

20 Abstract

21 **Background:** Leptospirosis, caused by *Leptospira* bacteria, is a common zoonosis worldwide
22 more prevalent in the tropics. Reservoir species and risk factors have been identified but
23 surveys for environmental sources of leptospirosis are rare. Furthermore, understanding of
24 environmental *Leptospira* containing pathogenic genes and possibly capable of causing disease
25 is incomplete and could result in some pathogenic strains evading detection, thereby
26 convoluting diagnosis, prevention, and epidemiology.

27 **Methodology/Principal Findings:** We collected environmental samples from 22 sites in Puerto
28 Rico during three sampling periods over 14-months (Dec 2018-Feb 2020); 10 water and 10 soil
29 samples were collected at each site. Samples were screened for pathogenic *Leptospira* DNA
30 using the *lipL32* PCR assay and positive samples were sequenced to assess genetic diversity.
31 One urban site in San Juan was sampled three times over 14 months to assess persistence in
32 soil; live leptospire were obtained during the last sampling period. Isolates were whole
33 genome sequenced and LipL32 expression was assessed *in vitro*.

34 We detected pathogenic *Leptospira* DNA at 15/22 sites; both soil and water were positive at
35 5/15 sites. We recovered *lipL32* sequences from 83/86 positive samples (15/15 positive sites)
36 and *secY* sequences from 32/86 (10/15 sites); multiple genotypes were identified at 12 sites.
37 These sequences revealed significant diversity across samples, including four novel *lipL32*
38 phylogenetic clades. Most samples from the serially sampled site were *lipL32* positive at each
39 time point. We sequenced the genomes of six saprophytic and two pathogenic *Leptospira*
40 isolates; the latter represent a novel pathogenic *Leptospira* species likely belonging to a new
41 serogroup.

42 **Conclusions/Significance:** Diverse and novel pathogenic *Leptospira* are widespread in the
43 environment in Puerto Rico. The disease potential of the novel lineages is unknown but several
44 persisted for >1 year in soil, which could contaminate water. This work increases understanding
45 of environmental *Leptospira* and should improve leptospirosis surveillance and diagnostics.

46 **Author Summary**

47 Leptospirosis is a common zoonotic disease worldwide, but more prevalent in the tropics. Cases
48 are more common following severe weather events, possibly due to flooding, which may more
49 readily distribute soil and/or water contaminated with *Leptospira* spp., the disease agents.
50 Human cases increased following the 2017 hurricanes that ravaged Puerto Rico (Maria and
51 Irma), prompting environmental sampling of soil and water to assess the presence, abundance,
52 and persistence of pathogenic leptospires in these environments. The goal was to better
53 understand these potential reservoirs of human and animal disease. Divergent and novel
54 groups of pathogenic *Leptospira* were abundant and widespread in soil and water in Puerto
55 Rico and sometimes persisted in these environments for >1 year. However, most groups we
56 identified have not previously been described from humans and/or other animals, so the
57 disease potential of these novel organisms is unknown. The results of this study reveal a
58 tremendous amount of previously uncharacterized *Leptospira* diversity in soil and water in
59 Puerto Rico, which could contribute to cryptic disease. The description and characterization of
60 these novel types improves our understanding of the genus *Leptospira*, and will aid in the
61 development of improved diagnostics and preventative tools to advance public health outcomes.

62 Introduction

63 Leptospirosis is a one-health zoonotic disease of worldwide importance (1) and the most
64 widespread zoonosis globally, affecting a wide variety of mammalian species (2, 3). It is caused
65 by pathogenic bacteria of the genus *Leptospira*. There are three broad groups of leptospires
66 within the genus that are categorized based upon 1) phylogenetic similarities, 2) *in vitro*
67 phenotypes, and 3) actual or presumed pathogenicity (*i.e.*, the presence of the *lipL32*
68 pathogenicity gene) (1, 4). To date, across four groups, there are 17 described species in the
69 pathogenic group (group P1), 21 species in the intermediate group (P2), and 26 in the
70 saprophytic groups (S1 and S2) (1).

71 Leptospires are obligate aerobic spirochetes that are highly motile and slender (0.15 ×
72 10–20µm) (5). Their genomes are comprised of two circular chromosomes and are larger than
73 other spirochetes and more diverse than those of many other bacterial genera (6) due to
74 horizontal gene transfer and gene duplication (7). This potentially facilitates an increased ability
75 to survive in a variety of hosts (humans, domestic and wild animals), environmental conditions
76 (soil and water), and climates (tropical, temperate, etc.) (6, 8).

77 Leptospirosis is maintained in animal reservoir hosts, such as rodents, cattle, and
78 domestic dogs (9, 10), with humans serving as incidental hosts. Leptospires colonize the
79 proximal renal tubules of infected animals and these reservoir hosts then excrete leptospires in
80 their urine, thereby contaminating the environment wherein leptospires can survive under
81 moist conditions (11) and infect other susceptible hosts (1, 12). Transmission of leptospirosis
82 typically occurs via direct contact with infected urine or indirectly through exposure to
83 contaminated soil and water (12). Leptospirosis infections in animals may become chronic and

84 can last for months or years (13). During chronic infection the shedding of leptospire in urine is
85 intermittent (14), which convolutes the identification of infected animals and disease mitigation
86 efforts.

87 Leptospirosis is a major global public health concern that disproportionately affects
88 resource-poor populations (15), especially developing countries with poor sanitation and dense
89 urban centers where rats, a common reservoir host, are abundant (12). Thus, people living in
90 urban centers and those directly exposed to infected livestock animals seem to be at highest
91 risk for leptospirosis, probably due to increased exposure rates (16). Recent estimates of global
92 human disease cases are 1.03 million annually, including 58,900 deaths (15); 73% of cases occur
93 in the tropics (17).

94 In the United States (US), leptospirosis is considered endemic with most human cases
95 arising in Puerto Rico and Hawaii (17). Leptospirosis was first reported in Puerto Rico in 1942
96 (18) and it is common in other parts of the Caribbean: 12,475 human cases were reported by
97 nine Caribbean countries between 1980 and 2005 (19). Despite this, it was removed from the
98 US reportable disease list in 1995 but later reinstated in 2014. During this time (1995-2014) and
99 despite the lack of requirement to report this disease, 759 cases of leptospirosis were reported
100 in Puerto Rico; 570 of these were confirmed, including 92 deaths. A retrospective analysis of
101 these cases by Santiago-Ramos *et al.* (18) identified a trend between annual case rates and
102 annual precipitation, although case rates were homogenous across Puerto Rico (*i.e.*,
103 geographical hot spots were not identified). Most cases occurred in males and in densely
104 populated urban areas (18).

