1   Genomic Analysis of Human-infecting *Leptospira borgpetersenii* isolates in Sri Lanka:

2   expanded PF07598 gene family repertoire, less overall genome reduction than bovine isolates

3

4   Indika Senavirathna,[1,2] Dinesha Jayasundara,[1,3] Janith Warnasekara,[1,4] Suneth Agampodi,[1,4]* Ellie J.

5   Putz,[5] Jarlath E. Nally,[5] Darrell O. Bayles,[5] Reetika Chaurasia,[6] and Joseph M. Vinetz[6]

6

7   [1]Leptospirosis Research Laboratory, Department of Community Medicine, Faculty of Medicine and

8   Allied Sciences, Rajarata University of Sri Lanka

9   [2]Department of Biochemistry, Faculty of Medicine and Allied Sciences, Rajarata University of Sri

10   Lanka

11   [3]Department of Microbiology, Faculty of Medicine and Allied Sciences, Rajarata University of Sri

12   Lanka

13   [4]Department of Community Medicine, Faculty of Medicine and Allied Sciences, Rajarata University

14   of Sri Lanka

15   [5]Infectious Bacterial Diseases Research Unit, National Animal Disease Center, Agricultural Research

16   Service, United States Department of Agriculture, Ames, IA

17   [6]Section of Infectious Disease, Department of Internal Medicine, School of Medicine, Yale

18   University, New Haven, CT

19

20

21   **Key words**

22   *Leptospira borgpetersenii,* genome reduction, virulence factor, evolution

23

24

25

26

27

28

Abstract

29

30   *Leptospira borgpetersenii* commonly causes human leptospirosis, including severe disease. The first

31   published analysis of *L. borgpetersenii*, performed on two strains of serovar Hardjo (L550 and

32   JB197), concluded that the *L. borgpetersenii* genome is in the process of genome decay with

33   functional consequences leading to a more obligately host-dependent life cycle. Yet whole genome

34   analysis has only been carried out on few strains of *L. borgpetersenii*, with limited closed genomes

35   and comprehensive analysis. Herein we report the complete, circularized genomes of seven non-

36   Hardjo *Leptospira borgpetersenii* isolates from human leptospirosis patients in Sri Lanka. These

37   isolates (all ST144) were found to be nearly identical by whole genome analysis; serotyping showed

38   they are a novel serovar. We show that the *L. borgpetersenii* isolated from humans in Sri Lanka are

39   less genomically decayed than previously reported isolates: fewer pseudogenes (N=141) and

40   Insertion Sequence (IS) elements (N=46) compared to N=248, N=270, and N=400 pseudogenes, and

41   N=121 and N=116 IS elements in published *L. borgpetersenii* Hardjo genomes (L550, JB197 and

42   TC112). Compared to previously published *L. borgpetersenii* whole genome analyses showing two to

43   three VM proteins in *L. borgpetersenii* isolates from cattle, rats and humans, we found that all of the

44   human *L. borgpetersenii* isolates from Sri Lanka, including previously reported serovar Piyasena,

45   have 4 encoded VM proteins, one ortholog of *L. interrogans* Copenhageni LIC12339 and 3 orthologs

46   of LIC12844. Our findings of fewer pseudogenes, IS elements and expansion of the LIC12844

47   homologs of the PF07598 family in these human isolates suggests that this newly identified *L.*

48   *borgpetersenii* serovar from Sri Lanka has unique pathogenicity. Comparative genome analysis and

49   experimental studies of these *L. borgpetersenii* isolates will enable deeper insights into the molecular

50   and cellular mechanisms of leptospirosis pathogenesis.

51

52

53

54

55

56

57 **Author Summary**

58 Leptospirosis is an emerging bacterial zoonosis worldwide. *Leptospira borgpetersenii* predominates

59 as the cause of human leptospirosis in some agricultural contexts. We address here the relatively

60 neglected comparative genome analysis of *L. borgpetersenii* . We show here that *L.*

61 *borgpetersenii* isolated from humans in Sri Lanka have less genome reduction compared to available

62 cattle isolates and have novel virulence characteristics compared to isolates from other animals

63 including cattle and rats.

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

Introduction

Leptospirosis, a globally important but neglected bacterial zoonosis [1-6], is caused by gram-negative spirochetes of the genus, *Leptospira*, and is an emerging zoonotic disease worldwide. Leptospirosis is conservatively estimated to affect approximately 1 million people with ~60,000 deaths per year [2, 4] with estimated Disability Adjusted Life Years (DALYs) annually, which is on par with cholera, typhoid fever and dengue [1, 4, 7-9]. The estimated number of cases of leptospirosis in humans exceeds an average of 500,000 per year, and the case fatality can be as high as 20% [2, 4, 6]. Leptospirosis incidence is strongly predicted to increase over coming years related to climate change [7, 10-14]. Therefore, cases of leptospirosis are likely to become more common as it has already been recognized as a reemerging infectious disease [2]. Identification and characterization of novel *Leptospira* species, which were discovered recently in both pathogen and intermediate lineages [3, 15], are critical for developing novel diagnostic tools for early detection of the disease, for making timely therapeutic decisions [10, 16-19], and to underpin vaccine development [20, 21].

Whole genome sequencing (WGS) has revolutionized in-depth understanding of infection and pathogenesis of leptospirosis at a molecular level [3, 22-24]. Whole genome analysis of new *Leptospira* isolates from different geographic locations has already advance our understanding of the pathogenic mechanisms [25], which may further facilitate the development of better treatment options [3, 20, 26]. The WGS approach has also become a powerful tool for bacterial strain classification and epidemiological typing [5][27, 28]. Leptospiral genome sequences published to date include at least 654 *Leptospira* sequences with most sequences (49%) belonging to *L .interrogans*, followed by *L . borgpetersenii* (7%), *L. santarosai* (6%), and *L. kirschneri* (5%). The size of these genomes varies from 3.9 to 4.6Mb [7]. This list continues to grow [3, 22].

The first whole genome sequence analysis of *L. borgpetersenii* was published by Bulach *et al. [29]*. A recent study published in 2018 reported the genome of *L. borgpetersenii* strain 4E, a highly virulent isolate obtained from *Mus musculus* in southern Brazil [10]. The above-referenced studies identified a total of 3,469 coding DNA sequences (CDSs), 37 transfer-RNAs (tRNAs), 4 ribosomal RNAs (rRNAs), one transfer-messenger RNA (tmRNA) and five riboswitch *loci* in *L. borgpetersenii*. Nevertheless, a fully closed complete genome of *L. borgpetersenii* was reported for

4

115   the first time based on the genome of laboratory-maintained reference strain, *L. borgpetersenii*

116   serogroup Sejroe serovar Ceylonica strain Piyasena isolated in 1964 (from a male patient in Colombo,

117   Sri Lanka). The complete genome sequences of four recent isolates of *L. borgpetersenii* serovar

118   Hardjo designated strains TC112, TC147, TC129, and TC273 were reported to have 3,345-3,495

119   coding sequences and 397 to 416 pseudo genes [12]. Recently, the PF07598 gene family that encodes

120   the Virulence Modifying Proteins was reported to encode secreted leptospiral exotoxins that may

121   contribute to the pathogenesis of leptospirosis [25]. While four VM proteins were reported in *L.*

122   *borgpetersenii* serovar Javanica, in contrast, two Hardjo strains have only three VM proteins [30].

123   In the present study, we performed whole-genome sequencing, *de novo* assembly, structural,

124   and functional annotation of seven pathogenic *L borgpetersenii* isolates recovered from humans in Sri

125   Lanka, tested the proposed genome reduction hypothesis and compared these isolates with others

126   isolated from different mammalian hosts for genomic content of the PF07598 gene family-encoded

127   VM proteins [29].

128

129   **Methods**

130   ***Leptospira* strains and genomic DNA extraction**

131   Isolates for this work were obtained from a large study conducted among febrile patients who

132   were clinically classified as 'probable' leptospirosis cases, five from the Teaching Hospital

133   Anuradhapura (FMAS_AP2, FMAS_AP3, FMAS_AP4, FMAS_AP8 and FMAS_AP9), and two

134   from the General Hospital Polonnaruwa (FMAS_PN1, FMAS_ PN4) [8, 16, 31, 32]. Details of patient

135   selection and culture isolation are reported in the original papers [8, 16, 31, 32]. These strains were

136   newly isolated from symptomatic patients and had few passages before genomic DNA extraction for

137   WGS. The organisms were first grown in semisolid EMJH media before being sub-cultured in liquid

138   EMJH medium. Cells were harvested in log phase growth, followed by DNA extraction carried out

139   using the gram-negative bacteria protocol from Qiagen's DNeasy Blood & Tissue Kit including an

140   RNase clean-up step after proteinase K^+ buffer ATL incubation [3]. Extracted DNA was quantified

141   using a Qubit 4 fluorometer (ThermoFisher).

142

Sample preparation

Genomic DNA (gDNA) size and integrity was assessed by pulsed field gel electrophoresis (PFGE) method before beginning library preparation. Multiplexed PacBio Single Molecule Real-Time (SMRT) bell libraries were prepared from extracted high quality gDNA using the SMRTbell® Express Template Prep Kit 2.0. To prepare 15-kb libraries, 1µg of genomic DNA was sheared using g-tubes™ from Covaris Woburn, MA, USA and AMPure PB Beads(Pacific Bioscience) were used for the concentration of DNA. The DNA was finally repaired by overnight ligation to the overhanging barcoded 8A adapter (Pacific Bioscience). Blue Pippin ™ size selection (Sage Science, Beverly, Massachusetts, USA) of 4 kb or more was performed according to the manufacturer's instructions. Conditions for annealing the sequencing primer and binding the polymerase to the purified The SMRTbell ™ template was evaluated using a calculator from RS Remote (Pacific Biosciences).

