
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/142022/

Deposited on: 09 June 2017

Enlighten – Research publications by members of the University of Glasgow

http://eprints.gla.ac.uk
Advances and challenges in barcoding pathogenic and environmental Leptospira

Vanina Guernier¹, Kathryn J. Allan², Cyrille Goarant³

1. Australian Institute of Tropical Health and Medicine, James Cook University, 1 James Cook Drive, Townsville, QLD 4811, Australia
2. Institute of Biodiversity Animal Health and Comparative Medicine, College of Medical Veterinary and Life Sciences, University of Glasgow, University Avenue, Glasgow, G12 8QQ, UK
3. Institut Pasteur in New Caledonia, Institut Pasteur International Network, Leptospirosis Research and Expertise Unit, Noumea, New Caledonia. 11 rue Paul Doumer, BP 61; 98845 Noumea cedex, New Caledonia

* Author for correspondence: cgoarant@pasteur.nc
Phone number: +687-27.75.31

Running title: Barcoding Leptospira
SUMMARY
Leptospirosis is a zoonotic bacterial disease of global importance. A large spectrum of asymptomatic animal hosts can carry the infection and contribute to the burden of human disease. Environmental sources in human contaminations also point to the importance of a hydro-telluric reservoir. Leptospirosis can be caused by as many as 15 different pathogenic or intermediate *Leptospira* species. However, classification of these bacteria remains complicated through the use of both serological and genetic classification systems that show poor correlation. With the advent of molecular techniques, DNA-based barcoding offers a conceptual framework that can be used for leptospirosis surveillance as well as source tracking. In this review, we summarize some of the current techniques, highlight significant successes and weaknesses and point to the future opportunities and challenges to successfully establish a widely applicable barcoding scheme for *Leptospira*.

KEYWORDS
Leptospira, epidemiology, zoonotic reservoir, barcoding, genotyping
INTRODUCTION

Leptospirosis is a zoonotic disease with a worldwide distribution. An estimated 1.03 million human cases and almost 60,000 deaths occur annually (Costa et al., 2015). Human leptospirosis is caused by infection with pathogenic *Leptospira* spp. bacteria through indirect exposure to water or moist environments contaminated with the urine of infected animals, or through direct contact with infected animals or their tissues. Although it is recognized as an emerging infectious disease, leptospirosis is also considered a neglected disease that places its greatest burden on impoverished populations from developing countries and tropical regions (McBride et al., 2005). Human leptospirosis is often diagnosed late, due to its broad spectrum of signs and symptoms that range from a flu-like syndrome to multi-organ failure, and because the clinical presentation of leptospirosis mimics that of many other diseases, including dengue fever, chikungunya and malaria. Laboratory confirmation of a clinical suspicion of leptospirosis is therefore essential to ensure optimal patient care (Goarant, 2016; Hartskeerl & Smythe, 2015).

The genus *Leptospira* belongs to the phylum of Spirochaetes, order Spirochaetales, family Leptospiraceae (Paster et al., 1991). Two classification schemes are used for leptospires, one of which is based on serology with the serovar as the basic taxon, and another which uses molecular taxonomy to identify the *Leptospira* species, sometimes referred to as genomospecies (Levett, 2001). Serological classification is based on antigenic determinants related to the structural heterogeneity of the lipopolysaccharide (LPS), the major component of the outer membrane of Gram-negative bacteria (Bharti et al., 2003). *Leptospira* have been classified serologically into 26 serogroups and over 300 serovars (both saprophytic and pathogenic) using Microscopic Agglutination Test (MAT) and Cross Agglutination Absorption Test (CAAT) respectively (Cerqueira & Picardeau, 2009; Hartskeerl & Smythe, 2015).

Phylogenetically, 22 species of the *Leptospira* genus have been described so far, based on 16S rRNA phylogeny and DNA-DNA hybridization (until recently the gold-standard for defining bacterial species) (Fouts et al., 2016). Those species are arranged into three large groups based on their pathogenicity: pathogenic species (*Leptospira interrogans*, *L. kirschneri*, *L. borgpetersenii*, *L. mayottensis*, *L. santarosai*, *L. noguchii*, *L. weilii*, *L. alexanderi*, *L. kmetyi*, and *L. alstonii*), intermediate (i.e. species of unclear or low pathogenicity: *L. broomii*, *L. fainei*, *L. inadai*, *L. licerasiae*, *L. wolffii*), and saprophytic species (i.e. free-living organisms found in water and soil and generally considered not to infect animal hosts: *L. biflexa*, *L. idonii*, *L. meyeri*, *L. terpstrae*, *L. vanthielli*, *L. wolbachii*, *L. yanagawae*) (Faine et al., 1999). Intermediate leptospires cause predominantly mild self-resolving illnesses without fatal complications, while pathogenic species produce disease of varying severity in both humans and animals, ranging from subclinical infections to severe...
disease and death. The most severe forms are frequently caused by pathogenic strains belonging to the evolutionarily-related species *L. interrogans*, *L. kirschneri*, and *L. noguchii* (Lehmann et al., 2014; Xu et al., 2016), although severe diseases might involve any other species (Levett, 2001).

Overall, there is poor correlation between the two *Leptospira* classification schemes, serological classification cannot be used to reliably predict the species of *Leptospira* isolates. For example, a single serogroup may contain serovars from different *Leptospira* species and similarly, a single genomic species may also contain representatives from several different serogroups (Levett, 2001). Furthermore, a number of *Leptospira* serovars are found in more than one *Leptospira* species (Levett, 2015). Although serological classification is still essential to support clinical diagnostics and surveillance, the development of *Leptospira* genomics is revealing new insights into the biology and pathogenesis of *Leptospira* infection (Fouts et al., 2016; Lehmann et al., 2014). Increasingly, genetic typing approaches for leptospires are being used to further understand the epidemiology of *Leptospira* infection in a range of clinical and research settings.