105 Heavy rains and floods have been identified as major risk factors for human
106 leptospirosis, particularly in tropical regions and on islands (20). Hurricanes are associated with
107 leptospirosis outbreaks (17, 21-23), and the severity of these major storms can be amplified
108 due to conditions associated with climate change (*e.g.*, increased ambient temperatures and
109 humidity). Hurricane Hortense in 1996 (24) and the 2017 hurricanes Irma and Maria caused
110 major damage to infrastructure across Puerto Rico, which limited access to potable water, as
111 well as electricity, communications, and transportation (25). In just the three months following
112 hurricanes Maria and Irma (26), the US Centers for Disease Control and Prevention (CDC)
113 identified 85 leptospirosis cases, including 14 deaths, which was greater than the reported
114 annual cases for any of the three previous years (2014-2016; 45-73 annual cases) (CDC personal
115 communication).

116 Although some current thinking suggest pathogenic leptospires are unable to multiply in
117 the environment (27, 28), leptospirosis has nonetheless been considered an environmentally-
118 borne infection (29) because indirect contamination via the environment is the most frequent
119 source of human infection (28). However, in recent years the use of the term “environmental
120 reservoirs” for leptospirosis has increased, particularly when referring to moist soil in endemic
121 regions (28), and the discovery of novel, pathogenic leptospires persisting in these
122 environments (this study and (1)) supports the adoption of this term. Although moist soil
123 conditions seem to be associated with environmental persistence of leptospires (30-32) little is
124 known about the duration of survival in the environment (28, 33). Environmental surveys are
125 thus of critical importance toward our understanding of these environmental reservoirs and
126 their role in the persistence and proliferation of pathogenic leptospires, as well as their

127 potential role in disease transmission. To this end, we conducted environmental surveys in
128 Puerto Rico after the two 2017 hurricanes to: 1) characterize the prevalence and geographic
129 distribution of pathogenic *Leptospira* spp. in soil and water, 2) understand the diversity of
130 pathogenic strains that occupy these environments, 3) investigate persistence of pathogenic
131 *Leptospira* spp. in moist soil, and 4) isolate and characterize pathogenic *Leptospira* spp.
132 persisting in soil in Puerto Rico.

133 **Methods**

134 **Sample collection**

135 The focus of our environmental sampling efforts, conducted over three time periods
136 (Dec 3-14, 2018; Feb 9-23, 2019; and Feb 13-14, 2020), were 12 municipalities in Puerto Rico
137 (**Figure 1**) in which there is a potentially higher leptospirosis risk to humans based on previous
138 human and canine leptospirosis cases and environmental factors, such as areas prone to
139 flooding and poor sanitation (18). We collected 440 samples (10 water and 10 soil per site) from
140 22 field sites within these municipalities (**Table S1**); we targeted urban rivers and canals as well
141 as runoff sites and/or flood zones near livestock farms where urine contamination from
142 livestock was likely. At each site we established five short (<3m) linear transects and collected
143 two soil samples (~50g each) and two water samples (>150mL each) per transect (sampling
144 strategy illustrated in **Figure S1**). For each transect, the first soil sample was collected from the
145 edge of the water and the second soil sample was 1m back and directly perpendicular to the
146 edge of the water. Likewise, the first water sample was collected at the edge of the water and
147 the second water sample was collected 1m beyond and perpendicular to the edge of the water

148 **(Figure S1)**. Soil samples were collected ≤ 10 cm below the surface and placed into 50mL
149 capacity conical tubes and, when possible, water samples were collected from stagnate or less
150 turbulent areas of the subsurface water (10-30cm deep) into a Whirl-Pak. For site 22, collection
151 in transects was not possible due to access constraints at this urban canal. At that site we
152 collected 10 water and 10 soil samples within ≤ 1 m of the edge of the water. For soil samples at
153 all sites we recorded approximate elevation of the samples relative to the water, in meters. Soil
154 elevation was categorized into four groupings: 0-0.33m, >0.33-0.66m, >0.66-1m, and >1m
155 **(Table S2)**. Finally, we revisited one pathogenic *Leptospira* spp. positive site in San Juan (site 16)
156 that was identified during collection period 1 (Dec 3-14) twice more, once in February 2019 and
157 again in February 2020 **(Figure 1)**. In February 2019 we collected 10 additional soil samples at
158 this site in a linear transect along the edge of the water in the same location where the
159 previous positive soil samples were collected. In February 2020, for the purpose of culturing
160 attempts to obtain live leptospire, we collected 20 soil samples at this site clustered at the
161 edge of the water (two clusters of 10 samples) where we detected a high abundance of
162 pathogenic *Leptospira* DNA from the previous two collection periods.

163 **Figure 1: Map of Puerto Rico indicating sites where environmental samples were collected for**
164 **the detection of pathogenic *Leptospira* spp. DNA from 2018-2020.** Black dots represent paired
165 water and soil sites (n=22); 20 samples were collected from each site (10 soil and 10 water).
166 Yellow triangles indicate sites where pathogenic *Leptospira* DNA was detected (15/22 sites).
167 One site in San Juan (site 16, red circle) was serially sampled three times over a 14-month
168 period to assess persistence of pathogenic *Leptospira* in soil.

169 **Soil and water processing**

170 Water samples (150mL) were filtered through a 250mL capacity 0.22µm nitrocellulose
171 filter as previously describe (34); afterward the filter was cut in half using sterile scissors and
172 half of the filter was placed in a 5mL DNeasy® PowerWater DNA bead tube (Qiagen, Valencia,
173 CA, USA) and refrigerated at 4°C until shipment to the laboratory. Prior to filtering, 1mL of each
174 homogenous water sample was placed in a 2mL capacity screw-cap tube for downstream pH
175 testing. We collected pH values directly from the water sample using a 6mm glass pH probe
176 (Cole-Parmer, Vernon Hills, IL, USA, Part # EW-55510) and from soil using methods outlined by
177 Hall *et al.*, 2019 (34) (**Table S2**).