Whole-Genome Sequencing and assembly

SMRTbell libraries were generated and sequenced on a PacBio RS II system (Maryland Genomics, Institute for Genome Sciences, University of Maryland School of Medicine). A minimum of 800X read coverage was obtained for all seven isolates. Raw read data were preprocessed using an in-house developed quality control pipeline. Genomes were assembled de novo using Canu 2.1 which were then circularized using Circlator[17] (http://sangerpathogens.github.io/circlator). Two overlapping contigs were recovered in all isolates after completion of the workflow. The annotation was completed in all 7 fully closed genomes using NCBI Prokaryotic Genome Annotation Pipeline with default settings.

Functional annotation and analysis

Genome-level functional annotation was performed using Prokka v1.13.3 (https://github.com/tseemann/prokka) [33] and the RAST server in our seven closed genomes . CRISPRs and Cas regions were predicted by the CRISPR Cas-finder tool (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index). CRISPRs and Cas regions were extracted from annotated data submitted to the RAST server [34]. The Virulence Factor of Bacterial Pathogen Database (VFDB)

171 was used to predict virulence factors in these *Leptospira* genomes [35]. Mobile elements of the seven

172 isolates were identified by screening using tools at http://www.genomicepidemiology.org/services.

173 BLAST search was performed against the IS finder database for the seven genomes at

174 https://isfinder.biotoul.fr [36]. VM proteins were identified by performing a BLAST search (RAST

175 server) against isolates with known VM proteins.

176

177 ***In silico* PubMLST, CG View and Multiple genome alignment**

178 Conventional Multi-locus Sequence Typing (MLST) for the seven isolates against the

179 PubMLST database was performed using seven standardized housekeeping genes

180 https://pubmlst.org/leptospira/ [37]. Fully circularized annotated genomes obtained from the RAST

181 server were uploaded to the CGView server [38], an interactive comparative genomics tool for

182 circular genomes. For identification and alignment of conserved genomic DNA in the presence of

183 rearrangements and horizontal gene transfer, the software package Mauve

184 (https://darlinglab.org/mauve/mauve.html) was used [39, 40]. For multiple alignments, three of our

185 isolates (FMAS_AP8, FMAS_AP9 and FMAS_PN1), strain Piyasena strain JB 197, and L550 were

186 used.

187

188 **Methods to identify PF07598 (VM) protein homolog in animal-infecting strains of *L.***

189 ***borgpetersenii.***

190 Several different approaches were used to identify which (or whether) any of the four Sri

191 Lanka isolate VM homologs (orthologs, paralogs) were present in different strains of *L.*

192 *borgpetersenii* obtained from animals including serovar Hardjo strains HB203, TC112, TC129,

193 TC147, TC273, serovar Ballum strain LR131, and serovar Tarassovi strain MN900 [30]. In the first

194 approach, the Hidden Markov Model (HMM) for the Conserved Protein Domain Family DUF1561

195 was obtained from NCBI (https://www.ncbi.nlm.nih.gov/). Currently, there is only one

196 Pfam, PF07598, associated with the this Domain of Unknown Function (DUF) (C.f. Pfam: Family:

197 DUF1561 (PF07598) (xfam.org)). Pfam currently uses 16 species and 83 protein sequences to define

198 DUF1561. The putative protein sequences for each genome were obtained from their respective

7

199   NCBI annotations. The program hmmscan (http://hmmer.org/) was used to search all the annotated

200   proteins against the DUF1561.hmm model. The hmmscan options "-E 0.001 --domE 0.001" were

201   specified for the searches. The hmmscan reported three proteins meeting these criteria in HB203,

202   TC112, TC129, TC147, and TC273 genomes, four proteins meeting these criteria in the LR131 strain,

203   and two proteins meeting these criteria in the MN900 strain. The second and third searching

204   approaches did not use the NCBI protein annotations. This was done to eliminate the possibility that a

205   homolog could have been missed due to an incorrect or missing protein annotation. For the second

206   approach, the liberal method of searching the translations from every ORF over 50 bp in all six

207   reading frames was utilized. These translations were searched against the DUF1516 HMM as

208   described for the NCBI annotations. In the third method, all four of the Sri Lanka protein sequences

209   were compared by tblastn (default parameters) to the nucleotide sequence of the genomes of all six

210   other *L. borgpetersenii strains*. Any hits with a bitscore > 50 was considered putative positive output.

211   This analysis identified exactly three regions in each of the HB203, TC112, TC129, TC147, and

212   TC273 genomes, and four regions in the LR131 strain, and two regions in the MN900 strain. Looking

213   at the annotations associated with those regions (within each strain) revealed that these were the same

214   three annotations found using method one above. Taken together, this leads us to the conclusion that

215   there are only three coding regions that are homologous to the four Sri Lankan proteins in HB203,

216   TC112, TC129, TC147, and TC273 genomes, and four regions in the LR131 strain, and two regions

217   in the MN900 strain.

218

219   **Results**

220   The GC content of the isolates were ranged from 39.36%-39.54% (**Table 1**).

221   Total coding regions predicted for the isolates ranged from 3368 - 3521. FMAS_AP8 and

222   FMAS_AP9 had same number of coding sequences (CDSs) (3521) while FMAS_AP4 had the lowest

223   number of coding sequences. According to the NCBI annotation, proteins with functional assignment

224   ranged from 3,226 to 3,380 (**Table 1**) while number of hypothetical proteins predicted in the strains

225   had a range of 136-146. FMAS_ PN4 had the lowest number of hypothetical proteins. Two different

226   genomic types were clearly observed based on the coding sequence. FMAS_AP2, FMAS_AP3, and

227   FMAS_AP4 (Group 01) can contain an average of approximately 3,370 protein coding sequences. On

228   the other hand, in FMAS_PN1, FMAS_PN4, FMAS_AP8 and FMAS_AP9 (Group 02) contain about

229   3,520 protein coding sequences, an increase of about 4.5%. The average protein coding sequences for

230   L550, JB197, 56604, and TC112 are approximately 3,280, representing a 2.7%

231   reduction compared to Group 1 and a 7.3% reduction compared to Group 2. Thirty-seven tRNAs

232   were identified except in FMAS_AP4 in which, only 36 tRNAs were observed. RAST server based

233   subsystem analysis identified 226 in all the strains except in FMAS_AP4 which had only 225. Based

234   on the RAST analysis, CDSs involved in amino acid biosynthesis appeared to be the most abundant

235   subsystem in all strains. The FMAS_AP2 (170) had the highest number of predicted subsystems

236   whereas, strain FMAS_AP4 (168) was predicted to have the least number of subsystems. The

237   subsystem distribution of predicted CDSs in each of the strains is shown in Figure 1.

238

239   **Figure 1. Subsystem analysis by RAST server**

240

241   The ST144 MLST profile and CRISPRs and Cas regions predicted by the CRISPR Cas-finder

242   tool and two Crisper-Cas systems were identified in all seven isolates.The circular representation of

243   the seven genomes (CG view) is given in **Figure 1** and the arrangement of the CRISPR system given

244   in **Figure 2**.

245

246   **Figure 2. Circular Genome (CG) View Plot showing organization of *Leptospira* genomes**

247   **with annotated features**

248

249   Several putative virulence factors were identified in these *Leptospira* genomes using the VFDB

250   database. Twelve virulence factors were identified in each of the seven isolates (**Table 2**). Five main

251   IS elements were identified in seven isolates such as SLbp8**,** ISLbp4, ISLbp6, IS1533 and ISLbp5

252   **(Table 3).**

253   To visualize the general organization of the genome and discover potential genome

254   rearrangements among strains, conserved regions were visualized using a Mauve genome aligner

255   (**Figure 3**).

256

257   **Fig 3: Multiple alignment using Mauve 2**

258

259   Large Collinear Blocks (LCBs) were identified. Colored rectangular and variant-specific

260   regions (genomic islands, GI) or white region spaces within or between LCBs were identified in both

261   chromosomes, in all strains. However, chromosome II was highly conserved in all strains (**Fig. 4**).

262   Dimensions and the location of the central LCB on chromosome I was significantly different in our

263   isolates compared to strain piyasena, JB197 and L550. However, FMAS_AP8 and JB 197 had

264   conserved regions throughout the genome. Genomic islands and major genome rearrangements,

265   insertion sequence (IS) elements are often located at the intersection of these rearrangements, which

266   can lead to recombination. The total number of IS elements identified in these strains were 46. The

267   number of pseudogenes identified varied from 136-146. All seven isolates had four VM proteins

268   (**Table 4**).

269

270   **Figure 4. Genomic alignments of multiple *L. borgpetersenii* genomes**

271

272   We found that *L. borgpeterseni* has fewer VM proteins than *L. interrogans*, as exemplified by

273   comparison to the *L. interrogans* serovar Copenhageni str. Fiocruz L1-130 reference genome [41].