In the same way that the black stripes of the “Universal Product Code” or “barcode” distinguish products in a supermarket, DNA barcoding was developed as a molecular identification technique in which a single short DNA sequence can be used for species identification (Hebert et al., 2003). DNA barcoding was originally proposed as an identification tool for animals, and used PCR amplification and sequence analysis of a ~650 bp (conserved) region of the mitochondrial cytochrome c oxidase subunit 1 (CO1) gene as a universal target. Barcoding has been very largely used in Arthropods, notably Lepidopterans, but was also developed in many other animals, as well as (frequently with other gene targets) in plants, algae, Protists, Fungi and Eubacteria. In Prokaryotes, sequencing of the small ribosomal 16S rRNA subunit gene (*rrs*) has been regarded as a standard for bacterial species identification for some time, and is analogous to the barcoding approach used in higher species. Unique sequence types could not always be attributed to a single bacterial species though, and reversely, some bacterial species could display sequence variations within the barcode. The term “molecular Operational Taxonomic Unit” or mOTU was therefore created to describe unique sequence types when studying complex communities. Prokaryote barcoding has been most notably used to describe microbial assemblages by directly amplifying *rrs* genes from environmental DNA extracts, cloning the PCR product and sequencing as many clones as possible (Pace, 1997), an ancestor form of what became “metabarcoding”, aimed at barcoding every individual from a complex community.

DNA-barcoding of a group of organisms requires a good understanding of the diversity of the genome in question, and the identification of suitable gene targets that can be used to
discriminate between different species within the group. For Leptospira species, the genomes range from 3.9 to 4.7 mega-bases (MB) in size and consist of two circular chromosomes: a large chromosome CI (~3.6 to 4.3 MB) and a smaller chromosome CII ~350 kilo-bases (KB) in length (Picardeau, 2015). Some saprophytic species such as Leptospira biflexa also have an extra-chromosomal element p74 of around 74 KB that has not been detected in any of the pathogenic Leptospira species (Fouts et al., 2016; Picardeau et al., 2008). Comparative genomic analysis suggests that pathogenic species have a common progenitor with a genome resembling that of L. biflexa. Many of the essential housekeeping genes are located on CI, as are most of the genes encoding virulence factors, such as lipL32 and ligB (Picardeau et al., 2008). In saprophytic species (e.g. L. biflexa), some housekeeping genes are also located on the p74 extra-chromosomal replicon. Leptospira show unusual mechanisms of gene regulation and patterns of genetic organization (Bulach et al., 2006; Saint Girons et al., 1992). In contrast to most bacteria where the rRNA genes are clustered and co-transcribed, in the Leptospira genome these genes are not linked to one another and are widely scattered along the CI chromosome (Picardeau, 2015). There also appears to be substantial amount of functional gene redundancy in Leptospira, particularly in pathogen-specific genes notably through gene duplication (Adler et al., 2011; Xu et al., 2016), and little synteny between pathogenic Leptospira species despite the short evolutionary distance between them (Picardeau et al., 2008). Pseudogenes and insertion sequences (IS) are common features in the Leptospira genomes (Picardeau, 2015). Comparative genomics revealed both overall genetic similarities and significant structural differences at the genus level, confirming genomic plasticity (Xu et al., 2016). IS-mediated sequence disruption and large chromosomal inversion or deletion are thought to be an important mechanisms in the evolution of Leptospira, and the number of IS-elements varies between species and serovars (Bulach et al., 2006). In general however, the Leptospira genome is considered relatively stable and Leptospira serovar identity can be maintained during in vitro cultures for more than 80 years in the absence of selective pressure (Picardeau, 2015).

In this article, we review the use of DNA-based ‘barcoding’ approaches to identify and explore the diversity of pathogenic Leptospira from clinical and environmental samples.

UTILITY OF LEPTOSPIRA BARCODING

Barcoding Leptospira in clinical and environmental samples is essential for a better understanding of the local epidemiology of infection and can provide information to inform the development of disease control strategies. Routine barcoding of infecting Leptospira supports surveillance of predominant strains and genotypes of Leptospira in local human
and animal populations and may indicate of changing trends of *Leptospira* infection within populations, such as new or atypical sources of infection or the emergence of a novel genotype. Studies of circulating *Leptospira* types can prove useful to monitor the evolution of the disease on a long-term perspective within a specific area and are of particular interest in countries where the disease is endemic and/or where recurrent epidemic outbreaks occur. As an example, a prospective study conducted in northern Thailand between 2000 and 2005, identified the emergence of a dominant clone of *L. interrogans* serovar Autumnalis which was a major cause of human disease during the outbreak unfolding in the early 2000s and has since diminished (Thaipadungpanit *et al.*, 2007).

Characterization of *Leptospira* spp. detected in different animal hosts or environmental sources is now a prerequisite for epidemiological and source attribution studies as it allows human *Leptospira* infection to be traced back to the probable source of contamination. A wide variety of mammals may become infected with pathogenic *Leptospira* and act as reservoirs of infection for people and other animals through excretion of infectious bacteria in their urine (Levett, 2001). Rodents are considered to be one of the main sources of *Leptospira* infection (Adler, 2015) but livestock, other domestic animals and many wild animal species can also carry and transmit pathogenic leptospires through urinary shedding (Gay *et al.*, 2014; Weekes *et al.*, 1997). Notable associations between animal reservoir host species and specific *Leptospira* serovars have been reported and are considered as a hallmark of leptospirosis epidemiology. Most of the knowledge of which animal hosts carry different *Leptospira* serovars relies on data from serological typing of isolates and seroprevalence surveys acquired over several decades. However, since the advent of the molecular taxonomy, studies that have used molecular analysis to explore host: pathogen relationships have revealed new insights into trends in host specificity (e.g. in bats and small mammals in the Western Indian Ocean islands (Dietrich *et al.*, 2014; Gomard *et al.*, 2016)).

