Current Applied Science and Technology Vol. 23 No. 6 (November-December 2023)

Review article

Sequencing Approaches for Understanding the Pathogenic Features in *Leptospira* Genome

Ramyadevi Veerabahu[1], Vigneshwaran Ravishankar[1,2]* and Eshaa Umashankar[1]

[1]*Department of Biotechnology, Mepco Schlenk Engineering College (Autonomous), Sivakasi 626005, Tamilnadu, India*
[2]*Foundation for Neglected Disease Research, Bengaluru 561203, Karnataka, India*

Received: 5 August 2022, Revised: 2 November 2022, Accepted: 27 March 2023

DOI: 10.55003/cast.2023.06.23.009

Keywords

Leptospira;
vaccine;
sequencing;
genome;
Illumina

Abstract

Leptospirosis is a zoonotic illness caused by a bacterial spirochaete belonging to the genus *Leptospira* that induces renal failure in humans. This bacterium has high pathogenicity and can even cause death. As a result, research into the pathogenic mechanism of this organism is essential. Sequencing can be performed to understand the evolution and characteristics of an organism. Sequencing technologies like Sanger sequencing, Illumina sequencing, and recent technologies like MinION nanopore sequencing can be performed to study the whole genome of leptospiral strains. Based on whole genome sequencing, pathogenic characteristics and pathways involved during disease infection can be investigated. Identification of pathogenic features in the whole genome of *Leptospira* sp. may enable certain gene modifications or the development of potential vaccines to eradicate the disease. This review is an overview of various sequencing approaches and highlights the comparison of genome features present in pathogenic leptospires and saprophytic leptospires.

1. Introduction

DNA sequencing is a sensitive approach for understanding the genetic information of a microorganism and the genes responsible for the regulatory instructions for turning on or off mechanisms [1]. Sequencing is performed to retrieve information about the genes responsible for causing diseases and specific mutations within the organism [2]. The main advantage of sequencing is the application in the areas of forensics and in the identification of extinct or endangered species. It can also be applied in agriculture to reduce pest attacks and increase productivity. Because of the advancement in science, sequencing technology has developed extensively and evolved from first, second, and even third-generation sequencing technologies, which help researchers and clinicians

*Corresponding author: Tel.: (+91) -9994828885
E-mail: rvigneswaran9490@yahoo.com

understand the genetic information in pathogenic strains [3]. One of the best ways to avoid communicable diseases is by vaccination. The development of vaccines has become easier because of high-throughput technologies like functional and structural genomics and full genome sequencing. The most recent advances in synthetic genomes offer greater opportunities for creating vaccines. Pan-genomic reverse vaccinology increases the potential for discovering new vaccine candidates by analyzing sequence data from diverse isolates of the same species of a pathogen. The time it takes for vaccine discovery research and development has been significantly lowered as a result of the prompt availability of comprehensive and accurate pathogenic genome sequences and the growth of analysis tools that can be used for mining biological information contained in genome sequences. Genome mining has made it possible to assign gene functions, predict a few features of encoded proteins such as cellular location, molecular weight, pI, or solubility, and predict genes that might encode factors that promote pathogenesis [4]. In this perspective reverse vaccinology can be seen as a particularly alluring method, delivering conserved potential antigens in the shortest amount of time, especially when combined with the knowledge from comparative and experimental genomics. The discovery of pathogenesis-related variables, which are crucial components in the production of novel vaccines, has also been accelerated by the development of postgenomics methods. With the potential to use genetic data to develop possible vaccine candidates, a new revolution in vaccine development may emerge from the application of genomic techniques that are innovative in their ability to recognize the molecular mechanisms of illness. The development of postgenomics techniques has also improved the identification of pathogenesis-related variables, which are vital elements in the design of novel vaccines. Sequencing may now be performed more quickly, with higher sensitivity and specificity, and without requiring skilled labour that can often be completed within a few days [5].

2. First Generation Sequencing

2.1 Chain termination method

The first breakthrough in DNA sequencing occurred in 1977 when Maxam and Gilbert introduced a method for sequencing based on the chemical modification of DNA. Their method involved breaking the DNA strands at specific sites, and it did not rely on DNA polymerase, unlike Sanger sequencing [6, 7]. The method involves radioactive labeling at 5'end of the DNA fragment to be sequenced. Chemical treatment is done to modify the DNA and cleave it at specific sites. Purines are depurinated using formic acid, guanines undergo methylation by dimethyl sulfate, and hydrazine is used to hydrolyze pyrimidines. This modified DNA is then cleaved by piperidine resulting in small fragments [8]. These small fragments are visualized using auto radiography. This technique suffers as it lacks automation, usage of hazardous chemicals, and is time-consuming [9, 10].

2.2 Sanger sequencing

In the same year, Fredrick Sanger introduced a method of first-generation sequencing called Sanger sequencing, which is also known as the "chain termination method". The method which utilized radiolabelled partially digested fragments, dominated the sequencing world for more than 30 years [11-13]. Sanger used this method to sequence the complete genome of the bacteriophage PhiX174, a virus that infects *Escherichia coli* [14]. This method involves dideoxyribonucleotides (ddNTPs) where 3'hydroxyl group is absent, a condition which is required for 5' to 3' extension of a DNA polynucleotide chain in a sequencing reaction [10]. When ddNTPs are incorporated, they terminate the chain reaction. The products are separated and resolved using electrophoresis. They can be

visualized using auto radiography in which the radioactive bands that are present fluoresce. Many improvements have been made over the years to this method, where the radiolabeling of phosphate or tritium is replaced by a fluorometric-based detection technique and capillary-based electrophoresis. This approach has made the sequencing of DNA faster, and it can also be done for complex species. Although Sanger's approach provides positive results with 99.9% accuracy and increased specificity [15], it suffers from drawbacks like a time-consuming process and another is that only one strand can be sequenced in a reaction.

2.3 ABI 370 sequencing

In 1984, Fritz Pohl invented a GATC 1500 sequencing platform where radioactive substances are not utilized for labeling nucleotides. In 1987, Hood and Hunkapiller improved the Sanger sequencing method by replacing the radioactive labels with fluorescent dyes, and the acquisition and analysis were interpreted by an ABI 370 machine [16, 17]. This approach was appreciated by the scientific community and paved the way for the world of second-generation sequencing.