178 **DNA extractions**

179 DNA extractions were conducted in a class II A2 biosafety cabinet (BSC) as required by
180 our USDA-APHIS-PPQ soil transport permit. DNA was isolated from water filters using DNeasy®
181 PowerWater DNA extraction kits (Qiagen, Valencia, CA, USA) following the recommendations of
182 the manufacturer, and from ~0.5g of soil using Qiagen PowerSoil DNA extraction kits (Qiagen,
183 Valencia, CA, USA) following previously published conditions (35) with one modification: after
184 the addition of solution C3, the sample was incubated at 4°C overnight. DNA extractions were
185 assessed for overall quality and bacterial abundance using a universal 16S real-time PCR assay
186 as previously described (35); this quality control step is necessary prior to conducting a
187 pathogen detection PCR assay because it provides confidence that a negative result truly
188 reflects the absence of pathogen DNA and not a technical failure associated with poor quality
189 or low yield DNA extracts.

190 ***Leptospira* qPCR detection and direct sequencing**

191 All DNA extracts were screened in triplicate 10 μ L PCRs to detect the presence of DNA
192 from pathogenic *Leptospira* spp. using the *lipL32* TaqMan[®] real-time PCR assay (4) containing
193 the following: 1x TaqMan[®] Environmental PCR master mix (Applied Biosystems, Foster City, CA,
194 USA), 0.9 μ M of each primer, 0.45 μ M of the MGB probe, and 1 μ L undiluted DNA template.
195 Quantitative PCRs were run interchangeably on Applied Biosystems QuantStudio[®] (QS) QS7 and
196 QS12 Real-Time PCR Systems with QS 12K Flex or QS Real-Time PCR software, as appropriate,
197 under the following conditions: 50°C for 2 minutes, 95°C for 10 minutes, and 45 cycles of 95°C
198 for 15 seconds and 58°C for 1 minute. Positive (*L. interrogans* strain Fiocruz L1-130) and non
199 template controls were included on all runs. *LipL32* amplicons from all positive samples were
200 subjected to direct Sanger sequencing to confirm that the amplified product was *Leptospira*,
201 using the same forward and reverse primers for the PCR (all replicates from a single sample
202 were sequenced independently). Treatment and sequencing conditions are described below.

203 ***secY* 203bp amplification**

204 All samples that yielded a positive PCR result with the *lipL32* assay (defined as
205 amplification of at least 1 of 3 replicates with a $C_t < 45$) were also subjected to PCR amplification
206 of the *secY* gene; we amplified a 203bp fragment using the “SecYIVF” and “SecYIVR” primers
207 designed to amplify pathogenic *Leptospira* spp. as described in Ahmed *et al.*, 2009 (36). PCRs
208 were carried out in 10 μ L volumes containing the following reagents (given in final
209 concentrations): 1 μ L of 1/25 diluted DNA template (none of the PowerWater[®] extractions were
210 diluted), 1x SYBR[®] Green Universal master mix (Applied Biosystems, Foster City, CA, USA), 1.2M

211 betaine, 0.2U Platinum® *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA) to improve
212 efficiency, and 0.4µM of each primer. The assay was run on an Applied Biosystems 7500 Fast
213 Real-Time PCR System with SDS 7500 software v2.0.6 under the following conditions: 50°C for 2
214 minutes, 95°C for 10 minutes, and 40 cycles of 95°C for 15 seconds and 57°C for 1 minute.
215 Positive and non template controls were included on all runs. Due to the low abundance of
216 *Leptospira* DNA in many of these samples, triplicate reactions were employed to increase the
217 number of *secY* sequences generated; frequently only a single reaction (out of three) would
218 yield an amplicon.

219 ***secY* 549bp amplification**

220 We also attempted to sequence a 549bp region of the *secY* gene for all *lipL32* positive
221 samples to gain more phylogenetic resolution for certain samples and increase the number of
222 *secY* sequences generated for this dataset. We used primer pair F-ATGCCGATCATTTTTGCTTC
223 and R-CCGTCCCTTAATTTTAGACTTCTTC, which were designed to amplify pathogenic *Leptospira*
224 spp. as previously described (37). Duplicate PCRs were carried out in 10µL volumes containing
225 the following reagents (given in final concentrations): 1µL of 1/25 diluted DNA template
226 (undiluted for all PowerWater extractions), 1x PCR buffer, 2.5mM MgCl₂, 0.2mM dNTPs, 0.8U
227 Platinum® *Taq* DNA polymerase, and 0.4µM of each primer. PCRs were thermocycled according
228 to following conditions: 95°C for 10 minutes to release the polymerase antibody, followed by 38
229 cycles of 94°C for 60 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, and a final
230 extension step of 72°C for 10 minutes to ensure completion of the fragments. Positive and non
231 template controls were included on all runs.

232 **Sanger sequencing**

233 For all three amplicons (*lipL32*-242bp, *secY*-203bp, and *secY*-549bp), PCR products were
234 visualized on a 2% agarose gel to ensure that the products were of the expected size and
235 treated with 1 μ L of ExoSAP-IT (Affymetrix, Santa Clara, CA, USA) added to 7 μ L of PCR product
236 under the following conditions: 37°C for 15 minutes followed by 80°C for 15 minutes. Treated
237 products were then diluted (based on amplicon intensity) and sequenced in both directions
238 using the same forward and reverse primers from the PCR in a BigDye[®] Terminator v3.1 Ready
239 Reaction Mix (Applied Biosystems, Foster City, CA, USA). We used 10 μ L volumes for sequencing
240 reactions containing the following reagents (given in final concentrations): 5x Sequencing
241 Buffer, 1 μ L BigDye[®] Terminator v3.1 Ready Reaction Mix, 1 μ M primer, and 5 μ L diluted PCR
242 product. The following thermocycling conditions were used: 96°C for 20 seconds, followed by
243 30 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Sanger sequences
244 were assembled and primers were manually removed using SeqMan Pro (DNASTAR Lasergene,
245 Madison, WI, USA), resulting in sequence lengths of 202bp (*lipL32*-242bp), 162bp (*secY*-203bp),
246 and 504bp (*secY*-549bp). FASTA database files for each gene (*lipL32* and *secY*) were generated
247 in BioEdit (38); sequences from both *secY* amplicons (162 and 504bp) were combined into a
248 single FASTA file.