Publicly available databases also allow the comparison of genotypes between regions and are important to trace the movement of different *Leptospira* types from a phylogeographic perspective.

Identifying sources of *Leptospira* infection is essential to develop evidence-based infection control and prevention strategies for both human and animal infection. Detecting animal or environmental reservoirs of infection can prove particularly useful to infer transmission routes, and identify ‘at-risk’ groups of people and high-transmission settings (Ganoza *et al.*, 2006; Viau & Boehm, 2011). Vaccinations against *Leptospira* are available but are serovar- (or at very best serogroup-) specific. Vaccine design and implementation policies require a good characterisation of predominant *Leptospira* in the target human or animal population and have been used in some settings as an effective method to reduce infection in animal
reservoirs of infection (e.g. vaccination of dairy cattle in New Zealand to reduce human infection (Marshall & Manktelow, 2002)).

LEPTOSPIRA BARCODING: TARGETS AND DATABASES

Identification and barcoding of pathogenic Leptospira species is usually performed on bacterial isolates obtained from infected human or animals. Increasingly, these approaches are also applied to DNA extracted directly from clinical or environmental samples following PCR-based pathogen detection. A variety of PCR assays have been described for use in the diagnosis of Leptospira infection in people and animals, and some of these can also be used for DNA-based typing of the infecting Leptospira. Leptospira-specific targets used for molecular diagnosis and typing target either conserved housekeeping genes where sequence-specific primers are used to differentiate between pathogenic and non-pathogenic species, or genes that are found only in the pathogenic species. One advantage of PCR assays designed for medical diagnosis is that they have usually been optimized for both sensitivity and specificity, assets that are useful when attempting to detect and characterize Leptospira in low numbers or in samples that contain high concentrations of non-Leptospira DNA. However, a drawback of these approaches is that diagnostic assays typically target highly conserved gene regions to ensure their sensitivity in detecting a variety of pathogenic Leptospira infections. Therefore, some of the gene targets commonly used for the diagnosis of Leptospira (e.g. lipL32 (Levett et al., 2005; Stoddard et al., 2009)) show poor discriminatory power in differentiating between Leptospira species when used alone. Some gene targets that have been used in the diagnosis and typing of Leptospira are shown in Table 1.

Table 1: PCR used or of potential use to generate Leptospira “barcodes”. The PCR published for diagnostic purpose are usually optimized for sensitivity and specificity.

Amongst the housekeeping genes, the secY gene has been used most widely for both the diagnosis and typing of Leptospira infections. The Leptospira secY gene is a housekeeping gene located on the CI chromosome that encodes a pre-protein translocase important for the export of proteins across the cytoplasmic membrane (Durack et al., 2015; Haake & Levett, 2015). Conserved regions of this gene were the target of one of the earliest diagnostic PCR assays (Gravekamp et al., 1993) and a real-time PCR was designed and validated for use in human clinical diagnosis (Ahmed et al., 2009). However, other regions of the secY gene demonstrate marked nucleotide polymorphism. A seminal study of 131 well-characterised Leptospira serovars by Victoria et al. (Victoria et al., 2008) demonstrated the
phylogenetic value of the variable regions of the secY gene. Sequence analysis of these regions can discriminate between known pathogenic Leptospira species and strains, and have been used in many studies to identify the infecting species of Leptospira (examples are listed in Table 2).

Table 2: Examples of successful typing of Leptospira from single short DNA sequences and corresponding discriminatory capacity.

An assay that is proving promising with regards to diagnosis and typing from a single assay is a SYBR Green I diagnostic qPCR developed to target fib1, a putative fibronectin-binding protein present in pathogenic leptospires (Merien et al., 2005). Interestingly, the PCR product generated has been shown to have a sequence polymorphism of epidemiological interest, and performs well in comparison to multi-locus sequence typing (MLST) for a selection of serovars belonging to the L. interrogans and L. borgpetersenii species (Perez & Goarant, 2010). Originally described for rapid identification of infecting Leptospira species from human clinical cases in New Caledonia, this approach has also revealed novel insights into the epidemiology and diversity of Leptospira infection in rodents in Madagascar (Moseley et al., in preparation) and livestock in Tanzania (K. Allan, unpublished data).

Another target that has been used to discriminate between Leptospira species is the flaB gene. This gene target, which encodes a flagellum protein, has been used successfully to identify other bacterial species including Campylobacter (Harrington et al., 2003) and Borrelia (Lin et al., 2004). Restriction Fragment Length Polymorphism of the PCR product (PCR-RFLP) of this gene has been successfully used to discriminate between Leptospira species in the laboratory (Kawabata et al., 2001) and was recently used in field settings to study the molecular epidemiology of canine leptospirosis in Japan (Koizumi et al., 2013) as well as to study the phylogenetic diversity in fruit-bat kidney specimens from Congo, revealing unique genotypes (Ogawa et al., 2015).