274   The PF07598 gene family encodes a newly identified leptospiral virulence factor family, the

275   Virulence Modifying (VM) proteins. There are four encoded VM proteins, one that is an ortholog of

276   LIC12339 and three that are orthologs of LIC12844. The sequence similarity ranged from 68.34% to

277   71.24%. The coding region encodes for 638, 632, 629, and 536 amino acids, respectively **(Table 5).**

278   We investigated whether the new Sri Lankan isolates shared VM homologs with strains of *L.*

279   *borgpetersenii* isolated from animal hosts, including bovine isolates of serovar Hardjo strains HB203,

280   TC112, TC129, TC147, TC273, a rodent isolate of serogroup Ballum strain LR131, and a bovine

281   isolate of serovar Tarassovi strain MN900. Three different methods were utilized to identify VM

282   homologs as described, including an hmmscan search of annotated protein sequences of the DUF1561

283   protein family, searching translations from all ORFs over 50 bp against the DUF1561 HMM, and

284   using tblastn to search the genomes' nucleotide sequences for any high scoring pairs returned from

285   querying with the known VM protein sequences. Compared to the four VM proteins present in the Sri

286   Lankan strains, collectively, all methods describe the presence of three VM homologs in the serovar

287   Hardjo strains (HB203, TC112, TC129, TC147, and TC273), four VM homologs in the rodent

288   serogroup Ballum strain LR131, and only two VM homologs in the serovar Tarassovi strain MN900.

289
290   **Discussion**

291   Here we present the whole genome analysis of new isolates of a novel, non-serotypable

292   *Leptospira borgpetersenii* isolated from humans in Sri Lanka. The main findings are that 1) these

293   isolates from humans, are essentially genomically identical; 2) the level of genome reduction appears

294   to be substantially less than originally proposed for reference *L. borgpetersenii* Hardjo-Bovis strains

295   [29] and therefore these new ex-human Sri Lanka isolates are not simply genomically degenerated

296   parasitic bacteria; and 3) the genomic analysis reflects emergence of a predominant leptospiral strain

297   (ST144), with Sri Lankan bovines as the likely source of human infection.

298   An increased incidence of human leptospirosis due to *L. borgpetersenii* has been reported

299   worldwide. In a study carried out in the Caribbean archipelago of Guadeloupe during an outbreak, the

300   isolates showed the emergence of the Ballum serogroup (*L. borgpetersenii*), serogroup

301   Icterohaemorrhagiae(*L. interrogans*) [8]. Another report from Malaysia also identified *L.*

302   *borgpetersenii* serovar Bataviae transmitted by two dominant rat species, *Rattus rattus* and *R.*

303   *norvegicus* [2]. *L. borgpetersenii* has been reported to cause severe human disease [42-44].

304   Previous studies have reported genome reduction in *L. borgpetersenii* serovar Hardjo strains

305   L550, JB197 and *L. borgpetersenii* serogroup Ballum serovar Ballum strain 56604 [5,9]. These

306   studies drew a general conclusion that the leptospiral species, *L. borgpetersenii*, has undergone IS-

307   mediated genome shrinkage due to inter-host transmission (not requiring environmental mediated

308   transmission). IS elements are also thought to be important features of the *L. borgpetersenii* genome

11

309 and mechanisms of genomic decay, contributing to multiple chromosomal rearrangements and

310 pseudogene formation. The total number of coding sequences reported in the three strains were

311 serovar Ballum 56604 (N=2618), serovar Hardjo strains L550 (N=2832) and serovar Hardjo strains

312 JB197 (N=2770) [5,9].

313   All seven isolates reported in this study belongs to MLST sequence type 144. In the

314 PubMLST database, seven isolates recovered both locally and globally have already been listed under

315 this ST. The first one was the *L. borgpetersenii* serovar Ceylonica isolated from a human in 1964

316 from Sri Lanka. Other local isolates include human samples from Gampaha, Giradurukotte ,

317 Bogammana and a rodent isolate from a black rat in Sri Lanka [32]. The other two are global isolates

318 each from Thailand and Laos. Since all seven isolates from the present study isolated from the dry

319 zone belong to the same ST 144, it might have emerged as the predominant sequence type in that

320 particular geographical region. cgMLST of these seven isolates revealed their clonal group as 267.

321 However, cgMLST data for the previous seven isolates aren't available for more comprehensive

322 analysis. According to the Mauve alignment genome, strain Piyasena (a previous Sri Lankan isolate)

323 is significantly different from our isolates. FMAS_AP8 and JB 197 had the significant number of

324 conserved regions. The pathogenesis of *L. borgpetereseni* strain Hardjo JB197 is an anomaly [45]; as

325 a laboratory isolate obtained from cattle, this strain is fairly unique for causing acute, lethal disease in

326 hamsters; its chromosome is rearranged significicantly compared to other Hardjo-Bovis strains.

327 However, we have not found this rearrangement of Chromosome II in our *L. borgpetereseni* isolates.

328   Previous studies have reported genome reduction in *L. borgpetersenii* serovar Hardjo strains

329 L550, JB197 and *L. borgpetersenii* serogroup Ballum serovar Ballum strain 56604 [5,9]. These

330 studies drew a general conclusion that the leptospiral species, *L. borgpetersenii*, has undergone IS-

331 mediated genome shrinkage due to inter-host transmission (not requiring environmental mediated

332 transmission). IS elements are also thought to be important features of the *L. borgpetersenii* genome

333 and mechanisms of genomic decay, contributing to multiple chromosomal rearrangements and

334 pseudogene formation. The total number of coding sequences reported in the three strains were

335 serovar Ballum 56604 (N=2,618), serovar Hardjo strains L550 (N=2,832) and serovar Hardjo strains

336 JB197 (N=2,770) [5,9].

337   Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated protein

338   systems are found in bacterial genomes, which are important to generate adaptive immunity against

339   invading exogenous genetic elements such as plasmid and phage infection [33,34]. The Cas gene

340   clusters are quite diverse, and they are frequently encoded by a diverse family of proteins with a wide

341   range of functional domains involved in nucleic acid interaction. Two main classes with six types and

342   numerous subtypes were identified in CRISPR Cas systems based on protein families and features of

343   the architecture of cas loci [35]. In pathogenic and intermediate *Leptospira*, three subtypes subtype I-

344   B, subtype I-C and subtype I-E were recognized. CRISPR Cas systems are not present in non-

345   infectious, saprophytic species [35].

346   CRISPR Cas Finder tool analysis revealed the presence of two CRISPR-Cas systems in all

347   seven isolates of this study. The CRISPR-Cas systems identified in these Sri Lankan isolates closely

348   resembles the sub type 1E with CRISPR array which was previously reported in *Leptospira*

349   *borgpetersenii* 56604. It was also identified in other group 1 species like *L. alexanderi, L. alstoni* and

350   *L. mayottensis, L. noguchii, L. santarosai, L. weilii* and *L. fainei* [35]. *L. borgpetersenii* serovar

351   Ballum reported to contain three crisper repeats GGTTCAACCCCACGCATGTGGGGAATAGGCT

352   between 2938442–2938534 [34]. In JB197 and L550 these repeats were not detected [34]. In our

353   seven isolates 6-7 repeats were detected. This shows the variability of our strains compared to

354   reported data in global literature. A recent study conducted in Malaysia has shown the presence of 10

355   to 16 loci with 1 to 13 spacers in the CRISPR arrays in six *L. interrogans* strains [13]. However, the

356   same study suggested further work was needed before making inferences on this observation with

357   relevance to pathogenicity and environmental adaptation of pathogenic *Leptospira*. [12].

358   The protein secretory systems that export proteins from the cytoplasm in *L. borpetersenii*

359   were found to be Type I and Type II [5]. However, the VF analyzer earch identified the presence of

360   VAS type VI secretion system in all seven of these isolates. The proteins, that were identified as

361   virulent factors were those coding for adherence, anti-phagocytosis, chemotaxis, mortality (invasion),

362   enzyme, lipid and fatty acid metabolism and stress adaptation. These proteins have also been

363   previously reported as virulence factors in other pathogenic *Leptospira* species. However, the number

13

364   of virulence factors identified in these seven isolates were comparatively low compared to other

365   pathogenic *Leptospira* [7,36].

366   Mobile elements (IS Elements) insertion can interrupt coding sequences and lead to

367   pseudogene formation in *Leptospira* [4,8]. The number of IS elements varies not just within species

368   but even within serovars. In *L. borgpetersenii,* a total of approximately 54 ISs scattered among

369   chromosomes of strain 56604 have been identified. This includes 31 copies of IS1533, 15 copies of

370   ISLin1, 4 copies of IS1502, 2 copies of IS1500, 1 copy of IS1501 and 1 copy of ISLin2 [9]. Strains

371   L550 and JB197 have been reported to have 121 and 116 IS elements copies, respectively. In

372   contrast, we found a lower number of IS elements in our isolates, N=46, a comparatively low number

373   [37]. In parallel to this observation, a relatively low number of pseudogenes were observed in human-

374   obtained Sri Lankan isolates (136-146**)** compared to published genomes of cattle-obtained Hardjo-

375   bovis L550, JB197, 56604, and TC112: N=248, 270, 231, and 400 pseudogenes respectively [12,38].

376   This could be attributed to relatively high number of mobile elements reported in those three strains

377   which may be related to host-pathogen or pathogen-environment interactions. Five types of mobile

378   elements (ISLbp4) belonging to the IS50 family, were identified in the seven Sri Lankan isolates

379   using a web-based mobile element finder. Strains L550 and JB197 were found to have 9 mobile

380   elements types belonging to different IS families. JB197 was isolated from cattle at slaughterhouses in

381   the United States and the L550 strain was isolated from a human with leptospirosis acquired

382   zoonotically from cattle in Australia. The strain 56604 of the serovar Ballum was isolated from a rat

383   in the west region of China. While genome reduction was observed in above strains, which were

384   probably having exclusive host-to-host transmission, our isolates from human cases, among whom the

385   transmission was probably environment-mediated and a lesser degree genome reduction was

386   observed. However this observation needs to be confirmed by further studies targeting animals,

387   humans, and the environment simultaneously [39]. The amplification of VM proteins in all seven

388   isolates isolated from humans compared with other animal-obtained *L. borgpetersenii* isolates which

389   have 2-3 VM proteins may be relevant to mechanisms of human infectivity and pathogenesis.

