenicity of the organism. Further, studies on the role of conserved genes across the pathogenic species using *in-vitro* validations will aid in unraveling the pathogenicity of *Leptospira* and in designing better diagnostic and vaccine candidates that can be applied globally.

Chapter 4

Summary and Outlook

The focus of the present study was to validate MLST method for *Leptospira* classification which can be applied globally and replace the traditional tedious serological classification, and to underpin the genetic differences between pathogenic and saprophytic strains of the *Leptospira* genus.

Traditional serological classification classified *Leptospira* into serovar and serogroups. In the aftermath of molecular typing methods *Leptospira* was classified into species using DNA-DNA hybridization studies. Several myriad of molecular tools were employed for classification of *Leptospira* but because of lack of sensitivity, specificity and reproducibility many methods could not be used globally with an ease. With the development of MLST typing method for *Leptospira*, with its inherent advantages of simplicity, easy of performance, reproducibility, affordability and result archival possibility with proven sensitivities, it has become method of choice for molecular typing. With availability of more than one MLST method for typing *Leptospira* isolates, it became arduous for the research fraternity to which one to use. Hence, our aim was to validate the MLST scheme across all the pathogenic species irrespective of their geography and host specificity.

MLST proposed by Ahmed et al in 2008, when applied to the global collection of pathogenic *Leptospira* isolates has resulted in phylogeny according to their species classification. This MLST result when analyzed with MCMC approach has revealed their possible ancestral relationships. When Ahmed's MLST scheme was compared with other popular genome based molecular typing method FAFLP, MLST stood out as more reliable method, conserving isolate relationships across the genus. When analyzed the Indian subcontinent isolates, it has pointed out that isolates from Andaman and Nicobar Islands serve as ancestors for the sub-continent isolates. In addition when Ahmed's MLST scheme was compared to other *Leptospira* MLST schemes it revealed that Ahmed's MLST scheme has superior discriminatory power that too by using lesser number of MLST loci. Comparison with the whole genome

derived phylogenetic relationships has proven that Ahmed's MLST scheme phylogenies were well conserved. This has confirmed that Ahmed's MLST scheme can be applied globally to all pathogenic species of the genus and as the scheme utilizes lesser number of loci, it is also considered as an economically viable method for *Leptospira* characterization.

As the whole genome provides better understanding of the organism's characters in terms of its evolution, pathogenic mechanism, host adaptation and survival, this has become method of choice for studying many of the bacteria especially for difficult to grow bacteria in laboratories. Given these advantages of whole genome sequences, we have adopted this approach to understand the pathogenicity of *Leptospira.*

Analysis of the whole genome of *Leptospira interrogans* for possible genomic inserts based on conserved motifs and variation in GC content, identified regions of genomic islands confirming their acquisition during course of evolution from environment by lateral transfer.

Two pathogenic *Leptospira* strains isolated from different host and geographies were sequenced, assembled, annotated and submitted to NCBI genome data base. With the success of sequencing of genome, we then aimed at looking into genome encoded properties for pathogenicity of *Leptospira* by using comparative genomics approach. To decipher pathogenic potential, core genomes for all pathogenic species was extracted and found that a total of 2708 proteins were conserved in all the pathogenic isolates of the genus.

These genes were then compared with saprophytic *L biflexa* isolates genetic content and found that a total of 427 proteins were specifically present only in pathogenic isolates. This has focused our study on deciphering possible underlying mechanisms of *Leptospira* pathogenicity.

When these proteins were analyzed for their function, it was found that a significant portion (almost 67%) of them are of hypothetical in nature with unknown function. Prediction

servers predicted that few of these hypothetical proteins are secretory in nature, lipoproteins and membrane proteins. Further investigations into their role in pathogen dynamics and causing fitness advantage to different niches will be helpful to better understand the pathogen evolution.

In conclusion, this work has validated Ahmed's MLST scheme against the backdrop of different molecular typing methods using parts of genome and whole genome, and found to be better applicable for typing *Leptospira* isolates. Analysis carried out with multiple genomes of various pathogenic species of *Leptospira* has revealed conserved proteins across all the pathogenic species and proteins specifically present in pathogenic species. This work thus would serve as a guide to understand pathogenicity of *Leptospira* at population genetic level. In future, characterizing these pathogen conserved proteins will help in designing better diagnostics and vaccines that can be used globally for controlling worldwide leptospirosis and understanding underlying mechanisms of *Leptospira* pathogenicity.

References

Abela-Ridder, B., Sikkema, R. and Hartskeerl, R. A. (2010) 'Estimating the burden of human leptospirosis', *International Journal of Antimicrobial Agents*, 36, pp. S5–S7. doi: 10.1016/j.ijantimicag.2010.06.012.

Adler, B. and de la Peña Moctezuma, A. (2010) 'Leptospira and leptospirosis', *Veterinary Microbiology*, pp. 287–296. doi: 10.1016/j.vetmic.2009.03.012.

Agampodi, S. B. *et al.* (2012) 'Utility of quantitative polymerase chain reaction in leptospirosis diagnosis: association of level of leptospiremia and clinical manifestations in Sri Lanka.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 54(9), pp. 1249–55. doi: 10.1093/cid/cis035.

Agren, J. *et al.* (2012) 'Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups.', *PloS one*, 7(6), p. e39107. doi: 10.1371/journal.pone.0039107.

Ahmed, A. and P. Grobusch, M. (2012) 'Molecular Approaches in the Detection and Characterization of Leptospira', *Journal of Bacteriology & Parasitology*. OMICS Group, 03(02). doi: 10.4172/2155-9597.1000133.

Ahmed, A. *et al.* (2009) 'Development and validation of a real-time PCR for detection of pathogenic leptospira species in clinical materials.', *PloS one*. Public Library of Science, 4(9), p. e7093. doi: 10.1371/journal.pone.0007093.

Ahmed, A. *et al.* (2011) 'Comparison of two multilocus sequence based genotyping schemes for Leptospira species', *PLoS Neglected Tropical Diseases*. Edited by M. Picardeau. Public Library of Science, 5(11), p. e1374. doi: 10.1371/journal.pntd.0001374.

Ahmed, N. *et al.* (2006) 'Multilocus sequence typing method for identification and genotypic classification of pathogenic Leptospira species', *Annals of Clinical Microbiology and Antimicrobials*, 5(1), p. 28. doi: 10.1186/1476-0711-5-28.

Ahmed, N. *et al.* (2008) 'Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention', *Nat Rev Microbiol*, 6(5), pp. 387–394. doi: 10.1038/nrmicro1889.

Ahrenfeldt, J. *et al.* (2017) 'Bacterial whole genome-based phylogeny: construction of a new benchmarking dataset and assessment of some existing methods', *BMC Genomics*. BioMed Central, 18(1), p. 19. doi: 10.1186/s12864-016-3407-6.

Arean, V. M. (1962) 'The pathologic anatomy and pathogenesis of fatal human leptospirosis (Weil's disease).', *The American journal of pathology*, 40, pp. 393–423. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/13862141 (Accessed: 23 September 2019).

Bandelt, H. J. *et al.* (1995) 'Mitochondrial portraits of human populations using median networks.', *Genetics*, 141(2), pp. 743–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8647407 (Accessed: 15 July 2019).

Bandelt, H. J., Forster, P. and Rohl, A. (1999) 'Median-joining networks for inferring intraspecific phylogenies', *Molecular Biology and Evolution*, 16(1), pp. 37–48. doi: 10.1093/oxfordjournals.molbev.a026036.

Barker, F. A. (1926) 'Leptospirosis. With Special Reference to the Existence of Spirochaetosis Ictero-Haemorrhagica, or Weil's Disease, in the Andaman Islands.', *The Indian medical gazette*. Pandeya Publications, 61(10), pp. 479–488. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29011249 (Accessed: 18 April 2019).

Barocchi, M. A. *et al.* (2001) 'Identification of New Repetitive Element in Leptospira interrogans Serovar Copenhageni and Its Application to PCR-Based Differentiation of Leptospira Serogroups', *Journal of Clinical Microbiology*, 39(1), pp. 191–195. doi: 10.1128/JCM.39.1.191-195.2001.

Bashiru, G. and Bahaman, A. (2018) 'Advances & challenges in leptospiral vaccine development', *Indian Journal of Medical Research*, 147(1), p. 15. doi: 10.4103/ijmr.IJMR_1022_16.

Bendtsen, Jannick D *et al.* (2005) 'Non-classical protein secretion in bacteria.', *BMC Microbiology*, 5(1), p. 58. doi: 10.1186/1471-2180-5-58.

Bendtsen, Jannick Dyrløv *et al.* (2005) 'Prediction of twin-arginine signal peptides.', *BMC bioinformatics*, 6, p. 167. doi: 10.1186/1471-2105-6-167.

Bharti, A. R. *et al.* (2003) 'Leptospirosis: A zoonotic disease of global importance', *Lancet Infectious Diseases*, pp. 757–771. doi: 10.1016/S1473-3099(03)00830-2.

Bikandi, J. *et al.* (2004) 'In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction', *Bioinformatics*, 20(5), pp. 798–799. doi: 10.1093/bioinformatics/btg491.

Bolin, C. A. and Koellner, P. (1988) 'Human-to-human transmission of Leptospira interrogans by milk.', *The Journal of infectious diseases*, 158(1), pp. 246–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3392418 (Accessed: 24 April 2019).

Boonsilp, S. *et al.* (2013) 'A Single Multilocus Sequence Typing (MLST) Scheme for Seven Pathogenic Leptospira Species', *PLoS Neglected Tropical Diseases*. Public Library of Science, 7(1), p. e1954. doi: 10.1371/journal.pntd.0001954.

Bourhy, P. *et al.* (2011) 'Comparison of real-time PCR assays for detection of pathogenic Leptospira spp. in blood and identification of variations in target sequences.', *Journal of clinical microbiology*, 49(6), pp. 2154–60. doi: 10.1128/JCM.02452-10.

Branger, C. *et al.* (2005) 'Polymerase chain reaction assay specific for pathogenic Leptospira based on the gene hap1 encoding the hemolysis-associated protein-1', *FEMS Microbiology Letters*, 243(2), pp. 437–445. doi: 10.1016/j.femsle.2005.01.007.

Brenner, D. J. *et al.* (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 Bacteriology*, 49(2), pp. 839–858. doi: 10.1099/00207713-49-2-839.

Bulach, D. M. *et al.* (2006) 'Genome reduction in Leptospira borgpetersenii reflects limited transmission potential.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 103(39), pp. 14560–5. doi: 10.1073/pnas.0603979103.

Busse, H.-J. *et al.* (2010) 'Notes on the characterization of prokaryote strains for taxonomic purposes', *International Journal of Systematic and Evolutionary Microbiology*, 60(1), pp. 249–266. doi: 10.1099/ijs.0.016949-0.

Carlos Guimaraes, L. *et al.* (2015) 'Inside the Pan-genome - Methods and Software Overview', *Current Genomics*. Bentham Science Publishers Ltd., 16(4), pp. 245–252. doi: 10.2174/1389202916666150423002311.

Carver, T. *et al.* (2012) 'Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data', *Bioinformatics*. Oxford University Press, 28(4), pp. 464–469. doi: 10.1093/bioinformatics/btr703.

Cerqueira, G. M. and Picardeau, M. (2009) 'A century of Leptospira strain typing.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 9(5), pp. 760–8. doi: 10.1016/j.meegid.2009.06.009.

Chalayon, P. *et al.* (2011) 'Leptospirosis serodiagnosis by ELISA based on recombinant outer membrane protein', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 105(5), pp. 289–297. doi: 10.1016/j.trstmh.2011.01.008.

Chandrasekaran, S. and Gomathi, S. (2004) 'A standard screening test for the early and rapid diagnosis of leptospirosis.', *Indian journal of medical microbiology*. Medknow Publications, 22(1), pp. 23–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17642681 (Accessed: 2 October 2019).

Chapman, A. J., Faine, S. and Adler, B. (1990) 'Antigens recognized by the human immune response to vaccination with a bivalent *hardjo/pomona* leptospiral vaccine', *FEMS Microbiology Letters*, 64(2), pp. 111–118. doi: 10.1111/j.1574-6968.1990.tb03508.x.

Chirathaworn, C. and Kongpan, S. (2013) 'Immune responses to Leptospira infection: roles as biomarkers for disease severity.', *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases*. Elsevier Editora Ltda, 18(1), pp. 77–81. doi: 10.1016/j.bjid.2013.08.002.

Choy, H. A. *et al.* (2007) 'Physiological osmotic induction of Leptospira interrogans adhesion: LigA and LigB bind extracellular matrix proteins and fibrinogen', *Infection and Immunity*, 75(5), pp. 2441–2450. doi: 10.1128/IAI.01635-06.

Chu, K. M. *et al.* (1998) 'Identification of Leptospira Species in the Pathogenesis of Uveitis and Determination of Clinical Ocular Characteristics in South India', *The Journal of Infectious Diseases*. Narnia, 177(5), pp. 1314–1321. doi: 10.1086/515273.

Ciceroni, L. *et al.* (2002) 'Differentiation of leptospires of the serogroup Pomona by monoclonal antibodies, pulsed-field gel electrophoresis and arbitrarily primed polymerase chain reaction.', *Research in microbiology*, 153(1), pp. 37–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11881897 (Accessed: 29 May 2019).

Coghlan, J. D. and Bain, A. D. (1969) 'Leptospirosis in Human Pregnancy followed by Death of the Foetus', *BMJ*, 1(5638), pp. 228–230. doi: 10.1136/bmj.1.5638.228.

Corwin, A. *et al.* (1990) 'A waterborne outbreak of leptospirosis among United States military personnel in Okinawa, Japan.', *International journal of epidemiology*, 19(3), pp. 743–8. doi: 10.1093/ije/19.3.743.

Costa, F. *et al.* (2015) 'Global Morbidity and Mortality of Leptospirosis: A Systematic Review.', *PLoS neglected tropical diseases*. Public Library of Science, 9(9), p. e0003898. doi: 10.1371/journal.pntd.0003898.

Cullen, P. A., Haake, D. A. and Adler, B. (2004) 'Outer membrane proteins of pathogenic spirochetes.', *FEMS microbiology reviews*, 28(3), pp. 291–318. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15449605 (Accessed: 3 April 2018).

Dellagostin, O. A. *et al.* (2011) 'Recombinant vaccines against Leptospirosis', *Human Vaccines*, 7(11), pp. 1215–1224. doi: 10.4161/hv.7.11.17944.

Dikken, H. and Kmety, E. (1978) 'Serological Typing Methods of Leptospires', *Methods in Microbiology*.

Academic Press, 11, pp. 259–307. doi: 10.1016/S0580-9517(08)70493-8.

Drummond, A. J. *et al.* (2012) 'Bayesian phylogenetics with BEAUti and the BEAST 1.7.', *Molecular biology and evolution*, pp. 1969–73. doi: 10.1093/molbev/mss075.

Drummond, A. J. and Rambaut, A. (2007) 'BEAST: Bayesian evolutionary analysis by sampling trees', *BMC Evolutionary Biology*. BioMed Central, 7(1), p. 214. doi: 10.1186/1471-2148-7-214.

Ellis, W. A. (2015) 'Animal Leptospirosis', in *Current topics in microbiology and immunology*, pp. 99–137. doi: 10.1007/978-3-662-45059-8_6.

Ellis, W. A. and Michno, S. W. (1976) 'Bovine leptospirosis: a serological and clinical study.', *The Veterinary record*, 99(20), pp. 387–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/997182 (Accessed: 18 April 2019).

Emanuelsson, O. *et al.* (2007) 'Locating proteins in the cell using TargetP, SignalP and related tools.', *Nature protocols*. Nature Publishing Group, 2(4), pp. 953–71. doi: 10.1038/nprot.2007.131.

Enright, M. C. and Spratt, B. G. (1999) 'Multilocus sequence typing.', *Trends in microbiology*, 7(12), pp. 482–7. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10603483.

Eshghi, A. *et al.* (2012) 'Leptospira interrogans Catalase Is Required for Resistance to H 2 O 2 and for Virulence', *Infection and Immunity*. Edited by S. R. Blanke, 80(11), pp. 3892–3899. doi: 10.1128/IAI.00466-12.

Van Eys, G. J. *et al.* (1991) 'Characterization of serovars of the genus Leptospira by DNA hybridization with hardjobovis and icterohaemorrhagiae recombinant probes with special attention to serogroup sejroe.', *Journal of clinical microbiology*. American Society for Microbiology, 29(5), pp. 1042–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2056039 (Accessed: 4 April 2018).

Faine, S. (1999) *Leptospira and leptospirosis*. MediSci. Available at: https://books.google.co.in/books/about/Leptospira_and_Leptospirosis.html?id=MU1WAAAAYAAJ (Accessed: 3 April 2018).

Faine, S. and World Health Organization (1982) *Guidelines for the control of leptospirosis*. Geneva: World Health Organization. Available at: https://apps.who.int/iris/handle/10665/37219.

Faisal, S. M. *et al.* (2016) 'Leptospira surface adhesin (Lsa21) induces Toll like receptor 2 and 4 mediated inflammatory responses in macrophages', *Scientific Reports*. Nature Publishing Group, 6.

doi: 10.1038/srep39530.

Farr, R. W. (1995) 'Leptospirosis', *Clinical Infectious Diseases*. Oxford University Press, pp. 1–6. doi: 10.2307/4458710.

Feigin, R. D. and Anderson, D. C. (1975) 'Human leptospirosis.', *CRC critical reviews in clinical laboratory sciences*, 5(4), pp. 413–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/237740 (Accessed: 4 May 2018).

Flores-Díaz, M. *et al.* (2016) 'Bacterial Sphingomyelinases and Phospholipases as Virulence Factors', *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 80(3), pp. 597–628. doi: 10.1128/mmbr.00082-15.

Foster, J. T. *et al.* (2009) 'Whole-genome-based phylogeny and divergence of the genus Brucella.', *Journal of bacteriology*. American Society for Microbiology Journals, 191(8), pp. 2864–70. doi: 10.1128/JB.01581-08.

Fox, G. E. *et al.* (1980) 'The phylogeny of prokaryotes.', *Science (New York, N.Y.)*, 209(4455), pp. 457–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6771870 (Accessed: 4 April 2018).

Galloway, R. L. and Levett, P. N. (2010) 'Application and Validation of PFGE for Serovar Identification of Leptospira Clinical Isolates', *PLoS Neglected Tropical Diseases*. Edited by S. Lukehart. Public Library of Science, 4(9), p. e824. doi: 10.1371/journal.pntd.0000824.

Gamberini, M. *et al.* (2005) 'Whole-genome analysis of *Leptospira interrogans* to identify potential vaccine candidates against leptospirosis', *FEMS Microbiology Letters*, 244(2), pp. 305–313. doi: 10.1016/j.femsle.2005.02.004.

Goarant, C. (2014) 'Leptospirosis: Time to move to molecular epidemiology', *Infection, Genetics and Evolution*, 21, pp. 484–485. doi: 10.1016/j.meegid.2013.10.018.

Goris, M. G. A. *et al.* (2013) 'Prospective Evaluation of Three Rapid Diagnostic Tests for Diagnosis of Human Leptospirosis', *PLoS Neglected Tropical Diseases*. Edited by J. N. Ricaldi, 7(7), p. e2290. doi: 10.1371/journal.pntd.0002290.