With regards to Leptospira barcoding, it is worth considering the so-called “universal” highly conserved genes used in large population or metagenomics studies of bacteria. In particular, the ribosomal 16S rRNA gene (rrs) has a recognized phylogenetic value for bacterial species classification, and was the first DNA sequence available for most Leptospira species as well as being used as a target for many diagnostic PCR assays (Merien et al., 1992; Smythe et al., 2002). Of note, among the published diagnostic PCRs, these 2 latter ones targeting rrs are currently the only ones which can detect pathogen as well as intermediate species, in spite of a single nucleotide polymorphism in the sequence of intermediates matching the LeptoF primer (Smythe et al., 2002) or by using a combination of the Forward primer A and the Reverse primer D (Merien et al., 1992). The 16S rRNA variable regions 3 &
4 sequence frequently used in metagenomics (Klindworth et al., 2013) has a good capacity to discriminate between the three Leptospira clusters, i.e. pathogens, intermediates and saprophytes. However, these regions have a low discriminatory power to differentiate between Leptospira species within a given clade. As an example, the 16S rRNA regions V3-V4 cannot be used to differentiate between the pathogenic species L. interrogans, L. noguchii and L. kirschnerii, which are classified within the same mOTU. Similarly, all saprophytic Leptospira species but Leptospira idonii belong to a single mOTU (Figure 1).

More recently, the RNA polymerase β-subunit gene rpoB, widely used in phylogenetic analyses of a number of bacterial genera, was proposed for Leptospira typing to circumvent the limitations of the 16S rRNA discrimination (La Scola et al., 2006). Bioinformatics studies later confirmed its high value for discrimination (Cerqueira et al., 2010) and its utility in epidemiological studies was demonstrated by work in Brazil and India that used this scheme to type Leptospira isolates from people and animals (Balamurugan et al., 2013; Jorge et al., 2012).

Another bacterial gene that is widely used in phylogenetic studies is the gyrB gene, which is reported to have a higher nucleotide divergence in Leptospira species than the 16S RNA rrs gene (Huang, 1996; Slack et al., 2006). This gene target has also been used for Leptospira diagnostic assays (Slack et al., 2006) and typing of isolates carried by rodents in Japan and the Philippines (Kawabata et al., 2006; Saito et al., 2015; Villanueva et al., 2014).

Other highly conserved genes have been proposed as targets to discriminate bacterial species by sequencing, but have rarely, if ever, been used in Leptospira studies. This includes cpn60, a chaperonin gene (also known as HSP60 in many bacteria or GroEL in Leptospira) proposed as a preferred universal barcode for bacteria compared to the 16S rRNA gene (Links et al., 2012). Such highly conserved genes might be of interest for improved discrimination between Leptospira species in microbial communities.

MULTI LOCUS SEQUENCE TYPING OF LEPTOSPIRA

Multilocus sequence typing (MLST) is an unambiguous procedure for characterising isolates of bacterial species using the sequences of internal fragments of several genes, including housekeeping genes (Maiden et al., 1998; Urwin & Maiden, 2003). For each gene/locus within a particular MLST scheme, a ~500 bp fragment is sequenced and analysed. The sequences present at each locus are assigned a distinct allele code and the combination of those alleles defines an allelic profile or sequence type (ST) for each isolate. Whilst MLST is not a barcoding method per se as it does not focus on a single marker, each locus considered separately can be considered as a barcode and the combination of multiple barcodes is considered one of the most robust genetic methods to identify the infecting
Leptospira strain currently available. Three major MLST schemes exist for Leptospira spp. typing and Leptospira sequence types for all schemes are publically available at the PubMLST website (http://pubmlst.org/leptospira/), which also hosts molecular typing databases for a variety of bacterial pathogens. Through this online portal, allelic profiles of Leptospira isolates can be easily compared to those from reference serovars compiled in the database. Presently, 13 Leptospira genes (including the highly discriminatory secY gene discussed above) from the three different MLST schemes have been included in the online database. MLST scheme #1, referred to as the 7L scheme, uses seven genes (glmU, pntA, sucA, tpiA, pfkB, mreA, caiB) to discriminate between the seven major Leptospira pathogenic species (Boonsilp et al., 2013). MLST scheme #3, denoted as the 6L scheme, includes three housekeeping genes (adk, icdA, secY), two genes encoding outer membrane proteins (lipL32, lipL42) and the 16S rRNA gene (rrs) that can be used for typing pathogenic and intermediate species (Ahmed et al., 2006). MLST scheme #2 combines the “best” loci (adk, glmU, icdA, lipL32, lipL41, mreA and pntA) from schemes #1 and #3 (Varni et al., 2014) but to date, has been used less commonly in the literature. Other MLST schemes have been suggested for leptospirosis research, but they are not yet supported by a website with referenced alleles or sequence types. As an example, in silico studies have proposed a reduced 4-loci scheme with high discriminatory power (Cerqueira et al., 2010). This scheme would be usable in L. interrogans and L. kirschneri with currently validated PCRs, but would need to be further adapted for application with a wider range of pathogenic species. In the future, it is anticipated that the MLST technique will significantly contribute to gaining insights into the evolution and phylogeographic affinities of leptospires. However, the current multiplicity of available schemes leads to some confusion in the field of leptospirosis research that has limited the comparison between isolates obtained from different studies and geographic areas.

BARCODING SUCCESSES / LIMITATIONS

A major feature of leptospirosis is the great diversity of etiological agents leading to the “leptospirosis” disease. As many as 15 different pathogenic and intermediate Leptospira species are infectious and have been implicated in human or animal infections (Balamurugan et al., 2013; Levett et al., 2006; Matthias et al., 2008; Petersen et al., 2001; Schmid et al., 1986; Slack et al., 2008; Tsuboi et al., 2017). However, as mentioned before, most diagnostic PCR only detect Leptospira from the pathogenic cluster and fail to detect intermediate species (Bourhy et al., 2011). Therefore, there is currently a strong bias toward the pathogenic cluster of Leptospira spp, and most frequently, only the pathogenic species might be considered with the current techniques.
A major advantage of using sequence-based genotyping of *Leptospira* is the ability for the technique to be standardised and compared between different laboratories and geographic sites. In comparison to serotyping, which is only performed at reference laboratories and requires considerable specialist expertise (Hartskeerl & Smythe, 2015), *Leptospira* barcoding techniques can be performed relatively quickly and cheaply, without the need for specialist equipment and extensive panels of reference antisera or monoclonal antibodies. The increasing availability of molecular diagnostic and sequencing facilities around the world also means that the technique can be performed in a wide variety of settings. Even where facilities are not available in-country, the fact that DNA-based typing does not require propagation or maintenance of live *Leptospira* cultures makes typing of *Leptospira* infecting animals and people in resource-limited settings more feasible (e.g. in Kenya and Zambia (Halliday *et al.*, 2013; Ogawa *et al.*, 2015)).