249 ***lipL32* high fidelity amplicon sequencing**

250 Sanger sequences from a subset of samples displayed multiple genotypes from replicate
251 amplicons and/or heterogeneous nucleotide calls in a single sequence, suggesting multiple
252 *Leptospira* spp. genotypes (*i.e.*, mixtures). Also, within some *lipL32* positive samples the

253 concentration of *Leptospira* spp. DNA was too low to sequence using traditional Sanger
254 sequencing methods. Thus, to recover sequences from these low-level positives and identify
255 and characterize individual genotypes within these mixtures, we employed a high fidelity
256 amplicon sequencing approach (AmpSeq) that enables differentiation of alleles without the
257 need for traditional cloning and sequencing of the PCR amplicon while providing confidence in
258 the detection and sequencing accuracy of low level positives (39). We amplified the same
259 242bp fragment of the *lipL32* gene described above in triplicate reactions using modified
260 primers that incorporated universal tails (39). The PCR setup and conditions for this modified
261 *lipL32* AmpSeq assay were the same as described above. Replicates were pooled and sequence
262 libraries were prepared for all positive samples as previously described (39), except a 1x
263 Agencourt AMPure XP (Beckman Coulter, Indianapolis, IN, USA) bead cleanup was used.
264 Uniquely indexed sample libraries were pooled together in equimolar amounts and sequenced
265 on an Illumina MiSeq instrument using a 500 cycle (2 x 250) MiSeq Reagent Kit v2 Nano with
266 PhiX control (Illumina, San Diego, CA, USA, part# MS-103-1003). We sequenced all *lipL32*
267 positive soil and water samples (n=86) in parallel on a single run.

268 **Illumina sequence processing and *lipL32* phylogenetic analysis**

269 Paired-end Illumina reads were processed within QIIME 2 (v2019.7 and 2020.6) (40);
270 primers were trimmed with cutadapt (41). The DADA2 (42) QIIME 2 plugin was used for quality
271 trimming (--p-trunc-len-f 115 --p-trunc-len-r 110), denoising, dereplication of sequences, and
272 chimera removal. *LipL32* sequences were identified with BLASTN (43) searches of
273 representative sequences against the NCBI nr/nt database (NCBI Resource Coordinators 2016)
274 and BLAST databases consisting of *lipL32* Sanger sequences generated from *Leptospira* positive

275 samples as part of this study. *LipL32* Illumina and Sanger sequences were combined with 538
276 publically available *lipL32* sequences that were extracted from whole genomes (GenBank
277 accession numbers in **Table S3**) using BLASTN v2.9.0+ (44) (*lipL32* sequence from *L. interrogans*
278 strain Fiocruz L1-130 was used as a reference) and aligned with MUSCLE (v3.8.31) (45).
279 Phylogenetic analysis was conducted using MEGA version 7 (46) and a maximum likelihood
280 phylogeny was inferred using the Tamura-Nei model; bootstrap values were calculated using
281 1,000 replicates.

282 ***secY* phylogenetic analysis**

283 Near full-length *secY* gene sequences (~1,300bp) were extracted from publicly available
284 genome assemblies (**Table S3**) with BLASTN (v2.2.29) (47). These nucleotide sequences were
285 combined with Sanger sequences generated from environmental samples in this study (162bp
286 and 504bp amplicons) and aligned with MUSCLE (v3.8.31) (45). A maximum likelihood
287 phylogeny was inferred with IQ-TREE (v1.6.12) (48) using the best-fit model (TIM+F+I+G4)
288 identified by ModelFinder (49) and the UFBoot2 ultrafast bootstrapping (50) and SH-aLRT (51)
289 options.

290 **Statistical analyses**

291 We tested for associations between environmental and sampling variables and the
292 detection of pathogenic *Leptospira spp.* in soil and water (**Table S2**). Sampling variables were
293 not available for 13 soil samples and site 22 was removed entirely from this analysis because
294 the sampling strategy for this site was different from the other sites (see above). Sample sizes
295 used for each test are listed in **Table S4**. For categorical variables (sample type, location, and

296 elevation), we used chi-square tests of independence, whereas for pH (a continuous variable)
297 we used Wilcoxon Rank Sum (W) test because these data were not normally distributed.
298 Categories for each environmental and sampling variable are described above under “Sample
299 collection” and “Soil and water processing” and listed in **Table S2**; *p*-values of <0.05 were
300 considered significant. All analyses were conducted in R studio v3.6.1 with tidyverse and
301 ggplot2 packages (52).

302 **Soil culturing**

303 Ten grams of soil samples were resuspended in 20mL of sterile water in 50mL conical
304 tubes, vigorously vortexed, and allowed to settle for 15 minutes for sedimentation of solid
305 matter. Supernatants were filtered through a 0.45µm filter as a selection step to remove larger
306 bacteria, and 2mL of filtrate was used to inoculate either a 2x HAN (53) or 2x EMJH media
307 (Difco, BD, Franklin Lakes, NJ, USA), both containing STAFF (54). EMJH cultures were grown at
308 29°C and HAN cultures were prepared in duplicate for growth at both 29 and 37°C. Each culture
309 was assessed daily by dark-field microscopy and, upon reaching high densities of growth (>10⁸
310 spirochetes/mL), an aliquot was removed for fluorescent antibody testing prior to storage in
311 liquid nitrogen for downstream analysis.

312 **Fluorescent Antibody Test (FAT)**

313 Ten microliters of cultured samples were placed on a glass slide with a 7mm well as
314 previously described (55). Briefly, slides were air dried overnight and fixed in acetone for
315 15 minutes and then placed in a humid chamber; 50µL of rabbit anti-LipL32 sera (1:250) was
316 added to each spot and then incubated for 1 hour at 37°C. Next, slides were washed for

317 10 minutes in PBS with gentle rocking and incubated with secondary Alexa Fluor 488 F(ab')₂
318 goat anti-rabbit IgG (Invitrogen, Waltham, MA, USA) for 1 hour at 37°C in the dark. After
319 extensive washing, slides were dried and counterstained for 15 seconds with Flazo Orange
320 (1:50, National Veterinary Services Laboratory). Slides were then rinsed with PBS and mounted
321 using ProLong™ Diamond Antifade Mountant with DAPI (Thermo Fisher, Waltham, MA, USA).
322 Microscopic examination was done using a Nikon Eclipse E800 microscope and B2-A filter
323 (excitation, 450–490 nm; emission, 520 nm) at 400× magnification (56).