Goris, M. G. A. and Hartskeerl, R. A. (2014) 'Leptospirosis serodiagnosis by the microscopic agglutination test.', *Current protocols in microbiology*, 32, pp. 12E.5.1-12E.5.18. doi: 10.1002/9780471729259.mc12e05s32.

Gravekamp,', C. *et al.* (1993) *Detection of seven species of pathogenic leptospires by PCR using two sets of primers*, *Journal of General Microbiology*. Available at: www.microbiologyresearch.org (Accessed: 16 May 2019).

Green, E. R. and Mecsas, J. (2016) 'Bacterial Secretion Systems: An Overview', *Microbiology Spectrum*. American Society for Microbiology, 4(1). doi: 10.1128/microbiolspec.vmbf-0012-2015.

Grimont, F. and Grimont, P. A. (1986) 'Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools.', *Annales de l'Institut Pasteur. Microbiology*, 137B(2), pp. 165–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2446639 (Accessed: 28 May 2019).

Guidugli, F., Castro, A. A. and Atallah, Á. N. (2000) 'Antibiotics for preventing leptospirosis', in Guidugli, F. (ed.) *The Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd, p. CD001305. doi: 10.1002/14651858.CD001305.

Haake, D. A. *et al.* (2004) 'Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer.', *Journal of bacteriology*, 186(9), pp. 2818–28. doi: 10.1128/jb.186.9.2818-2828.2004.

Haake, D. A. and Levett, P. N. (2015) 'Leptospirosis in humans.', *Current topics in microbiology and immunology*, 387, pp. 65–97. doi: 10.1007/978-3-662-45059-8_5.

Hall (1999) 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT', *Nucleic Acids Symposium Series*, 41.

Harrison, N. A. and Fitzgerald, W. R. (1988) 'Leptospirosis--can it be a sexually transmitted disease?', *Postgraduate medical journal*. BMJ Publishing Group, 64(748), pp. 163–4. doi: 10.1136/pgmj.64.748.163.

Hartskeerl, R. A. *et al.* (2004) 'Classification of Leptospira from the Eyes of Horses Suffering from Recurrent Uveitis', *Journal of Veterinary Medicine Series B*, 51(3), pp. 110–115. doi: 10.1111/j.1439-0450.2004.00740.x.

Haubold, B. and Hudson, R. R. (2000) 'LIAN 3.0: detecting linkage disequilibrium in multilocus data', *Bioinformatics*. Oxford University Press (OUP), 16(9), pp. 847–849. doi: 10.1093/bioinformatics/16.9.847.

Henderson, B. *et al.* (2011) 'Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins', *FEMS Microbiol Rev*. 2010/08/11, 35(1), pp. 147–200. doi: 10.1111/j.1574-6976.2010.00243.x.

Henneberry, R. C. and Cox, C. D. (1970) 'Beta-oxidation of fatty acids by Leptospira.', *Canadian journal of microbiology*, 16(1), pp. 41–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5415967 (Accessed: 18 April 2019).

Herrmann, J. L. *et al.* (1992) 'Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification.', *Journal of clinical microbiology*, 30(7), pp. 1696–702. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1629323 (Accessed: 28 May 2019).

Hiller, K. *et al.* (2004) 'PrediSi: Prediction of signal peptides and their cleavage positions', *Nucleic Acids Research*, 32(WEB SERVER ISS.). doi: 10.1093/nar/gkh378.

Hunter, P. R. and Gaston, M. A. (1988) 'Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity', *Journal of Clinical Microbiology*, 26(11), pp. 2465–2466.

Huson, D. H. and Bryant, D. (2006) 'Application of Phylogenetic Networks in Evolutionary Studies', *Molecular Biology and Evolution*. Oxford University Press, 23(2), pp. 254–267. doi: 10.1093/molbev/msj030.

John, T. J. (2005) 'The prevention and control of human leptospirosis.', *Journal of postgraduate medicine*. Medknow Publications, 51(3), pp. 205–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16333194 (Accessed: 26 April 2019).

Johnson, R. C. and Rogers, P. (1964) 'DIFFERENTIATION OF PATHOGENIC AND SAPROPHYTIC LEPTOSPIRES WITH 8-AZAGUANINE.', *Journal of bacteriology*, 88, pp. 1618–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14244050 (Accessed: 23 May 2019).

Jolley, K. A. *et al.* (2001) 'Sequence type analysis and recombinational tests (START)', *Bioinformatics*, 17(12), pp. 1230–1231. doi: 10.1093/bioinformatics/17.12.1230.

Jolley, K. A., Bray, J. E. and Maiden, M. C. J. (2018) 'Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications', *Wellcome Open Research*, 3. doi: 10.12688/wellcomeopenres.14826.1.

Jolley, K. A. and Maiden, M. C. (2013) 'Automated extraction of typing information for bacterial pathogens from whole genome sequence data: Neisseria meningitidis as an exemplar.', *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 18(4), p. 20379. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23369391 (Accessed: 3 April 2014).

Kao, R. R. *et al.* (2014) 'Supersize me: how whole-genome sequencing and big data are transforming epidemiology', *Trends in Microbiology*, 22(5), pp. 282–291. doi: 10.1016/j.tim.2014.02.011.

Karande, S. *et al.* (2005) 'Concurrent outbreak of leptospirosis and dengue in Mumbai, India, 2002.', *Journal of tropical pediatrics*, 51(3), pp. 174–81. doi: 10.1093/tropej/fmh100.

Karlin, S. (2001) 'Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes', *Trends in Microbiology*, 9(7), pp. 335–343. doi: 10.1016/S0966-842X(01)02079-0.

Kassegne, K. *et al.* (2014) 'Identification of Collagenase as a Critical Virulence Factor for Invasiveness and Transmission of Pathogenic Leptospira Species', *The Journal of Infectious Diseases*, 209(7), pp. 1105–1115. doi: 10.1093/infdis/jit659.

Katz, A. R. *et al.* (2001) 'Assessment of the Clinical Presentation and Treatment of 353 Cases of Laboratory- Confirmed Leptospirosis in Hawaii, 1974–1998', *Clinical Infectious Diseases*, 33(11), pp. 1834–1841. doi: 10.1086/324084.

Kmety, E. and Dikken, H. (1993) *Classification of the species Leptospira interrogans and history of its serovars*. Uinversity Press Groningen.

Ko, A. I., Goarant, C. and Picardeau, M. (2009) 'Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen.', *Nature reviews. Microbiology*. Nature Publishing Group, 7(10), pp. 736–47. doi: 10.1038/nrmicro2208.

Kobayashi, Y. (2001) 'Discovery of the causative organism of Weil's disease: historical view.', *Journal of infection and chemotherapy : official journal of the Japan Society of Chemotherapy*, 7(1), pp. 10–5. doi: 10.1007/s1015610070010.

Koboldt, D. C. *et al.* (2013) 'The next-generation sequencing revolution and its impact on genomics', *Cell*. Cell Press, p. 27. doi: 10.1016/j.cell.2013.09.006.

Koizumi, N. *et al.* (2012) 'A new loop-mediated isothermal amplification method for rapid, simple, and sensitive detection of Leptospira spp. in urine.', *Journal of clinical microbiology*. American Society for Microbiology (ASM), 50(6), pp. 2072–4. doi: 10.1128/JCM.00481-12.

Koizumi, N. and Watanabe, H. (2005) 'Leptospirosis vaccines: past, present, and future.', *Journal of postgraduate medicine*, 51(3), pp. 210–214. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16333195 (Accessed: 3 October 2019).

Kositanont, U. *et al.* (2007) 'Assessment of Southern blot ribotyping for differentiation of Leptospira strains isolated from field rats.', *Journal of microbiological methods*, 69(2), pp. 288–97. doi: 10.1016/j.mimet.2007.01.012.

Kryazhimskiy, S. and Plotkin, J. B. (2008) 'The Population Genetics of dN/dS', *PLoS Genetics*. Edited by T. Gojobori, 4(12), p. e1000304. doi: 10.1371/journal.pgen.1000304.

Kumar, N. *et al.* (2015) 'Comparative genomic analysis of Helicobacter pylori from Malaysia identifies

three distinct lineages suggestive of differential evolution.', *Nucleic acids research*, 43(1), pp. 324–35. doi: 10.1093/nar/gku1271.

Kumar, S. *et al.* (2008) 'MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences', *Briefings in Bioinformatics*, 9(4), pp. 299–306. doi: 10.1093/bib/bbn017.

Kuriakose, M. *et al.* (2008) *Leptospirosis in a midland rural area of Kerala State*, *The Indian journal of medical research*.

Langille, M. G. I. and Brinkman, F. S. L. (2009) 'IslandViewer: an integrated interface for computational identification and visualization of genomic islands', *Bioinformatics*, 25(5), pp. 664–665. doi: DOI 10.1093/bioinformatics/btp030.

Larkin, M. A. *et al.* (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23(21), pp. 2947–2948. doi: 10.1093/bioinformatics/btm404.

Larsen, M. V *et al.* (2012) 'Multilocus sequence typing of total-genome-sequenced bacteria.', *Journal of clinical microbiology*, 50(4), pp. 1355–61. doi: 10.1128/JCM.06094-11.

Laurichesse, H. *et al.* (2007) 'Safety and immunogenicity of subcutaneous or intramuscular administration of a monovalent inactivated vaccine against Leptospira interrogans serogroup Icterohaemorrhagiae in healthy volunteers', *Clinical Microbiology and Infection*, 13(4), pp. 395–403. doi: 10.1111/j.1469-0691.2007.01662.x.

Lehmann, J. S. *et al.* (2014) 'Leptospiral pathogenomics', *Pathogens*. Multidisciplinary Digital Publishing Institute, 3(2), pp. 280–308. doi: 10.3390/pathogens3020280.

Leon, A. *et al.* (2010) 'Multilocus Sequence Analysis for Typing Leptospira interrogans and Leptospira kirschneri', *Journal of Clinical Microbiology*, 48(2), pp. 581–585. doi: 10.1128/JCM.00543-09.

Leonard, F. *et al.* (1992) 'Duration of urinary excretion of leptospires by cattle naturally or experimentally infected with Leptospira interrogans serovar hardjo', *Veterinary Record*, 131(19), pp. 435–439. doi: 10.1136/vr.131.19.435.

Levett, P. N. (2001) 'Leptospirosis.', *Clinical microbiology reviews*, 14(2), pp. 296–326. doi: 10.1128/CMR.14.2.296-326.2001.

Li, L., Stoeckert, C. J. and Roos, D. S. (2003) 'OrthoMCL: identification of ortholog groups for eukaryotic genomes.', *Genome research*, 13(9), pp. 2178–89. doi: 10.1101/gr.1224503.

Liang, W. *et al.* (2012) 'Pan-genomic analysis provides insights into the genomic variation and evolution of Salmonella Paratyphi A.', *PloS one*, 7(9), p. e45346. doi: 10.1371/journal.pone.0045346.

Lingappa, J. *et al.* (2004) 'HLA-DQ6 and ingestion of contaminated water: possible gene–environment interaction in an outbreak of Leptospirosis', *Genes & Immunity*, 5(3), pp. 197–202. doi: 10.1038/sj.gene.6364058.

Lo, M. *et al.* (2006) 'Effects of temperature on gene expression patterns in Leptospira interrogans serovar Lai as assessed by whole-genome microarrays.', *Infection and immunity*, 74, pp. 5848–5859. doi: 10.1128/IAI.00755-06.

Louvel, H. *et al.* (2006) 'Comparative and Functional Genomic Analyses of Iron Transport and Regulation in Leptospira spp.', *Journal of Bacteriology*, 188(22), pp. 7893–7904. doi: 10.1128/JB.00711-06.

Lukjancenko, O., Wassenaar, T. M. and Ussery, D. W. (2010) 'Comparison of 61 sequenced Escherichia coli genomes.', *Microbial ecology*, 60(4), pp. 708–20. doi: 10.1007/s00248-010-9717-3.

Lux, R., Moter, A. and Shi, W. (2000) 'Chemotaxis in pathogenic spirochetes: directed movement toward targeting tissues?', *Journal of molecular microbiology and biotechnology*, 2(4), pp. 355–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11075906 (Accessed: 23 September 2019).

Maiden, M. C. *et al.* (1998) 'Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms.', *Proceedings of the National Academy of Sciences of the United States of America*, 95, pp. 3140–3145. doi: 10.1073/pnas.95.6.3140.

Maiden, M. C. J. (2006) 'Multilocus sequence typing of bacteria.', *Annual review of microbiology*. Annual Reviews, 60, pp. 561–88. doi: 10.1146/annurev.micro.59.030804.121325.

Majed, Z. *et al.* (2005) 'Identification of Variable-Number Tandem-Repeat Loci in Leptospira interrogans Sensu Stricto', *Journal of Clinical Microbiology*, 43(2), pp. 539–545. doi: 10.1128/JCM.43.2.539-545.2005.

Marchler-Bauer, A. *et al.* (2014) 'CDD: NCBI's conserved domain database.', *Nucleic acids research*, 43(D1), pp. D222-226. doi: 10.1093/nar/gku1221.

Marquez, A. *et al.* (2017) 'Overview of laboratory methods to diagnose Leptospirosis and to identify and to type leptospires.', *International microbiology : the official journal of the Spanish Society for Microbiology*, 20(4), pp. 184–193. doi: 10.2436/20.1501.01.302.

Marshall, R. B., Wilton, B. E. and Robinson, A. J. (1981) 'Identification of leptospira serovars by restriction-endonuclease analysis', *Journal of Medical Microbiology*, 14(1), pp. 163–166. doi: 10.1099/00222615-14-1-163.

Masuzawa, T. *et al.* (1988) 'Classification of Leptospira interrogans serovar lai strain 017 by using monoclonal antibodies.', *Journal of clinical microbiology*. American Society for Microbiology (ASM), 26(11), pp. 2332–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3235658 (Accessed: 28 May 2019).

Matsunaga, J. *et al.* (2007) 'Response of Leptospira interrogans to physiologic osmolarity: relevance in signaling the environment-to-host transition.', *Infection and immunity*, 75(6), pp. 2864–74. doi: 10.1128/IAI.01619-06.

McBride, A. J. A. *et al.* (2009) 'Genetic diversity of the Leptospiral immunoglobulin-like (Lig) genes in pathogenic Leptospira spp.', *Infection, Genetics and Evolution*, 9(2), pp. 196–205. doi: 10.1016/j.meegid.2008.10.012.

Medini, D. *et al.* (2005) 'The microbial pan-genome.', *Current opinion in genetics & development*, 15(6), pp. 589–94. doi: 10.1016/j.gde.2005.09.006.

Merien, F. *et al.* (1998) 'In vivo apoptosis of hepatocytes in guinea pigs infected with *Leptospira interrogans* serovar icterohaemorrhagiae', *FEMS Microbiology Letters*, 169(1), pp. 95–102. doi: 10.1111/j.1574-6968.1998.tb13304.x.

Merien, F. *et al.* (2000) 'Identification of a 36-kDa fibronectin-binding protein expressed by a virulent variant of Leptospira interrogans serovar icterohaemorrhagiae.', *FEMS microbiology letters*, 185(1), pp. 17–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10731601 (Accessed: 19 April 2018).

Merien, F. *et al.* (2005) 'A rapid and quantitative method for the detection of *Leptospira* species in human leptospirosis', *FEMS Microbiology Letters*, 249(1), pp. 139–147. doi: 10.1016/j.femsle.2005.06.011.

Metzker, M. L. (2010) 'Sequencing technologies - the next generation.', *Nature reviews. Genetics*, 11(1), pp. 31–46. doi: 10.1038/nrg2626.

Milne, I. *et al.* (2010) 'Tablet-next generation sequence assembly visualization', *Bioinformatics*, 26(3), pp. 401–402. doi: DOI 10.1093/bioinformatics/btp666.

Mohan, A. and Harikrishna, J. (2015) 'Biomarkers for the diagnosis of bacterial infections: in pursuit of the "Holy Grail".', *The Indian journal of medical research*, 141(3), pp. 271–3. doi: 10.4103/0971-5916.156551.

Morey, R. E. *et al.* (2006) 'Species-specific identification of Leptospiraceae by 16S rRNA gene sequencing.', *Journal of clinical microbiology*, 44(10), pp. 3510–6. doi: 10.1128/JCM.00670-06.

Musso, D. and La Scola, B. (2013) 'Laboratory diagnosis of leptospirosis: A challenge', *Journal of Microbiology, Immunology and Infection*, 46(4), pp. 245–252. doi: 10.1016/j.jmii.2013.03.001.

Nalam, K. *et al.* (2010) 'Genetic Affinities within a Large Global Collection of Pathogenic Leptospira: Implications for Strain Identification and Molecular Epidemiology', *PLoS ONE*. Edited by R. K. Aziz, 5(8), p. e12637. doi: 10.1371/journal.pone.0012637.

Nally, J. E. *et al.* (2004) 'Alveolar Septal Deposition of Immunoglobulin and Complement Parallels Pulmonary Hemorrhage in a Guinea Pig Model of Severe Pulmonary Leptospirosis', *The American Journal of Pathology*, 164(3), pp. 1115–1127. doi: 10.1016/S0002-9440(10)63198-7.

Narayanavari, S. A. *et al.* (2012) 'Multiple leptospiral sphingomyelinases (or are there?).', *Microbiology (Reading, England)*. Microbiology Society, 158(Pt 5), pp. 1137–46. doi: 10.1099/mic.0.057737-0.

Nascimento, A. L. T. O. *et al.* (2004) 'Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis.', *Journal of bacteriology*. American Society for Microbiology, 186(7), pp. 2164–72. doi: 10.1128/JB.186.7.2164-2172.2004.

Natarajaseenivasan, K. *et al.* (2004) 'Human leptospirosis in Erode, South India: serology, isolation, and characterization of the isolates by randomly amplified polymorphic DNA (RAPD) fingerprinting.', *Japanese journal of infectious diseases*, 57(5), pp. 193–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15507774 (Accessed: 29 May 2019).

National Guidelines Diagnosis, Case Management Prevention and Control of Leptospirosis Programme for Prevention and Control of Leptospirosis (2015). Available at: https://ncdc.gov.in/WriteReadData/l892s/File558.pdf (Accessed: 26 April 2019).

Nicholas, KB *et al.* (1997) 'GeneDoc: a tool for editing and annotating multiple sequence alignments'. Available at: https://www.scienceopen.com/document?vid=c8a87cd1-255f-4129-802b-d2382bb0fb37 (Accessed: 2 January 2019).

Parkhill, J. (2002) 'The importance of complete genome sequences.', *Trends in microbiology*, 10(5), pp. 219–20; author reply 220. doi: 10.1016/s0966-842x(02)02353-3.

Pérez-Losada, M. *et al.* (2006) 'Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data', *Infection, Genetics and Evolution*, 6(2), pp. 97–112. doi: 10.1016/j.meegid.2005.02.003.

Pérez-Losada, M. *et al.* (2013) 'Pathogen typing in the genomics era: MLST and the future of molecular epidemiology.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 16, pp. 38–53. doi: 10.1016/j.meegid.2013.01.009.