390   According to the literature, paralogous (PF07598) exists in all group I pathogens and the number

391   ranges from 2 to 12. Some proteins, such as LA1402, LA 0589, have been shown to be upregulated

392   during infection[40]. *L. borgpetersenii* sv. Javanica strain UI09931 was found to have four distinct

393   types of VM orthologs, including LA0591, LA0769, LA0835, and LA1402 [40]. However, only two

394   distinct orthologs LA1402 (N=1) and LA 0589 (N=3) were found in all seven Sri Lankan isolates.

395   The strains JB197, L550, 203, and L49 had identical VM orthologs in a 1:2 ratio to LA1402 and LA

396   0589. The strain *L. borgpetersenii* sv. Ceylonica strain Piyasena recovered from human subjects in Sri

397   Lanka had similar number of VM proteins as our seven isolates. However, the number of amino acids

398   coded for in one VM protein is relatively low (452) when compared to the seven isolates (536 aa), but

399   genome accuracy still remains to be validated regarding the true VM protein sequences.

400   While homologs of the four VM proteins found in the Sri Lankan isolates were identified in

401   alternate animal isolates of *L. borgpetersenii*, they were not consistent across serovar with only three

402   ortho/paralogs found in the serovar Hardjo strains (HB203, TC112, TC129, TC147, and TC273), four

403   homologs in the LR131 serovar Ballum strain, and only two in the MN900 serovar Tarassovi strain.

404   Further, while the number of ortho/paralogs may vary between strains, expression patterns of those

405   VM proteins may also vary by strain and between environmental conditions. For instance, recent

406   analysis of the transcriptome of HB203 (causes chronic disease in the hamster model of leptospirosis)

407   and JB197 serovar (causes severe acute) Hardjo strains cultured at 29°C and 37°C, shows that two of

408   the three VM Hardjo homologs were differentially expressed between strains at both 29°C and 37°C;

409   none of the VM proteins were differentially expressed at the transcriptomic level within strains

410   between temperatures under conditions tested, and *ex vivo* analysis or changes in the NaCl

411   concentrartion mimicking *in vivo* conditions may be required to see upregulation [41][26, 46, 47]. It is

412   notable that between strains, VW expression was higher in the severe disease causing JB197 strain

413   compared to the chronic HB203, which broadly suggests VM gene expression may be associated with

414   acute disease presentation in the hamster. In a proteomic data set looking at the highly similar HB203,

415   TC129, and TC273 strains, there is also evidence of strain-to-strain variation for VM proteins[42].

416   Collectively these results emphasize the need to further characterize expression of these unique

417   proteins and their role in promoting virulence of pathogenic leptospires. These datta provides some

418   indication that there may be less genome reduction and a larger PF07598 gene family repertoire in

419   human-infecting *L. borgpetersenii* strains, independent of serovar.

15

420 **Conclusion**

421   We isolated seven essentially identical *L. borgpetersenii* strains from humans with acute

422 febrile over a three year period. We show that a single Sequence Type, ST144, became the dominant

423 strain to cause human infection in the dry zone of Sri Lanka during this period. Genome reduction,

424 described for *L. borgpetersenii* Hardjo strains L550, JB197 and *L. borgpetersenii* serovar Ballum

425 strain 56604, was observed to a lesser degree in these seven Sri Lankan human isolates. Mauve

426 alignment indicates the presence of conserved regions and genome rearrangement within our isolates.

427 VM protein expansion in human-infecting *L. borgpetersenii* in Sri Lanka may contribute to adaptive

428 mechanisms for survival in the environment leading to human infection.

429

430

16

Author contributions

Indika Senavirathna: Conceptualization, Methodology, Software Data curation, Writing- Original draft preparation. **Dinesha Jayasundara**: Data curation, Visualization, Investigation, Writing- Reviewing and Editing. **Janith Warnasekara**: Visualization, Investigation, Editing. **Suneth Agampodi**: Funding acquisition, Supervision, Project Administration, Software, Validation, Writing- Reviewing and Editing. **Ellie J. Putz**: Investigation, Data curation, Methodology, Writing- Reviewing and Editing. **Jarlath E. Nally**: Investigation, Writing- Reviewing and Editing. **Darrell O. Bayles**: Investigation, Data curation, Methodology, Writing- Reviewing and Editing. **Joseph M. Vinetz**: Supervision, Reviewing and Editing, Project administration, funding acquisition.

Declaration of Competing Interest

Some of work reported here has been filed in patent applications from Yale University. JMV and spouse have an equity interest in Luna Bioscience, Inc, which may have a future interest in licensing this work. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

Funding

17

455   **Data availability**

456   Annotated assemblies are available in GenBank under accession numbers:

457   CP072630:CP072631(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072630:CP072631[accn])

458   CP072628:CP072629(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072628:CP072629[accn])

459   CP072626:CP072627(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072626:CP072627[accn])

460   CP072624:CP072625(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072624:CP072625[accn])

461   CP072622:CP072623(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072622:CP072623[accn])

462   CP072620:CP072621(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072620:CP072621[accn])

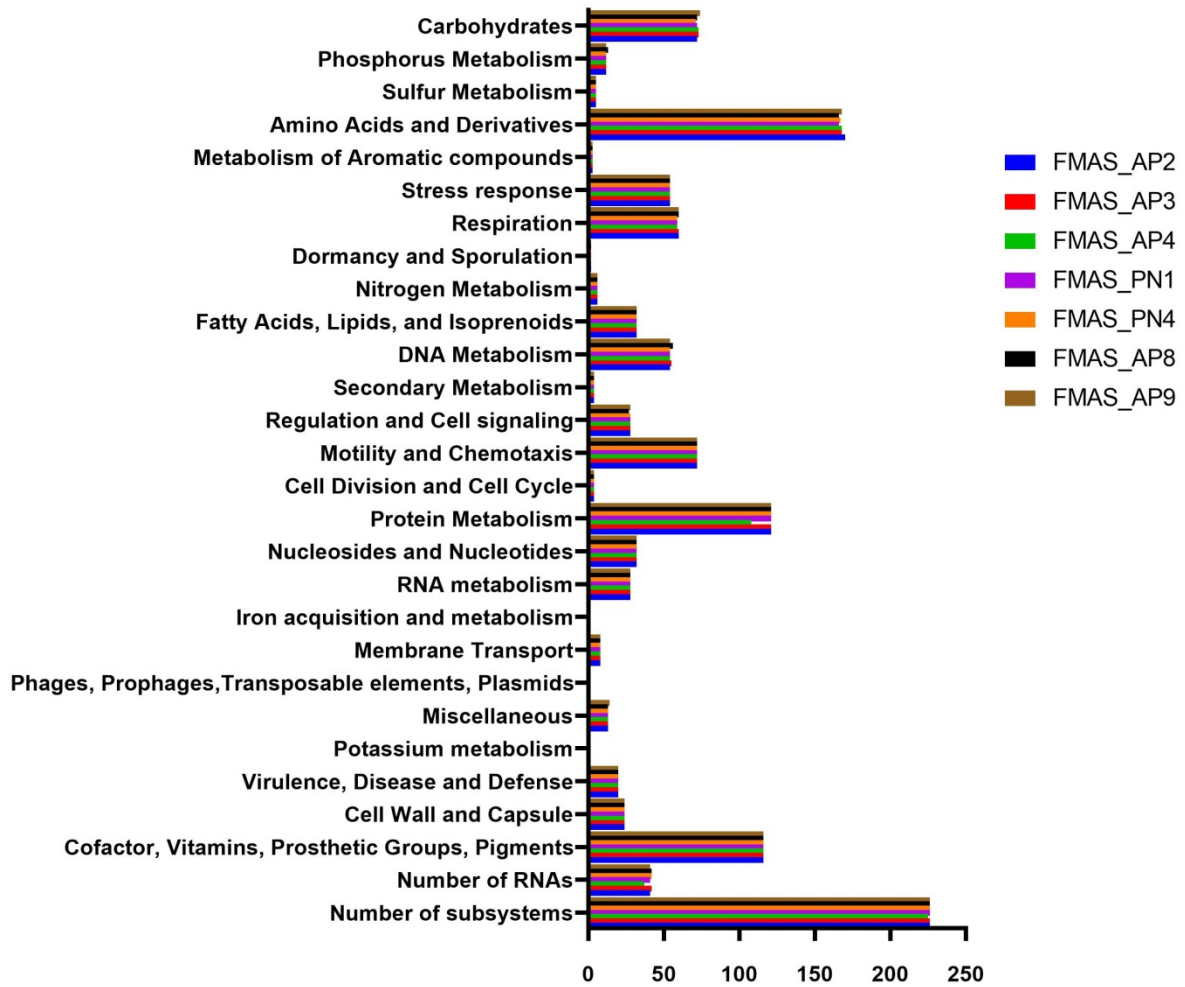463   CP072618:CP072619(https://www.ncbi.nlm.nih.gov/nuccore/?term=CP072618:CP072619[accn])