The ability to directly compare *Leptospira* sequences from different parts of the world has the potential to reveal new insights into the epidemiology of the infection. The application of genotyping techniques has demonstrated outbreaks of human disease caused by a single clonal complex of *L. interrogans* in Thailand (Thaipadungpanit *et al.*, 2007) and revealed intriguing patterns of animal host-specificity in Madagascar (Dietrich *et al.*, 2014; Gomard *et al.*, 2016). As yet, comparison of sequences from different geographic regions on a global scale has not been explored. However, the increasing application of standard barcoding or typing approaches has the potential for large-scale patterns of infection to be investigated.

One of the major criticisms about the use of barcoding in diversity analyses is that it relies on a single marker and hence is often unable to discriminate between strains of a single *Leptospira* species (Dupuis *et al.*, 2012; Mallo & Posada, 2016). The use of multiple barcoding loci to type *Leptospira* is a way to avoid misinterpretations because of large scale sequence changes such as horizontal gene transfer, which may occur between different *Leptospira* strains (Bulach *et al.*, 2006; Haake *et al.*, 2004; Llanes *et al.*, 2016; Picardeau *et al.*, 2008; Victoria *et al.*, 2008). MLST (see above) is based on genotyping and establishing phylogenetic relationships between bacterial isolates using concatenated sequences derived from several loci, thus minimizing the possible biases originating from horizontally acquired DNA. MLST is considered one of the most robust and efficient methods in identifying ancestral relationships between *Leptospira* and segregating strains isolated from an outbreak, and in identifying the source(s) of human contamination (Nalam *et al.*, 2010). Though not new, this method is not widely used yet for leptospirosis epidemiologic studies, mostly because it still requires bacterial isolates, in spite of ongoing efforts to type directly from non-cultured material.
The first objective of barcoding is to identify leptospires at the species level in patients with clinical leptospirosis. Among the genes and barcoding schemes described in this paper, the 16S rRNA rrs gene and the translocase preprotein secY have been most frequently used for typing of *Leptospira* species from patient samples. Increasingly, these typing schemes are being directly applied to biological specimens without strain isolation. In acute cases of human leptospirosis, *Leptospira* DNA can be detected and typed in whole blood, serum or urine samples. In animal carriers of infection, the kidneys or urine have been most frequently used. Similar approaches have been applied to type *Leptospira* detected in environmental surface water samples (Ganoza et al., 2006), a field of research that has recently gained renewed interest (Mason et al., 2016; Muñoz-Zanzi et al., 2014; Thibeaux et al., 2017). Examples of successful genotyping from non-cultured clinical or environmental material are listed in Table 2.

Direct MLST typing of *Leptospira* from non-isolate clinical specimens has had limited success (Agampodi et al., 2013; Perez & Goarant, 2010), whatever the primers used. A revision of the MLST scheme #3 has recently been proposed, using newly designed nested primers to improve the sensitivity to make it usable directly from clinical specimens (Weiss et al., 2016). However, even this optimized procedure also only proved successful in a subset of clinical specimens. The success of MLST when performed directly on human or animal clinical samples has been shown to be correlated with the bacterial load in the specimen. In one study performed during an outbreak of acute human leptospirosis in Sri Lanka, Agampodi and colleagues demonstrated a substantial difference in the mean bacterial load between specimens with complete MLST (2.2 x 10^5 *Leptospira*/mL) compared to specimens where the full MLST scheme could not be completed (1.3 x 10^4 *Leptospira*/mL), as measured by real-time PCR (Agampodi et al., 2013). A relatively high threshold of 4.9 x 10^4 *Leptospira*/mL was observed for specimens with complete MLST profiles. In addition to bacterial load, failure of sequence-based typing has also been reported in association with divergent or novel *Leptospira* types, particularly in areas that have been relatively poorly characterised for circulating *Leptospira* diversity. Standard typing approaches may fail to amplify target gene sequences due to sequence polymorphism of local serovars at primer binding sites. In a study of *Leptospira* infection in small mammals in Madagascar (Dietrich et al., 2014), modification of standard secY primer sequences with inclusion of degenerate bases to account for sequence polymorphisms greatly improved the efficiency of single locus secY typing on *Leptospira* in non-isolate samples. Similarly, modified primers also facilitated direct typing of *Leptospira* in clinical samples from livestock.
in Tanzania (Allan, 2016) or from rodents in New Caledonia (Perez et al., 2011) where standard-typing assays had failed.