Perolat, P. *et al.* (1993) 'Diversity of ribosomal DNA fingerprints of Leptospira serovars provides a database for subtyping and species assignation', *Research in Microbiology*, 144(1), pp. 5–15. doi: 10.1016/0923-2508(93)90210-S.

Petersen, T. N. *et al.* (2011) 'SignalP 4.0: discriminating signal peptides from transmembrane regions', *Nature Methods*, 8(10), pp. 785–786. doi: 10.1038/nmeth.1701.

Picardeau, M. *et al.* (2008) 'Genome sequence of the saprophyte Leptospira biflexa provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis.', *PloS one*, 3(2), p. e1607. doi: 10.1371/journal.pone.0001607.

Priya, C. G. *et al.* (2003) 'Identification and evaluation of LPS antigen for serodiagnosis of uveitis associated with leptospirosis', *Journal of Medical Microbiology*, 52, pp. 667–673. doi: 10.1099/jmm.0.05120-0.

Qin, J.-H. *et al.* (2008) 'Identification of a Novel Prophage-Like Gene Cluster Actively Expressed in Both Virulent and Avirulent Strains of Leptospira interrogans Serovar Lai', *Infection and Immunity*, 76(6), pp. 2411–2419. doi: 10.1128/IAI.01730-07.

Quail, M. A. *et al.* (2008) 'A large genome center's improvements to the Illumina sequencing system.', *Nature methods*. Nature Publishing Group, 5(12), pp. 1005–10. doi: 10.1038/nmeth.1270.

Rahman, O. *et al.* (2008) 'Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria', *World Journal of Microbiology and Biotechnology*. Springer Netherlands, 24(11), pp. 2377–2382. doi: 10.1007/s11274-008-9795-2.

Raja, V. and Natarajaseenivasan, K. (2013) 'Pathogenic, diagnostic and vaccine potential of leptospiral outer membrane proteins (OMPs).', *Critical reviews in microbiology*, 7828, pp. 1–17. doi: 10.3109/1040841X.2013.787387.

Rajapakse, S. *et al.* (2015) 'Current immunological and molecular tools for leptospirosis: diagnostics, vaccine design, and biomarkers for predicting severity.', *Annals of clinical microbiology and antimicrobials*, 14(1), p. 2. doi: 10.1186/s12941-014-0060-2.

Ralph, D. *et al.* (1993) 'Leptospira species categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes.', *Journal of bacteriology*, 175(4), pp. 973–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8094390 (Accessed: 5 April 2018).

Ramadass, P. *et al.* (1992) 'Genetic Characterization of Pathogenic Leptospira Species by DNA Hybridization', *International Journal of Systematic Bacteriology*, 42(2), pp. 215–219. doi:

10.1099/00207713-42-2-215.

Ramakrishnan, R. *et al.* (2003) 'An institutional outbreak of leptospirosis in Chennai, South India.', *The Journal of communicable diseases*, 35(1), pp. 1–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15239298 (Accessed: 24 April 2019).

Rathinam, S. R. (2002) 'Ocular leptospirosis.', *Current opinion in ophthalmology*, 13(6), pp. 381–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12441841 (Accessed: 26 April 2019).

Ren, S.-X. *et al.* (2003) 'Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing.', *Nature*, 422(6934), pp. 888–93. doi: 10.1038/nature01597.

Ricaldi, J. N. *et al.* (2012) 'Whole Genome Analysis of Leptospira licerasiae Provides Insight into Leptospiral Evolution and Pathogenicity', *PLoS Neglected Tropical Diseases*. Edited by M. Picardeau, 6(10), p. e1853. doi: 10.1371/journal.pntd.0001853.

Romero, E. C., Blanco, R. M. and Galloway, R. L. (2011) 'Analysis of Multilocus Sequence Typing for Identification of Leptospira Isolates in Brazil', *Journal of Clinical Microbiology*, pp. 3940–3942. doi: 10.1128/JCM.01119-11.

Romero, E. C. and Yasuda, P. H. (2006) 'Molecular characterization of Leptospira sp. strains isolated from human subjects in São Paulo, Brazil using a polymerase chain reaction-based assay: a public health tool.', *Memorias do Instituto Oswaldo Cruz*, 101(4), pp. 373–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16951806 (Accessed: 5 April 2018).

Routray, A. *et al.* (2018) 'Leptospirosis-A Review on its Zoonosis and Related Aspects', *International Journal of Livestock Research*, 36(11), pp. 29–37. doi: 10.5455/ijlr.20171020111742.

Salaun, L. *et al.* (2006) 'Application of Multilocus Variable-Number Tandem-Repeat Analysis for Molecular Typing of the Agent of Leptospirosis', *Journal of Clinical Microbiology*, 44(11), pp. 3954–3962. doi: 10.1128/JCM.00336-06.

Sehgal, S. C. *et al.* (1999) 'LEPTO Dipstick: a rapid and simple method for serodiagnosis of acute leptospirosis', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 93(2), pp. 161–164. doi: 10.1016/S0035-9203(99)90293-6.

Sehgal, S. C. (2006) 'Epidemiological patterns of leptospirosis.', *Indian journal of medical microbiology*, 24(4), pp. 310–1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17185863 (Accessed: 3 May 2018).

Sethi, S. *et al.* (2010) 'Increasing trends of leptospirosis in northern India: a clinico-epidemiological

study.', *PLoS neglected tropical diseases*. Public Library of Science, 4(1), p. e579. doi: 10.1371/journal.pntd.0000579.

Sharma, K. *et al.* (2011) 'Leptospirosis-induced still birth and postpartum sepsis', *Indian Journal of Pathology and Microbiology*, 54(2), p. 426. doi: 10.4103/0377-4929.81617.

Shivakumar, S. (2008) 'Leptospirosis-Current Scenario in India', in S K bichile (ed.) *Medicine Update*. 18th edn. www.apiindia.org. Available at: http://www.apiindia.org/pdf/medicine_update_2008/chapter_106.pdf (Accessed: 26 April 2019).

Silva, J. J. P. *et al.* (2002) 'Clinicopathological and immunohistochemical features of the severe pulmonary form of leptospirosis.', *Revista da Sociedade Brasileira de Medicina Tropical*, 35(4), pp. 395–9. doi: 10.1590/s0037-86822002000400017.

Skilbeck, N. W. and Chappel, R. J. (1987) 'Immunogold silver staining for visualization of leptospires in histologic sections.', *Journal of clinical microbiology*, 25(1), pp. 85–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2432082 (Accessed: 24 September 2019).

Smythe, L. *et al.* (2013) 'Classification of Leptospira genomospecies 1, 3, 4 and 5 as Leptospira alstonii sp. nov., Leptospira vanthielii sp. nov., Leptospira terpstrae sp. nov. and Leptospira yanagawae sp. nov., respectively', *International Journal of Systematic and Evolutionary Microbiology*, 63(PART 5), pp. 1859–1862. doi: 10.1099/ijs.0.047324-0.

Smythe, L. D. *et al.* (2009) 'Short Report: The Microscopic Agglutination Test (MAT) Is an Unreliable Predictor of Infecting Leptospira Serovar in Thailand', *Am. J. Trop. Med. Hyg*, 81(4), pp. 695–697. doi: 10.4269/ajtmh.2009.09-0252.

Sridhar, V. *et al.* (2008) 'Diagnostic potential of an iron-regulated hemin-binding protein HbpA that is widely conserved in Leptospira interrogans', *Infection, Genetics and Evolution*. Elsevier, 8(6), pp. 772–776. doi: 10.1016/J.MEEGID.2008.07.004.

Sritharan, M. (2012) 'Insights into Leptospirosis, a Neglected Disease', in Lorenzo-Morales, J. (ed.) *Zoonosis*. Rijeka: IntechOpen. doi: 10.5772/38838.

Stewart, P. E. *et al.* (2012) 'Characterization of the Bat proteins in the oxidative stress response of Leptospira biflexa', *BMC Microbiology*. BioMed Central, 12(1), p. 290. doi: 10.1186/1471-2180-12-290.

Stuart M. Brown (no date) *Next-Generation DNA Sequencing Informatics, Second Edition*. Available at: https://www.cshlpress.com/default.tpl?cart=153868150558613794&fromlink=T&linkaction=full&linksortby=oop_title&--eqSKUdatarq=1041 (Accessed: 5 October 2018).

Tamura, K. *et al.* (2011) 'MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.', *Molecular biology and evolution*, 28(10), pp. 2731–9. doi: 10.1093/molbev/msr121.

Tatusova, T. *et al.* (2016) 'NCBI prokaryotic genome annotation pipeline', *Nucleic Acids Research*. Oxford University Press, 44(14), pp. 6614–6624. doi: 10.1093/nar/gkw569.

Terpstra, W. J. *et al.* (1985) 'The classification of Sejroe group serovars of Leptospira interrogans with monoclonal antibodies.', *Zentralblatt für Bakteriologie, Mikrobiologie, und Hygiene. Series A, Medical microbiology, infectious diseases, virology, parasitology*, 259(4), pp. 498–506. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3901617 (Accessed: 30 July 2015).

Tettelin, Hervé *et al.* (2005) 'Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome"', *Proceedings of the National Academy of Sciences*, 102(39), pp. 13950–13955. doi: 10.1073/pnas.0506758102.

Tettelin, H *et al.* (2005) 'Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome" (vol 102, pg 13950, 2005)', *Proceedings of the National Academy of Sciences of the United States of America*, 102(45), p. 16530. doi: DOI 10.1073/pnas.0508532102.

Tettelin, H. *et al.* (2008) 'Comparative genomics: the bacterial pan-genome', *Curr Opin Microbiol*, 11(5), pp. 472–477. doi: DOI 10.1016/j.mib.2008.09.006.

Thaipadungpanit, J. *et al.* (2007) 'A dominant clone of Leptospira interrogans associated with an outbreak of human leptospirosis in Thailand.', *PLoS neglected tropical diseases*. Edited by M. Picardeau. Public Library of Science, 1(1), p. e56. doi: 10.1371/journal.pntd.0000056.

Tiengrim, S. *et al.* (2009) 'The Microscopic Agglutination Test (MAT) Is an Unreliable Predictor of Infecting Leptospira Serovar in Thailand', *The American Journal of Tropical Medicine and Hygiene*, 81(4), pp. 695–697. doi: 10.4269/ajtmh.2009.09-0252.

Tulsiani, S. M. *et al.* (2010) 'High-resolution melt-curve analysis of random-amplified-polymorphic-DNA markers, for the characterisation of pathogenic *Leptospira*', *Annals of Tropical Medicine & Parasitology*, 104(2), pp. 151–161. doi: 10.1179/136485910X12607012374037.

Urwin, R. and Maiden, M. C. J. J. (2003) 'Multi-locus sequence typing: a tool for global epidemiology.', *Trends in Microbiology*, 11(10), pp. 479–87. doi: 10.1016/j.tim.2003.08.006.

Varni, V. *et al.* (2014) 'Reassessment of MLST schemes for Leptospira spp. typing worldwide', *Infection, Genetics and Evolution*. Elsevier B.V., 22, pp. 216–222. doi: 10.1016/j.meegid.2013.08.002.

Venkatesha, M. D. and Ramadass, P. (2001) 'Identification of leptospiral isolates by bacterial restriction endonuclease analysis (Brenda).', *Indian journal of medical microbiology*. Medknow Publications, 19(2), pp. 26–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17664802 (Accessed: 28 May 2019).

Verma, A., Stevenson, B. and Adler, B. (2013) 'Leptospirosis in horses', *Veterinary Microbiology*, 167(1–2), pp. 61–66. doi: 10.1016/j.vetmic.2013.04.012.

Victoria, B. *et al.* (2008) 'Conservation of the S10-spc-alpha locus within otherwise highly plastic genomes provides phylogenetic insight into the genus Leptospira.', *PloS one*. Public Library of Science, 3(7), p. e2752. doi: 10.1371/journal.pone.0002752.

Vijayachari, P. *et al.* (2004) 'Use of fluorescent amplified fragment length polymorphism for molecular epidemiology of leptospirosis in India.', *Journal of clinical microbiology*, 42(8), pp. 3575–80. doi: 10.1128/JCM.42.8.3575-3580.2004.

Vijayachari, P. (2007) *Leptospirosis Laboratory Manual*. Available at: https://apps.who.int/iris/bitstream/handle/10665/205429/B2147.pdf?sequence=1&isAllowed=y (Accessed: 29 April 2019).

Vijayachari, P. *et al.* (2015) 'Leptospirosis among the self-supporting convicts of Andaman Island during the 1920s - the first report on pulmonary haemorrhage in leptospirosis?', *Indian Journal of Medical Research*. Medknow Publications and Media Pvt. Ltd., 142(1), p. 11. doi: 10.4103/0971-5916.162087.

Vijayachari, P., Sugunan, A. P. and Shriram, A. N. (2008) 'Leptospirosis: an emerging global public health problem.', *Journal of biosciences*, 33(4), pp. 557–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19208981 (Accessed: 15 September 2019).

Vinetz, J. M. *et al.* (1996) 'Sporadic Urban Leptospirosis', *Annals of Internal Medicine*. American College of Physicians, 125(10), pp. 794–798. doi: 10.7326/0003-4819-125-10-199611150-00002.

Vinetz, J. M. (2001) 'Leptospirosis.', *Current opinion in infectious diseases*, 14(5), pp. 527–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11964872 (Accessed: 24 September 2019).

Wild, C. J. *et al.* (2002) 'An improved immunohistochemical diagnostic technique for canine leptospirosis using antileptospiral antibodies on renal tissue.', *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*. NIH Public Access, 14(1), pp. 20–4. doi: 10.1177/104063870201400105.

Wilson, J. W. (2002) 'Mechanisms of bacterial pathogenicity', *Postgraduate Medical Journal*, 78(918), pp.

216–224. doi: 10.1136/pmj.78.918.216.

Woese, C. R. (1987) 'Bacterial evolution.', *Microbiological reviews*. American Society for Microbiology (ASM), 51(2), pp. 221–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2439888 (Accessed: 4 April 2018).

World Health Organisation (2003) *Human Leptospirosis: Guidance for diagnosis, surveillance and control*, *Geneva, Switzerland*.

Xue, F. *et al.* (2010) 'Transcriptional Responses of Leptospira interrogans to Host Innate Immunity: Significant Changes in Metabolism, Oxygen Tolerance, and Outer Membrane', *PLoS Neglected Tropical Diseases*. Edited by S. J. Peacock. Public Library of Science, 4(10), p. e857. doi: 10.1371/journal.pntd.0000857.

Yang, C.-W. *et al.* (2006) 'Toll-like receptor 2 mediates early inflammation by leptospiral outer membrane proteins in proximal tubule cells', *Kidney International*, 69(5), pp. 815–822. doi: 10.1038/sj.ki.5000119.

Yasuda, P. H. *et al.* (1987) 'Deoxyribonucleic Acid Relatedness between Serogroups and Serovars in the Family Leptospiraceae with Proposals for Seven New Leptospira Species', *International Journal of Systematic Bacteriology*, 37(4), pp. 407–415. doi: 10.1099/00207713-37-4-407.

Zamocky, M., Furtmüller, P. G. and Obinger, C. (2008) 'Evolution of catalases from bacteria to humans', *Antioxidants and Redox Signaling*, pp. 1527–1547. doi: 10.1089/ars.2008.2046.

Zerbino, D. R. and Birney, E. (2008) 'Velvet: algorithms for de novo short read assembly using de Bruijn graphs.', *Genome research*, 18(5), pp. 821–9. doi: 10.1101/gr.074492.107.

Zhao, Y. *et al.* (2014) 'PanGP: a tool for quickly analyzing bacterial pan-genome profile.', *Bioinformatics (Oxford, England)*, 30(9), pp. 1297–9. doi: 10.1093/bioinformatics/btu017.

Zuerner, R. *et al.* (2000) 'Technological advances in the molecular biology of Leptospira.', *Journal of molecular microbiology and biotechnology*, 2(4), pp. 455–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11075918 (Accessed: 24 July 2014).

Zuerner, R. L. and Bolin, C. A. (1997) 'Differentiation of Leptospira interrogans isolates by IS1500 hybridization and PCR assays.', *Journal of clinical microbiology*. American Society for Microbiology, 35(10), pp. 2612–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9316917 (Accessed: 5 April 2018).

OPEN ACCESS Freely available online

PLoS one

Genetic Affinities within a Large Global Collection of Pathogenic *Leptospira*: Implications for Strain Identification and Molecular Epidemiology

Kishore Nalam[1☯], Ahmed Ahmed[2☯], Sundru Manjulata Devi[3☯], Paolo Francalacci[4], Mumtaz Baig[5], Leonardo A. Sechi[6], Rudy A. Hartskeerl[2], Niyaz Ahmed[1]*

1 Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad, India, **2** WHO/FAO/OIE and National Collaborating Centre for Reference and Research on Leptospirosis, Department of Biomedical Research, Royal Tropical Institute (KIT), Amsterdam, The Netherlands, **3** Central Food Technological Research Institute, Mysore, India, **4** Department of Zoology and Evolutionary Genetics, University of Sassari, Sassari, Italy, **5** Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India, **6** Department of Biomedical Sciences, University of Sassari, Sassari, Italy

Abstract

Leptospirosis is an important zoonosis with widespread human health implications. The non-availability of accurate identification methods for the individualization of different *Leptospira* for outbreak investigations poses bountiful problems in the disease control arena. We harnessed fluorescent amplified fragment length polymorphism analysis (FAFLP) for *Leptospira* and investigated its utility in establishing genetic relationships among 271 isolates in the context of species level assignments of our global collection of isolates and strains obtained from a diverse array of hosts. In addition, this method was compared to an in-house multilocus sequence typing (MLST) method based on polymorphisms in three housekeeping genes, the *rrs* locus and two envelope proteins. Phylogenetic relationships were deduced based on bifurcating Neighbor-joining trees as well as median joining network analyses integrating both the FAFLP data and MLST based haplotypes. The phylogenetic relationships were also reproduced through Bayesian analysis of the multilocus sequence polymorphisms. We found FAFLP to be an important method for outbreak investigation and for clustering of isolates based on their geographical descent rather than by genome species types. The FAFLP method was, however, not able to convey much taxonomical utility sufficient to replace the highly tedious serotyping procedures in vogue. MLST, on the other hand, was found to be highly robust and efficient in identifying ancestral relationships and segregating the outbreak associated strains or otherwise according to their genome species status and, therefore, could unambiguously be applied for investigating phylogenetics of *Leptospira* in the context of taxonomy as well as gene flow. For instance, MLST was more efficient, as compared to FAFLP method, in clustering strains from the Andaman island of India, with their counterparts from mainland India and Sri Lanka, implying that such strains share genetic relationships and that leptospiral strains might be frequently circulating between the islands and the mainland.