3. Second Generation Sequencing

3.1 Pyrosequencing

In 1996, Mathias Uhlen and Nyren introduced a new approach called pyrosequencing, also known as 454 pyrosequencing, which utilizes DNA polymerase, ATP sulfurylase, and luciferase enzyme to synthesize pyrophosphate for sequencing the DNA element [18, 19]. This method does not require the usage of fluorescence dyes, ddNTPs and gels. The emulsion PCR (oil-water) method, which is used for amplification, isolates and categorizes the template DNA to prevent unproductive artifacts between the DNA sequences. The DNA is first isolated, fragmented and an adapter is attached to the fragments, which are made into a single separate strand [20]. The capture beads contain a single oligonucleotide primer that is complementary to the adapter. After vigorous agitation of the oil-water system, emulsification occurs, and amplification of template DNA takes place. The concentration is optimized to ensure that only one template DNA and one bead are present in each droplet. The emulsion is then broken to release the beads that contain the single-stranded DNA, and they are deposited into the well where the pyrosequencing reaction occurs. As a result of the reaction, light is produced and is captured by a camera. The peaks are detected using a pyrogram. Eventually, the nucleotide sequences are determined. This method is quicker than Sanger and Maxam-Gilbert sequencing methods and can be easily automated and processed parallelly. This method is mainly used for sequencing the human microbiome to study its role in digestion, metabolism, and even the pathogens. The cons of this method are that the lengths of individual reads are around 300-500 nucleotides only, and it lacks proofreading, thus reducing accuracy.

3.2 Illumina sequencing

In 1998, Balasubramanian and his research group developed a sequencing-by-synthesis-based approach known as Illumina sequencing, which utilizes Solexa technology including fluorescent dyes [21, 22]. The DNA is initially purified, fragmented and attached to adapters, which can be used as reference points. The processed DNA mixture is loaded into a flow cell, where the first amplification step starts. The flow cell contains a nanowell, which can arrange the DNA fragments to avoid overcrowding. The nanowell contains an oligonucleotide that helps attach the adapters to the well. After the attachment of adapters, cluster generation begins whereby thousands of copies of

the DNA are created, and this is done using bridge amplification PCR. This process is carried out to ensure that the signal intensity of the bases can be achieved with standard signal generation. The primers and the nucleotides are added in which the nucleotides contain reversible dye terminators so that the DNA polymerase is made sure to be attached with only one nucleotide with a fragment. A chemical solution is added to remove the 3'OH fluorescent terminal group, and the fluorescence is captured using a camera or other equipment and the bases are analyzed. This process is continued until the whole DNA molecule is sequenced. Illumina sequencing leverages clonal amplification and sequencing by synthesis technology to produce accurate results. This technique is successful as it can sequence multiple strands and possess a huge data of sequencing information. Even though the read lengths are ~ 150 bp, the cost of the run is cheaper compared with other sequencing technologies [23]. However, there are disadvantages like the high substitution error rate and the high instrumentation cost; furthermore, only highly trained personnel can perform this technique.

4. Third Generation Sequencing

One spectacular entry into the third-generation sequencing was the approach of Single-Molecule Real-Time (SMRT) sequencing technology introduced by Pacific Biosciences in 2010 [24, 25]. They developed a zero-mode waveguide (ZMW) technique that utilizes nanohole containing a single DNA polymerase. When a nucleotide labeled with fluorescent dye is passed through the nanohole, it detects the light emitted by the molecule and is captured by the detector placed below the ZMW. In 2011, the Ion Torrent sequencing method (by Life Technology) used sequencing–by–synthesis technology which can detect hydrogen ions when new DNA strands are synthesized. The latest version of SMRT-GridION was introduced in 2012, and is significantly reduced in size, and utilizes the electric current produced when the DNA strand passes through the biological nanopore to identify the nucleotide sequence. This GridION (Oxford Nanopore Technology, UK) combines 5 flow cells that can be run individually or together. Each flow cell has 512 nanopores and can generate more than 100 GB of data from a single run with increased sensitivity and accuracy [26]. It is used to sequence various smaller genomes and transcriptional applications and enables users to stop and start the sequencing as desired. The data can be retrieved using MinKNOW software [27], and the processing can be done based on either EPI2ME software or any other pipeline.

5. Fourth Generation Sequencing

5.1 MinION sequencing

In 2015, MinION, a portable device, was introduced and made possible for sequencing reactions to be performed directly at sample collection sites [28]. The device is an example of a fourth generation sequencing technique and operates following the Sanger sequencing principle. But the main difference between these techniques is that the MinION sequencing uses array-based technologies, whereas the latter uses separate steps for sequencing, separation and detection. These whole-genome sequencing methods have made a huge development in basic research, applied sciences, diagnostics and many other fields [16]. They have contributed to knowledge of bacterial genome sequences and genes responsible for diseases, and this data has in turn led to the development of vaccines.

MinION sequencing has made library preparation more accessible, and preparation can be done within 10 min, according to the instructions mentioned by the manufacturer. Then the DNA is loaded into the flow cell comprising nanopores in an electrically resistant membrane, where the DNA is disrupted, resulting in the generation of electrical signals that can be eventually converted

into nucleotide base pairs. The sequence is further trimmed to remove the adapters and errors as they run at a faster rate. They are then made to go through various steps like finding the overlaps, assembling based on de novo assembly and sequence polishing to reduce the error rates. The final sequence can be annotated to know about the genes present and their associated functions. These can be performed with available tools running in real-time [29].

Another advanced sequencing technology that is an extended version of MinION is PromethION [30]. This bench top instrument provides 48 individual flow cells with over 3000 nanopore channels. It also provides user accessibility, which helps operators make any changes during the sequencing. This instrument runs 500 base pairs per second and achieves the highest sequencing accuracy. Voltrax is an automated library preparation that reduces human errors [31]. Research in developing novel sequencing technologies is currently performed, which enable the study of various microorganisms and their mechanisms.

The first complete genomes of *Haemophilus influenza* [32] and *Mycoplasma genitalium* [33] were sequenced in 1995. In earlier days, the genomes could not be sequenced in a single reaction and required more time to sequence individual fragments, leading to read mapping and de novo assembly. But the error rate was high because of the repetitive sequences and accuracy problems. That is why the next generation of sequencing started to dominate the world. New techniques can be used to sequence long-read sequence in a shorter time, are of low cost and do not require an amplification step before the sequencing procedure. Many bacterial genomes have been sequenced over the years and the sequencing data were submitted to GenBank.