464

465

466

467
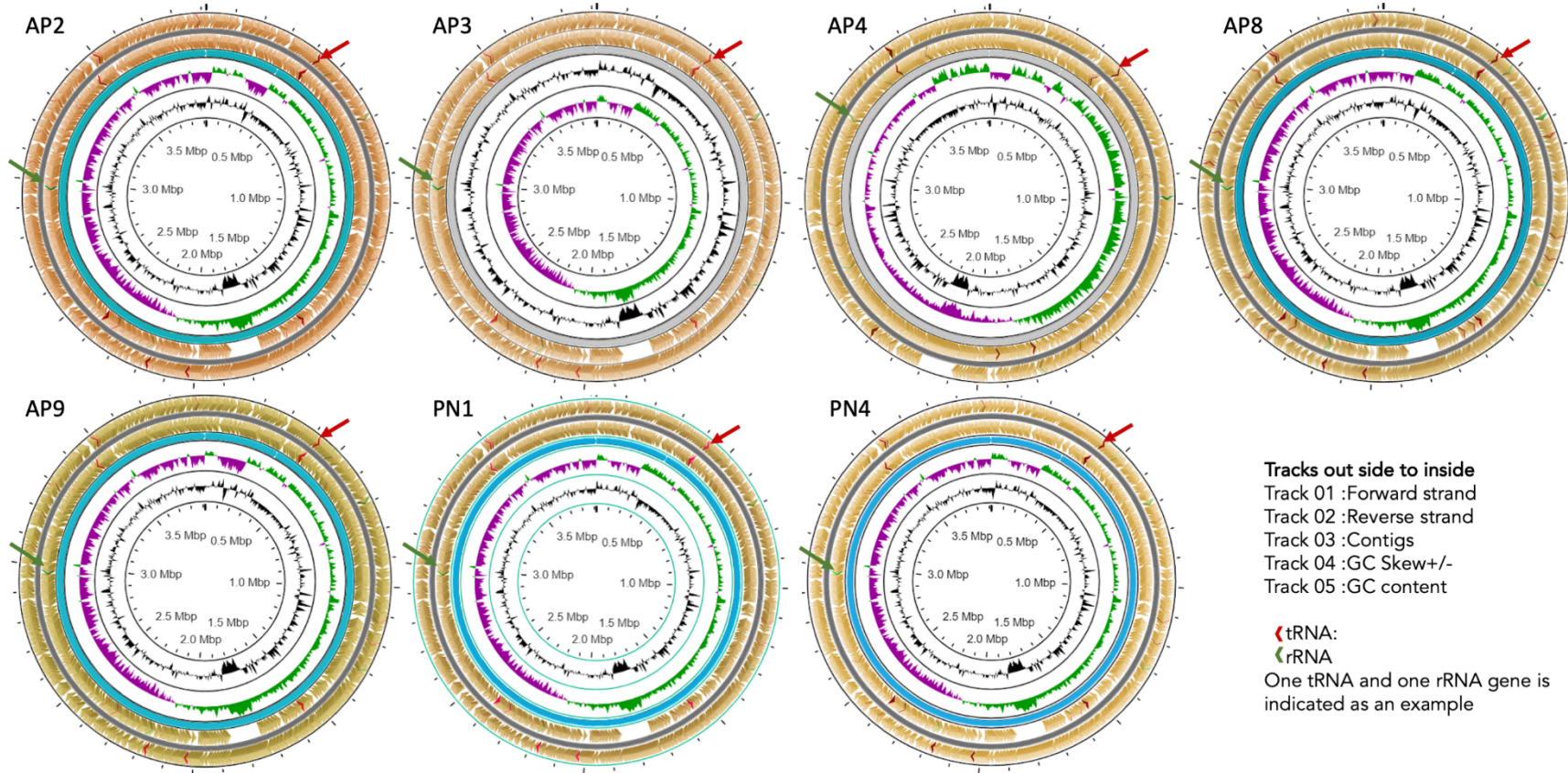
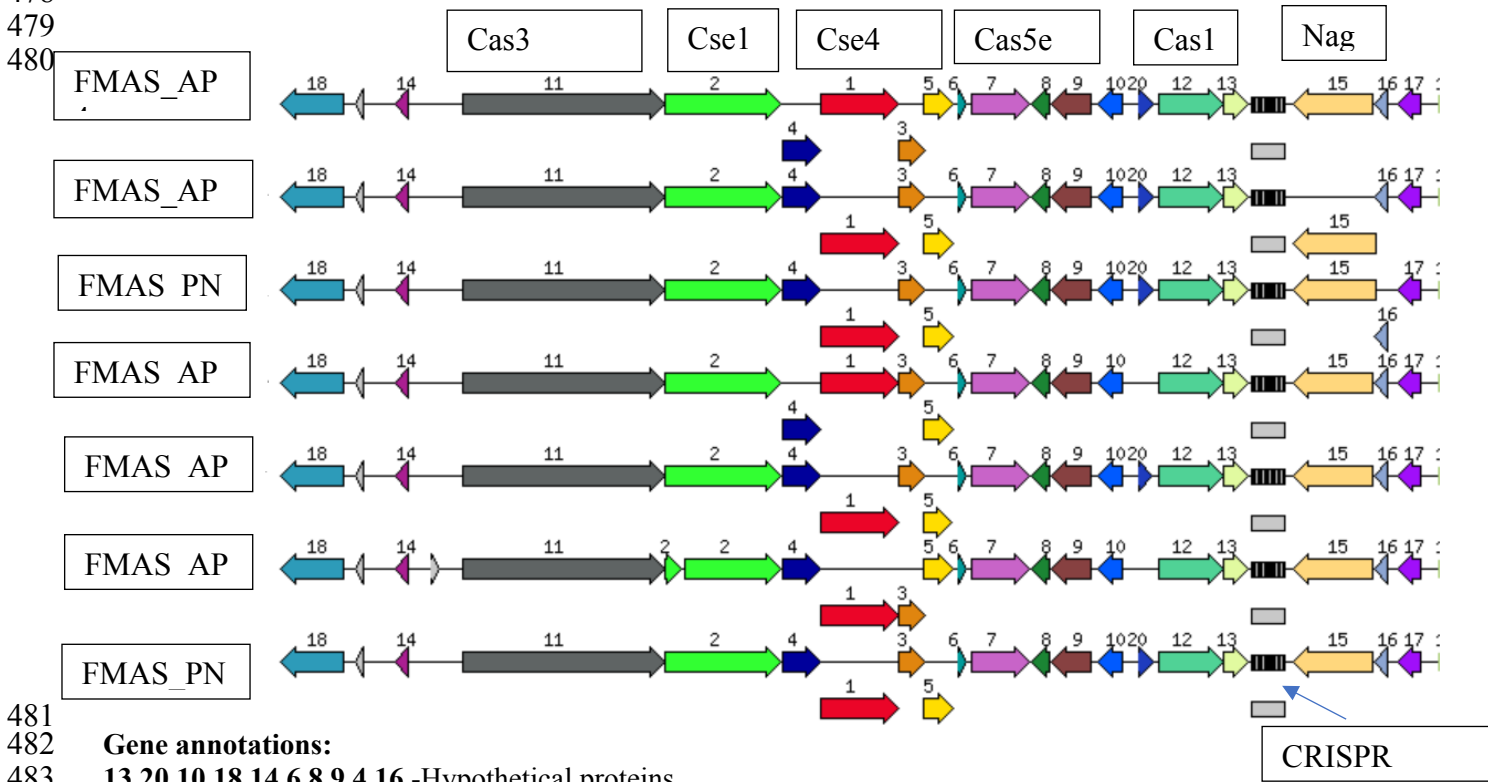468   **Figure 1. Genomic functional analysis by functional category**

469

Figure 2. Annotated organization of seven -ex human, Sri Lankan *Leptospira borgpetersenii* genomes using Circular Genome View Plot

Figure 3. Arrangement of CRISPR/Cas systems in seven -ex human, Sri Lankan *Leptospira borgpetersenii* genomes generated using RAST subsystem analysis

Gene annotations:

13,20,10,18,14,6,8,9,4,16 -Hypothetical proteins

17-SMR protein family:Small multidrug resistance family

11-Cas3- CRISPR associated helicase, **2**-Cse1-CRISPR associated protein, **1**-Cse4- CRISPR associated protein, **5**-Cas5e- CRISPR associated protein,

 12-Cas1- CRISPR associated protein, NagX- N-acetylglucosamine related transporter