Citation: Nalam K, Ahmed A, Devi SM, Francalacci P, Baig M, et al. (2010) Genetic Affinities within a Large Global Collection of Pathogenic *Leptospira*: Implications for Strain Identification and Molecular Epidemiology. PLoS ONE 5(8): e12637. doi:10.1371/journal.pone.0012637

Editor: Ramy K. Aziz, Cairo University, Egypt

Received July 21, 2010; **Accepted** August 15, 2010; **Published** August 27, 2010

Funding: This study did not receive any designated funding and the work was carried out from the routine laboratory supplies and resources available to the investigators in the form of a core support. Kishore Nalam would like to acknowledge the Junior Research Fellowship of the Council for Scientific and Industrial Research (CSIR) of India. Rudy Hartskeerl would like to acknowledge support from the EC grant ICA4-CT-2001-10086. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Niyaz Ahmed, Leonardo Sechi and Rudy Hartskeerl are members of the PLoS ONE Editorial Board. Niyaz Ahmed is a member of the PLoS International Advisory Group.

* E-mail: niyazSL@uohyd.ernet.in

☯ These authors contributed equally to this work.

Introduction

Leptospirosis caused by the pathogenic spirochetes of the genus *Leptospira* is the most widespread zoonosis in the world [1–5] where the number of severe cases probably exceeds 500,000 per year. Case-fatality rates are >10% and >50% in patients who develop acute hepato-renal failure or pulmonary hemorrhage syndrome, respectively. Pathogenic *Leptospira* consist of about 300 distinct antigenic types referred to as serovars, which vary with their carrier animal species [5–7]. *Leptospira* are maintained in the genital tract and renal tubules of wild and domestic animals and are excreted with urine into the environment where they can survive for several months depending on favorable conditions such as warm, humid environment with a neutral to slightly alkaline pH [2,5,7]. Infection of accidental hosts occurs by direct contact with the infected animals or their urine or indirectly via urine-contaminated environment. Accidental hosts develop clinical manifestations with a varying degree of severity and potentially leading to death [4]. *Leptospira* is a genus within the order Spirochaetales, an early branch in eubacterial evolution that, as a group, has unusual patterns of genetic organization. Analyses based on DNA composition have identified 20 *Leptospira* species with seven pathogenic species, which are *L. interrogans*, *L. borgpetersenii*, *L. santarosai*, *L. noguchi*, *L. weilii*, *L. kirschneri* and *L. alexanderi* comprising the main agents of leptospirosis [1,2,6]. The pathogenic *Leptospira* spp. form a common branch in evolution,

distinct from saprophytic *Leptospira* [8]. Recent reports identified an increasing intermediate group of *Leptospira* isolated from animals and humans with no or mild clinical symptoms [8–12]. The significance of this intermediate group in leptospirosis is yet unknown.

Genome sequencing has revealed a high-level plasticity of *Leptospira* genomes [13,14]. It has been proposed that *Leptospira* had a common progenitor with a genome resembling to that of *L. biflexa*. Mammalian infection potential could be associated with the acquisition of genes [15] expanding *Leptospira*'s capacity to survive host-determined environmental conditions while subsequent genome reduction increased host dependence.

Considering its unusual high antigenic and genetic flexibility, the genus *Leptospira* presents an extremely important research model for the understanding of pathogen evolution. However, focused *Leptospira* evolution research is scarce up to date.

The enormous repertoire of *Leptospira* serovars is mainly based on ever-changing surface antigens, notably the LPS. This presents an unreliable scenario of strain diversity and makes the serological approach difficult to track strains whose molecular identity keeps changing according to the host and environmental niches they inhabit and cross through. Multilocus sequence typing (MLST) [16,17], fluorescent amplified fragment length polymorphism (FAFLP) [18] and multilocus variable number of tandem repeats analysis (MLVA) [19] are the first genome sequence based molecular approaches having already established promise in unraveling *Leptospira* phylogeny, albeit in studies on limited strain panels or strains with restricted geographic prevalence. These methods have their advantages and disadvantages: MLST makes use of sequences deduced from PCR amplified DNA segments and thus depends on the success of amplification, which in turn depends on the annealing efficiency of the PCR primers. Sequence drift between *Leptospira* species will thus limit the applicability of MLST, particularly to the strains that fall in genetically distant branches. Amplification in FAFLP does not depend on the bacterial sequence composition and thus has a wide applicability. The drawback is that FAFLP requires high quality reagents and purified, concentrated genomic DNA. MLVA methods generally do not expand beyond *L. interrogans* or have limited flexibility to extend to all pathogenic and non-pathogenic species [20]. Given these issues, it would be relevant to test these methods in conjunction on a defined, global collection of strains and to see how they complement and supplement each other.

In the present study, we describe the genetic affinities and ancestral origins among the members of a strong 271 strains collection representing global dispersal and corresponding to a diverse array of hosts (Table S2). We applied both MLST and FAFLP with a focus on all pathogenic species. We further dissected diversity and composition of *L. interrogans* (being the largest subgroup within our collection), by a fluorescent MLVA technique. In addition, we studied genetic linkages among strains obtained from geographically close regions; such as the gene flow among *L. interrogans* within the Indian sub-continent.

Materials and Methods

Bacterial strains and genomic DNA samples

We included 271 *Leptospira* strains and isolates in the phylogenetic study. The strains and their sources are listed in supplementary information, tables S1 and S2. All the strains were cultured by the WHO reference laboratory at the KIT Biomedical Research Centre at The Royal Tropical Institute, Amsterdam, The Netherlands and at the Veterinary Sciences Division (VSD), The Queen's University of Belfast, United Kingdom and the WHO reference centre at Port Blair, India. The bacterial isolates were obtained over the last few years as a part of routine diagnostic/epidemiological investigations and they do not correspond to any cohorts or recruited patients/individuals (supplementary information, tables S1 and S2). Hence, they did not require any consents or ethics approval. Even then, the Institutional Biosafety Committee of the University of Hyderabad approved the study protocols. The study also has approvals from the Institutional Review Boards of all the participating institutions. *Leptospira* were grown to late log phase, harvested by centrifugation and genomic DNA was extracted using a QIAamp DNA mini kit (Qiagen, Germany) following the manufacturer's instructions.

Gene loci, nucleotide sequences and data access

The rational for the choice of the 6 candidate gene loci, their coordinates and amplification conditions etc. as needed for the design, testing and validation of the MLST scheme have all been detailed previously by us [8,16]. The *secY* gene and its resolution power in comparison with other signature loci such as *rrs* have been already determined in a previous study from our extended group [8]. The relevant sequence records are available via GenBank accession numbers EU365895-EU365966 and EU357938–EU358070. For other gene loci, prototype sequences needed for the design of MLST were obtained from the genomes of *L. interrogans* serovar Lai (NC_004342 and NC_004343) and *L. interrogans* serovar Copenhageni (NC_005823 and NC_005824) The multi locus sequences of all the 271 isolates obtained as a part of this study are available in full under supporting information (Table S2).

FAFLP method and phylogenetic analysis

Whole genome fingerprinting based on FAFLP genotyping was performed as described previously [18]. Briefly, the profiling of whole genome micro-restriction fingerprints with *Eco*RI/*Mse*I enzymes using fluorescence tagged primer pairs *Eco*RI+A/*Mse*I+0 and *Eco*RI+G/*Mse*I+0 was performed for all the strains. The PCR amplified fragments for each of the strains were then subjected to electrophoretic separation on a 5% acrylamide gel on an ABI Prism automated DNA sequencer and scoring of the fluorescent markers was done using the same DNA analysis workstation (ABI Prism 3100 DNA sequencer). Cluster analysis of DNA profiles was conducted on the basis of fingerprint characteristics scored in the form of a binary table for the presence and absence of alleles within the bins generated for fragment sizing [18]. Phylogenetic tools within MEGA 3.0 were used to generate Neighbor-joining trees with bootstrapping as described earlier [21,22].

MLST method and phylogenetic analysis

Six 600 bp-long regions from six genes spread throughout the genome were amplified by PCR and sequenced exactly as described previously [3]. Sequencing was performed on the two strands, using the DNA sequencer (see above). PCR and direct sequencing were performed at least twice to determine and confirm the DNA sequences for each isolate. Consensus sequence for each of the samples was generated using Genedoc (version 2.6.002). Multiple alignments of sequenced nucleotides were carried out using Clustal X (version 1.81). Bifurcating Neighbor-joining trees were constructed in MEGA 3.0 using bootstrapping at 10000 bootstrap trials and through Kimura-2 parameter [21,22].

Network analysis based on FAFLP data

Network analysis using the program Network 4.5.0.0 (http://www.fluxus-engineering.com) was performed on MLST sequences

and on FAFLP data. In particular, the median-joining algorithm, which can handle large data sets and multistate characters, was used [23]. Because of a program limitation, that it cannot handle more than 1000 polymorphic sites at once, we performed the analysis separately on two exact halves of the concatenated product (comprising of the multilocus sequences). This partition was neither necessary for FAFLP data nor for the *L. interrogans* subsample, which presented reduced numbers of polymorphisms; consequently, these datasets were analyzed in a complete form.

Phylogenetic reconstruction by Bayesian Morkov Chain Monte Carlo (MCMC) approach using MLST data

The MLST data were subjected to Bayesian MCMC analysis using BEAST version 1.5.2 [24]. The most important feature of Bayesian MCMC analysis using BEAST is that it offers rooted phylogeny. While constructing the phylogeny we used relaxed molecular clock approach [25]. Both coalescent constant population size and Yule speciation tree prior were employed. Two independent runs for each model were achieved for 30000000 steps and sampled every 1000 steps. The first 100000 steps of each run were discarded as burn-in. The tree was annotated using the program TreeAnnotator v1.5.2 (http://tree.bio.ed.ac.uk/). Finally, the annotated tree thus obtained was viewed and saved using the program FigTree v1.2.2 (http://tree.bio.ed.ac.uk/).

Automated MLVA analysis of the *L. interrogans* subsample

Multilocus variable number of tandem repeats analysis (MLVA) was carried out essentially as described previously by Majed *et al.* [19] except that the method was adopted for automated sequencer(s) by incorporating fluorescent labels into the reverse primers corresponding to all the loci previously tested [19]. Samples were analyzed on an automated DNA sequencer (ABI Prism 3100) and allele calling/binning was performed in a binary format as described previously for the FAFLP analysis [18]. MLVA data were used in MEGA 3.0 to generate phylogenetic trees.

Results and Discussion

FAFLP and MLVA as applied to the leptospiral genomes

Whole genome micro restriction patterns as scanned by the FAFLP method have revealed quite complex and confusing genetic affinities among various *Leptospira* species (Figure 1). This was not surprising given the resolution power of the FAFLP technique [18]. Nevertheless, plausible ancestral associations were found in terms of co-clustering of *L. kirschneri* and *L. interrogans*; *L. borgpetersenii* with *L. santarosai* and *L. noguchi* with *L. kirschneri*. Further analysis of the bifurcating Neighbor-joining tree revealed broadly species-specific clusters although clade-switching by a few strains in almost each of the clusters was clearly evident with respect to their projected genomic DNA-based species status. Thus, more than one cluster were observed for *L. interrogans*, *L. kirschneri*, and *L. borgpetersenii*. Further, this splitting of the clusters was not in agreement with geographical descent or the host species types, although we believe that discreet genetic associations arising due to recombinational events could lead to such subclustering. We tested this phenomenon by sub classifying the *L. interrogans* group by MLVA analysis. The tree based on MLVA (Figure 2) revealed embranchment broadly conforming to the serotypical positions and corresponding to the serogroups Icterohaemorrhagiae, Djasiman, Autumnalis, Australis, Canicola, Sejroe, Pyrogenes, Hebdomadis, Pomona, and Grippotyphosa. Containing this great diversity of serogroups within just 2 clusters should be seen as an appreciable specificity of the FAFLP method, which could be tapped for the investigation of small, regional outbreaks as previously shown by some of us [18]. However, in our opinion, when we seek replacement of serotyping as a tool for individualization, this sensitivity is not sufficient and perhaps FAFLP is not capable of distinguishing between the outcomes of genetic recombination or erroneous serotyping. Employing therefore a method that uses housekeeping genes, such as MLST appears to be inevitable.

MLST analysis as applied to the *Leptospira* gene pool

MLST analysis of all the 271 isolates revealed a highly organized phylogenetic tree (Figure 3) with no split clusters or widespread clade switching as seen with FAFLP. Analysis with MEGA 3.0 (N-J) or Median Joining by Network did not change the tree topology or the composition of various branches. However, in order to reproduce and validate the associations in another robust manner and to know the extent of genetic affinities among various branches we performed Bayesian MCMC analysis (see the Materials and Methods section).

Both coalescent constant population size and Yule speciation tree prior yielded similar topology. The Bayesian inference (BI) tree (Figure 3) reveals that the global pathogenic *Leptospira* in current study split into twelve embranchment. These twelve embranchments show five distinct major clades, namely, A or A1, D, G or G1, J or J1, K and their minor (sub) clades designated as A2, A3, A4, G2, J2 and J3. All these major and minor clades were well supported by 95% highest posterior density (HPD) intervals. Of note, the BI tree was rooted with *L. inadai*, whose intermediate nature in the *Leptospira* systematics is well supported by earlier studies [8,26]. In the BI tree constructed from 271 global isolates, the *L. interrogans* designated as clade A in Figure 3 comprises the largest collection of isolates and emerged as a tight cluster comprising of 123 isolates. Further, four visible sub-branching within the *L. interrogans* clade could perhaps suggest strain specific or host/environment specific genetic alterations. The other major clades found were *L. santarosai* and *L. kirschneri* designated as J and K respectively. Interestingly, the clade containing *L. borgpetersenii* isolates split into one major clade designated as G1 and a minor clade, G2. Both these major and minor clades show clustering with *L. weilii* isolates. The clade D with sole inclusion of *L. inadai* comprises entirely of *L. noguchi* strains. The minor clades A2 and A4 include four and three *L. interrogans* isolates respectively and behave as outliers. Taking into account the genetic distance, the minor clade A4 shows affinities towards *L. kirschneri* and not towards *L. interrogans*. In terms of genetic distance, the position of a sole *L. interrogans* isolate depicted as branch A3 in the tree lies intermediate between the *L. interrogans* and *L.kirschneri*. Thus the minor clade A2 which harbors two isolates from rodents (one each from a bandicoot and a necked field mouse) seems to be the most recent plausible ancestor of the present day *L. interrogans* found in varied hosts. The role of rodents as the reservoir of *Leptospira* needs no explanation. Further, *L. noguchi* shows closer genetic affinity towards *L. kirschneri* than *L. interrogans*. The basal position of *L. noguchi* cluster in the vicinity of the *L. kirschneri* and *L. interrogans* clusters is suggestive of the ancestral nature of *L. noguchi* or in a simpler way, *L. kirschneri* and *L.interrogans* may have both originated from *L. noguchi* or *L. noguchi* like ancestor. In contrast to the less distinct evolution of *L. kirschneri* and *L. interrogans*, the evolution of *L. santarosai*, *L. borgpetersenii and L. weilii* is marked by more distinct speciation events along the intermediate *L. inadai* or *L. inadai* like ancestor. The minor clade G2, comprises of two *L. weilii* and one *L. borgpetersenii* and *L.interrogans* isolate each. From the basal position of *L. weilii* in the minor clade G2 and in the major clade

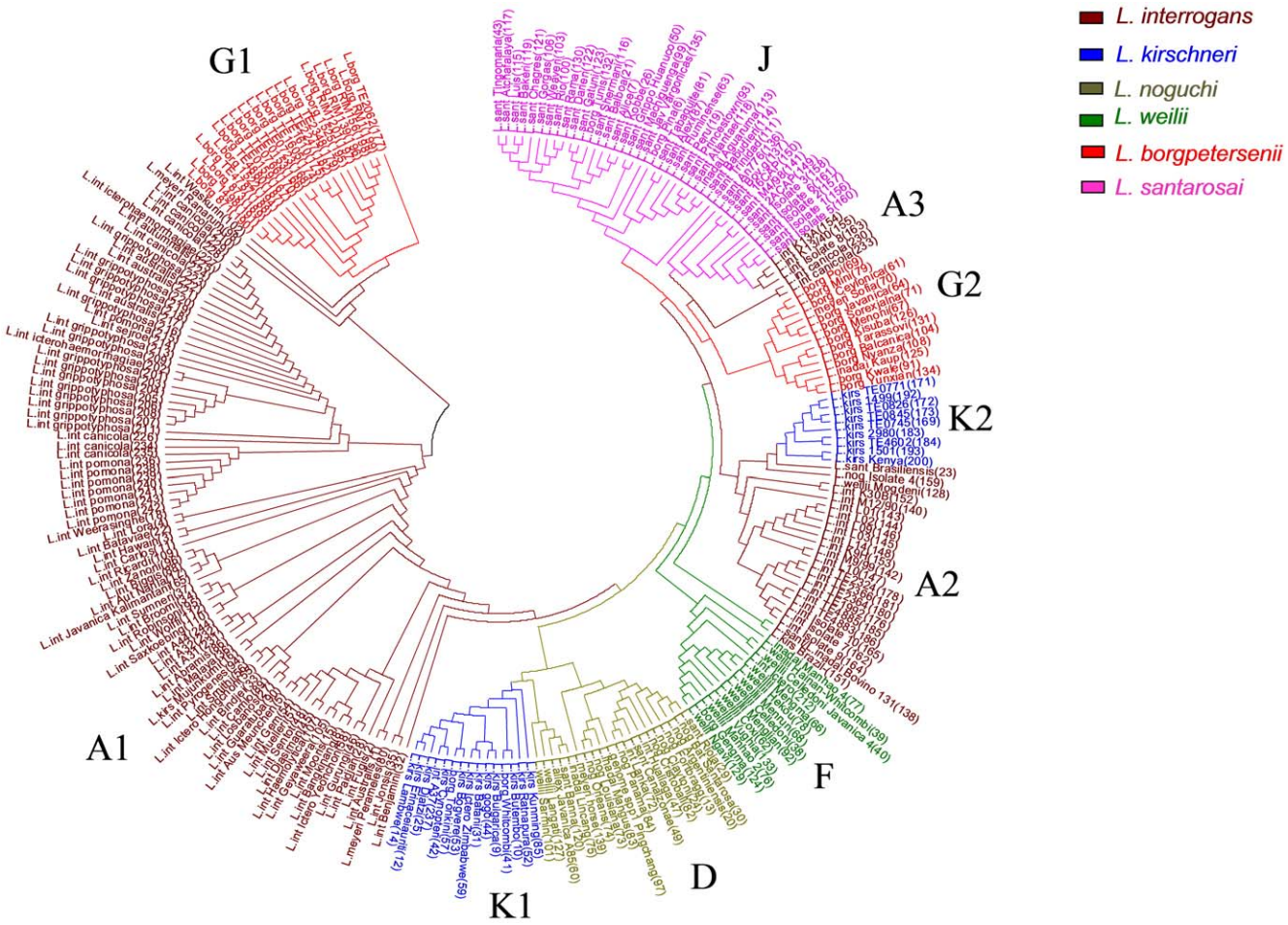

Figure 1. Genetic relatedness among *Leptospira* isolates based on FAFLP analysis. (See supplementary information, tables S1 and S2 for details.) Clades roughly corresponding to different species types have been marked in different colors. Major clades and minor clades have been identified by code names (A to K) which overlap with the designation of similar clades identified by MLST technique. Identities of individual isolates need not be comprehensible in the tree itself, but they can be read clearly in the supplementary Table S2. The phylogenetic tree was rendered and visualized by MEGA3.0 software.
doi:10.1371/journal.pone.0012637.g001

G1, it seems that *L. weilii* is the most recent common ancestor of the *L. santarosai* and *L. borgpetersenii* or *L. santarosai* and *L.borgpetersenii* originated from *L. weilii* or *L. weilii* like ancestor.