6. Sequencing of Various Strains of *Leptospira* sp. with Special Reference to Genome Features

The genome of *Leptospira* sp. comprises a large chromosome I and a mini chromosome II [34, 35]. Nearly 450 serovars of *Leptospira* sp. have been sequenced, and the whole genome sequence information is available in the public NCBI database [36, 37]. Whole-genome sequencing was performed with Illumina technology on a DNA isolate of a patient in Yunnan, China and sequencing data revealed that >95% similarity was obtained with *L. weilii* serovar Heyan strain L 231 [38]. Similarly, *L. interrogans* serovar Icterohaemorrhagiae strain Lai was collected from the National Reference Laboratory for *Leptospira* at NIFDC, China [39]. After processing the sequenced data, the draft genomes were assembled using *L. interrogans* gene sequence as a reference standard for studying their pathogenic mechanisms. The virulence factors present in these strains were analyzed using various methods and it was found that lipopolysaccharides play a significant role in causing the disease, and their composition varies in pathogenic, intermediate and saprophytic leptospires [40]. The analysis showed that 2617 genes were orthologous with *L. interrogans*, 2417 were orthologous with *L. borgpetersenii* and 1907 were orthologous with *L. biflexa* [38]. Furthermore, researchers found three riboswitches for vitamin B12 biosynthesis transport, indicating that these operons were fully complete only in pathogenic and not in saprophytic, and might help in causing disease and adaptation to different environments and hosts. More than 30 coding sequences (CDSs) were found for motility and chemotaxis, and many secretory pathways were also found. The type II secretory system helps transport toxins and hydrolytic enzymes across the cell membrane facilitating pathogenesis. It was found that even within the same serovar groups, these strains were distinct, indicating that the disease caused by the pathogenic species were strain dependent. Also, the comparison of specific genome features among the pathogenic, intermediate, and saprophytic strains were identified.

6.1 Genome features in saprophytic leptospires

Saprophytic species are non-pathogenic, and they live in aquatic environments. Among the saprophytic leptospires, *Leptospira biflexa* serovar Patoc strain Patoc1 (Ames) was first sequenced by the whole-genome shotgun sequencing method [41]. The sequence information of strain Patoc 1 (Ames) was compared with two other pathogenic strains, namely *L. interrogans* and *L. borgpetersenii*, and it was found that more than one-third of the features present in pathogenic species were absent in saprophytic species, indicating that they could not be involved in causing leptospirosis. Biofilm formation is also observed in *L. biflexa* and *L. interrogans,* both thriving for survival in the environment with solid biofilms [42], whereas biofilm formation was not observed in strains of *L. borgpetersenii*. Since *L. biflexa* is a fast-growing species, it was believed that the number of tRNAs would be high when compared with other strains [43]. But the assumption was incorrect, and the tRNA numbers were lower than other species, confirming that the growth depends not on tRNAs but on the metabolism used for their survival [44]. Even though this strain contains a putative hemolysin, the genome that codes for the enzyme to degrade the tissues is absent. But the genes responsible for DNA, RNA metabolism, intermediary metabolism, protein secretions and cell structure are all the same in every species, indicating that the virulence gene might have evolved at later stages.

6.2 Genome features in intermediate or semi-pathogenic species

Intermediate strains of *Leptospira* sp. are found between the saprophytic and pathogenic strains, indicating that the pathogenic strains share the significant genes responsible for virulence [45]. These species that are found to be in between saprophytes and pathogens share most of the genes with both species. The genomes of two intermediate strains, *L. licerasiae* serovar Varillala strain VAR010 and *L. licerasiae* strain MMD0835, were sequenced using the Illumina Solexa genome analyzer platform [46], and the genome features were studied and found to have maximum similarity and comprising several proteins related to nitrogen, amino acid and carbohydrate metabolism which might enable the bacteria to grow better in artificial media than the pathogenic species. For instance, the gene *ilvA* encoding threonine ammonia-lyase is present only in saprophytic leptospires and not in pathogenic leptospires [46]. These genes were further examined for riboswitches and other regulatory elements. A cis-regulatory element presents only in an intermediate strain of *L. licerasiae*, ydaO-yuaA complex, was found to be expressed during osmotic shock. It was found that more than an average number of the proteins found in *L. licerasiae* were similar to those in *L. interrogans*, suggesting that the intermediate species were closely related to pathogenic strains, rather than saprophytic strains. Recent findings have proven that O antigen synthesis is unique in *L. licerasiae* and no orthologs have been identified in *L. interrogans* and *L. biflexa*, confirming its nature of semi-pathogenicity. The *rfb* locus is located in pathogenic strains and a few intermediate strains of *L. inadai*, *L. broomii* and *L. fainei* between a copper-binding protein on the left and the ribosomal protein S6 on the right [47]. Hence, the researchers concluded that intermediate strains are moving towards pathogenic and semi-pathogenic or intermediate leptospiral strains.

6.3 Genome features in pathogenic leptospires

Pathogenic species are responsible for causing infectious disease, and they induce many clinical complications in humans. Sequencing was performed with Illumina and Single-Molecule Real-Time (SMRT) sequencing technologies on six strains of *L. interrogans* obtained from Malaysia [48]. The strains chosen were *L. interrogans* serovar Lai strain Langkawi, *L. interrogans* serovar Copenhageni strain 898, *L. interrogans* serovar Bindjei strain 782, *L. interrogans* serovar Paidjan strain 1489, *L.*

interrogans serovar Bataviae strain 1548 and *L. interrogans* serovar Yeoncheon strain 1530. The genes present in these pathogenic strains were analyzed and the virulence factors responsible for causing infection were found. Predominantly, the components responsible for causing virulence like sialic acid, biofilm formation and lipid A biosynthesis were identified in pathogenic strains [49]. The researchers discovered that the alginate biosynthesis pathway was complete only in pathogenic leptospires and not in saprophytic and intermediate leptospires.

The sialic acid biosynthesis with unique features like O-acetyltransferase is present only in *L. alexanderi* and *L. borgpetersenii* and these species contain nucleoside-bisphosphate-sugar epimerase [48]. Only the pathogenic species *L. santarosai* has N-acetylneuraminic acid with the phosphoglyceric dehydrogenase domain whereas intermediate and saprophytic species lack this gene. The genes encoding for molecules that adhere to the extracellular matrix suggest that these genes are based on survival and mammalian adaptation to the environment. Three genes encoded for adhesive molecules present only in pathogenic species are *Lsa30*, *Lsa44* and *Mfn6*. Moreover, virulence proteases like collagenase and thermolysins that avoid complement and extracellular matrix degradation via metalloproteases were found. These proteases are found in neither intermediate nor saprophytic species.