Figure 4. Genomic alignments of seven -ex human, Sri Lankan *Leptospira borgpetersenii* isolates and comparison with reference *Leptospira interrogans* serovar Copenhageni genome using Mauve 2
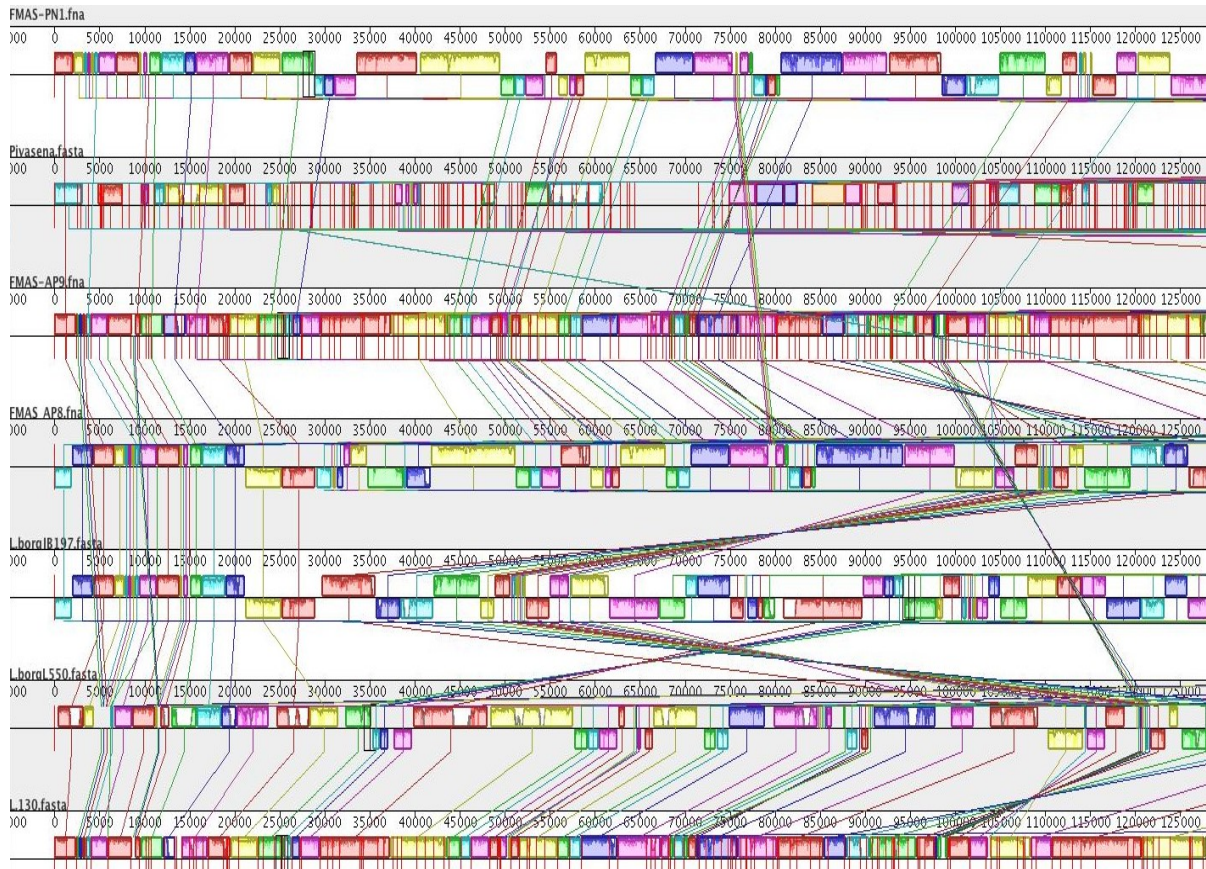
Snap shot comparing the genomic organization of three Sri Lankan *L. borgpetersenii* isolates with other genomes. Genomes of these strains were aligned and arranged using the Mauve genome aligner in the following order: Top, FMAS_PN1, *L. borgpetersenii* serovar Piyasena, FMAS_AP9, FMAS_AP8, JB 197, L550; at very botton, the reference genome, *L. interrogans* serovar Copenhageni, strain L1-130. Large collinear blocks (LCBs) correspond mainly to conserved syntenic regions, as represented by colored boxes. The lines between the genomes connect the blocks that are conserved between two strains and larger scale rearrangements.

516 **Table 1. Genome features of Seven _L.borgpetersenii_ isolates**
517

	FMAS_AP2	FMAS_AP3	FMAS_AP4	FMAS_PN1	FMAS_PN4	FMAS_AP8	FMAS_AP9	
Size(Mbp)	3.91	3.91	3.91	3.91	3.91	3.91	3.91	
Chr I size (Mb)	3.59	3.59	3.59	3.59	3.59	3.59	3.59	
Chr II size (Mb)	0.32	0.32	0.32	0.32	0.32	0.32	0.32	
G+C(%)	39.36	39.36	39.37	39.43	39.4	39.38	39.54	
CDS								
Hypothetical Protein	143	144	139	140	136	146	141	
Protein with functional assignment	3232	3226	3229	3375	3373	3375	3380	
Total	3375	3370	3368	3515	3509	3521	3521	
tRNA genes	37	37	36	37	37	37	37	
CRISPER's(#repeats)	2	2	2	2	2	2	2	
rRNA								
23s	2	2	2	2	2	2	2	
16s	2	2	2	2	2	2	2	
5s	1	1	1	1	1	1	1	
Pseudo genes	143	144	139	140	136	146	141	
Pseudo Genes (ambiguous residues)	0	0	0	0	0	0	0	
Pseudo Genes (frameshifted)	108	109	104	104	101	111	107	
Pseudo Genes (incomplete)	57	57	58	57	57	57	58	
Pseudo Genes (internal stop)	33	33	33	33	33	33	31	
Pseudo Genes (multiple problems)	45	45	45	45	45	45	45	

518   *Details were given in the table were generated from NCBI annaotation pipeline
519
520   NA-Not available
521
522   TC112, TC147, TC129 and TC273 _Leptospira borgpetersenii_ serovar Hardjo isolated cattle from
523   USA. Only TC112 details were given in the table since all four of these strains have similar features.
524
525
526

23

527 **Table 2. Virulence factors identified in Seven Sri Lankan *Leptospira borgpetersenii* Isolates**

Virulence class	Virulence factor	Gene	Length
Adherence	Mannose-sensitive hemagglutinin (MSHA type IV pilus)	pilB	1674
	GroEL(Clostridium)	groEL	1641
Antiphagocytosis	Capsular polysaccharide	Capsule 1	1023
Invasion	Flagella	flhA	2118
		cheA	3195
		cheB	1047
Enzyme	Streptococcal enolase(Streptococcus)	eno	1299
Secretion system	VAS type VI secretion system	clpV	2232
		clpV	2544
Lipid and fatty acid metabolism	Pantothenate synthesis(Mycobacterium)	panD	351
Stress adaptation	Catalases	katA	1458

547 *All seven isolates had similar distribution of virulence factors. VF analyzer:automatic pipeline was
548 used for the systematic analysis and was accessed September 1, 2024.
549

24

Table 3. Mobile Elements

Main Mobile elements from seven isolates Sequences producing significant alignment	IS Family	Group	Organism
ISLbp8	ISNCY	ISLbi1	*Leptospira borgpetersenii*
ISLbp4	IS4	IS50	*Leptospira kirschneri*
ISLbp6	IS630		*Leptospira borgpetersenii*
IS1533	IS110	IS1111	*Leptospira borgpetersenii*
ISLbp5	IS5	IS427	*Leptospira borgpetersenii*
***L. borgpetersenii* serovar Hardjo, strains L550**			
Sequences producing significant alignment	**IS Family**	**Group**	**Organism**
ISLbp2	IS256		*Leptospira borgpetersenii*
ISLbp1	IS4	IS4	*Leptospira borgpetersenii*
IS1533	IS110	IS1111	*Leptospira borgpetersenii*
ISLbp4	IS4	IS50	*Leptospira kirschneri*
ISLbp3	IS982		*Leptospira borgpetersenii*
ISLbp8	ISNCY	ISLbi1	*Leptospira borgpetersenii*
ISLbp6	IS630		*Leptospira borgpetersenii*
ISLbp5	IS5	IS427	*Leptospira borgpetersenii*
IS1501	IS3	IS3	*Leptospira interrogans*
***L. borgpetersenii serovar* Hardjo, strains JB197**			
Sequences producing significant allinment	**IS Family**	**Group**	**Organism**
SLbp2	IS256		*Leptospira borgpetersenii*
ISLbp1	IS4	IS4	*Leptospira borgpetersenii*
IS1533	IS110	IS1111	*Leptospira borgpetersenii*
ISLbp4	IS4	IS50	*Leptospira kirschneri*
ISLbp3	IS982		*Leptospira borgpetersenii*
ISLbp8	ISNCY	ISLbi1	*Leptospira borgpetersenii*
ISLbp6	IS630		*Leptospira borgpetersenii*
ISLbp5	IS5	IS427	*Leptospira borgpetersenii*
IS1501	IS3	IS3	*Leptospira interrogans*

Mobile elements were identified via BLAST search of the seven genomes against the IS Finder Database (https://isfinder.biotoul.fr).