Apart from systematics, the inferred tree also reveals instances of unusual clustering. One of the isolates serologically typed as *L. interrogans* clustered with the *L. kirschneri* clade. Similarly, three *L. interrogans* serovars clustered with the *L. santarosai* clade. The tree also shows tight clustering of single *L. borgpetersenii and L. meyeri* isolates with the *L. interrogans* cluster. Also, two serologically classified *L. kirschneri* isolates clustered with the *L. interrogans* clade. Taking into consideration the genetic distances between these five major clades, clade switching seems to be feasible between *L. interrogans* and *L. kirschneri* and not between *L. interrogans*, *L. santarosai* and *L. borgpetersenii*. Such unusual clustering(s) could only be explained by either incorrect serological typing or clade switching due to horizontal gene transfer (HGT). A further mechanistic dissection of these isolates based on whole genome sequence analysis could shed some light on their genome evolution trends and the extent of HGT occurring within the *rfb* cluster that encodes much of the surface antigens responsible for determination of their serotype.

In summation of our analyses above (supplementary information, tables S1 and S2), it is possible to espouse that the *L. interrogans*, *L. kirschneri* and probably *L. borgpetersenii* are ubiquitous species. Within this, *L. interrogans* seems to be the most frequently isolated species from humans and is largely reported from South Asia. This trend perhaps reflects endemicity and the maintenance of this species in that region. Whereas *L. weilii* is largely confined to Asia, *L. santarosai* and *L. noguchi* are found to be adapted to the Americas. So far, this trend mostly holds for the isolates that have been received over the years for typing at the KIT, Amsterdam laboratory and certainly, this would have implications for understanding the transmission dynamics and evolution of the *Leptospira* or the (ubiquitous, or confined hosts) of the pathogenic species.

Genetic affinities within the *L. interrogans* subsample

Of the 271 rigorously sampled isolates comprising of our collection (Table S2), 134 were identified as *L. interrogans* (clade A, Figure 3). These 134 isolates were further sub classified based on MLVA to reveal that they were in fact diverse, belonging to various different serogroups, and representing different sub-ecotypes of the same species spread over the entire South Asian region (Figure 2). This was indirectly a proof that the collection did not represent convenient sampling. Further, most of these isolates

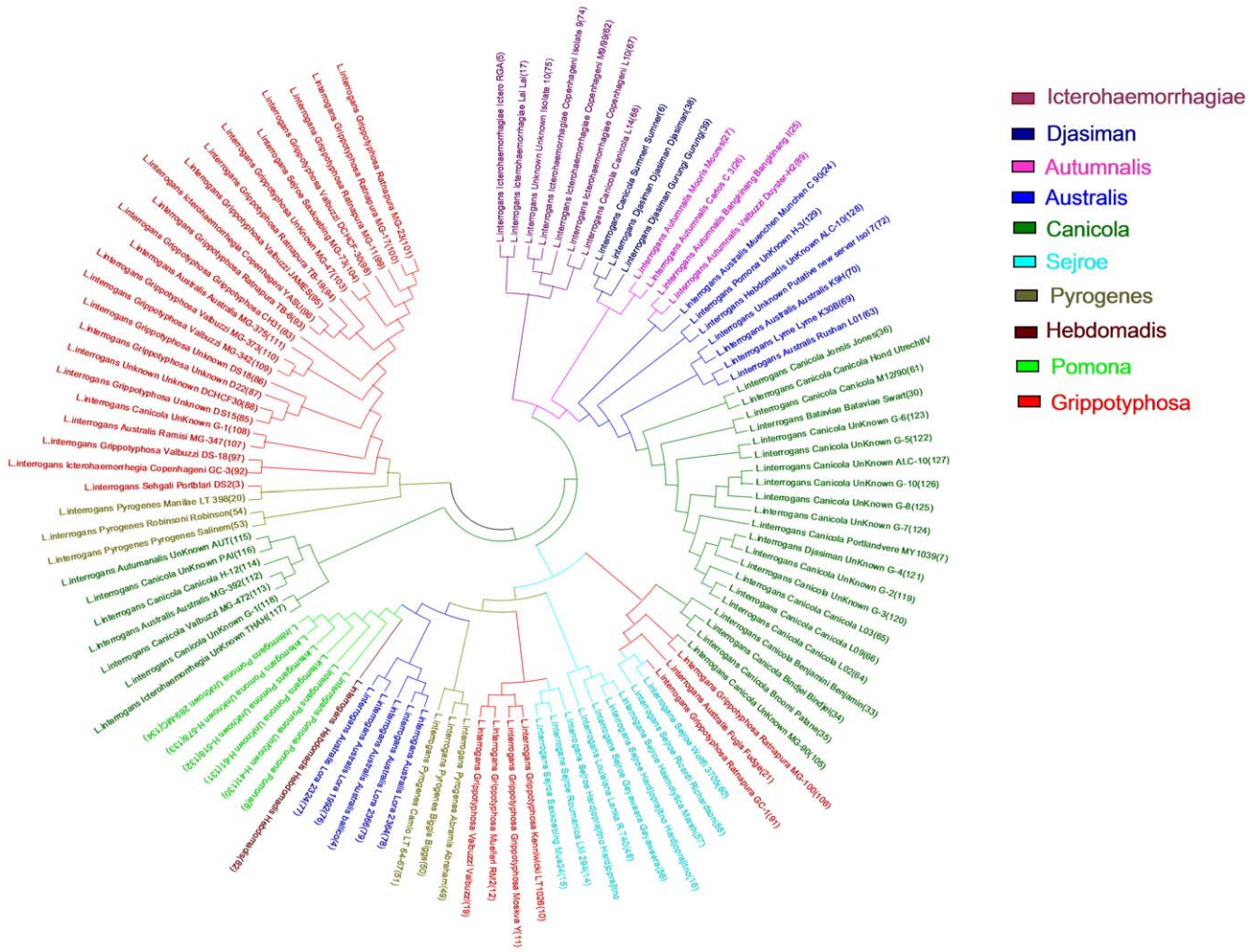

Figure 2. Genetic analysis of *L. interrogans* subsample based on MLVA analysis. Clusters corresponding to individual serogroups have been identified by different colors which have no correspondence with the color code of clusters shown in Figures 1 and 3. Please refer to supplementary information (Table S2) for details of the *L. interrogans* samples analyzed here. The phylogenetic tree was generated and visualized by using MEGA3.0 software.
doi:10.1371/journal.pone.0012637.g002

(33%) were cultured in India (many from the Andaman and Nicobar islands) and were all obtained from human clinical cases as a part of routine outbreak investigations, over a period. As indicated, a majority of *L. interrogans* from our collection belonged to South Asia and were isolated from human cases of leptospirosis. Despite being geographically distinct, they formed a tight cluster. This shows their possible clonal origins and perhaps recent dispersal within the South Asian countries with less opportunity to diversify or accumulate substitutions within the candidate gene loci. Therefore, such loci appear to be conserved and stable within the clade A. In fact, it is a desirable property of the loci included in a MLST scheme that they should be more static within a particular species [27]. Given the above, it would perhaps be tempting to espouse that the *L. interrogans* ss could have been the most dominant and fittest ecotype/species to cause disease in humans in this region. A majority of these isolates were from the Andaman and Nicobar islands, where, as previously determined, leptospirosis has been traditionally endemic and has caused fatal outbreaks [28]. Not all the three schemes, MLST, FAFLP and MLVA could split the *L. interrogans* cluster based on geographic or ecological basis and thus it would be difficult to highlight its routes of spread in and out of Andaman or out of other main lands. Nevertheless, a careful exploration of the Indian isolates, based on MLST data, when analyzed by Network, revealed that a few sequences from the region were divergent, possibly because of recombination. The enlargement of the core sequence network (Figure 4) revealed four sequences from the Andaman Islands in central position, posing as a possible source of the Indian *L. interrogans* variability with distinct, derived clusters corresponding to South India, Central India and Andaman islands. Given this, it is possible to think of an early spread to and from Andaman to mainland India and to other adjoining countries, possibly through rodents that travelled in vessels and ships to Andaman from India and *vice-versa*.

MLST as a gold standard for *Leptospira* strain typing

The six MLST loci selected and previously tested by us [8,16] on a limited set of isolates from present collection were suitable for strain individualization. These loci could be amplified and sequenced in all the isolates (irrespective of their taxonomic status) representing pathogenic as well as saprophytic species; nonetheless, they required greater standardization for non-pathogenic

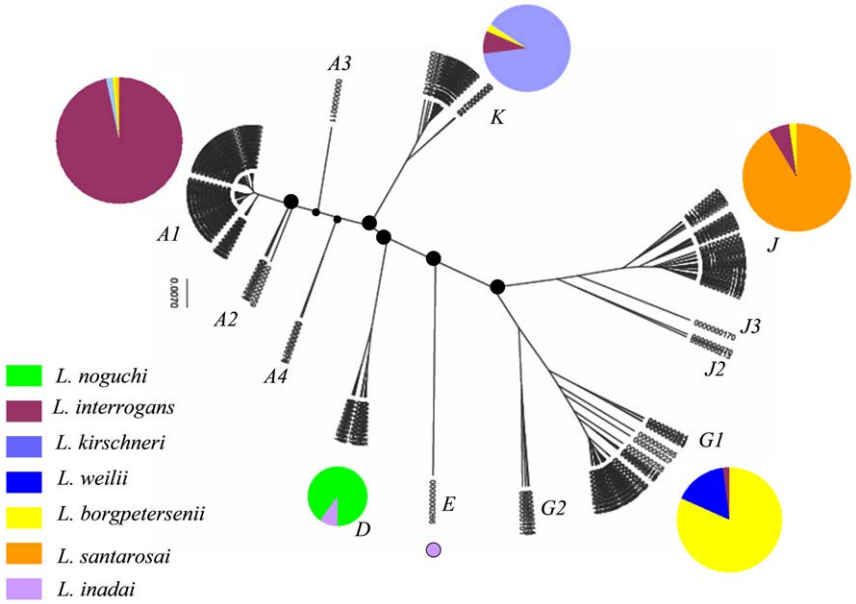

Figure 3. Genetic clustering and Bayesian inferred phylogeny of *Leptospira* isolates and strains based on MLST analysis. (See supplementary information, tables S1 and S2 for detailed information on all the isolates/strains analyzed.) Clades or embranchment corresponding to different species types have been marked in different colors. Major clades and minor clades have been identified by code names (A to K) which overlap with the designation of similar clades identified by FAFLP technique. Identities of individual isolates need not be comprehensible in the tree itself, but they can be read clearly in the supplementary Table S2. The analysis was performed using BEAST version 1.5.2, the tree was annotated using TreeAnnotator v1.5.2 and was visualized through FigTree v1.2.2.
doi:10.1371/journal.pone.0012637.g003

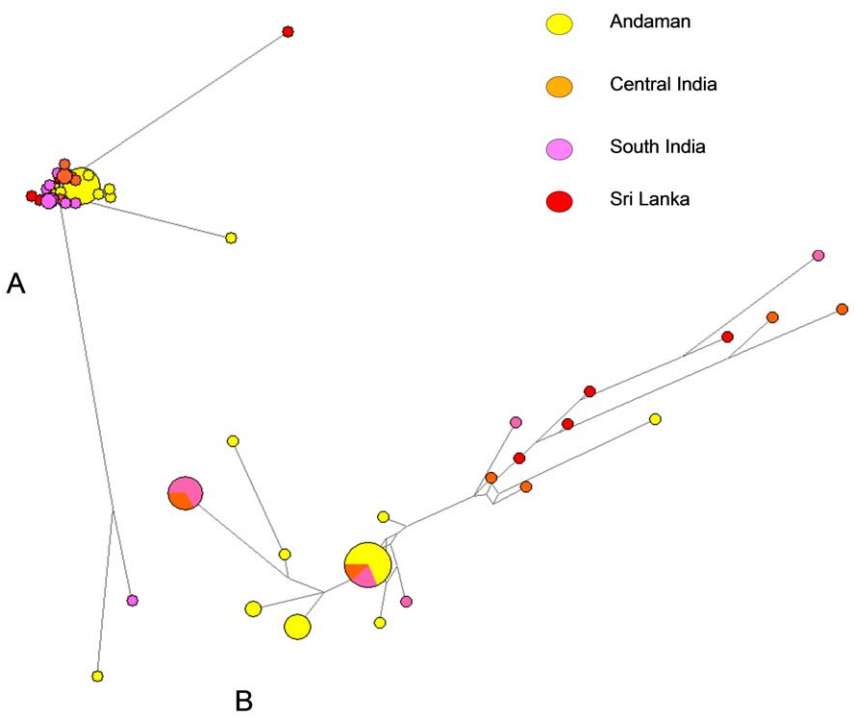

Figure 4. Median-joining network analysis of the MLST data obtained for the Indian (sub continent) subsample of *L. interrogans*. As evident from the trees A and B, a few sequences were highly divergent, possibly because of recombination. The enlargement of the core sequence network (B) revealed four sequences from the Andaman Islands in the central position, suggestive of a possible source of the Indian *L. interrogans* variability, followed by distinct derived clusters from South India, Central India and Andaman islands. Such a geographic structure is not apparent in the FAFLP network (Figure 1) of the Indian *L. interrogans*, confirming the reduced phylogenetic resolution of FAFLP in respect to MLST analysis. The color codes have no overlaps with those shown in Figures 1, 2 and 3. The tree/associations were deduced using Network 4.5.0.0 package.
doi:10.1371/journal.pone.0012637.g004

variants. Among the species, these loci exhibited a high degree of sequence diversity and resolution.

Several molecular tools that have so far been described for the diagnosis of *Leptospira* are associated with drawbacks, either in the form of technical complications or the difficulties of interpretation, portability and reproducibility. Some of the methods need live organisms or a very high purity and concentration of genomic DNA. Our MLST approach overcomes all these disadvantages as the technique is simple and requires an automated DNA sequencer that is more widely available in most of the laboratories and the sequence data generated is unambiguous and explicit. The main advantage of MLST is the transfer of data that can be shared and compared between different laboratories easily through the Internet. To date a large number of organisms have been typed by MLST, which proved to be a highly discriminatory technique [29–31]. MLST analysis of the *Leptospira* strains confirms earlier findings [26] that the serovars and the serogroups are not clustered together but according to the species. This method is more suitable in identifying the species of leptospires as indicated by the clustering patterns up to genome species level. Due to the greater sequence diversity observed in all the six genes except *rrs*2, the dendrogram generated could differentiate effectively the *L. interrogans* ss, *L. kirschneri*, *L. noguchi*, *L. weilii*, *L. santarosai* and *L. borgpetersenii*. Thus, our MLST technique and its analysis by Bayesian MCMC were capable of individualizing *Leptospira* up to species level with flexibility to type isolates with many different taxonomic identities as compared to another MLST scheme [17] which has been limited to outbreak investigation(s) over small epidemiological territories and could not type isolates beyond *L. interrogans* ss. Having said this, we should also consider the obvious limitation of MLST: its failure to resolve the horizontal variome [32], but this really depends on the extent and impact of HGT in different bacterial species. With this issue in mind, we already included targets other than the housekeeping genes, namely, the envelope proteins, LipL32 and LipL41 in our MLT scheme [8,16]; this may allow sampling of variation beyond the core genome and which might be relevant in epidemiologic/taxonomic resolution of the strains.

The future of *Leptospira* genotyping

With our extensive evaluation of the MLST technique and its comparison with FAFLP, we believe that the issues related to strain diversity as well as the taxonomic organization and accuracy of the reference collection(s) were set to rest in a best possible way. This will help understand population genetic structure of this pathogen with diverse host range and under different ecological conditions and will provide a scope for genotype-phenotype correlation to be established. Analyses based on the allelic profiles generated by MLST could be successfully used to gain insights into the evolution and phylogeographic affinities of leptospires as it has been done for many other organisms. Given the associations and affinities within our collection, it will be possible in the foreseeable future to develop a sophisticated database of the genomic profiles based on all the three typing techniques. Finally, our rigorous categorization of the ecotypes and genotypes herein may be seen as the first, needed step under the mandate of the post genomic profiling of *Leptospira* from different hosts [33]. This will help the leptospirosis community in planning for future whole genome sequencing [34] of Leptospirae or establishing their metagenome. Such approaches will be able to generate extremely valuable information in the form of diagnostic markers, vaccine candidates, and strain specific co-ordinates relevant in re-constructing the evolutionary history of the organisms emerging or reemerging in a particular epidemiological catchment area. This reality ultimately holds promise for strengthening the cause of 'functional molecular infection epidemiology (FMIE)' of *Leptospira*. FMIE is an emerging area of medical microbiology that entails correlation of genetic variation in a pathogen, with a unique function in corresponding host(s) related to disease severity, disease progression, or host susceptibility. This kind of functional epidemiology is likely to explain not only the genome level, descriptive, host-pathogen associations, but also the global juxtaposition of pathogen and host variations with a prospective impact on our understanding of pathogen/infection biology.

In conclusion, our integrated genotyping approach provides evidence that *Leptospira* represent a globally distributed zoonotic agent and their gene pool being diverse and somewhat geographically compartmentalized. In addition, *L. interrogans* appears to be the single, most prevalent *Leptospira* species, which inflicts rodents, livestock and humans in different continents but predominantly in the South Asian countries. While we found FAFLP to be an important method for outbreak investigation and for clustering of isolates based on their geographical descent rather than by genome species types, it was not able to convey much taxonomical utility sufficient to replace tedious serotyping procedures, currently in practice, worldwide. By contrast, MLST was observed to be highly robust and efficient in identifying ancestral relationships and segregating the outbreak associated strains according to their genome species status. We believe that this large-scale evaluation of different genotyping methods sets stage for the implementation of MLST as a highly sought-after replacement of serotyping, for accurate identification and classification of *Leptospira*.

Supporting Information

Table S1 Characteristics, distribution and phylogenetic affiliation of *Leptospira* isolates. Species short names *L. int*, *L. borg*, *L. kirsch*, *L. sant*, and *L. nog* refer to *L. interrogans ss*, *L. borgpetersenii*, *L. kirschneri*, *L. santarosai* and *L. noguchi*, respectively. Short names of serogroups, Gripp., Ict., Aust., and Cani. refer to Grippotyphosa, Icterohaemorrhagiae, Australis, and Canicola respectively.
Found at: doi:10.1371/journal.pone.0012637.s001 (0.15 MB DOC)

Table S2 Full details and multilocus sequences of the *Leptospira* isolates and strains used in this study.
Found at: doi:10.1371/journal.pone.0012637.s002 (0.39 MB XLS)

Acknowledgments

We acknowledge with thanks the contribution of various *Leptospira* typing and reference laboratories worldwide who have contributed their strains and DNA for our reference collection of the *Leptospira* organisms. We are thankful to P. Vijayachari for the DNA samples from Andaman Island and Seyed E. Hasnain for helpful discussions and suggestions. We acknowledge with thanks the technical and lab management support of G. Navamallika. Niyaz Ahmed is the Speaker (India) of *Internationales Graduiertenkolleg – Functional Molecular Infection Epidemiology* (GRK1673), established by the German Research Foundation. Niyaz Ahmed is also the Professor (Adjunct) of Molecular Biosciences at the University of Malaya, Kuala Lumpur.