Three enzymes responsible for dealing with the oxidative stress in the host are catalases, peroxidases and superoxide dismutases. The catalases are present only in pathogenic species, indicating that they play a role in helping the bacteria to live inside the mammalian host. Superoxide dismutase is absent in pathogenic species and is present only in saprophytic species indicating that these species might not have been exposed to oxygen radicals in the environment and may have developed an alternate strategy to remove the oxygen radicals.

CRISPR/cas systems, which are only present in pathogenic species, are used for the immune system in bacteria to protect them from foreign antigens, and because of the cas system presence, the prophages are incomplete in pathogenic and intermediate species [50]. The *MtrD* efflux pump, a potential defensive mechanism present in *Neisseria gonorrhea* [51], is also present in the pathogenic species, indicating that this gene can support growth under hostile conditions. *LipL41*, a major outer membrane protein highly expressed in pathogenic species, is recognized as a heme-binding protein [52]. *LipL32* is also highly conserved in pathogenic, and it is expressed during infection. Ramli *et al*. [53] concluded that *L. interrogans* is a dynamic pathogen and there are still many virulence features to be discovered.

One among the pathogenic strains is *Leptospira santarosai* which is comparatively less virulent than *L. interrogans*. This strain was reported to cause a leptospirosis outbreak in Taiwan from 2001 to 2006, and 291 patients suffered from renal dysfunctions and systemic dissemination. The *L. santarosai* serovar Shermani strain LT821 was obtained from ATCC and sequenced using Illumina sequencing technology [54]. The raw data were processed using tools like MAQ for assembling the sequence [55], and with annotation using PGAP. The genome feature of this strain consists of 111 contigs and 3,936,333 bp, with an N50 read of 97,507 bp [56]. The GC content of the sequence was 41.82% and had 4033 CDS, 37 tRNAs and 5 rRNAs. Then this strain was typed using multilocus sequence typing (MLST) based on the seven housekeeping genes, *fadD*, *sucA*, *mreA*, *pfkB*, *pntA*, *tpiA* and *glmU* [57]. The LT821 strain was compared for genomic analysis with other strains, including *L. interrogans* serovar Lai str. 56601, *L. interrogans* serovar Lai str. IPAV, *L. interrogans* serovar Copenhageni str. Fiocruz L (1-130), *L. borgpetersenii* serovar Hardjo-bovis L 550, *L. borgpetersenii* serovar Hardjo-bovis JB 197 and *L. santarosai* serovar Shermani strain LT 821. The primers used for *pntA*, *mreA* and *tpiA* failed to provide results for PCR amplification for *L. santarosai* and this meant other ways to distinguish the strain were needed. Similarly, many virulence factors were identified, such as sphingomyelinase, motility-associated genes, and chemotaxis, ankyrin repeat domains that are involved in antibiotic resistance, nutrient acquisition, and leucine-rich repeat, which are involved in penetration of host and survival. The findings also

revealed that *L. borgpetersenii* and *L. santarosai* were closely related after mapping and genotyping, whereas *L. interrogans* and *L. santarosai* were closely related via genomics and blastX analysis.

Leptospirosis can be severe in some cases, and the bacteria can cause infection in the host immune system, leading to the failure in adaptive and innate immune response in the host and even leading to fatal conditions. Researchers have developed many potential diagnostic biomarkers to identify the disease to avoid these conditions. Biomarkers like RANTES, and CHI3L1 can be used in which RANTES serves as a chemokine, and they recruit immune cells to the site of infection. Recent findings prove that a high nitric oxide level in patients causes severe infection, and tissue and organ damage [58]. Pentraxins are some conserved proteins that are known to be involved in acute responses, and if these protein levels are found to be higher, they may cause severe disease, and so can be used to monitor the disease severity level [59]. Another biomarker is cathelicidin, which is conserved in the immune system and shows antimicrobial activity. It is involved in inducing neutrophil and phagocytosis, and if the level of cathelicidin decreases, the severity of the leptospires increases, perhaps leading to death. So, the cathelicidin can be used as an indicator or a potential therapeutic agent for this disease [60].

6.3.1 Outbreak of leptospirosis in Malaysia

During the outbreak of leptospirosis in Malaysia, seven intermediate and saprophyte species from the environment were isolated, namely *L. semungkisensis* strain SSS9*, L. fletcheri* strain SSW15, *L. langatensis* strain SSW18, *L. selangorensis* strain SSW17, *L. jelokensis* strain L5S1, *L. perdikensis* strain HP2, *L. congkakensis* strain SCS9 and three pathogenic species namely *L. weilii* strain SC295, *L. interrogans* strain HP358 and *L. borgpetersenii* strain HP364 from humans. These strains were sequenced using Illumina sequencing technology, and the raw data were processed using various platforms. The clinical infection of these strains was observed in the hamster, which depicts that only the *L. interrogans* caused fatal illness in the hamster, whereas *L. weilii* showed weight loss and became inactive, and *L. borgpetersenii* showed no clinical infection. The virulence factors present in the pathogenic species were found using VFDB and BLAST. The *LipL32*, Lig A-like immunoglobulin, putative lipoprotein and sphingomyelinase C were not found in intermediate and saprophytic species [61]. The presence of the *cirA* gene in *L. interrogans* revealed the impairment of antibiotic resistance in the species, and *SspH1,* also known as effector protein, helps in bacterial survival and accounts for proteasomal degradation in the host. It was concluded that *L. interrogans* was the most virulent among all other strains, and its virulence factors are yet to be experimentally studied [62].

6.3.2 Various gene elements present in pathogenic leptospires and their functions

After sequencing, the pathogenic strain *Leptospira* was then analyzed to know about the virulence factors present in it, and to determine the genes responsible for causing infection. Magnesium is required for the functioning of several organs and if there is any damage in the Loop of Henle, which is the primary site for magnesium absorption, the kidney can be damage and its function affected. The strain's virulent genes account for various factors like movement inside the host tissue, biofilm matrix formation, and invasion of the host tissues. The genes predominantly present in the pathogenic strains are CRISPR-associated genes, cas system genes, an adaptive immune system which can protect the bacteria from bacteriophages and foreign genetic elements. Included is the *algA* gene and its alginate biosynthesis protein, which is used for biofilm formation. The previous major effort to study the pathogenesis was solely based on analyzing the protein-coding sequences and developing diagnostic techniques. But later, the search for the virulence factors became more widespread, including research into non-coding sequences, CRISPR-associated regions, and riboswitches. Riboswitches are the non-coding tRNA motifs present in the 5’ untranslated regions

responsible for the activation of gene expression. They are involved in the synthesis of vitamins like cobalamin and thiamine. Cobalamin is transported using a Ton B-like receptor suitable for the potential vaccine target because this receptor presents many surface-exposed epitopes [34].