324 ELISA

325 Cultured spirochetes from soil were recovered from liquid media (4,000 x g for 15
326 minutes), washed twice with PBS, and resuspended in 500 μL of PBS. Cells were counted by dark
327 field microscopy, suspensions were brought to a concentration of 5x10⁷ spirochetes/mL, and
328 100 μL (~5x10⁶) was used to coat individual wells on ELISA plates in triplicate (Thermo Fisher,
329 Waltham, MA, USA) for 16 hours at room temperature. Plates were then washed three times
330 with PBS and blocked with PBS-containing 1% bovine serum albumin (BSA) for 1 hour at 29°C.
331 For detection of pathogenic leptospire within the sample population, wells were incubated for
332 1 hour with polyclonal rabbit anti-LipL32 (1:1,000 in blocking buffer) followed by secondary
333 horseradish peroxidase-conjugated goat anti-rabbit IgG (1:3,000). Wells were washed six times
334 and reactivity was revealed by the addition of 100 μL of TMB substrate (SeraCare, Milford, MA,
335 USA) for 15 minutes. Reactions were stopped with 50 μL of TMB STOP™ Solution and optical
336 densities were taken at 620 nm in a 96-well plate reader. Sample analysis was performed in
337 triplicate. Positive and negative controls were comprised of wells coated with 100 μL (containing

338 5×10^6 , 2.5×10^6 , 1.25×10^6 , or 6.25×10^5) of *L. interrogans* serovar Copenhageni (LipL32 positive)
339 and $100 \mu\text{L}$ (5×10^6) *L. biflexa* serovar Patoc (LipL32 negative), respectively.

340 **Selection and culture of pathogenic leptospire**

341 ELISA and FAT-positive mixed cultures were enumerated by dark field microscopy and
342 leptospire were diluted in HAN media to 10^4 cells/mL; $100 \mu\text{L}$ was spread onto HAN agar plates
343 supplemented with 0.4% rabbit serum (56) and incubated at 37°C until colonies were visible.
344 Individual colonies were then harvested from the plates, vigorously homogenized in $100 \mu\text{L}$ of
345 HAN for microscopic visualization and, after confirmation of motile leptospire, $10 \mu\text{L}$ of each
346 suspension was used for coating FAT slides for LipL32 detection, as described above. The
347 remaining sample was maintained at 29°C and, in the event of a positive FAT with anti-LipL32,
348 used to inoculate fresh HAN media. Cultures were grown at 37°C , harvested at mid-log phase,
349 and were further confirmed as pathogens by PCR with the *lipL32* and *secY*-549bp primer sets
350 described above. *SecY* amplicons were Sanger sequenced (as above) to identify and
351 differentiate individual colonies.

352 **WGS sequencing of isolated leptospire**

353 Genomic DNA was extracted from isolated colonies using DNeasy kits (Qiagen, Valencia,
354 CA, USA) according to the recommendations of the manufacturer, except the buffer AL
355 incubation step occurred at 80°C for 1 hour. The gDNA was assessed for quality and quantity on
356 a 0.7% agarose gel using λ DNA-HindIII Digest (New England Biolabs, Ipswich, MA, USA). WGS
357 library construction was performed using the KAPA Hyper Prep Kits for Illumina NGS platforms
358 per the manufacturer's protocol with double-sided size-selection performed after sonication

359 (KAPA Biosystems, Woburn, MA, USA, part# KK8504). The adapters and 8bp index oligos
360 purchased from IDT® (Integrated DNA Technologies, San Diego, CA, USA), based on Kozarewa
361 and Turner, 2011 (57), were used in place of those supplied in the KAPA preparation kit. The
362 final libraries were quantified on an Applied Biosystems™ QuantStudio™ 7 Flex Real-Time PCR
363 System using the KAPA SYBR® FAST ROX Low qPCR Master Mix for Illumina platforms (part#
364 KK4873). The libraries were then pooled together at equimolar concentrations and quality was
365 assessed with a Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA,
366 part# 5067-1504). Final quantification by qPCR preceded sequencing of the final library. The
367 samples were sequenced on an Illumina MiSeq using the 600-cycle v3 kit (part# MS-102-3003)
368 with the standard Illumina procedure. The appropriate sequencing primers were added to the
369 MiSeq kit as previously described (57).

370 **WGS phylogenetic analysis**

371 **External genomes**

372 Publically available *Leptospira* genomes were downloaded from the GenBank assembly
373 database (58) with the ncbi-genome-download tool ([https://github.com/kbclin/ncbi-genome-](https://github.com/kbclin/ncbi-genome-download)
374 [download](https://github.com/kbclin/ncbi-genome-download)) on September 9th, 2021. Genomes were renamed to include the assembly accession
375 and taxonomic information. The dataset resulted in a set of 802 reference genomes that
376 included pathogenic, intermediate, and saprophytic leptospires (**Table S3**).

377 **Genome assembly**

378 Raw reads were trimmed with `bbduk.sh` (v38.92)
379 (<https://sourceforge.net/projects/bbmap/>) and assembled with SPAdes (v3.13.0) (59). Reads

380 were mapped back against contigs with minimap2 (v2.22) (60) and the depth of coverage was
381 calculated with Samtools (v1.13) (61). Two hundred nucleotides from each contig were aligned
382 against the GenBank nt database with BLASTN (v.2.11.0) (44) and the taxonomy of the top hit
383 was recorded. Contigs with an anomalously low depth of coverage or aligned against
384 contaminants were manually removed.

385 ***Leptospira dendrogram***

386 Pairwise MASH distances (62) were calculated on all *Leptospira* genomes and a
387 dendrogram was generated with mashPy
388 (<https://gist.github.com/jasonsahl/24c7cb0fb78b4769521752193a43b219>), a tool that
389 incorporates SciPy (63) and Skbio (<http://scikit-bio.org>).

390 **Average nucleotide identity**

391 Average nucleotide identity (ANI) between the novel *Leptospira* spp. described herein
392 and *L. yasudae* strain 201601115 was calculated using the ANIb stat in PYANI (v0.2.11) as
393 described previously (64)

394 **SNP and indel discovery**

395 Genome assemblies were annotated with Prokka (v1.14.6) (65). Reads were aligned
396 against each reference and SNPs and indels were called with Snippy (v4.6.0)
397 (<https://github.com/tseemann/snippy>) using default parameters.