Author Contributions

Conceived and designed the experiments: NA. Performed the experiments: KN AAA SMD. Analyzed the data: KN PF MB NA. Contributed reagents/materials/analysis tools: PF LAS RAH NA. Wrote the paper: NA. Corrected the draft of the manuscript: RAH.

References

1. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, et al. (2003) Leptospirosis: a zoonotic disease of global importance. Lancet Infect Dis 3: 757–771.
2. McBride AJ, Athanazio DA, Reis MG, Ko AI (2005) Leptospirosis. Curr Opin Infect Dis 18: 376–386.
3. Gouveia EL, Metcalfe J, de Carvalho AL, Aires TS, Villasboas-Bisneto JC, et al. (2008) Leptospirosis-associated Severe Pulmonary Hemorrhagic Syndrome, Salvador, Brazil. Emerg Infect Dis 14: 505–508.
4. Segura ER, Ganoza CA, Campos K, Ricaldi JN, Torres S, et al. (2005) Clinical spectrum of pulmonary involvement in leptospirosis in a region of endemicity, with quantification of leptospiral burden. Clin Infect Dis 40: 343–351.
5. Levett PN (2001) Leptospirosis. Clin Microbiol Rev 14: 296–326.
6. Kmety E, Dikken H (1993) Classification of the species *Leptospira interrogans* and history of its serovars. Groningen, The Netherlands: University Press. pp 17–22.
7. Faine SB, Adler B, Bolin C, Perolat P (1999) *Leptospira* and leptospirosis. Melbourne, Australia: MediSci.
8. Victoria B, Ahmed A, Zuerner RL, Ahmed N, Bulach DM, et al. (2008) Conservation of the *S10-spc-alpha* locus within otherwise highly plastic genomes provides phylogenetic insight into the genus *Leptospira*. PLoS One 3: e2752.
9. Levett PN, Morey RE, Galloway RL, Steigerwalt AG (2006) *Leptospira broomii* sp. nov., isolated from humans with leptospirosis. Int J Syst Evol Microbiol 56: 671–673.
10. Matthias MA, Ricaldi JN, Cespedes M, Diaz MM, Galloway RL, et al. (2008) Human leptospirosis caused by a new, antigenically unique *Leptospira* associated with a *Rattus* species reservoir in the peruvian Amazon. PLoS Negl Trop Dis 2: e213.
11. Perolat P, Chappel RJ, Adler B, Baranton G, Bulach DM, et al. (1998) *Leptospira fainei sp. nov.*, isolated from pigs in Australia. Int J Syst Bacteriol 48(Pt 3): 851–858.
12. Gravekamp C, Van de Kemp H, Franzen M, Carrington D, Schoone GJ, et al. (1993) Detection of seven species of pathogenic leptospires by PCR using two sets of primers. J Gen Microbiol 139: 1691–1700.
13. Nascimento AL, Ko AI, Martins EA, Monteiro-Vitorello CB, Ho PL, et al. (2004) Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. Journal of Bacteriology 186: 2164–2172.
14. Picardeau M, Bulach DM, Bouchier C, Zuerner RL, Zidane N, et al. (2008) Genome Sequence of the Saprophyte *Leptospira biflexa* Provides Insights into the Evolution of *Leptospira* and the Pathogenesis of Leptospirosis. PLoS ONE 3: e1607.
15. Ahmed N, Dobrindt U, Hacker J, Hasnain SE (2008) Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. Nature Rev Microbiol 6: 387–394.
16. Ahmed N, Devi SM, Valverde Mde L, Vijayachari P, Machang'u RS, et al. (2006) Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. Ann Clin Microbiol Antimicrob 5: 28.
17. Thaipadungpanit J, Wuthiekanun V, Chierakul W, Smythe LD, Petkanchanapong W, et al. (2007) A Dominant Clone of *Leptospira interrogans* Associated with an Outbreak of Human Leptospirosis in Thailand. PLoS Negl Trop Dis 1: e56.
18. Vijayachari P, Ahmed N, Sugunan AP, Ghousunnissa S, Rao KR, et al. (2004) Use of fluorescent amplified fragment length polymorphism for molecular epidemiology of leptospirosis in India. J Clin Microbiol 42: 3575–80.
19. Majed Z, Bellenger E, Postic D, Pourcel C, Baranton G, et al. (2005) Identification of variable-number tandem-repeat loci in Leptospira interrogans sensu stricto. J Clin Microbiol 43: 539–545.
20. Cerqueira GM, Picardeau M (2009) A century of *Leptospira* strain typing. Infect Genet Evol 9: 760–768.
21. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinfor 5: 150–163.
22. Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, et al. (2007) Ancestral European roots of Helicobacter pylori in India. BMC Genomics 8: 184.
23. Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations. Genetics 141: 743–753.24.
24. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.
25. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4: e88.
26. Brenner DJ, Kaufmann AF, Sulzer KR, Steigerwalt AG, Rogers FC, et al. (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. Int J Syst Bacteriol 49(Pt 2): 839–858.
27. Enright MC, Spratt BG (1999) Multilocus sequence typing. Trends Microbiol 7: 482–487.
28. Vijayachari P, Sugunan AP, Sharma S, Roy S, Natarajaseenivasan K, et al. (2008) Leptospirosis in the Andaman Islands, India. Trans R Soc Trop Med Hyg 102: 117–122.
29. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, et al. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95: 3140–3145.
30. Maiden MC (2006) Multilocus sequence typing of bacteria. Annu Rev Microbiol 60: 561–588.
31. Ibarz Pavon AB, Maiden MC (2009) Multilocus sequence typing. Methods Mol Biol 551: 129–140.
32. Aziz RK, Nizet V (2010) Pathogen microevolution in high resolution. Sci Transl Med 2(16): 16ps4.
33. Levett PN (2007) Sequence-based typing of *Leptospira*: epidemiology in the genomic era. PLoS Negl Trop Dis 1: e120.
34. Ahmed N (2009) A flood of microbial genomes-do we need more? PLoS One 4: e5831.

OPEN ACCESS Freely available online

PLoS NEGLECTED TROPICAL DISEASES

Comparison of Two Multilocus Sequence Based Genotyping Schemes for *Leptospira* Species

Ahmed Ahmed[1*☯], Janjira Thaipadungpanit[2,3☯], Siriphan Boonsilp[2], Vanaporn Wuthiekanun[2], Kishore Nalam[4], Brian G. Spratt[5], David M. Aanensen[5], Lee D. Smythe[6], Niyaz Ahmed[4,7], Edward J. Feil[8], Rudy A. Hartskeerl[1], Sharon J. Peacock[2,9,10]

1 WHO/FAO/OIE and National Collaborating Centre for Reference and Research on Leptospirosis, Department of Biomedical Research, Royal Tropical Institute (KIT), Amsterdam, The Netherlands, **2** Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, **3** Medical Proteomics Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Bangkok, Thailand, **4** Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad, India, **5** Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom, **6** WHO/FAO/OIE Collaborating Centre for Reference and Research on Leptospirosis, Centre for Public Health Sciences, Queensland Health Scientific Services, Brisbane, Australia, **7** Institute of Biological Sciences, Universiti Malaya, Kuala Lumpur, Malaysia, **8** Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom, **9** Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, **10** Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom

Abstract

Background: Several sequence based genotyping schemes have been developed for *Leptospira* spp. The objective of this study was to genotype a collection of clinical and reference isolates using the two most commonly used schemes and compare and contrast the results.

Methods and Findings: A total of 48 isolates consisting of *L. interrogans* (n = 40) and *L. kirschneri* (n = 8) were typed by the 7 locus MLST scheme described by Thaipadungpanit et al., and the 6 locus genotyping scheme described by Ahmed et al., (termed 7L and 6L, respectively). Two *L. interrogans* isolates were not typed using 6L because of a deletion of three nucleotides in *lipL32*. The remaining 46 isolates were resolved into 21 sequence types (STs) by 7L, and 30 genotypes by 6L. Overall nucleotide diversity (based on concatenated sequence) was 3.6% and 2.3% for 7L and 6L, respectively. The D value (discriminatory ability) of 7L and 6L were comparable, i.e. 92.0 (95% CI 87.5–96.5) vs. 93.5 (95% CI 88.6–98.4). The dN/dS ratios calculated for each locus indicated that none were under positive selection. Neighbor joining trees were reconstructed based on the concatenated sequences for each scheme. Both trees showed two distinct groups corresponding to *L. interrogans* and *L. kirschneri*, and both identified two clones containing 10 and 7 clinical isolates, respectively. There were six instances in which 6L split single STs as defined by 7L into closely related clusters. We noted two discrepancies between the trees in which the genetic relatedness between two pairs of strains were more closely related by 7L than by 6L.

Conclusions: This genetic analysis indicates that the two schemes are comparable. We discuss their practical advantages and disadvantages.

Citation: Ahmed A, Thaipadungpanit J, Boonsilp S, Wuthiekanun V, Nalam K, et al. (2011) Comparison of Two Multilocus Sequence Based Genotyping Schemes for *Leptospira* Species. PLoS Negl Trop Dis 5(11): e1374. doi:10.1371/journal.pntd.0001374

Editor: Mathieu Picardeau, Institut Pasteur, France

Received March 21, 2011; **Accepted** September 13, 2011; **Published** November 8, 2011

Funding: This study was funded by the Wellcome Trust. SJP is supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: a.ahmed@kit.nl

☯ These authors contributed equally to this work.

Introduction

Leptospirosis is a common zoonotic disease worldwide, with a particularly high prevalence in warm humid countries [1–4]. About 350,000 severe cases of leptospirosis are estimated to occur annually, with case fatality reports up to 50% [5–7]. Reported cases are likely to be a gross under-estimate of global incidence rates, the result of a combination of factors including lack of surveillance, diagnostics and notification in those countries with the highest disease burden. Leptospirosis is currently considered a globally re-emerging disease, with frequent outbreaks in South East Asia (including Thailand, India, The Philippines and Sri Lanka) as well as in Latin America [3,8–14]. International travel also leads to presentation of leptospirosis cases in settings where incidence is low and clinicians are unfamiliar with its clinical manifestations [7,15].

Identification and typing of *Leptospira* species plays an important role in understanding disease epidemiology and pathogenicity, together with the development of diagnostic tools, effective vaccines and preventive strategies. During the last three decades many molecular typing methods have been proposed for *Leptospira* spp. These include DNA-DNA hybridization analysis [16–19],

Author Summary

Two independent multilocus sequence based genotyping schemes (denoted here as 7L and 6L for schemes with 7 and 6 loci, respectively) are in use for *Leptospira* spp., which has led to uncertainty as to which should be adopted by the scientific community. The purpose of this study was to apply the two schemes to a single collection of pathogenic *Leptospira*, evaluate their performance, and describe the practical advantages and disadvantages of each scheme. We used a variety of phylogenetic approaches to compare the output data and found that the two schemes gave very similar results. 7L has the advantage that it is a conventional multi-locus sequencing typing (MLST) scheme based on housekeeping genes and is supported by a publically accessible database by which genotypes can be readily assigned as known or new sequence types by any investigator, but is currently only applicable to *L. interrogans* and *L. kirschneri*. Conversely, 6L can be applied to all pathogenic *Leptospira* spp., but is not a conventional MLST scheme by design and is not available online. 6L sequences from 271 strains have been released into the public domain, and phylogenetic analysis of new sequences using this scheme requires their download and offline analysis.

randomly amplified polymorphic DNA (RAPD) fingerprinting [20], arbitrarily primed PCR (AP-PCR) [21,22], pulsed field gel electrophoresis (PFGE) [23,24], restriction fragment length polymorphism (RFLP) analysis [25,26], bacterial typing methods based on insertion sequences (IS) [27], detection of variable number of tandem repeats (VNTR) [28,29], *rrs* sequencing [30–32], and sequencing of specific genes or gene fragments including *rpoB*, *gyrB*, *secY* and *ligB* [33–37].

Multilocus sequencing typing (MLST) has been widely adopted for the study of bacterial evolution and population biology of a large number of microbial species [38], and represents the leading molecular method for bacterial genotyping. MLST based on 7 housekeeping loci has been developed for *Leptospira* [39], and is supported by a publically accessible database by which genotypes can be readily assigned as known or new sequence types. An alternative sequence based genotyping scheme of 6 loci including housekeeping genes, a 16S rRNA gene and genes encoding surface-expressed proteins has also been developed and used by several groups. This has led to uncertainty as to which scheme should be adopted. The aim of the current study was to compare the two schemes in terms of their discriminatory ability, both within and between species, by generating data using both schemes for a single set of isolates. We also discuss the practical aspects relating to each scheme.

Materials and Methods

Leptospira isolates and DNA isolation

The *Leptospira* isolates used in this study and their providers are shown in Table 1. Genomic DNA was extracted from laboratory bacterial cultures as described previously [39,40].

Genotyping

All isolates were evaluated using both genotyping schemes [39,40]. The MLST scheme described by Thaipadungpanit et al. (2007), is based on *pntA, sucA, pfkB, tpiA, mreA, glmU and fadD* [39], and the scheme described by Ahmed et al. (2006) is based on *adk, icdA, secY, rrs2, lipL41, and lipL32* [40]. The terms 7L and 6L have been adopted throughout to refer to the 7 and 6 gene schemes, respectively. No modifications were made to the published primers or cycling conditions of 7L. Table 2 lists the primer pairs used for 6L. Four of the 12 primers (*adk*-F, *adk*-R, *secY*-R and *icdA*-R) were modified compared with the published 6L scheme, and used in a repeat PCR reaction in the event that the original primers failed to generate an amplicon. Cycling conditions were as described previously for 6L, with the exception that reactions using the four new 6L primers had a reduced annealing temperature of 54°C. Sequence data were edited using SeqMan software contained within the DNASTAR package (DNASTAR Inc., Wisconsin, USA). The region of sequence used to define each locus of 7L was as described previously [39], but the region used to define each locus of 6L was altered as follows. Three loci (*secY*, *lipL32* and *lipL41*) were changed because the published PCR product and the region of sequence used to define the locus were either identical (*secY* and *lipL32*) or different by just two bp [40]. This meant that we were unable to obtain high quality sequence traces for the first 10–20 bases of the amplicon, and so trimmed the sequence in frame by approximately 20 bp at either end for all three genes. The other 3 published loci of 6L (*adk*, *icdA* and *rrs2*), were trimmed by one or two bases to put them in frame, which simplifies the analysis of synonymous and non-synonymous substitutions. The sequence start and end points for the 6 loci of 6L are shown in Table 2.

The alleles at each of the 7L loci were assigned and the sequence type (ST) defined using the publically accessible *Leptospira* MLST website (http://leptospira.mlst.net/). Allelic numbers, profiles and STs were not generated for the 6L data.

Sequence analysis

Sequence alignment, nucleotide diversity and reconstruction of phylogenetic trees were performed using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 [41]. Mean pairwise distances (p distance) were calculated using the Kimura Two Parameter nucleotide substitution model. Synonymous (dS) and non-synonymous (dN) nucleotide substitutions were calculated based on the Modified Nei-Gojobori method with Jukes Cantor correction using MEGA 4. Neighbor joining trees were reconstructed based on concatenated sequences of each scheme using the Kimura Two-Parameter substitution model. Gene order of the concatenated sequences were *glmU*, *pntA*, *sucA*, *fadD*, *tpiA*, *pfkB*, and *mreA* for 7L, and *adk*, *icdA*, *lipL32*, *lipL41*, *rrs2*, *and secY* for 6L. Discriminatory ability (D value) and 95% confidence intervals (CI) were estimated as described previously [42,43]. These values were verified using the LIAN web tool housed on pubmlst.org [44]. A sliding window analysis of within- and between-species variation was carried out using DNAsp v. 5.0 [45]. An initial "window" of 400-bp was selected, as this is roughly equivalent to a single allele. The first window was thus from base 1 to base 400 of the concatenated sequence. From this we took each species in turn and calculated the average number of nucleotide differences per site over all pairwise comparisons (π), to give the within species polymorphism. Similarly, we calculated the number of fixed differences between species (substitutions) per site to gauge the divergence between *L. interrogans* and *L. kirschneri*. The window region was then moved along 50-bp and these parameters recalculated. GenBank accession numbers of 6L generated sequences are JF509178–JF509357.

Results

Discriminatory power of the two genotyping schemes

A total of 48 strains and isolates belonging to *L. interrogans* (n = 40) and *L. kirschneri* (n = 8) were included in this study, of which 17 were reference strains and 31 were clinical isolates –

Table 1. *Leptospira* isolates used in this study.

Species	Serovar	Strain	ST (7 loci scheme)[#]	Origin	Source[*]
L. interrogans	Copenhageni	M 20	17	Reference	Aus& KIT
L. interrogans	Guaratuba	An 7705	37	Reference	Aus
L. interrogans	Hardjo	Hardjoprajitno	20	Reference	Aus& KIT
L. interrogans	Icterohaemorrhagiae	RGA	17	Reference	Aus& KIT
L. interrogans	Kenniwicki	LT1026	37	Reference	KIT
L. interrogans	Kuwait	136/2/2	26	Reference	MORU
L. interrogans	Lai	Lai	1	Reference	GenBank[†]
L. interrogans	Pomona	Pomona	37	Reference	Aus& KIT
L. interrogans	Portlandvere	MY1039	37	Reference	ND
L. interrogans	Schueffneri	Vleermuis90C	3	Reference	Aus
L. interrogans	Sumneri	Sumner	7	Reference	Aus& KIT
L. interrogans	Valbuzzi	Valbuzzi	61	Reference	Aus& KIT
L. interrogans	Autumanlis	3	34	Thailand	MORU
L. interrogans	Autumnalis	86	34	Thailand	MORU
L. interrogans	Autumnalis	L0020	34	Thailand	MORU
L. interrogans	Autumnalis	L0661	34	Thailand	MORU
L. interrogans	Autumnalis	L1151	34	Thailand	MORU
L. interrogans	Autumnalis	UT227	34	Thailand	MORU
L. interrogans	Autumnalis	548	34	Thailand	MORU
L. interrogans	Autumnalis	729	34	Thailand	MORU
L. interrogans	Autumnalis	LP101	22	Thailand	MORU
L. interrogans	Bataviae	L1111	42	Thailand	MORU
L. interrogans	Bataviae	UT229	46	Thailand	MORU
L. interrogans	Bataviae	UT234	46	Thailand	MORU
L. interrogans	Medanensis	L0448	46	Thailand	MORU
L. interrogans	Medanensis	L0887	46	Thailand	MORU
L. interrogans	Medanensis	L0941	46	Thailand	MORU
L. interrogans	Pomona	UT364	38	Thailand	MORU
L. interrogans	Pyrogenes	UD009	37	Thailand	MORU
L. interrogans	Pyrogenes	L0443	49	Thailand	MORU
L. interrogans	Pyrogenes	L0374	49	Thailand	MORU
L. interrogans	Unknown	654	33	Thailand	MORU
L. interrogans	Unknown	M04	34	Thailand	MORU
L. interrogans	Unknown	M08	34	Thailand	MORU
L. interrogans	Unknown	UT126	40	Thailand	MORU
L. interrogans	Unknown	L1085	42	Thailand	MORU
L. interrogans	Unknown	L0996	46	Thailand	MORU
L. interrogans	Unknown	UT053	46	Thailand	MORU
L. interrogans	Unknown	M10	49	Thailand	MORU
L. interrogans	Unknown	L1207	26	Thailand	MORU
L. kirschneri	Grippotyphosa	Moskva V	110	Reference	KIT
L. kirschneri	Mozdok	5621	117	Reference	KIT
L. kirschneri	Ratnapura	Wumalasena	116	Reference	KIT
L. kirschneri	Tsaratsovo	B 81/7	115	Reference	KIT
L. kirschneri	Vanderhoedeni	Kipod 179	110	Reference	KIT
L. kirschneri	Grippotyphosa	UT130	68	Thailand	MORU

Table 1. Cont.