The sphingomyelinase gene is also a major virulence factor that can lyse the erythrocytes and cause destruction in the host tissue. The *sph_2* gene is expressed in humans during infection and the iron in RBC cells is absorbed by bacteria for their growth and survival within the host. Another possibility is that if iron is depleted in the blood, sphingomyelinase activity increases. Among the various paralogs of sphingomyelinases, sphingomyelinase C activity has been confirmed only for *sph2*, which can cleave into ceramide and phosphocholine. The genes encoding for sphingomyelinase are absent in saprophytic species, confirming their role in pathogenesis [63]. Hemolysins are also assumed to assist the pathogenic species in acquiring the iron from the host. They are responsible for producing cytokines during infection in humans, inducing Interleukin - 1β, and TNF-α, which are inflammatory cytokines produced during acute infection, and can cause apoptosis and necrosis. Heme is used as a growth source for *L. interrogans*, and heme oxygenase is also involved in utilizing oxygen and acts as a virulence-promoting factor [64].

The presence of the heat shock protein (*hspA*) also reveals that the survival of the bacteria within the host can protect the cell from any attack, and it is not influenced by environmental conditions [65]. Many chaperone proteins are present, which assist in protein folding and the avoidance of non-specific aggregation. Biofilm consists in surface-associated cells synthesized from some extracellular materials, consisting of some exopolymeric substances. The formation helps bacteria neutralize antibiotics, and the cell in biofilm can attach to both biotic and abiotic surfaces and multiply to form a microcolony. It also helps in the protection from the non-specific host defense during infection and adherence and contributes to long-term survival in environmental water resources. The *Pnp* gene is an indirect regulator of cyclic monophosphate nucleotides, and the key secondary messenger in biofilm formation [42].

The *LepA* gene helps in the survival of both pathogenic and saprophytic species in environmental conditions [66]. *LigA*, which belongs to the class of high molecular weight immunoglobulin molecules, is only present in pathogenic species, and is responsible for the adherence to host cells as it binds and interacts with fibronectin and collagenase. So, either the *LigA* or the *LigB* gene is required to cause infection. Calcium is an essential nutrient for binding with fibronectin using the *LipL* gene. Vitamin B12 and thiamine are required for bacterial growth and these genes were also found in this strain [54].

The outer membrane is speculated to be the cause of pathogenicity and is responsible for the adherence to the host. So, by studying the outer membrane, potential targets and vaccines are being developed. Some outer membrane components are lipopolysaccharides, lipoproteins and peptidoglycan, which are used for cell wall synthesis and account for pathogenesis. Some lipopolysaccharides are primary virulence factors due to their adherence and they can activate macrophages through the Toll-like receptor pathway in humans. But the LPS composition of pathogenic and intermediate species may vary. Peptidoglycan, which is present in large amounts in leptospires, is antigenic component that is tightly cross-linked, helping the leptospire to survive. So, these can be selected as targets for antibiotics [67].

Iron is a micronutrient responsible for host and pathogen survival for immune function and infectious disease. Neutrophils contain myeloperoxidase, an iron-dependent enzyme responsible for killing the bacteria, and a reduction in iron level causing a decrease in enzyme level [68]. The outer membrane proteins like *gpsA* and metXA are involved in the process of leptospiral infection. metXA targets three human proteins, namely AKAP8L (kinase anchor protein 8-like), B2M (beta-2-microglobulin), which is involved in the antibacterial humoral response, and TACC1 (transforming acidic coiled-coil containing protein 1), which aids in cell replication, gspA targets human protein HIVEP1-zinc which is involved in signal transduction [69]. Membrane proteins facilitate the adherence to the endothelial cell by binding with VE-cadherin, leading to endothelial damage and

causing hemorrhagic manifestation. Periplasmic proteins like *flaB* and *glyA* gene are conserved only in pathogenic and intermediate species. Some inner membrane proteins like *FeoB*, *htpX*, *uppP* and outer membrane proteins like *xseA*, *pgi* genes are only present in pathogenic species but not in intermediate and saprophytic species. Some bacterial proteins like *clpB*, *trpB*, *secA*, *leuB*, *carA*, *fusA* and *purC* are involved in pathogen-host interaction that inactivates and damages host T cells. The *clpB* gene is responsible for survival under stress conditions such as oxidative and thermal conditions, and is involved in developing an infection inside the host [69].

The genes encoding for chemotaxis are highly conserved and amino acid sequence are more similar in pathogenic organisms than in the other two species. Each species behaves differently towards the chemotactic signals and MCPs (methyl accepting chemotaxis) protein-containing species that might trigger intracellular signal transduction. Some chemotaxis regulators, *cheA, cheR, cheB* and *cheY* have been identified, suggesting that these proteins might affect leptospiral survival and ability to infect the host [47]. The *hslU*, *hslV* genes are ATP-dependent protease subunits that contribute to survival and virulence during infection and disease transmission. The *lpxD* gene plays an important role in temperature adaptation and virulence [70]. Vitamins like cobalamin are only synthesized from L-glutamate present in pathogenic leptospiral strains because these strains are autotrophic, whereas the saprophytes are auxotrophic, which helps in iron absorption when lysing the red blood cells. *Loa22* has been proven to cause infection in the host, and is absent in saprophytic species since they do not have genes that encode tissue degrading capabilities [71]. The *fadA* gene responsible for adhesion in the host cell was also found. Still, many virulence factors responsible for the infection are being experimentally investigated and knowledge of these factors will be invaluable in the development of potential vaccines against leptospirosis.

7. Conclusions

Infection caused by *Leptospira* sp. can progress to multi-organ failure or even lead to death in humans due to the progression of acute infection to an asymptomatic level. Hence, there is a need to identify the genetic information of various strains of *Leptospira* sp. and to study the genetic characteristics of the pathogenic leptospires. These sequencing approaches can help clinicians investigate the virulence factors present in the pathogenic leptospires, and the genes responsible for causing disease. Currently, as there is a lot of advancement in sequencing technologies, the most accurate form of sequencing is essential with less processing time and fewer biological inputs. Among the sequencing approaches described in this review, and considering the advancements in sequencing technology, the best possible approach requires less availability of reagents, less sequencing run-time, less utilization of consumables, and more user-friendly instrumental approaches. Furthermore, advancements in the sequencing platforms should minimize the processing of the raw data sequences with minimal storage.