398 **Core genome SNP phylogenies**

399 For pathogenic genomes, assemblies were aligned against *Leptospira kmetyi* strain LS-
400 001/16 (GCA_003722295.1) with NUCmer (v3.1) (66) and SNPs were called with NASP (v1.2.0)

401 (67); SNPs that fell within duplicated regions, based on a reference self-alignment with
402 NUCmer, were filtered from downstream analyses. A maximum likelihood phylogeny was
403 inferred on the concatenated SNP alignment with IQ-TREE (v2.0.3) (68) in conjunction with
404 ModelFinder (49) and rooted with *Leptospira interrogans* FDAARGOS 203 (GCA_002073495.2).
405 The reference genome, *Leptospira montravelensis* strain 201800278 (GCA_004770045.1), was
406 used for the phylogeny including saprophytic genomes, which was rooted by *Leptospira*
407 *noumeaensis* strain 201800287 (GCA_004770765.1).

408 **Hamster infection with isolated colonies**

409 To assess virulence of the newly identified pathogenic isolate described herein, Golden
410 Syrian hamsters were inoculated by intraperitoneal (IP) injection with 10^8 leptospires in 1mL as
411 previously described (69). Negative control hamsters received 1mL of media alone. Animals
412 were assessed for acute disease through day 11 post infection and euthanized on day 21 to
413 provide ample time for renal carriage to develop. All animal experimental procedures were
414 performed in accordance with relevant guidelines and regulations, and as approved by USDA
415 institutional guidelines.

416 **Microscopic agglutination test (MAT)**

417 The MAT was performed according to OIE guidelines (70) at two-fold dilutions from an
418 initial dilution of 1:50 to 1:6400 using 18 serovars of *Leptospira* spp. listed in **Table S5** and the
419 newly identified pathogenic isolates described herein.

420 **Culture**

421 At three weeks post-infection, kidneys were harvested using aseptic techniques for
422 culture of leptospires in T80/40/LH semi-solid media incubated at 29°C (71) and HAN semi-solid
423 media incubated at both 29 and 37°C (53).

424 **Data sharing**

425 All WGS reads and assemblies generated during this study have been deposited in NCBI
426 BioProject database under accession # PRJNA766613. The associated BioSample accession
427 numbers for these isolates are sequentially assigned beginning with # SAMN21833208 and
428 ending with # SAMN21833215. Likewise, SRA accession numbers for these isolates are
429 sequentially assigned beginning with # SRR16134453 and ending with # SRR16134460. Partial
430 gene sequences for *lipL32* and *secY* have been deposited in NCBI (accession numbers
431 sequentially assigned beginning with # OK345072 and ending with # OK345268) and are also
432 provided in Supplemental File 1 (**Table S6 and Table S7**).

433 **Results**

434 **Pathogenic *Leptospira* detection**

435 Pathogenic *Leptospira* spp. were common and distributed widely throughout Puerto
436 Rico. Using *lipL32* PCR, we detected pathogenic *Leptospira* spp. DNA at 15 of 22 paired soil and
437 water exploratory sites (68.2%) (**Figure 1**); 8/15 sites yielded positive soil samples only, 2/15
438 positive water samples only, and 5/15 both positive soil and water samples (**Figure S2**). From
439 these exploratory sites, 63 samples (out of 440) were positive, including 15 water and 48 soil
440 samples. At the serially sampled soil site (site 16), an additional 23 (out of 30 total) soil samples

441 collected during the second and third visits to this site also were positive. In total, 86 positive
 442 samples from Puerto Rico were detected. At five paired soil and water sites (sites 7, 12, 17, 21,
 443 and 22) we detected pathogenic *Leptospira* spp. in both sample types (**Table 1** and **Figure S2**).
 444 At four of these five sites (12, 17, 21, and 22) a higher proportion of soil samples were positive
 445 than water samples and the phylogenetic clades represented in the water samples were a
 446 subset of those identified in the corresponding soil samples (**Table 1**). This observation was not
 447 possible at site 7 because only a single soil and a single water sample were *lipL32* positive and
 448 *lipL32* sequence was not recovered from the positive soil sample (**Table S1**).

449 **Table 1: Table of all positive sites in Puerto Rico illustrating the relationship between *lipL32***
 450 **genotypes found in soil and water.** In total, 15 sites were positive for pathogenic *Leptospira*
 451 DNA; eight sites wherein only soil was positive, two sites wherein only water was positive, and
 452 five sites wherein both soil and water were positive (highlighted in gray). At four of the latter
 453 sites, the clades present in water were a subset of the clades identified in soil (bolded text),
 454 suggesting that soil reservoirs may lead to the contamination of water.

Site ID	Site Type	Proportion Positive	Clades
1	Water	0.1	4b
2	Soil	0.2	4a
5	Water	0.7	1, 6
7	Water	0.1	6
7	Soil	0.1	unknown
10	Soil	0.1	4a
11	Soil	0.2	4a
12	Water	0.1	4b
12	Soil	0.6	3, 4a, 4b , 6, 9
13	Soil	0.1	6
14	Soil	0.4	3, 4a, 4b, 9
15	Soil	0.3	4b
16	Soil	0.7	4b, 6
17	Water	0.1	6
17	Soil	0.4	3, 4b, 6
19	Soil	0.3	4b, 6

21	Water	0.1	6
21	Soil	0.9	4a, 4b, 6 , 9
22	Water	0.3	4b , 6
22	Soil	0.5	4b , 6 , 9

455

456 **Gene sequencing *lipL32***

457 Sequencing of the *lipL32* amplicons revealed extensive and previously undescribed
458 diversity in pathogenic *Leptospira* spp. at this locus. Using AmpSeq and Sanger sequencing
459 methods, we generated 147 *lipL32* sequences from 83 of the 86 positive samples representing
460 all 15 positive sites (**Table S1**) to: 1) confirm that the generated amplicons were from *Leptospira*
461 DNA, and 2) assess genetic diversity of pathogenic *Leptospira* spp. within and across sites and
462 samples; information on generated nucleotide sequences is provided in **Table S6**. Thirty-four
463 alleles were identified in this dataset from Puerto Rico of which 32 were novel (*i.e.*, were not a
464 100% nucleotide match to any available sequences in GenBank). We assigned either a high or
465 moderate level of confidence to each newly identified allele. If a sequence was observed in
466 more than one sample or was generated via Sanger sequencing methods, this allele was
467 assigned a high level of confidence; 20 of 34 alleles met this criterion. The remaining 14 alleles
468 were only observed in a single sample, only in the Illumina dataset, and/or with a low depth of
469 coverage (3-42 paired end reads) and thus were assigned a moderate level of confidence. As a
470 quality control step, we also translated all the *lipL32* sequences in frame to confirm that they
471 resulted in functional amino acid sequences. Although the dual indexing and paired end
472 analysis approach described above provides additional confidence in the accuracy of these
473 sequences, we acknowledge that these putative alleles could benefit from further validation. As
474 such, no major phylogenetic conclusions are inferred from these 14 lower confidence alleles