Species	Serovar	Strain	ST (7 loci scheme)#	Origin	Source*
L. kirschneri	Unknown	M06	68	Thailand	MORU
L. kirschneri	Unknown	M07	71	Thailand	MORU

#STs are not shown for the 6 loci scheme because this is not supported by a MLST website, and allelic numbers, profiles and STs have not been assigned to the sequence data.

*MORU, Mahidol-Oxford Tropical Medicine Research Unit, Bangkok, Thailand (MORU); KIT, KIT Biomedical Research, WHO/FAO/OIE Collaborating Center for Reference & Research on Leptospirosis, Amsterdam, Netherlands; Aus, WHO/FAO/OIE Collaborating Center for Reference & Research on Leptospirosis, Brisbane, Australia. Isolates from two different sources were identified using one of two MLST schemes only.

†in silico analysis was performed on this isolate.

doi:10.1371/journal.pntd.0001374.t001

further referred to as strains (Table 1). Nine strains had been evaluated previously by both schemes [39,40], and 39 strains typed previously by only one of the two schemes were typed by the other scheme during this study. Two strains (a Thai clinical isolate strain L1207 of unknown serovar and a reference strain of serovar Kuwait strain 136/2/2) could not be typed using 6L as both had a deletion of three nucleotides in the *lipL32* sequence. These two strains were excluded from further analysis.

7L resolved the 46 strains into 21 STs, shown in Table 1. 6L data were analysed off line, and the alleles at the six loci given arbitrary allelic numbers to construct an allelic profile and determine the number of genotypes. This demonstrated a total of 30 genotypes (data not shown). Overall levels of diversity (D) were comparable for the 7L and 6L schemes (92.0 (95% CI 87.5–96.5) and 93.5 (95% CI 88.6–98.4), respectively). The discriminatory ability per locus ranged from 59% (*sucA*) to 87% (*glmU* and *mreA*) for 7L and 66% (*rrs2*) to 92% (*secY*) for 6L (Table 3). All D values were verified using the LIAN web tool housed at pubmlst.org and found to be identical to the values shown. The majority of alleles of both schemes were species specific (that is, found in either *L. interrogans* or *L. kirscheri* but not both). There were three exceptions where alleles were found in both species, as follows: 7L, allele 1 of *sucA*; 6L, one allele of *lipL32* and one allele of *rrs2*.

Nucleotide diversity of genetic loci

Overall nucleotide diversity (based on concatenated sequences) for the 46 isolates was 3.6% and 2.3% for 7L and 6L, respectively (Table 3). The diversity within *L. interrogans* was lower than that within *L. kirschneri* (0.5% and 1.1% for 7L, and 0.4% and 0.8% for 6L, respectively). Table 3 also details the nucleotide diversity by locus. This ranged from 3.6% to 6.1% for 7L, and 0.5% to 6.7% for 6L. The lowest diversity was observed for *lipL32* and *rrs2* of 6L. The dN/dS ratios calculated for each locus indicated that none were under positive selection (that is, all values were lower than 1) (Table 3).

A sliding window analysis of the concatenated sequences was performed to provide a visual comparison of the degree of polymorphism within both species, and the level of divergence between them. This revealed a generally higher level of variation within *L. kirschneri* compared to *L. interrogans*, particularly at *sucA* (7L) and to a lesser extent *lipL41* (6L), although the sample size for the former species was very small (n = 8) (Figure 1). This analysis confirmed that the degree of within species polymorphism showed very little difference between the 7L and 6L scheme. However, 7L tended to provide better resolution between species, which was largely accounted for by the low level of divergence for *lipL32* and *rrs2* of 6L.

Relatedness of *Leptospira* spp. inferred from the two genotyping schemes

Neighbor joining trees were reconstructed for 7L and 6L based on the concatenated sequences of their respective loci (Figure 2). Both trees showed two distinct groups corresponding to *L. interrogans* and *L. kirschneri*. There were also several obvious

Table 2. Primers for 6 locus genotyping scheme used during this study [39].

Gene	Published primers (5′- 3′)	New primers (5′- 3′)	Location of sequence used to define MLST locus#	Size of MLST locus (bp)
adk	F-gggctggaaaaggtacacaa	F-acattatcttcatgggacctcc	3458789–3458361	429
	R-acgcaagctccttttgaatc	R-ttacacaagctccctttgaat		
icdA	F-gggacgagatgaccaggat		3980926–3980372	555
	R-ttttttgagatccgcagcttt	R-ctttttttgagatctccggcttt		
lipL32	F-atctccgttgcactctttgc		1667072–1666641	432
	R-accatcatcatcatcgtcca			
lipL41	F-taggaaattgcgcagctaca		3603644–3604120	477
	R-gcatcgagaggaattaacatca			
rrs2	F-catgcaagtcaagcggagta		1862535–1862984	450
	R-agttgagcccgcagttttc			
secY	F-atgccgatcatttttgcttc		3459402–3458902	501
	R-ccgtcccttaattttagacttcttc	R-ccttcctttaattttagacttttttc		

doi:10.1371/journal.pntd.0001374.t002

Table 3. Discriminatory ability of two genotyping schemes and their respective loci.

	Number of alleles	p-distance#	dN/dS*	Discriminatory ability (%) (95% confidence intervals)
7 loci scheme (21 STs)				92.0 (87.5–96.5)
glmU	11	2.30%	0.073	86.9 (82.9–90.8)
pntA	11	3.60%	0.012	64.3 (49.0–79.5)
sucA	7	4.70%	0.007	59.3 (45.2–73.5)
fadD	7	4.00%	0.066	76.3 (69.1–83.5)
tpiA	10	6.10%	0.093	84.7 (79.1–90.4)
pfkB	14	4.70%	0.048	83.4 (76.0–90.7)
mreA	12	4.20%	0.007	86.9 (82.1–91.6)
Concatenated sequence of 6 loci (2,844 nt)		3.60%		
6 loci scheme (30 genotypes)				93.5 (88.6–98.4)
adk	10	6.70%	0.057	70.2 (57.2–83.2)
icdA	12	2.50%	0.022	74.8 (62.7–86.8)
lipL32	7	0.50%	0.154	71.9 (62.3–81.5)
lipL41	7	2.70%	0.01	81.9 (77.4–86.5)
rrs2	6	0.40%	ND	66.3 (58.1–74.4)
secY	20	5.50%	0.019	91.8 (87.4–96.2)
Concatenated sequence of 7 loci (3,165 nt)		2.30%		

#p distances were estimated based on the Kimura Two Parameter nucleotide substitution model.
*dN/dS were estimated based on the Modified Nei-Gojobori Method with Jukes Cantor correction using MEGA 4. The values shown represent a combined value for *L. interrogans* and *L. kirschneri*. dN/dS was not estimated for *rrs2* as this does not encode a protein.
doi:10.1371/journal.pntd.0001374.t003

similarities within *L. interrogans* between the two trees. For example, the clonal structure of ST34 and ST46 as defined by 7L was maintained by 6L. A common finding, however, was that 6L had a tendency to split single STs as defined by 7L into closely related clusters. For example, the three isolates designed as ST49 by 7L were split into three different genotypes by 6L. Further examples of splitting of a clone by the 6L scheme were 7L ST42, ST37, ST68 and ST17. A number of discrepancies were noted between the two trees. Two strains of *L. kirschneri* (strains Moskva V and Kipod 179) were designated by 7L as ST110, but these were resolved into different genotypes by 6L. These two strains differed by 9 nucleotides over 3 loci, with *secY* accounting for 7 of these. A difference was also noted for *L. interrogans* strain 654 (a Thai clinical isolate), which was closely related to *L. interrogans* strain Hard-

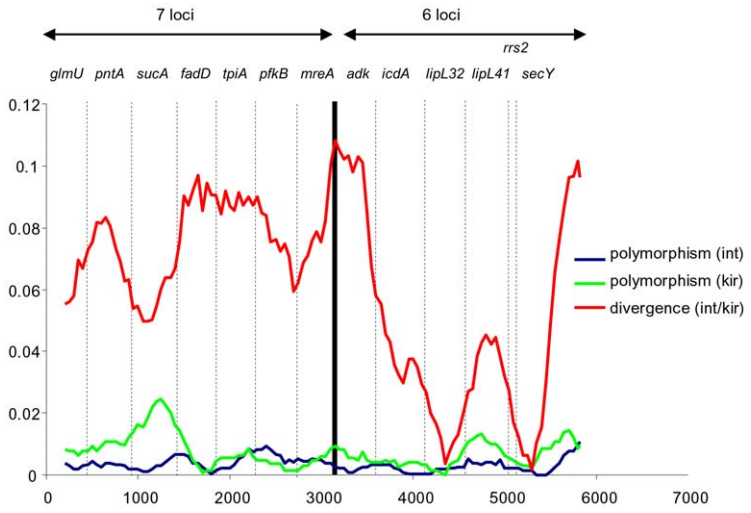

Figure 1. Sliding window analysis of concatenated sequence of all 13 loci. Sliding window analysis of concatenated sequence of all 13 loci, carried out using DNAsp v 5 using a window size of 400-bp, a step size of 50-bp, and points based on the mid-point of each window (i.e. the first point is at position 200). The names of the individual loci are shown. Three plots are given to represent the level of polymorphism within each of the two species, and the level of diversity between them. In terms of the within species variation, there is little difference between the two schemes and both point to generally higher levels of variation within *L. kirschneri* than *L. interrogans*. However, there are two loci used in the 6L scheme that are highly conserved between species (*lipL32* and *rrs2*), which means that in general the 7L scheme provides better between-species resolution.
doi:10.1371/journal.pntd.0001374.g001

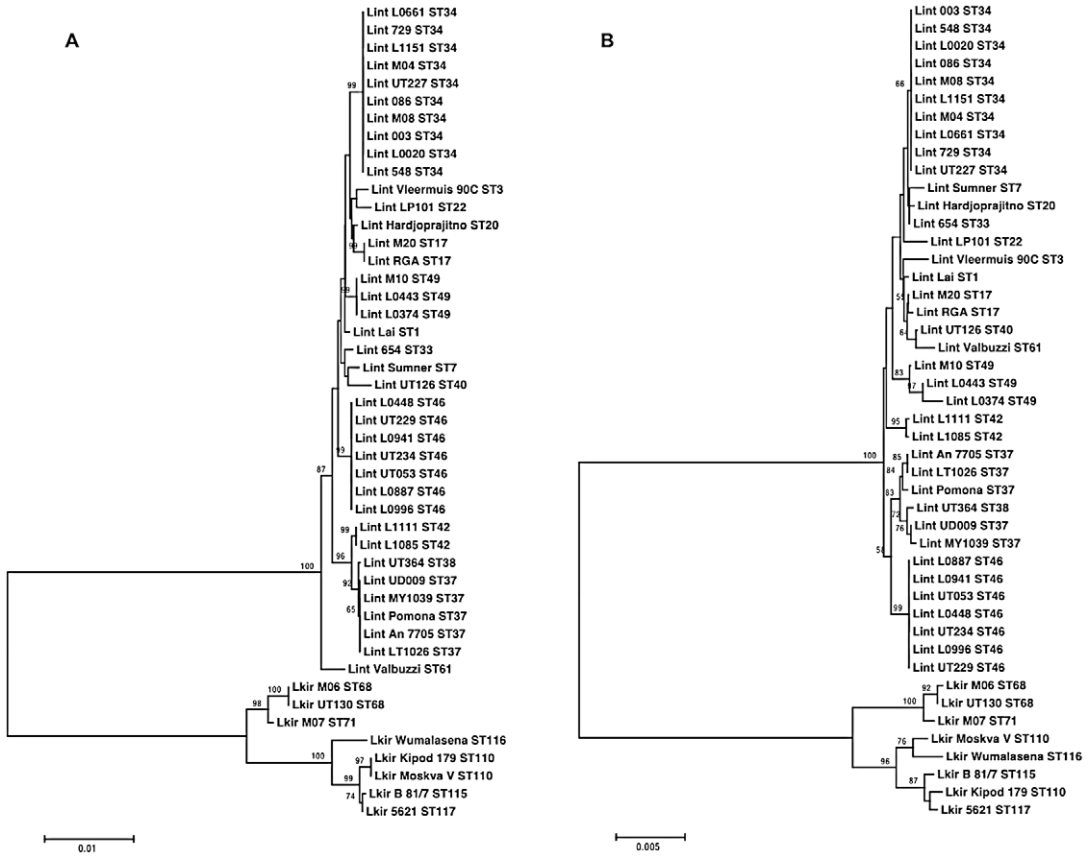

Figure 2. Neighbor joining trees of the 7L scheme and the 6 loci scheme. Neighbor joining trees reconstructed based on concatenated sequences of the 7L scheme (3,165 bp) (A), and the 6 loci scheme (2,844 bp) (B). Each bacterial strain is labeled by the following string: abbreviation of species name (Lint- *L. interrogans*, Lkir- *L. kirschneri*), strain name, and (for the 7L scheme only) sequence type (ST).
doi:10.1371/journal.pntd.0001374.g002

joprajitno by 6L (differing by only 1 nucleotide), but was more distantly related by 7L (differing by 11 nucleotides over 6 loci).

Discussion

The authors of this paper include representatives of the scientific groups that reported two independent genotyping schemes for *Leptospira* spp. Here, we provide the scientific community with the findings of a study that compared and contrasted the two schemes, together with a discussion of the practical aspects related to undertaking each.

The two schemes are unrelated and different by design. 7L was founded on a conventional strategy for MLST of selecting 7 housekeeping genes that were distributed around the genome and were not under positive selection. The design of 6L varied from this in that 6 loci were selected from different functional categories. For example, *lipL41* and *lipL32* encode surface expressed proteins that would be expected to be under positive selection as a result of being immunogenic and a target for the host response. At the other end of the spectrum, *rrs2* is one of two 16S rRNA genes that would be predicted to be highly conserved.

Contrary to our expectations, we did not find that any of the 6L genes were under positive selection. More genotypes were resolved by 6L than by 7L, in part a function of the high number of alleles for *secY*. Analysis of genetic diversity indicated that there was little difference in within-species variation difference between the two schemes, both pointing to generally higher levels of variation within *L. kirschneri* than *L. interrogans*. The conserved nature of two loci used in 6L (*lipL32* and *rrs2*), resulted in the finding on sliding window analysis that 7L provided better between-species resolution. Interestingly we noted that *rrs2* of 6L showed a higher D value than the housekeeping gene *sucA* of 7L. Although this is an exception to the general rule that housekeeping metabolic genes provide more discrimination than conserved genes such as those encoding ribosomal RNA, such an observation is not unprecedented [46].

6L has been applied to six pathogenic *Leptospira* spp. [40], which compares favorably with 7L which was designed for the two closely related species *L. interrogans and L. kirschneri*. However, this disadvantage of 7L will be resolved within the next 12 months; the scheme has already been extended to *L. borpetersenii* (manuscript in preparation), and the laboratory work to extend this to all pathogenic species is now completed. These improvements will be made publicly available by the end of 2011.

Conversely, the 6L scheme does not conform to the original concept of MLST as it includes a non-housekeeping gene (*rrs2*), and genes that encode cell surface proteins. Furthermore, the sequence start and stop sites used to define the allele for each locus were not provided in the original description of 6L scheme and so could not be performed based on the published methodology alone, although these have been detailed in this study. Minor changes were necessary to the start and stop sites, but we think it unlikely that this led to a change in the performance of the scheme.

The 6L scheme is not associated with a publically accessible website that allows an investigator to compare new data with existing sequence data. 6L has recently been applied to an

extended set of strains and isolates (n = 271) encompassing a wide diversity of hosts and geographic regions [47], providing a rich source of sequence data that has been released into the public domain (GenBank). Comparative phylogenetic analysis by individual investigators will require downloading and storage of these data. In contrast, a website for 7L was launched at the time of publication and is regularly maintained and curated. At least one representative of each ST is recorded in a downloadable spreadsheet, providing a mechanism by which a picture of global bacterial diversity can be developed over time. This is easy to use, provides tools for comparison of a given strain with all of the other strains in the database, is more suited to investigators with limited phylogenetic training and experience, and so has the power to reach a wider audience.

In conclusion, we have provided detailed comparisons of two major genotyping schemes for *Leptospira* spp., and have described their advantages and disadvantages. 7L complies with the philosophy of MLST (housekeeping genes only supported by website), but will not be ready for use for the study of all pathogenic *Leptospira* spp. until the end of 2011. In the meantime, a bioinformatics analysis of the discriminatory power of 4 genes (three of which are not present in either scheme) as well as a new scheme with 7 loci both limited to *L. interrogans* and *L. kirschneri* have been reported [48,49], adding further diversity to the tools available for the phylogenetic study of *Leptospira* spp. There is a pressing need for consensus within the leptospirosis community as to the preferred genotyping scheme, an essential step if the wealth of knowledge gathered for other bacterial species based on detailed analysis within a single scheme is to be replicated for *Leptospira* spp. Both schemes contain highly discriminative and less discriminative loci. While it is feasible to formulate a consensus MLST combining the most discriminative housekeeping genes from both schemes, we have resisted the temptation of presenting an interim scheme that has not been extensively validated. Instead, we aim to expedite the release of the 7L MLST scheme for all the major pathogenic species, and recommend its use for the study of the global epidemiology of pathogenic *Leptospira* spp.

Author Contributions

Conceived and designed the experiments: RAH SJP. Performed the experiments: AA JT SB VW EJF. Analyzed the data: NA BGS DMA LDS EJF. Contributed reagents/materials/analysis tools: NA BGS DMA LDS EJF. Wrote the paper: AA JT KN EJF RAH SJP.