25

583 **Table 4. Comparison of VM Protein Profile Among Seven Ex-human Sri Lankan *Leptospira***

584 ***borgpetersenii* New Isolates and Historical Data**

Isolate	Species	Serovar	No. VM Proteins	Host
FMAS_AP2	*L. borgpetersenii*	No agglutination	4	Human
FMAS_AP3	*L. borgpetersenii*	No agglutination	4	Human
FMAS_AP4	*L. borgpetersenii*	No agglutination	4	Human
FMAS_PN1	*L. borgpetersenii*	No agglutination	4	Human
FMAS_PN4	*L. borgpetersenii*	No agglutination	4	Human
FMAS_AP8	*L. borgpetersenii*	No agglutination	4	Human
FMAS_AP9	*L. borgpetersenii*	No agglutination	4	Human
Strain 4E	*L. borgpetersenii*	No agglutination	2	Rat
55604	*L. borgpetersenii*	Ballum	2	Rat
JB197	*L. borgpetersenii*	Hardjo-Bovis	3	Cattle
L550	*L. borgpetersenii*	Hardjo-Bovis	3	Human
Piyasena	*L. borgpetersenii*	Ceylonica	4	Human
L49	*L. borgpetersenii*	Hardjo-Bovis	3	Cattle
203	*L. borgpetersenii*	Hardjo-Bovis	3	Cattle

585

586
587

588

589

590

591
592

26

593   **Table 5. Comparison of the PF07598 (VM Protein) Homologs (Orthologs, Paralogs) in the New**

594   **Seven Ex-Human *Leptospira borgpetersenii* New Isolates and Historical Sri Lankan Isolates to**

595   **Ex-Animal Isolates**

Serovar	Strain	Number of VM proteins	Number of Amino acids	*Sequence similarity	Accession numbers	Locus tag
Hardjo-bovis	JB197	3	638	67.75%	AAS70908.1	LI
			629	69.14%	AAS71397.1	LI
			626	69.58%	AAS71397.1	LI
Hardjo-bovis	L550	3	638	67.81%	AAS70908.1	LI
			629	69.14%	AAS71397.1	LI
			626	69.90%	AAS71397.1	LI
Ballum	4E	2	638	68.12%	AAS70908.1	LI
			632	68.60%	AAS71397.1	LI
Ballum	56604	2	638	68.12%	AAS70908.1	LI
			632	68.60%	AAS71397.1	LI
Hardjo-bovis	203	3	638	67.81%	AAS70908.1	LI
			629	69.31%	AAS71397.1	LI
			626	69.90%	AAS71397.1	LI
Hardjo-bovis	L49	3	638	67.81%	AAS70908.1	LI
			629	69.31%	AAS71397.1	LI
			626	69.90%	AAS71397.1	LI
Ceylonica	Piyasena	4	638	68.43%	AAS70908.1	LI
			632	70.37%	AAS71397.1	LI
			536	68.70%	AAS71397.1	LI
			452	71.40%	AAS71397.1	LI
No agglutination	***FMAS_AP2	4	638	68.43%	AAS70908.1	LI
			632	70.37%	AAS71397.1	LI
			629	71.24%	AAS71397.1	LI
			536	68.70%	AAS71397.1	LI

596   * Sequence similarity is given to compared *Leptospira interrogans* serovar Copenhageni str. Fiocruz
597   L1-130*** FMAS_AP3 ,FMAS_AP4,FMAS_AP8,FMAS_AP9,FMAS_PN1 and FMAS_PN4
598   identical to FMAS_AP2.
599
600
601
602

27

References

603

604

605 1.   Vinetz JM, Chetrick AN, Steimel KE, Devillers CE, Excler J-L, Kim JH, et al. Getting
606 leptospirosis onto the lists of neglected tropical diseases. Open Access Government
607 . 2024;https://www.openaccessgovernment.org/article/getting-leptospirosis-onto-the-lists-of-
608 neglected-tropical-diseases/177928:52-4.
609 2.   Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, et al. Global
610 Morbidity and Mortality of Leptospirosis: A Systematic Review. PLoS Negl Trop Dis.
611 2015;9(9):e0003898. Epub 2015/09/18. doi: 10.1371/journal.pntd.0003898. PubMed PMID:
612 26379143; PubMed Central PMCID: PMCPMC4574773.
613 3.   Fouts DE, Matthias MA, Adhikarla H, Adler B, Amorim-Santos L, Berg DE, et al. What
614 Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira.
615 PLoS Negl Trop Dis. 2016;10(2):e0004403. Epub 2016/02/20. doi:
616 10.1371/journal.pntd.0004403. PubMed PMID: 26890609; PubMed Central PMCID:
617 PMCPMC4758666.
618 4.   Torgerson PR, Hagan JE, Costa F, Calcagno J, Kane M, Martinez-Silveira MS, et al. Global
619 Burden of Leptospirosis: Estimated in Terms of Disability Adjusted Life Years. PLoS Negl Trop Dis.
620 2015;9(10):e0004122. Epub 2015/10/03. doi: 10.1371/journal.pntd.0004122. PubMed PMID:
621 26431366; PubMed Central PMCID: PMCPMC4591975.
622 5.   Warnasekara J, Agampodi S. Neglecting the neglected during the COVID-19 pandemic: the
623 case of leptospirosis in Sri Lanka. Epidemiol Health. 2022;44:e2022015. Epub 2022/01/19. doi:
624 10.4178/epih.e2022015. PubMed PMID: 35038829.
625 6.   Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, Lovett MA, et al. Leptospirosis: A
626 zoonotic disease of global importance. Lancet Infect Dis. 2003;3:757-71.
627 7.   Kim CL, Agampodi S, Marks F, Kim JH, Excler JL. Mitigating the effects of climate change
628 on human health with vaccines and vaccinations. Front Public Health. 2023;11:1252910. Epub
629 20231012. doi: 10.3389/fpubh.2023.1252910. PubMed PMID: 37900033; PubMed Central
630 PMCID: PMCPMC10602790.
631 8.   Warnasekara J, Srimantha S, Kappagoda C, Jayasundara D, Senevirathna I, Matthias M, et al.
632 Diagnostic method-based underestimation of leptospirosis in clinical and research settings; an
633 experience from a large prospective study in a high endemic setting. PLoS Negl Trop Dis.
634 2022;16(4):e0010331. Epub 2022/04/05. doi: 10.1371/journal.pntd.0010331. PubMed PMID:
635 35377883; PubMed Central PMCID: PMCPMC9009773.
636 9.   Warnasekara J, Srimantha SM, Agampodi SB. Estimating the burden of leptospirosis: global
637 lessons from Sri Lanka. BMJ Global Health. 2021;6(10):e006946. doi: 10.1136/bmjgh-2021-
638 006946.
639 10.   Athukorala N, Wickramasinghe S, Yatawara L, Weerakoon K. A Systematic Review of
640 Zoonotic Disease Prevalence in Sri Lanka (2000-2022). Vector Borne Zoonotic Dis.
641 2024;24(8):453-72. Epub 20240522. doi: 10.1089/vbz.2023.0141. PubMed PMID: 38775108.
642 11.   Baharom M, Ahmad N, Hod R, Ja'afar MH, Arsad FS, Tangang F, et al. Environmental and
643 Occupational Factors Associated with Leptospirosis: A Systematic Review. Heliyon.
644 2024;10(1):e23473. Epub 20231210. doi: 10.1016/j.heliyon.2023.e23473. PubMed PMID:
645 38173528; PubMed Central PMCID: PMCPMC10761560.
646 12.   Stark AM, Nohrenberg M, Draper AD, McMahon KE, Hewitt TA, Lomas K, et al. A cluster
647 of leptospirosis cases associated with crocodile workers in the Northern Territory, Australia, 2022.
648 Commun Dis Intell (2018). 2023;47. Epub 20231116. doi: 10.33321/cdi.2023.47.70. PubMed
649 PMID: 37968070.
650 13.   Rees EM, Lotto Batista M, Kama M, Kucharski AJ, Lau CL, Lowe R. Quantifying the
651 relationship between climatic indicators and leptospirosis incidence in Fiji: A modelling study. PLOS
652 Glob Public Health. 2023;3(10):e0002400. Epub 20231011. doi: 10.1371/journal.pgph.0002400.
653 PubMed PMID: 37819894; PubMed Central PMCID: PMCPMC10566718.