The ability to perform sequence-based typing of *Leptospira* spp. in the absence of bacterial isolation has helped to improve the utility of these techniques in a clinical setting, but risks neglecting the role of some strains or indeed species in the local epidemiology of leptospirosis. This is especially true in relatively unexplored geographic regions. For example, the most recently described *Leptospira* species, *L. mayottensis* (Bourhy et al., 2014) was first detected in human clinical cases on the tropical island of Mayotte, Indian ocean (Bourhy et al., 2012). However, qPCR detection of this species was challenging with low diagnostic sensitivity for standard qPCR diagnostic assays (Bourhy et al., 2011; Bourhy et al., 2012). Following culture and isolation of the infecting *Leptospira*, more extensive phenotypic and genotypic characterisation was performed leading to the description of a new species, which would have been impossible without a bacterial isolation. Similarly, a major limitation mentioned above is the failure of most current PCR designs to detect *Leptospira* belonging to the intermediate cluster (Bourhy et al., 2011). Because all species in this cluster have already proven pathogenic and been isolated from clinical specimens, there is growing recognition that these species will need to be considered and new PCR designs will be needed to more extensively address the contribution of these species to animal and human leptospirosis (Tsuboi et al., 2017). These two cases point to the current limitations of approaches only based on PCR with specific primers. Considering that the biodiversity of Leptospira is insufficiently recognized, they show the need for continued efforts of *Leptospira* isolation.

**WHAT’S NEXT?**

The rapid rise and development of next generation sequencing (NGS) has exciting applications to the *Leptospira* field. The first *Leptospira* genome was fully reported 15 years ago (Ren et al., 2003), and since then, whole genome sequencing has been made increasingly easier by the rise of NGS technologies. These massively parallel DNA sequencing methods provide high throughput genomic data faster and cheaper than first generation sequencing.

A large-scale sequencing project entitled “Leptospira Genomics and Human Health” ([http://gsc.jcvi.org/projects/gsc/leptospira/](http://gsc.jcvi.org/projects/gsc/leptospira/)) led by the Craig Venter Institute and initiated in 2011 has generated whole genome sequences for numerous strains belonging to 20 *Leptospira* species from diverse origins and geographical areas (Lehmann et al., 2014). The global analysis of these genomes has identified a core genome of more than 1760 genes, of which 737 are specific to *Leptospira*, 369 are specific to species with some degree of
pathogenicity (Pathogens and Intermediates), and 416 are specific to pathogenic species (Fouts et al., 2016). Beside all basic knowledge that is (and will still be) gained from these wide comparative genomics studies (Xu et al., 2016), the identification of these genes offers opportunities to identify highly specific targets for *Leptospira* barcoding but also improve the feasibility of a universal, multilocus molecular typing system (Gerth & Bleidorn, 2013). The availability of new 'single-molecule' sequencing technologies (generally referred to as third-generation sequencing) that can produce longer reads and highly accurate de novo assemblies of hundreds of microbial genomes has enabled greatly improved analysis of genome structure (Koren et al., 2013; Loman et al., 2015). When the price and computational challenges are overcome, these new technologies, with the prospect of readily available full genomic sequences data, may make core genome MLST (cgMLST) a new gold standard and trigger the downfall of DNA barcoding, especially for bacteria (Taylor & Harris, 2012).

The combination of DNA barcoding with NGS technologies has also facilitated the taxonomic profiling of complex communities through the simultaneous sequencing of many thousands of DNA barcodes from each sample. "Metabarcoding" is distinguished from conventional barcoding by operating on the collective DNA rather than the isolated DNA of individual organisms (Baker et al., 2016). This broader community approach has been used by health scientists to investigate animals for zoonotic pathogens, allowing the taxonomic classification of all infecting ones (Razzauti et al., 2015), or by microbial ecologists to investigate genetic materials from environmental samples on a tremendous scale and without cloning. Because of this ‘inventory’ approach, the leptospirosis scientific community may gain unexpected information about *Leptospira* spp. through the active data mining of the sequence databases fed by metabarcoding studies with no specific focus on leptospirosis.

As an emerging and re-emerging infectious disease, a future challenge of *Leptospira* typing will be to continue to develop and adapt barcoding, typing and *Leptospira* classification schemes to deal with the increasing array of pathogen diversity described in the literature. New *Leptospira* species are still being discovered, more than a century after the first isolation of pathogenic *Leptospira* (Bourhy et al., 2014; Ido et al., 1917), and additional species may still be waiting to be discovered. In relatively poorly characterised regions of the world such as Africa, studies that are performed often detect new or divergent *Leptospira* strains (Allan, 2016; Allan et al., 2015; Mgode et al., 2015). The challenge remains to identify the best DNA targets and amplification techniques for this purpose. These should fulfil two hardly compatible needs: (i) a need to be highly sensitive to account for the frequently very low bacterial burden of clinical specimens and (ii) a need to generate DNA products with a
sequence polymorphism of epidemiological relevance. Additionally, these should still take into account the wide diversity of the genus *Leptospira*. Estimating the true diversity of *Leptospira* is tied to how well the genomic diversity of the community is represented by the genomes available in sequence databases. Whilst barcoding and multi-locus typing of infectious *Leptospira* allow genetic classification of the infecting leptospires, diagnosis and surveillance of infection as well as vaccination are still heavily dependent upon serology and the serological classification schemes. As yet, a major challenge for the leptospirosis community is to reconcile the serological and molecular *Leptospira* classification schemes. The different molecular typing methods described in this review do not identify serovars but at best point to “putative serovars” based on correlations determined previously. The effective marriage of serovars and genotypes might be achieved through a molecular-based serovar typing system. The nucleotide sequence of the LPS biosynthetic operon *rfb* has therefore been pointed as a potential target for genotyping leptospires at the serovar level (Ahmed *et al.*, 2012; Bezerra da Silva *et al.*, 2011). Although the performance of the method did not provide optimal results yet, the approach in itself proved very useful and promising (Bezerra da Silva *et al.*, 2011), and whole genome sequences might be used in the future to infer serological features. For now, the ongoing challenge of a double non-convergent taxonomy - serovar vs genotype - looks set to continue. However, the exciting developments in the field of WGS may yet offer new insights into understanding the epidemiology, pathogenesis and taxonomy of this complex but important zoonotic pathogen.