Considering the genome feature of the *Leptospira* genome, the virulence genes responsible for causing the infection were identified only in pathogenic leptospires and not in saprophytic leptospires. The genes that are mainly responsible for causing the disease like *hslU* - chaperone protein that can help in the survival of the bacteria inside the host and helps in the transmission of the disease, *tlyA gen*e, which encodes hemolysin that induces cytolysis in cells and causes death, and *sphingomyelinase C* that lyses the erythrocytes and causes cell membrane damage can be targeted, and inactivating the genes might be possible for reducing the virulence nature of the bacteria. Since this bacterium can also attach to the host, genes that encode for biofilm formation like *algA*, outer membrane protein such as *loa22*, *lipL32* that adhere to the host cell, and immunoglobulin-like proteins like *ligA*, *ligB*, *ligC* that can cause infection, can be targeted to reduce their attachment to host thus reducing infection. Antibiotics are also a good option for preventing

the spread of disease and slowing the growth of bacteria. Antibiotic resistance genes like *cirA*, tetracycline and many multidrug resistance genes are present, making the narrowing the target genes more challenging. Hence, there is a need to improve the functional study of genes and develop medicines against leptospirosis.

8. Acknowledgements

The authors thank Mepco Schlenk Engineering College (Autonomous), Sivakasi, for providing permission to conduct the present study.

References

[1] Shen, Y., Nie, J., Kuang, L., Zhang, J. and Li, H., 2021. DNA sequencing, genomes and genetic markers of microbes on fruits and vegetables. *Microbial Biotechnology,* 14(2), 323-362.

[2] Pareek, C.S., Smoczynski, R. and Tretyn, A., 2011. Sequencing technologies and genome sequencing. *Journal of applied genetics*, 52 (4), 413-435.

[3] Verma, S. and Gazara, R.K., 2021. Next-generation sequencing: an expedition from workstation to clinical applications. In: K. Raza and N. Dey, eds. *Translational Bioinformatics in Healthcare and Medicine*. London: Academic Press, pp. 29-47.

[4] Bambini, S. and Rappuoli, R., 2009. The use of genomics in microbial vaccine development. *Drug Discovery Today*, 14(5), 252-260.

[5] Kukurba, K.R. and Montgomery, S.B., 2015. RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 2015(11), 951-969.

[6] Ungelenk, M., 2021. Sequencing approaches. In: T. Liehr, ed. *Cytogenomics*. London: Academic Press, pp. 87-122.

[7] Maxam, A.M. and Gilbert, W., 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), 560-564.

[8] Lehmann, A.R., 1997. Technologies for Detection of DNA Damage and Mutations. *Journal of Medical Genetics*, 34(5), 438-438.

[9] Vilgis, S. and Deigner, H.P., 2018. Sequencing in precision medicine. In: H.P. Deigner and M. Kohl, eds. *Precision Medicine*. London: Academic Press, pp. 79-101.

[10] Borsting, C. and Morling, N., 2015. Next generation sequencing and its applications in forensic genetics. *Forensic Science International Genetics*, 18, 78-89.

[11] Berg, P., 2014. Fred Sanger: a memorial tribute. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3), 883-884.

[12] Sanger, F. and Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Bioliogy*, 94(3), 441-448.

[13] Sanger, F., Nicklen, S. and Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467.

[14] Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Wang, J., Bao, Q., Kong, H. and Hu, S., 2005. Complete genome sequence of bacteriophage T5. *Virology*, 332(1), 45-65.

[15] Crossley, B.M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M.L., Clement, T. and Toohey-Kurth, K., 2020. Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, 32(6), 767-775, DOI: 10.1177/1040638720905833.

[16] Heather, J.M. and Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1-8.
[17] Hood, L.E., Hunkapiller, M.W. and Smith, L.M., 1987. Automated DNA sequencing and analysis of the human genome. *Genomics*, 1(3), 201-212.
[18] Shetty, P.J., Amirtharaj, F. and Shaik, N.A., 2019. Introduction to nucleic acid sequencing. In: N.A. Shaik, K.R. Hakeem, B. Banaganapalli and R. Elango, eds. *Essentials of Bioinformatics*. Volume I. Cham: Springer, pp. 97-126.
[19] Nyrén, P. and Lundin, A., 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical Biochemistry*, 151(2), 504-509.
[20] Gansauge, M.-T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A. and Meyer, M., 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research*, 45(10), 79-79.
[21] Shendure, J. and Ji, H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135-1145.
[22] Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11(1), 31-46.
[23] Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M., 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012, DOI: 10.1155/2012/251364.
[24] Schadt, E.E., Turner, S. and Kasarskis, A., 2010. A window into third-generation sequencing. *Human Molecular Genetics*, 19(2), 227-240.
[25] Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S. and Turner, S.W., 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research*, 38(15), DOI: 10.1093/nar/gkq543.
[26] Koren, S., Harhay, G.P., Smith, T.P., Bono, J.L., Harhay, D.M., McVey, S.D., Radune, D., Bergman, N.H. and Phillippy, A.M., 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biology*, 14(9), DOI: 10.1186/gb-2013-14-9-r101.
[27] Petersen, L.M., Martin, I.W., Moschetti, W.E., Kershaw, C.M. and Tsongalis, G.J., 2019. Third-generation sequencing in the clinical laboratory: Exploring the advantages and challenges of nanopore sequencing. *Journal of Clinical Microbiology*, 58(1), 01315-01319.
[28] Jain, M., Olsen, H.E., Paten, B. and Akeson, M., 2016. The Oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1), DOI: 10.1186/s130593-16-1103-0.
[29] Wang, Y., Zhao, Y., Bollas, A., Wang, Y. and Au, K.F., 2021. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348-1365.
[30] De Coster, W., De Rijk, P., De Roeck, A., De Pooter, T., D'Hert, S., Strazisar, M., Sleegers, K. and Van Broeckhoven, C., 2019. Structural variants identified by Oxford nanopore PromethION sequencing of the human genome. *Genome Research*, 29(7), 1178-1187.
[31] Hess, J.F., Kohl, T.A., Kotrová, M., Rönsch, K., Paprotka, T., Mohr, V., Hutzenlaub, T., Brüggemann, M., Zengerle, R., Niemann, S. and Paust, N., 2020. Library preparation for next generation sequencing: A review of automation strategies. *Biotechnology Advances*, 41, DOI: 10.1016/j.biotechadv.2020.107537.
[32] Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. and Venter, J.C., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science*, 269(5223), 496-512.

[33] Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A. and Venter, J.C., 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science*, 270(5235), 397-403.