475 and none of these alleles define the novel *lipL32* clades described herein. However, all 34
476 alleles were used to illustrate the diversity of pathogenic *Leptospira* spp. genotypes present in
477 Puerto Rico soil and water.

478 **Gene sequencing *secY***

479 Unlike *lipL32*, which has only been documented in pathogenic leptospires, the *secY* gene
480 is present in both pathogenic (P1 and P2) and saprophytic (S1 and S2) *Leptospira* groups (1, 72).
481 It is also commonly used to obtain *Leptospira* spp. identification from clinical samples. Thus, to
482 compliment the *lipL32* sequence data and provide support for phylogenetic membership to the
483 pathogenic clade and the relationship to other pathogenic leptospires, we generated *secY*
484 sequences for these environmental samples. This was an important quality control step
485 because divergent alleles that have not previously been described from the *lipL32* gene were
486 observed in these samples. Twenty-one *secY*-203bp sequences from 16 samples (3 water and
487 13 soil) and 29 *secY*-549bp sequences from 23 samples (2 water and 21 soils) were generated
488 (**Table S1**). In combination, *secY* sequences were generated from 32/86 positive samples
489 representing 10/15 sites. All sequences fell within the pathogenic group of the *secY* phylogeny
490 (**Figure S3**), thereby confirming the *lipL32* results; nucleotide sequences are provided in **Table**
491 **S7**.

492 **Mixtures**

493 Sequences generated for the *lipL32* and *secY* genes revealed mixtures (multiple
494 pathogenic genotypes) in 39 soil and two water samples from 12 sites; up to five unique *lipL32*
495 genotypes in a single sample were identified (**Table S1 and Table S6**).

496 Phylogenetics

497 Phylogenetic analyses of *lipL32* sequences revealed a high level of genetic diversity
498 among these samples, including four previously undescribed pathogenic *Leptospira*
499 phylogenetic clades (3, 5, 8, and 9; **Figure 2 and Figure S4**); 18 sequences obtained from the
500 environment in Puerto Rico assigned to these four clades. Clades 4 and 6 contain 128 *lipL32*
501 sequences from Puerto Rico as well as environmentally acquired isolates from other locations
502 (1). The two most common *lipL32* genotypes from soil and water in Puerto Rico were a perfect
503 match to *L. gomenensis* in clade 6 (n=56) or a close match to *L. yasudae* (previously *L.*
504 *dzianensis*) in clade 4b (n=28) (**Figure 2 and Figure S4**). Interestingly the *L. gomenensis*
505 genotype was recently described in New Caledonia and *L. yasudae* (*L. dzianensis*) in Mayotte
506 (1), suggesting that these environmental *Leptospira* spp. are distributed globally. The *secY*
507 phylogeny provided additional confidence in the phylogenetic placement of these novel types
508 because those genotypes were also a close match to *L. yasudae* and *L. gomenensis* (**Figure S3**).
509 Interestingly, there also was sequence similarity to *L. santarosai* for some *secY* genotypes,
510 including those from the newly acquired LGVF01 and LGVF02 isolates that are described herein
511 (**Figure S3**). Clades 1, 2, and 7 of the *lipL32* phylogeny (**Figure 2 and Figure S4**) contain
512 described pathogenic leptospires that are commonly found in human and animal infections: *L.*
513 *interrogans*, *L. noguchii*, and *L. kirschneri* (clade 1); *L. borgpetersenii* and *L. santarosai* (clade 2);
514 and *L. alexanderi*, *L. weilii*, and *L. mayottensis* (clade 7). A single sequence from site 5 assigned
515 to clade 1 (**Figure 2 and Figure S4**) and no sequences from Puerto Rico assigned to clades 2 and
516 7.

517 **Figure 2: Genetically diverse and previously undescribed lineages of pathogenic *Leptospira***
518 **are widespread in Puerto Rico.** A) Maximum likelihood phylogeny constructed using a 202bp
519 segment of the *lipL32* gene, including sequences obtained from soil and water samples
520 collected in this study (blue text) and those available in public databases (black text). Nine
521 major genetic clades are represented, seven of which were present in Puerto Rico, including
522 four not previously described. Clades are labelled 1 through 9 on the tree and color coded. B)
523 Sites where each major clade was collected, with colors within the pie charts corresponding to
524 those in the phylogeny; superscript indicates if soil (S) and/or water (W) samples were positive
525 at each site. Fifteen positive sites are included on this map.

526 **Detection Statistics**

527 We observed a significant association between the distribution (median) of pH values
528 for samples collected from water that were positive (median pH=8.4) and negative (median
529 pH=8.2) for pathogenic *Leptospira* spp. (W=550, $p=0.0015$), but not from soil, where positive
530 and negative median pH values were ~ 7.5 regardless of pathogenic *Leptospira* status (W=3057,
531 $p=0.9399$). We also detected pathogenic *Leptospira* spp. in soil more commonly than in water:
532 48/210 (22.9%) of soil samples compared to 15/210 (7.1%) of water samples ($\chi^2=15.087$,
533 $p<0.001$). No other significant associations were observed among environmental/sampling
534 variables and the detection of pathogenic *Leptospira* spp. (**Table S4**).