**FINANCIAL SUPPORT**

This review work was supported by the author’s institutions and was not funded by any external grant.
Table 1: PCR used or of potential use to generate *Leptospira* “barcodes”. The PCR published for diagnostic purpose are usually optimized for sensitivity and specificity.

<table>
<thead>
<tr>
<th>Initial purpose</th>
<th>Gene target</th>
<th>Technology</th>
<th>Product size</th>
<th>Reference</th>
</tr>
</thead>
</table>
| Diagnostic      | 16S rRNA *rrs* | Conventional nested PCR | 331 bp       | (Merien et al., 1992)  
Primers A & D detect all *Leptospira* species                      |
| Diagnostic      | *lipL32*    | SYBR Green I qPCR   | 423 bp       | (Levett et al., 2005)                                                    |
| Diagnostic      | *lipL32*    | TaqMan qPCR         | 242 bp       | (Stoddard et al., 2009)                                                  |
| Diagnostic      | *secY*      | SYBR Green I qPCR   | 202 bp       | (Ahmed et al., 2009)                                                     |
| Diagnostic      | *lfb1*      | SYBR Green I qPCR   | 331 bp       | (Merien et al., 2005)                                                    |
| Diagnostic      | *flaB*      | Conventional PCR    | 793 bp       | (Kawabata et al., 2001)                                                  |
| Diagnostic      | *secY*      | Conventional PCR    | 285 bp       | (Gravekamp et al., 1993)                                                 |
| Typing          | *gyrB*      | Conventional PCR or SYBR Green I qPCR | 504 bp | (Slack et al., 2006)                                                    |
| Typing          | *rpoB*      | Conventional PCR    | 600 bp       | (La Scola et al., 2006)                                                  |
| Typing (MLST)   | 16SrRNA *rrs* | Conventional PCR   | 541 bp       | (Ahmed et al., 2006)                                                     |
| Typing (MLST)   | *secY*      | Conventional PCR    | 549 bp       | (Ahmed et al., 2006)                                                     |
| Typing (MLST)   | *lipL32*    | Conventional PCR    | 474 bp       | (Ahmed et al., 2006)                                                     |
| Metagenomics    | 16SrRNA *rrs* regions V3-V4 | Conventional PCR | ~450 bp       | (Klindworth et al., 2013)                                               |
| Metagenomics    | *cpn60* (*hsp60*) | Conventional PCR  | ~550-600 bp  | (Goh et al., 1996; Klindworth et al., 2013)                               |
| Barcoding       | COI-1       | Conventional PCR    | ~550 bp      | (Smith et al., 2012)                                                    |
Table 2: Examples of successful typing of *Leptospira* from single short DNA sequences and corresponding discriminatory capacity