[34] Jorge, S., Kremer, F.S., Oliveira, N.R.d., Navarro, G.d.O.S.V., Guimarães, A.M., Sanchez, C.D., Woloski, R.D.D.S., Ridieri, K.F., Campos, V.F., Pinto, L.d.S. and Dellagostin, O.A., 2018. Whole-genome sequencing of *Leptospira interrogans* from southern Brazil: genetic features of a highly virulent strain. *Memorias do Instituto Oswaldo Cruz*, 113(2), 80-86.

[35] Nascimento, A.L., Verjovski-Almeida, S., Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E., Digiampietri, L.A., Harstkeerl, R.A., Ho, P.L., Marques, M.V., Oliveira, M.C., Setubal, J.C., Haake, D.A. and Martins, E.A., 2004. Genome features of *Leptospira interrogans* serovar Copenhageni. *Brazilian Journal of Medical and Biological Research*, 37(4), 459-477.

[36] Guglielmini, J., Bourhy, P., Schiettekatte, O., Zinini, F., Brisse, S. and Picardeau, M., 2019. Genus-wide *Leptospira* core genome multilocus sequence typing for strain taxonomy and global surveillance. *PLoS Neglected Tropical Diseases*, 13(4), DOI: 10.1371/journal.pntd.0007374.

[37] Fraga, T.R., Carvalho, E., Isaac, L. and Barbosa, A.S., 2015. *Leptospira* and Leptospirosis. In: Y.-W. Tang, M. Sussman, D. Liu, I. Poxton and J. Schwartzman, eds. *Molecular Medical Microbiology*. Boston: Academic Press, pp. 1973-1990.

[38] Xu, Y., Zheng, H., Zhang, Y., Wang, Y., Zhang, J., Li, Z., Cui, S., Xin, X., Ye, Q., Chang, Y.F. and Wang, J., 2017. Genomic analysis of a new serovar of *Leptospira weilii* serogroup Manhao. *Frontiers in Microbiology*, 8, DOI: 10.3389/fmicb.2017.00149.

[39] He, P., Sheng, Y.Y., Shi, Y.Z., Jiang, X.G., Qin, J.H., Zhang, Z.M., Zhao, G.P. and Guo, X.K., 2007. Genetic diversity among major endemic strains of *Leptospira interrogans* in China. *BMC Genomics*, 8, DOI: 10.1186/1471-2164-8-204.

[40] Evangelista, K.V. and Coburn, J., 2010. *Leptospira* as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. *Future Microbiology*, 5(9), 1413-1425.

[41] Picardeau, M., Bulach, D.M., Bouchier, C., Zuerner, R.L., Zidane, N., Wilson, P.J., Creno, S., Kuczek, E.S., Bommezzadri, S., Davis, J.C., McGrath, A., Johnson, M.J., Boursaux-Eude, C., Seemann, T., Rouy, Z., Coppel, R.L., Rood, J.I., Lajus, A., Davies, J.K., Médigue, C. and Adler, B., 2008. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis. *PloS ONE*, 3(2), DOI: 10.1371/journal.pone.0001607.

[42] Ristow, P., Bourhy, P., Kerneis, S., Schmitt, C., Prevost, M.C., Lilenbaum, W. and Picardeau, M., 2008. Biofilm formation by saprophytic and pathogenic leptospires. *Microbiology*, 154(5), 1309-1317.

[43] Picardeau, M., 2018. Toolbox of molecular techniques for studying *Leptospira* spp. *Current Topics in Microbiology and Immunology*, 415, 141-162.

[44] Wei, Y., Silke, J.R. and Xia, X., 2019. An improved estimation of tRNA expression to better elucidate the coevolution between tRNA abundance and codon usage in bacteria. *Scientific Reports*, 9(1), DOI: 10.1038/s41598-019-39369-x.

[45] Vincent, A.T., Schiettekatte, O., Goarant, C., Neela, V.K., Bernet, E., Thibeaux, R., Ismail, N., Mohd Khalid, M.K.N., Amran, F., Masuzawa, T., Nakao, R., Amara Korba, A., Bourhy, P., Veyrier, F.J. and Picardeau, M., 2019. Revisiting the taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. *PLoS Neglected Tropical Diseases*, 13(5), DOI: 10.1371/journal.pntd.0007270.

[46] Ricaldi, J.N., Fouts, D.E., Selengut, J.D., Harkins, D.M., Patra, K.P., Moreno, A., Lehmann, J.S., Purushe, J., Sanka, R., Torres, M., Webster, N.J., Vinetz, J.M. and Matthias, M.A., 2012.

Whole genome analysis of *Leptospira licerasiae* provides insight into leptospiral evolution and pathogenicity. *PLoS Neglected Tropical Diseases*, 6(10), DOI: 10.137/journal.pntd.000 1853.

[47] Fouts, D.E., Matthias, M.A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D.E., Bulach, D., Buschiazzo, A., Chang, Y.-F., Galloway, R.L., Haake, D.A., Haft, D.H., Hartskeerl, R., Ko, A.I., Levett, P.N., Matsunaga, J., Mechaly, A.E., Monk, J.M., Nascimento, A.L.T., Nelson, K.E., Palsson, B., Peacock, S.J., Picardeau, M., Ricaldi, J.N., Thaipandungpanit, J., Wunder, E.A., Jr., Yang, X.F., Zhang, J.-J. and Vinetz, J.M., 2016. What makes a bacterial species pathogenic?: Comparative genomic analysis of the genus *Leptospira*. *PLoS Neglected Tropical Diseases*, 10(2), DOI: 10.137/journal.pntd.0004403.

[48] Ramli, S.R., Moreira, G.M.S.G., Zantow, J., Goris, M.G.A., Nguyen, V.K., Novoselova, N., Pessler, F. and Hust, M., 2019. Discovery of *Leptospira* spp. seroreactive peptides using ORFeome phage display. *PLoS Neglected Tropical Diseases*, 13(1), DOI: 10.137/journal.pntd.0007131.

[49] Sudhakara, P., Sellamuthu, I. and Aruni, A.W., 2019. Bacterial sialoglycosidases in virulence and pathogenesis. *Pathogens*, 8(1), DOI: 10.3390/pathogens8010039.

[50] Louwen, R., Staals, R.H.J., Endtz, H.P., van Baarlen, P. and van der Oost, J., 2014. The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiology and Molecular Biology Reviews*, 78(1), 74-88.