References

1. World Health Organization (2003) Human leptospirosis: Guidance for diagnosis, surveillance and control. Geneva: World Health Organization & International Leptospirosis Society.
2. Faine SB, Adler B, Bolin C, Perolat P (1999) *Leptospira* and leptospirosis. Melbourne, Australia: MediSci.
3. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, et al. (2003) Leptospirosis: a zoonotic disease of global importance. Lancet Infect Dis 3: 757–771.
4. Levett PN (2001) Leptospirosis. Clin Microbiol Rev 14: 296–326.
5. World Health Organization (WHO) and International Leptospirosis Society (ILS), 2006; online at www.leptonet.net.
6. McBride AJ, Athanazio DA, Reis MG, Ko AI (2005) Leptospirosis. Curr Opin Infect Dis 18: 376–386.
7. Hartskeerl RA (2006) Leptospirosis: current status and future trends. Indian J Med Microbiol 24: 309.
8. Agampodi S, Peacock SJ, Thevanesam V (2009) The potential emergence of leptospirosis in Sri Lanka. Lancet Infect Dis 9: 524–526.
9. Koizumi N, Gamage CD, Muto M, Kularatne SA, Budagoda BD, et al. (2009) Serological and genetic analysis of leptospirosis in patients with acute febrile illness in Kandy, Sri lanka. Jpn J Infect Dis 62: 474–475.
10. McCurry J (2009) Philippines struggles to recover from typhoons. Lancet 374: 1489.
11. World Health Organization (2000) Leptospirosis, India: report of the investigation of a post-cyclone outbreak in Orissa, November 1999. Wkly Epidemiol Rec 75: 217–23.
12. Sanders EJ, Rigau-Perez JG, Smits HL, Deseda CC, Vorndam VA, et al. (1999) Increase of leptospirosis in dengue-negative patients after a hurricane in Puerto Rica in 1996. Am J Trop Med Hyg 61: 935–941.
13. Ko AI, Galvao RM, Ribeiro Dourado CM, Johnson WD, Jr., Riley LW (1999) Urban epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group. Lancet 354: 820–825.
14. Zaki SR, Shieh WJ (1996) Leptospirosis associated with outbreak of acute febrile illness and pulmonary haemorrhage, Nicaragua, 1995. The Epidemic Working Group at Ministry of Health in Nicaragua. Lancet 347: 535–536.
15. Lau C, Smythe L, Weinstein P (2010) Leptospirosis: an emerging disease in travellers. Travel Med Infect Dis 8: 33–39.
16. Yasuda PH, Steigerwalt AG, Sulzer KR, Kaufmann AF, Rogers FC, et al. (1987) Deoxyribonucleic acid relatedness between serogroups and serovars in the Family *Leptospiraceae* with proposals for seven new *Leptospira* species. Int J Syst Bacteriol 37: 407–415.
17. Ramadass P, Jarvis BD, Corner RJ, Penny D, Marshall RB (1992) Genetic characterization of pathogenic *Leptospira* species by DNA hybridization. Int J Syst Bacteriol 42: 215–219.
18. Brenner DJ, Kaufmann AF, Sulzer KR, Steigerwalt AG, Rogers FC, et al. (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*. Int J Syst Bacteriol 49 Pt 2: 839–858.
19. Perolat P, Chappel RJ, Adler B, Baranton G, Bulach DM, et al. (1998) *Leptospira fainei* sp. nov., isolated from pigs in Australia. Int J Syst Bacteriol 48 Pt 3: 851–858.
20. Corney BG, Colley J, Djordjevic SP, Whittington R, Graham GC (1993) Rapid identification of some *Leptospira* isolates from cattle by random amplified polymorphic DNA fingerprinting. J Clin Microbiol 31: 2927–2932.
21. Ralph D, McClelland M, Welsh J, Baranton G, Perolat P (1993) *Leptospira* species categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes. J Bacteriol 175: 973–981.
22. Perolat P, Merien F, Ellis WA, Baranton G (1994) Characterization of *Leptospira* isolates from serovar Hardjo by ribotyping, arbitrarily primed PCR, and mapped restriction site polymorphisms. J Clin Microbiol 32: 1949–1957.
23. Herrmann JL, Baril C, Bellenger E, Perolat P, Baranton G, et al. (1991) Genome conservation in isolates of *Leptospira interrogans*. J Bacteriol 173: 7582–7588.
24. Herrmann JL, Bellenger E, Perolat P, Baranton G, Saint Girons I (1992) Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification. J Clin Microbiol 30: 1696–1702.
25. Zuerner RL, Bolin CA (1990) Nucleic acid probe characterizes *Leptospira interrogans* serovars by restriction fragment length polymorphisms. Vet Microbiol 24: 355–366.
26. Zuerner RL, Ellis WA, Bolin CA, Montgomery JM (1993) Restriction fragment length polymorphisms distinguish *Leptospira borgpetersenii* serovar Hardjo type Hardjo-bovis isolates from different geographical locations. J Clin Microbiol 31: 578–583.
27. Zuerner RL, Alt D, Bolin CA (1995) IS*1533*-based PCR assay for identification of *Leptospira interrogans* sensu lato serovars. J Clin Microbiol 33: 3284–3289.
28. Majed Z, Bellenger E, Postic D, Pourcel C, Baranton G, et al. (2005) Identification of variable-number tandem-repeat loci in *Leptospira interrogans* sensu stricto. J Clin Microbiol 43: 539–545.
29. Slack A, Symonds M, Dohnt M, Smythe L (2006) An improved multiple-locus variable number of tandem repeats analysis for *Leptospira interrogans* serovar Australis: a comparison with fluorescent amplified fragment length polymorphism analysis and its use to redefine the molecular epidemiology of this serovar in Queensland, Australia. J Med Microbiol 55: 1549–1557.
30. Hookey JV, Bryden J, Gatehouse L (1993) The use of 16S rDNA sequence analysis to investigate the phylogeny of *Leptospiraceae* and related spirochaetes. J Gen Microbiol 139: 2585–2590.
31. Morey RE, Galloway RL, Bragg SL, Steigerwalt AG, Mayer LW, et al. (2006) Species-specific identification of *Leptospiraceae* by 16S rRNA gene sequencing. J Clin Microbiol 44: 3510–3516.
32. Cerqueira GM, McBride AJ, Queiroz A, Pinto LS, Silva EF, et al. (2010) Monitoring *Leptospira* strain collections: the need for quality control. Am J Trop Med Hyg 82: 83–87.
33. La SB, Bui LT, Baranton G, Khamis A, Raoult D (2006) Partial *rpoB* gene sequencing for identification of *Leptospira* species. FEMS Microbiol Lett 263: 142–147.
34. Slack AT, Symonds ML, Dohnt MF, Smythe LD (2006) Identification of pathogenic *Leptospira* species by conventional or real-time PCR and sequencing of the DNA gyrase subunit B encoding gene. BMC Microbiol 6: 95.
35. Victoria B, Ahmed A, Zuerner RL, Ahmed N, Bulach DM, et al. (2008) Conservation of the *S10-spc-alpha* locus within otherwise highly plastic genomes provides phylogenetic insight into the genus *Leptospira*. PLoS One 3: e2752.

36. Ahmed A, Engelberts MF, Boer KR, Ahmed N, Hartskeerl RA (2009) Development and validation of a real-time PCR for detection of pathogenic *Leptospira* species in clinical materials. PLoS One 4: e7093.
37. Cerqueira GM, McBride AJ, Picardeau M, Ribeiro SG, Moreira AN, et al. (2009) Distribution of the leptospiral immunoglobulin-like (*lig*) genes in pathogenic *Leptospira* species and application of *ligB* to typing leptospiral isolates. J Med Microbiol 58: 1173–1181.
38. Maiden MC (2006) Multilocus sequence typing of bacteria. Annu Rev Microbiol 60: 561–588.
39. Thaipadungpanit J, Wuthiekanun V, Chierakul W, Smythe LD, Petkanchanapong W, et al. (2007) A dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis in Thailand. PLoS Negl Trop Dis 1: e56. 10.
40. Ahmed N, Devi SM, Valverde ML, Vijayachari P, Machang'u RS, et al. (2006) Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. Ann Clin Microbiol Antimicrob 5: 28.
41. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.
42. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465–2466.
43. Grundmann H, Hori S, Tanner G (2001) Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. J Clin Microbiol 39: 4190–4192.
44. Haubold B, Hudson RR (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16: 847–848.
45. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
46. Cooper JE, Feil EJ (2006) The phylogeny of *Staphylococcus aureus* - which genes make the best intra-species markers? Microbiology 152: 1297–1305.
47. Nalam K, Ahmed A, Devi SM, Francalacci P, Baig M, et al. (2010) Genetic affinities within a large global collection of pathogenic *Leptospira*: implications for strain identification and molecular epidemiology. PLoS One 5: e12637.
48. Cerqueira GM, McBride AJ, Hartskeerl RA, Ahmed N, Dellagostin OA, et al. (2010) Bioinformatics describes novel Loci for high resolution discrimination of *Leptospira* isolates. PLoS One 5: e15335.
49. Leon A, Pronost S, Fortier G, Andre-Fontaine G, Leclercq R (2009) Multilocus sequence analysis for typing *Leptospira interrogans* and *Leptospira kirschneri*. J Clin Microbiol 48: 581–585.

POPULATION GENETICS AND GENOMICS OF PATHOGENIC LEPTOSPIRA

by S. A. Vardhan Kishore Nalam

Submission date: 27-Dec-2019 04:53PM (UTC+0530)
Submission ID: 1238493578
File name: Thesis_NK_271219.pdf (1M)
Word count: 27273
Character count: 153749

POPULATION GENETICS AND GENOMICS OF PATHOGENIC LEPTOSPIRA

ORIGINALITY REPORT

12%	9%	11%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	journals.plos.org Internet Source	4%
2	link.springer.com Internet Source	1%
3	Gustavo M. Cerqueira, Mathieu Picardeau. "A century of Leptospira strain typing", Infection, Genetics and Evolution, 2009 Publication	<1%
4	itswellington.massey.ac.nz Internet Source	<1%
5	whoindia.org Internet Source	<1%
6	Current Topics in Microbiology and Immunology, 2015. Publication	<1%
7	www.omicsonline.org Internet Source	<1%

Kishore Nalam, Ahmed Ahmed, Sundru

8	Manjulata Devi, Paolo Francalacci et al. "Genetic Affinities within a Large Global Collection of Pathogenic Leptospira: Implications for Strain Identification and Molecular Epidemiology", PLoS ONE, 2010 Publication	<1%
9	Mathieu Picardeau. "Genome Sequence of the Saprophyte Leptospira biflexa Provides Insights into the Evolution of Leptospira and the Pathogenesis of Leptospirosis", PLoS ONE, 02/13/2008 Publication	<1%
10	Submitted to Universiti Sains Malaysia Student Paper	<1%
11	Ana S. Ferreira, Ahmed Ahmed, Teresa Rocha, Maria L. Vieira et al. "Genetic diversity of pathogenic leptospires from wild, domestic and captive host species in Portugal", Transboundary and Emerging Diseases, 2019 Publication	<1%
12	Submitted to American Public University System Student Paper	<1%
13	molecular.biosciences.wsu.edu Internet Source	<1%
14	ja.scribd.com Internet Source	<1%

15	preview-ann-clinmicrob.biomedcentral.com Internet Source	<1%
16	Lehmann, Jason, Michael Matthias, Joseph Vinetz, and Derrick Fouts. "Leptospiral Pathogenomics", Pathogens, 2014. Publication	<1%
17	Marcos Pérez-Losada, Patricia Cabezas, Eduardo Castro-Nallar, Keith A. Crandall. "Pathogen typing in the genomics era: MLST and the future of molecular epidemiology", Infection, Genetics and Evolution, 2013 Publication	<1%
18	www.pagepress.org Internet Source	<1%
19	ncdc.gov.in Internet Source	<1%
20	Chin VK, Lee TY, Lim WF, Wan Shahriman YWY, Syafinaz AN, Zamberi S, Maha A. "Leptospirosis in human: Biomarkers in host immune responses", Microbiological Research, 2018 Publication	<1%
21	"Bioinformatics for High Throughput Sequencing", Springer Nature, 2012 Publication	<1%
	espace.library.uq.edu.au	

22	Internet Source	<1%
23	Submitted to Ross University Student Paper	<1%
24	Submitted to J S S University Student Paper	<1%
25	Akinola Adekoya Alafiatayo, Kok-Song Lai, Syahida Ahmad, Maziah Mahmood, Noor Azmi Shaharuddin. "RNA-Seq analysis revealed genes associated with UV-induced cell necrosis through MAPK/TNF-α pathways in human dermal fibroblast cells as an inducer of premature photoaging", Genomics, 2019 Publication	<1%
26	Ahmed, A.. "A simple and rapid molecular method for Leptospira species identification", Infection, Genetics and Evolution, 201010 Publication	<1%
27	Submitted to The University of Manchester Student Paper	<1%
28	hdl.handle.net Internet Source	<1%
29	"The Prokaryotes", Springer Nature, 2014 Publication	<1%
30	www.reseausega-coi.org	

	Internet Source	<1%
31	repub.eur.nl Internet Source	<1%
32	VANINA GUERNIER, KATHRYN J. ALLAN, CYRILLE GOARANT. "Advances and challenges in barcoding pathogenic and environmental Leptospira", Parasitology, 2017 Publication	<1%
33	Sedigheh Zakeri, Nargess Khorami, Zahra F. Ganji, Neda Sepahian, Abdol-Ali Malmasi, Mohammad Mehdi Gouya, Navid D. Djadid. "Leptospira wolffii, a potential new pathogenic Leptospira species detected in human, sheep and dog", Infection, Genetics and Evolution, 2010 Publication	<1%
34	Submitted to Universiti Putra Malaysia Student Paper	<1%
35	academic.oup.com Internet Source	<1%
36	Murugesan Kanagavel, Alphonse Asirvatham Princy Margreat, Manivel Arunkumar, Shanmugarajan Gnanasekaran Prabhakaran et al. "Multilocus sequence typing (MLST) of leptospiral strains isolated from two geographic	<1%

locations of Tamil Nadu, India", Infection, Genetics and Evolution, 2016
Publication

---

37 Rafael García-González, Angélica Reyes-Torres, María del Rocío Reyes-Montes, Esperanza Duarte-Escalante et al. "Genotyping of Leptospira interrogans isolates from Mexican patients", Revista do Instituto de Medicina Tropical de São Paulo, 2019 <1%
Publication

---

38 Ramani Baddam, Narender Kumar, Sabiha Shaik, Aditya Kumar Lankapalli, Niyaz Ahmed. "Genome dynamics and evolution of Salmonella Typhi strains from the typhoid-endemic zones", Scientific Reports, 2014 <1%
Publication

---

39 Ben Adler. "The Genus Leptospira", The Prokaryotes, 2006 <1%
Publication

---

40 Vanina Varni, Paula Ruybal, Juan José Lauthier, Nicolás Tomasini, Bibiana Brihuega, Ariel Koval, Karina Caimi. "Reassessment of MLST schemes for Leptospira spp. typing worldwide", Infection, Genetics and Evolution, 2014 <1%
Publication

---

41 M. A. Barocchi. "Rapid Translocation of

	Polarized MDCK Cell Monolayers by Leptospira interrogans, an Invasive but Nonintracellular Pathogen", Infection and Immunity, 12/01/2002 Publication	<1%
42	Siriphan Boonsilp, Janjira Thaipadungpanit, Premjit Amornchai, Vanaporn Wuthiekanun et al. "A Single Multilocus Sequence Typing (MLST) Scheme for Seven Pathogenic Leptospira Species", PLoS Neglected Tropical Diseases, 2013 Publication	<1%
43	espace.cdu.edu.au Internet Source	<1%
44	cmr.asm.org Internet Source	<1%
45	teses.usp.br Internet Source	<1%
46	Luciane A. Santos, Haritha Adhikarla, Xiting Yan, Zheng Wang et al. "Genomic Comparison Among Global Isolates of L. interrogans Serovars Copenhageni and Icterohaemorrhagiae Identified Natural Genetic Variation Caused by an Indel", Frontiers in Cellular and Infection Microbiology, 2018 Publication	<1%
47	T. R. Fraga. "Leptospirosis: Aspects of Innate	<1%

	Immunity, Immunopathogenesis and Immune Evasion From the Complement System : Leptospiral Immunity", Scandinavian Journal of Immunology, 05/2011 Publication	
48	Submitted to Mahidol University Student Paper	<1%
49	Perkins, Timothy T., Chin Yen Tay, Fanny Thirriot, and Barry Marshall. "Choosing a Benchtop Sequencing Machine to Characterise Helicobacter pylori Genomes", PLoS ONE, 2013. Publication	<1%
50	Submitted to University of Mauritius Student Paper	<1%
51	Submitted to Indiana University Student Paper	<1%
52	Christian A. Ganoza. "Asymptomatic Renal Colonization of Humans in the Peruvian Amazon by Leptospira", PLoS Neglected Tropical Diseases, 02/23/2010 Publication	<1%
53	"Handbook of Disease Burdens and Quality of Life Measures", Springer Nature, 2010 Publication	<1%
54	eprints.adm.unipi.it	

Internet Source

<1%

55 Hamond, C., M. Pinna, M. A. Medeiros, P. Bourhy, W. Lilenbaum, and M. Picardeau. "A Multi Locus VNTR Analysis assay provides high discrimination for genotyping Leptospira santarosai strains", Journal of Medical Microbiology, 2015.
Publication

<1%

56 The Prokaryotes, 2006.
Publication

<1%

57 C. Gravekamp. "Detection of seven species of pathogenic leptospires by PCR using two sets of primers", Microbiology, 08/01/1993
Publication

<1%

58 www.pyiephyomaung.yolasite.com
Internet Source

<1%

59 Fouts, Derrick E., Michael A. Matthias, Haritha Adhikarla, Ben Adler, Luciane Amorim-Santos, Douglas E. Berg, Dieter Bulach, Alejandro Buschiazzo, Yung-Fu Chang, Renee L. Galloway, David A. Haake, Daniel H. Haft, Rudy Hartskeerl, Albert I. Ko, Paul N. Levett, James Matsunaga, Ariel E. Mechaly, Jonathan M. Monk, Ana L. T. Nascimento, Karen E. Nelson, Bernhard Palsson, Sharon J. Peacock, Mathieu

<1%

Picardeau, Jessica N. Ricaldi, Janjira Thaipandungpanit, Elsio A. Wunder, X. Frank Yang, Jun-Jie Zhang, and Joseph M. Vinetz. "What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira", PLoS Neglected Tropical Diseases, 2016.
Publication

60	pinnacle.allenpress.com Internet Source	<1%
61	www.labome.org Internet Source	<1%
62	Bisen, . "Leptospirosis", Emerging Epidemics Management and Control, 2013. Publication	<1%
63	Toshiyuki Masuzawa, Mitsumasa Saito, Ryo Nakao, Yasuhiko Nikaido et al. " Molecular and phenotypic characterization of sp. nov., sp. nov. and sp. nov. isolated from soil and water in Japan ", Microbiology and Immunology, 2019 Publication	<1%
64	Rebekah A. Oomen, Matthew W. Reudink, Joseph J. Nocera, Christopher M. Somers, M. Clay Green, Christopher J. Kyle. "Mitochondrial Evidence for Panmixia despite Perceived Barriers to Gene Flow in a Widely Distributed Waterbird", Journal of Heredity, 2011	<1%

Publication

65 Vanina Guernier, Vaea Richard, Tuxuan Nhan, Eline Rouault, Anita Tessier, Didier Musso. "Leptospira diversity in animals and humans in Tahiti, French Polynesia", PLOS Neglected Tropical Diseases, 2017 <1%

Publication

66 P. Vijayachari. "Leptospirosis: an emerging global public health problem", Journal of Biosciences, 11/2008 <1%

Publication

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		