<table>
<thead>
<tr>
<th>Article</th>
<th>Molecular target(s)</th>
<th>Starting material</th>
<th>Finding(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ganoza <em>et al.</em>, 2006)</td>
<td>Nested 16SrRNA (Ganoza <em>et al.</em>, 2006) (product cloned in <em>E. coli</em>)</td>
<td>Surface waters</td>
<td>Presumptive <em>Leptospira</em> species and comparison with human isolates, unknown <em>Leptospira</em> clade</td>
</tr>
<tr>
<td>(Agampodi <em>et al.</em>, 2013)</td>
<td>MLST (Thaipadungpanit <em>et al.</em>, 2007)</td>
<td>Human serum or whole blood</td>
<td>Determination of partial to complete ST in 12/58 samples</td>
</tr>
<tr>
<td>(Koizumi <em>et al.</em>, 2013)</td>
<td><em>flaB</em> (Koizumi <em>et al.</em>, 2008)</td>
<td>Blood or urine from dogs</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>)</td>
</tr>
<tr>
<td>Study</td>
<td>Target Genes</td>
<td>Source</td>
<td>Identification</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------</td>
<td>-------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>(Goarant et al., 2014)</td>
<td><em>lfb1</em> (Merien et al., 2005), <em>secY</em> (Ahmed et al., 2009), <em>lipL32</em> (Levett et al., 2005), 16SrRNA (Merien et al., 1992)</td>
<td>Human blood</td>
<td>Identification of an exotic <em>Leptospira</em> species (<em>L. weilii</em>)</td>
</tr>
<tr>
<td>(Gay et al., 2014)</td>
<td><em>lfb1</em> (Merien et al., 2005)</td>
<td>Kidney or urine from dogs, pigs, deer</td>
<td>Identification of putative serovar</td>
</tr>
<tr>
<td>(Mayer-Scholl et al., 2014)</td>
<td><em>secY</em> (Gravekamp et al., 1993)</td>
<td>Kidneys from small Mammals</td>
<td>Identification of <em>Leptospira</em> species (<em>L. kirschneri</em>)</td>
</tr>
<tr>
<td>(Muñoz-Zanzi et al., 2014)</td>
<td><em>lipL32</em> (Stoddard et al., 2009)</td>
<td>Surface waters</td>
<td>Confirmation of pathogenic leptospires</td>
</tr>
<tr>
<td>(Dietrich et al., 2014)</td>
<td><em>secY</em>, <em>adk</em>, <em>lipL32</em>, <em>lipL41</em>, 16SrRNA (Ahmed et al., 2009)</td>
<td>Kidneys from small Mammals and bats</td>
<td>Identification of <em>Leptospira</em> species (<em>L. borgpetersenii</em>, <em>L. kirschneri</em> and <em>L. mayottensis</em>)</td>
</tr>
<tr>
<td>(Cosson et al., 2014)</td>
<td><em>secY</em>, <em>adk</em>, 16SrRNA (Ahmed et al., 2006)</td>
<td>Kidneys from rodents</td>
<td>Identification of <em>Leptospira</em> species (<em>L. borgpetersenii</em>, <em>L. interrogans</em>, <em>L. kirschneri</em> and <em>L. weilii</em>) and subspecific genotypes</td>
</tr>
<tr>
<td>(Ogawa et al., 2015)</td>
<td><em>flaB</em> (Koizumi et al., 2008), <em>secY</em> (Gravekamp et al., 1993), 16SrRNA (Ogawa et al., 2015)</td>
<td>Kidneys from bats</td>
<td>Identification of <em>Leptospira</em> species (<em>L. kirschneri</em> and <em>L. borgpetersenii</em>) as well as novel genotypes (probable novel pathogenic species)</td>
</tr>
<tr>
<td>(Pagès et al., 2015)</td>
<td><em>lfb1</em> (Merien et al., 2005), <em>secY</em> (Ahmed et al., 2006)</td>
<td>Human urine</td>
<td>Identification of <em>Leptospira</em> species (<em>L.</em></td>
</tr>
<tr>
<td>Publication</td>
<td>Methodology</td>
<td>Sample Type</td>
<td>Identification of Leptospira Species</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------------------------------</td>
<td>------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>(Verma et al., 2015)</td>
<td>MLST (Ahmed et al., 2006)</td>
<td>Kidneys from pigs</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>). All genes but <em>adk</em> amplified and sequenced.</td>
</tr>
<tr>
<td>(Gomard et al., 2016)</td>
<td><em>secY, adk, 16SrRNA</em> (Ahmed et al., 2006)</td>
<td>Pool of kidney, spleen and lung from bats</td>
<td>Identification of <em>Leptospira</em> species (<em>L. kirschneri</em> and <em>L. borgpetersenii</em>) as well as novel genotypes (probable novel pathogenic species).</td>
</tr>
<tr>
<td>(Guernier et al., 2016)</td>
<td>MLST (Ahmed et al., 2006), 16SrRNA (Fenner et al., 2010; Merien et al., 1992), <em>secY</em> (Gravekamp et al., 1993)</td>
<td>Human sera, Kidney or urine from Mammals</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>, <em>L. kirschneri</em>, <em>L. borgpetersenii</em> and <em>L. mayottensis</em>) Identification of complete or partial ST from MLST</td>
</tr>
<tr>
<td>(Hamond et al., 2016)</td>
<td><em>secY</em> (Ahmed et al., 2006)</td>
<td>Urine from cattle</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>, <em>L. borgpetersenii</em>, <em>L. noguchii</em> and <em>L. santarosai</em>) and genotype diversity</td>
</tr>
<tr>
<td>(Obiegala et al., 2016)</td>
<td><em>gyrB</em> (Slack et al., 2006), MLST (Thaipadungpanit et al., 2007)</td>
<td>Kidneys from small Mammals</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>, <em>L. borgpetersenii</em> and <em>L. kirschneri</em>). ST for some samples.</td>
</tr>
<tr>
<td>(Weiss et al., 2016)</td>
<td>MLST adapted from (Boonsilp et al., 2013)</td>
<td>Human samples (serum, whole blood, buffy)</td>
<td>Identification of <em>Leptospira</em> species (<em>L. borgpetersenii</em>, <em>L. kirschneri</em>, <em>L. santarosai</em> and <em>L. weilii</em>). Partial or complete ST.</td>
</tr>
<tr>
<td>Study</td>
<td>Method/Marker Used</td>
<td>Sample Type</td>
<td>Evaluation</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ayral et al., 2016</td>
<td>16SrRNA (Merien et al., 1992), Multispacer Sequence Typing (Zilber et al., 2014)</td>
<td>Kidneys from small Mammals</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>, <em>L. borgpetersenii</em> and <em>L. kirschneri</em>). Putative serogroup or serovar from Multispacer ST.</td>
</tr>
<tr>
<td>Lagadec et al., 2016</td>
<td>16SrRNA (Merien et al., 1992), MLST (Ahmed et al., 2006; Dietrich et al., 2014)</td>
<td>Pool of kidney, spleen and lung or urine from Mammals</td>
<td>Identification of <em>Leptospira</em> species (<em>L. borgpetersenii</em>, <em>L. mayottensis</em>, <em>L. interrogans</em>, <em>L. kirschneri</em>).</td>
</tr>
<tr>
<td>Muller et al., 2016</td>
<td>16SrRNA (Mgode et al., 2005)</td>
<td>Human blood</td>
<td>Confirmation of pathogenic <em>Leptospira</em> sp.</td>
</tr>
<tr>
<td>Mason et al., 2016</td>
<td>secY (Ahmed et al., 2009)</td>
<td>Environmental waters</td>
<td>Identification of <em>Leptospira</em> species (<em>L. interrogans</em>, <em>L. kirschneri</em> and <em>L. weilii</em>) as well as putative novel species.</td>
</tr>
<tr>
<td>Thibeaux et al., 2017</td>
<td>lfb1 (Merien et al., 2005)</td>
<td>Soils from contamination areas</td>
<td>Identification of a particular <em>Leptospira</em> interrogans strain and identity with the clinical strain. Evidence of novel genotypes</td>
</tr>
</tbody>
</table>
References


Figure 1: Phylogeny deduced from (A) 1262 bp or (B) the variable regions V3 & V4 (440 bp) of the 16S rRNA rrs gene. The phylogenetic topology of the genus Leptospira is shown in the upper panel A. Unique sequences of the V3 & V4 regions of the rrs gene, frequently used in metagenomics studies, discriminate only 12 mOTUs from the 22 Leptospira species currently described. The corresponding resolution losses are highlighted in red in lower panel B. Reference sequences were retrieved from GenBank and used to establish phylogeny (Neighbour Joining). Scale bars are substitution rates.