28

654   14.   Jayaramu V, Zulkafli Z, De Stercke S, Buytaert W, Rahmat F, Abdul Rahman RZ, et al.
655   Leptospirosis modelling using hydrometeorological indices and random forest machine learning. Int J
656   Biometeorol. 2023;67(3):423-37. Epub 20230131. doi: 10.1007/s00484-022-02422-y. PubMed
657   PMID: 36719482.
658   15.   Vincent AT, Schiettekatte O, Goarant C, Neela VK, Bernet E, Thibeaux R, et al. Revisiting
659   the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics.
660   PLoS neglected tropical diseases. 2019;13(5):e0007270.
661   16.   Jayasundara D, Gamage C, Senavirathna I, Warnasekara J, Matthias MA, Vinetz JM, et al.
662   Optimizing the microscopic agglutination test (MAT) panel for the diagnosis of Leptospirosis in a low
663   resource, hyper-endemic setting with varied microgeographic variation in reactivity. PLoS Negl Trop
664   Dis. 2021;15(7):e0009565. Epub 2021/07/02. doi: 10.1371/journal.pntd.0009565. PubMed
665   PMID: 34197457; PubMed Central PMCID: PMCPMC8279374.
666   17.   Levett PN, Branch SL. Evaluation of two enzyme-linked immunosorbent assay methods for
667   detection of immunoglobulin M antibodies in acute leptospirosis. Am J Trop Med Hyg.
668   2002;66(6):745-8. PubMed PMID: 12224584.
669   18.   Yersin C, Bovet P, Merien F, Wong T, Panowsky J, Perolat P. Human leptospirosis in the
670   Seychelles (Indian Ocean): a population-based study. Am J Trop Med Hyg. 1998;59(6):933-40.
671   19.   Levett PN. Leptospirosis. Clin Microbiol Rev. 2001;14(2):296-326.
672   20.   Maia MAC, Bettin EB, Barbosa LN, de Oliveira NR, Bunde TT, Pedra ACK, et al.
673   Challenges for the development of a universal vaccine against leptospirosis revealed by the evaluation
674   of 22 vaccine candidates. Front Cell Infect Microbiol. 2022;12:940966. Epub 20221007. doi:
675   10.3389/fcimb.2022.940966. PubMed PMID: 36275031; PubMed Central PMCID:
676   PMCPMC9586249.
677   21.   Chaurasia R, Salovey A, Guo X, Desir G, Vinetz JM. Vaccination With Leptospira
678   interrogans PF07598 Gene Family-Encoded Virulence Modifying Proteins Protects Mice From
679   Severe Leptospirosis and Reduces Bacterial Load in the Liver and Kidney. Front Cell Infect
680   Microbiol. 2022;12:926994. Epub 20220628. doi: 10.3389/fcimb.2022.926994. PubMed PMID:
681   35837473; PubMed Central PMCID: PMCPMC9274288.
682   22.   Giraud-Gatineau A, Nieves C, Harrison LB, Benaroudj N, Veyrier FJ, Picardeau M.
683   Evolutionary insights into the emergence of virulent Leptospira spirochetes. PLoS Pathog.
684   2024;20(7):e1012161. Epub 20240717. doi: 10.1371/journal.ppat.1012161. PubMed PMID:
685   39018329; PubMed Central PMCID: PMCPMC11285912.
686   23.   Picardeau M. Virulence of the zoonotic agent of leptospirosis: still terra incognita? Nat Rev
687   Microbiol. 2017;15(5):297-307. Epub 2017/03/07. doi: 10.1038/nrmicro.2017.5. PubMed PMID:
688   28260786.
689   24.   Ko AI, Goarant C, Picardeau M. Leptospira: the dawn of the molecular genetics era for an
690   emerging zoonotic pathogen. Nat Rev Microbiol. 2009;7(10):736-47. Epub 2009/09/17. doi:
691   nrmicro2208 [pii]
692   10.1038/nrmicro2208. PubMed PMID: 19756012.
693   25.   Chaurasia R, Marroquin AS, Vinetz JM, Matthias MA. Pathogenic Leptospira Evolved a
694   Unique Gene Family Comprised of Ricin B-Like Lectin Domain-Containing Cytotoxins. Front
695   Microbiol. 2022;13:859680. Epub 20220329. doi: 10.3389/fmicb.2022.859680. PubMed PMID:
696   35422779; PubMed Central PMCID: PMCPMC9002632.
697   26.   Lehmann JS, Fouts DE, Haft DH, Cannella AP, Ricaldi JN, Brinkac L, et al. Pathogenomic
698   inference of virulence-associated genes in Leptospira interrogans. PLoS Negl Trop Dis.
699   2013;7(10):e2468. Epub 2013/10/08. doi: 10.1371/journal.pntd.0002468. PubMed PMID:
700   24098822; PubMed Central PMCID: PMCPMC3789758.
701   27.   Guglielmini J, Bourhy P, Schiettekatte O, Zinini F, Brisse S, Picardeau M. Genus-wide
702   Leptospira core genome multilocus sequence typing for strain taxonomy and global surveillance.
703   PLoS Negl Trop Dis. 2019;13(4):e0007374. Epub 20190426. doi: 10.1371/journal.pntd.0007374.
704   PubMed PMID: 31026256; PubMed Central PMCID: PMCPMC6513109.

705   28.   Grillova L, Picardeau M. Core Genome Multi-locus Sequence Typing Analyses of Leptospira
706   spp. Using the Bacterial Isolate Genome Sequence Database. Methods Mol Biol. 2020;2134:11-21.
707   doi: 10.1007/978-1-0716-0459-5_2. PubMed PMID: 32632855.
708   29.   Bulach DM, Zuerner RL, Wilson P, Seemann T, McGrath A, Cullen PA, et al. Genome
709   reduction in Leptospira borgpetersenii reflects limited transmission potential. Proc Natl Acad Sci U S
710   A. 2006;103(39):14560-5. Epub 2006/09/16. doi: 0603979103 [pii]
711   10.1073/pnas.0603979103. PubMed PMID: 16973745; PubMed Central PMCID: PMC1599999.
712   30.   Putz EJ, Bayles DO, Alt DP, Nally JE. Complete Genome Sequence of Four Strains of
713   Leptospira borgpetersenii serovar Hardjo isolated from Cattle in the Central United States. J
714   Genomics. 2022;10:45-8. Epub 20220214. doi: 10.7150/jgen.69822. PubMed PMID: 35300048;
715   PubMed Central PMCID: PMCPMC8922301.
716   31.   Agampodi S, Warnasekara J, Siribaddana S, Kularatna S, Gamage C, Jayasundara D, et al.
717   Demographic, exposure, clinical, biochemical and diagnostic data of febrile patients recruited for the
718   largest field study on leptospirosis in Sri Lanka. Data Brief. 2022;43:108378. Epub 20220612. doi:
719   10.1016/j.dib.2022.108378. PubMed PMID: 35770027; PubMed Central PMCID:
720   PMCPMC9234077.
721   32.   Jayasundara D, Senavirathna I, Warnasekara J, Gamage C, Siribaddana S, Kularatne SAM, et
722   al. 12 Novel clonal groups of Leptospira infecting humans in multiple contrasting epidemiological
723   contexts in Sri Lanka. PLoS Negl Trop Dis. 2021;15(5):e0009471. Epub 20210526. doi:
724   10.1371/journal.pntd.0009471. PubMed PMID: 34038414; PubMed Central PMCID:
725   PMCPMC8153494.
726   33.   Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
727   2014;30(14):2068-9. Epub 20140318. doi: 10.1093/bioinformatics/btu153. PubMed PMID:
728   24642063.
729   34.   Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server:
730   rapid annotations using subsystems technology. BMC Genomics. 2008;9:75. Epub 20080208. doi:
731   10.1186/1471-2164-9-75. PubMed PMID: 18261238; PubMed Central PMCID:
732   PMCPMC2265698.
733   35.   Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial
734   virulence factors. Nucleic Acids Res. 2005;33(Database issue):D325-8. doi: 10.1093/nar/gki008.
735   PubMed PMID: 15608208; PubMed Central PMCID: PMCPMC539962.
736   36.   Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for
737   bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32-6. doi:
738   10.1093/nar/gkj014. PubMed PMID: 16381877; PubMed Central PMCID: PMCPMC1347377.
739   37.   Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb
740   software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124. Epub
741   20180924. doi: 10.12688/wellcomeopenres.14826.1. PubMed PMID: 30345391; PubMed Central
742   PMCID: PMCPMC6192448.
743   38.   Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using
744   the CGView family of tools. Brief Bioinform. 2019;20(4):1576-82. doi: 10.1093/bib/bbx081.
745   PubMed PMID: 28968859; PubMed Central PMCID: PMCPMC6781573.
746   39.   Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved
747   genomic sequence with rearrangements. Genome Res. 2004;14(7):1394-403. doi:
748   10.1101/gr.2289704. PubMed PMID: 15231754; PubMed Central PMCID: PMCPMC442156.
749   40.   Darling AE, Treangen TJ, Messeguer X, Perna NT. Analyzing patterns of microbial evolution
750   using the mauve genome alignment system. Methods Mol Biol. 2007;396:135-52. doi:
751   10.1007/978-1-59745-515-2_10. PubMed PMID: 18025691.
752   41.   Nascimento AL, Ko AI, Martins EA, Monteiro-Vitorello CB, Ho PL, Haake DA, et al.
753   Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology
754   and pathogenesis. J Bacteriol. 2004;186(7):2164-72. PubMed PMID: 15028702.

755   42.   Rajaonarivelo JA, Desmoulin A, Maillard O, Collet L, Baudino F, Jaffar-Bandjee MC, et al.
756   Clinical manifestations of human leptospirosis: bacteria matter. Front Cell Infect Microbiol.
757   2023;13:1259599. Epub 20231025. doi: 10.3389/fcimb.2023.1259599. PubMed PMID:
758   37953799; PubMed Central PMCID: PMCPMC10635415.
759   43.   Desmoulin A, Rajaonarivelo A, Maillard O, Collet L, Jaffar-Bandjee MC, Moiton MP, et al.
760   A Comparative Study of Human Leptospirosis between Mayotte and Reunion Islands Highlights
761   Distinct Clinical and Microbial Features Arising from Distinct Inter-Island Bacterial Ecology. Am J
762   Trop Med Hyg. 2024;111(2):237-45. Epub 20240702. doi: 10.4269/ajtmh.23-0846. PubMed
763   PMID: 38955193; PubMed Central PMCID: PMCPMC11310607.
764   44.   …, Mariet J, Picardeau M, Agampodi S, ... 12 Novel clonal groups of Leptospira infecting
765   humans in multiple contrasting epidemiological contexts in Sri Lanka: journals.plos.org; 2021.
766   45.   Putz EJ, Sivasankaran SK, Fernandes LGV, Brunelle B, Lippolis JD, Alt DP, et al. Distinct
767   transcriptional profiles of Leptospira borgpetersenii serovar Hardjo strains JB197 and HB203
768   cultured at different temperatures. PLoS Negl Trop Dis. 2021;15(4):e0009320. Epub 20210407.
769   doi: 10.1371/journal.pntd.0009320. PubMed PMID: 33826628; PubMed Central PMCID:
770   PMCPMC8055020.
771   46.   Matsunaga J, Lo M, Bulach DM, Zuerner RL, Adler B, Haake DA. Response of Leptospira
772   interrogans to physiologic osmolarity: relevance in signaling the environment-to-host transition. Infect
773   Immun. 2007;75(6):2864-74. Epub 2007/03/21. doi: IAI.01619-06 [pii]
774   10.1128/IAI.01619-06. PubMed PMID: 17371863; PubMed Central PMCID: PMC1932867.
775   47.   Matsunaga J, Sanchez Y, Xu X, Haake DA. Osmolarity, a key environmental signal
776   controlling expression of leptospiral proteins LigA and LigB and the extracellular release of LigA.
777   Infect Immun. 2005;73(1):70-8. PubMed PMID: 15618142.
778