[51] Handing, J.W., Ragland, S.A., Bharathan, U.V. and Criss, A.K., 2018. The MtrCDE efflux pump contributes to survival of *Neisseria gonorrhoeae* from human neutrophils and their antimicrobial components. *Frontiers in Microbiology*, 9, DOI: 10.3389/fmicb.2018.02688.

[52] Hsu, S.H. and Yang, C.W., 2022. Insight into the Structure, Functions, and Dynamics of the Leptospira Outer Membrane Proteins with the Pathogenicity. *Membranes (Basel)*, 12(3), DOI: 10.3390/membranes12030300.

[53] Ramli, S.R., Bunk, B., Spröer, C., Geffers, R., Jarek, M., Bhuju, S., Goris, M., Mustakim, S. and Pessler, F., 2021. Complete genome sequencing of *Leptospira interrogans* isolates from Malaysia reveals massive genome rearrangement but high conservation of virulence-associated genes. *Pathogens*, 10(9), DOI: 10.3390/pathogens10091198.

[54] Chou, L.F., Chen, Y.T., Lu, C.W., Ko, Y.C., Tang, C.Y., Pan, M.J., Tian, Y.C., Chiu, C.H., Hung, C.C. and Yang, C.W., 2012. Sequence of *Leptospira santarosai* serovar Shermani genome and prediction of virulence-associated genes. *Gene*, 511(2), 364-370.

[55] Scheibye-Alsing, K., Hoffmann, S., Frankel, A., Jensen, P., Stadler, P.F., Mang, Y., Tommerup, N., Gilchrist, M.J., Nygård, A.B., Cirera, S., Jørgensen, C.B., Fredholm, M. and Gorodkin, J., 2009. Sequence assembly. *Computational Biology and Chemistry*, 33(2), 121-136.

[56] Chou, L.-F., Chen, T.-W., Ko, Y.-C., Pan, M.-J., Tian, Y.-C., Chiu, C.-H., Tang, P., Hung, C.-C. and Yang, C.-W., 2014. Potential impact on kidney infection: a whole-genome analysis of *Leptospira santarosai* serovar Shermani. *Emerging Microbes and Infection*, 3(1), DOI: 10.1038/emi.2014.78.

[57] Ahmed, N., Devi, S.M., Valverde Mde, L., Vijayachari, P., Machang'u, R.S., Ellis, W.A. and Hartskeerl, R.A., 2006. Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. *Annals of Clinical Microbiology and Antimicrobials*, 5, DOI: 10.1186/1476-0711-5-28.

[58] Chin, V.K., Lee, T.Y., Lim, W.F., wan Shahriman, Y.W.Y., Syafinaz, A.N., Zamberi, S., Maha, A., 2018. Leptospirosis in human: Biomarkers in host immune responses. *Microbiological Research*, 207, 108-115.

[59] Wagenaar, J.F., Goris, M.G., Gasem, M.H., Isbandrio, B., Moalli, F., Mantovani, A., Boer, K.R., Hartskeerl, R.A., Garlanda, C. and van Gorp, E.C., 2009. Long pentraxin PTX3 is associated with mortality and disease severity in severe Leptospirosis. *Journal of Infection*, 58(6), 425-432.

[60] Lindow, J.C., Wunder, E.A., Jr., Popper, S.J., Min, J.N., Mannam, P., Srivastava, A., Yao, Y., Hacker, K.P., Raddassi, K., Lee, P.J., Montgomery, R.R., Shaw, A.C., Hagan, J.E.,

Araújo, G.C., Nery, N., Jr., Relman, D.A., Kim, C.C., Reis, M.G. and Ko, A.I., 2017. Correction: Cathelicidin insufficiency in patients with fatal leptospirosis. *PLoS Pathogens*, 13(9), DOI: 10.1371/journalppat.1006646.

[61] Hauk, P., Macedo, F., Romero, E.C., Vasconcellos, S.A., de Morais, Z.M., Barbosa, A.S. and Ho, P.L., 2008. In LipL32, the major leptospiral lipoprotein, the C terminus is the primary immunogenic domain and mediates interaction with collagen IV and plasma fibronectin. *Infection and immunity*, 76(6), 2642-2650.

[62] Philip, N., Jani, J., Azhari, N.N., Sekawi, Z. and Neela, V.K., 2021. *In vivo* and *in silico* virulence analysis of *Leptospira* species isolated from environments and rodents in leptospirosis outbreak areas in Malaysia. *Frontiers in Microbiology*, 12, DOI: 10.3389/fmicb.2021.753328.

[63] Narayanavari, S.A., Lourdault, K., Sritharan, M., Haake, D.A. and Matsunaga, J., 2015. Role of *sph2* Gene Regulation in Hemolytic and Sphingomyelinase Activities Produced by *Leptospira interrogans*. *PLoS Neglected Tropical Diseases*, 9(8), DOI: 10.1371/journal.pntd.0003952.

[64] Murray, G.L., Srikram, A., Henry, R., Puapairoj, A., Sermswan, R.W. and Adler, B., 2009. *Leptospira interrogans* requires heme oxygenase for disease pathogenesis. *Microbes and Infection*, 11(2), 311-314.

[65] Guernier, V., Allan, K.J. and Goarant, C., 2018. Advances and challenges in barcoding pathogenic and environmental *Leptospira*. *Parasitology*, 145(5), 595-607.

[66] Ghazaei, C., 2018. Pathogenic *Leptospira*: Advances in understanding the molecular pathogenesis and virulence. *Open Veterinary Journal*, 8(1), 13-24.

[67] Haake, D.A. and Matsunaga, J., 2021. Leptospiral immunoglobulin-like domain proteins: roles in virulence and immunity. *Frontiers in Immunology*, 11, DOI: 10.3389/fimmu.2020.579907.

[68] Herman, H.S., Mehta, S., Cárdenas, W.B., Stewart-Ibarra, A.M. and Finkelstein, J.L., 2016. Micronutrients and leptospirosis: A review of the current evidence. *PLoS Neglected Tropical Diseases*, 10(7), DOI: 10.137/journal.pntd.0004652.

[69] Kumar, S., Lata, K.S., Sharma, P., Bhairappanavar, S.B., Soni, S. and Das, J., 2019. Inferring pathogen-host interactions between *Leptospira interrogans* and *Homo sapiens* using network theory. *Scientific Reports*, 9(1), DOI: 10.1038/s41598-018-38329-1.

[70] Eshghi, A., Henderson, J., Trent, M.S. and Picardeau, M., 2015. *Leptospira interrogans lpxD* homologue is required for thermal acclimatization and virulence. *Infection and Immunity*, 83(11), 4314-4321.

[71] 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*, 7(10), 736-747.