535 **Persistence in soil**

536 During our first sampling period (December 2018) we identified one exploratory site in
537 an urban district of San Juan where seven soil samples (out of ten) tested positive for the

538 presence of pathogenic *Leptospira* spp. DNA (**Table S1**), the highest single site detection rate
539 from this first period. We revisited the site two months later (February 2019) and collected an
540 additional ten soil samples in a linear transect along the edge of the water to test the
541 hypothesis that pathogenic *Leptospira* spp. persists in water-soaked soil at this site; all ten of
542 the new soil samples tested positive for pathogenic *Leptospira* spp. DNA (**Table S1**). We
543 revisited this site once again 14 months after the initial sampling period (February 2020) and
544 collected 20 samples clustered along the edge of the water to use for live *Leptospira* spp.
545 recovery and isolation attempts, as well as persistence analysis; 13 of these 20 soil samples
546 tested positive for pathogenic *Leptospira* spp. DNA (**Table S1**). Thus, at all three time points the
547 majority of soil samples collected at this site were positive for pathogenic *Leptospira* spp. DNA:
548 7/10, 10/10, and 13/20 of samples, respectively (**Figure 3**). Among these positive samples,
549 identical *LipL32* sequences from clades 4b and 6 were identified across all three time points
550 (**Figure 3A**), suggesting these genotypes persisted in soil at this site.

551 **Figure 3: Site 16 in Caño Martín Peña, San Juan, Puerto Rico where sampling was conducted**
552 **at three separate timepoints to assess environmental persistence of pathogenic *Leptospira* in**
553 **soil near the edge of the water.** A) Maximum likelihood phylogeny of all *lipL32* sequences
554 obtained from this soil site over the course of three sampling periods (color coded by date of
555 sampling period); two of five clades detected at this site were identified at all three timepoints,
556 suggesting environmental persistence. Clade colors correspond to Figure 2. Two isolates
557 representing a novel pathogenic *Leptospira* species were obtained from soil sample 27 (red
558 asterisk). B) Photo of exact location where samples 21-30 were collected and the isolates were
559 obtained. C) Transect where 40 soil samples were collected over three sampling periods.

560 Balloons are color coded to indicate sampling period as in the phylogeny (panel A) and the red
561 asterisk denotes the location where the *Leptospira* isolates originated. D) Location of site 16 in
562 Caño Martín Peña.

563 **Quantification of pathogenic *Leptospira* spp. in soil cultures**

564 Single aliquots of culture from FAT-positive soil sample #27 and FAT-negative soil
565 sample #40 (both from site 16) were defrosted in a water bath at 37°C and used for a
566 confirmative FAT and for coating onto ELISA plates. As expected, sample #27 presented with a
567 high proportion of LipL32-reactive leptospire (Figure S5A), which was quantified by ELISA
568 (Figure S5B); approximately 50% of the mixed population was pathogens. In contrast, only basal
569 reactivity was obtained when wells were coated with cells from sample #40, consistent with a
570 small number, or absence of, pathogenic leptospire. Cultures were seeded onto HAN plates
571 and incubated at 37°C in a 3% CO₂ atmosphere; colonies were typically observed in five days.
572 Colonies were picked from both plates (27 colonies from sample #27 and 54 from sample #40),
573 vigorously homogenized in liquid HAN, and the suspensions used to coat FAT slides. In strong
574 agreement with ELISA results, 14 out of 27 colonies from sample #27 were FAT positive (51.9%),
575 whereas none of the 54 colonies from sample #40 showed significant LipL32 reactivity.
576 Representative FAT images of positive (colony 9 and 17) and negative (colony 8) clonal
577 suspensions from sample #27 are presented in Figure S5C. FAT-positive cells were inoculated
578 into liquid HAN and harvested at mid-log phase for PCR confirmation, wherein colony 8 was
579 also employed as a negative control. All 14 FAT-positive cultures were also PCR-positive for
580 both *lipL32* and *secY* and, as expected, no bands were observed when cells from colony 8 were
581 used as a template (Figure S5D).

582 **Sequence identification of pathogenic colonies**

583 All *secY* amplicons were sequenced and the assembled sequences (446bp in length after
584 trimming) were submitted to nucleotide alignment by BLASTN (43). The 14 clones presented an
585 average identity of 95% to *L. santarosai* serovar Princetown strain TRVL 112499. We identified
586 two clones (colony 9 and 18) that were slightly divergent from the other 12 and contained
587 several SNPs within the *secY* sequence (**Figure S3 and Table S7**). One representative isolate of
588 each genotype, designated as strain LGVF01 and strain LGVF02, respectively, were re-grown in
589 liquid HAN for DNA extraction and whole genome sequencing. As a quality control step we also
590 sequenced the *lipL32* gene from these isolates and confirmed that they matched the sequences
591 we obtained from the originating soil samples.

592 **Isolation of other leptospire**

593 A total of 40 saprophytic isolates also were obtained during the purification process for
594 pathogenic colonies. We sequenced near full length 16S rRNA fragments (1,330bp) for these
595 isolates using universal primers (73) and identified two genotypes among them that differed by
596 four SNPs. One genotype (16S-01) was a perfect match to *L. mtsangambouensis*, *L. jelokensis*, *L.*
597 *noumeaensis*, *L. congkakensis*, *L. levettii*, and *L. macculloughii* and the other (16S-02) was a
598 perfect match to *L. bandrabouensis*, *L. kemamanensis*, *L. bouyouniensis*, and *L. meyeri*; three
599 representatives of each genotype were whole genome sequenced (**Table S8**).

600 **WGS**

601 In total, we generated whole genome sequences for two pathogenic and six saprophytic
602 isolates obtained from a single soil sample (#27) at the serially sampled site 16 (**Figure 3**); the

603 former are two distinct strains of a novel pathogenic *Leptospira* species that differ from each
604 other by 20,980 pairwise SNPs (4,982 non-synonymous) (**Figure 4B**). This novel pathogen
605 species falls within clade 4b of the *lipL32* phylogeny (**Figure 2** and **Figure S4**) and is most closely
606 related to, but is clearly distinct from, *L. yasudae*; the ANI between these species is 92.8%.
607 Samples from five other positive soil sites and one water site from eastern Puerto Rico (sites 14,
608 15, 17, 19, 21, and 22) share the same *lipL32* genotype found in this novel pathogenic species,
609 suggesting it may be fairly common and widespread in Puerto Rico (**Figure 2**). The six
610 saprophytic isolates represent three previously known species of *Leptospira*: they were
611 genomically identified as *L. levettii* (n=1), *L. bandrabouensis* (n=4), and *L. mtsangambouensis*
612 (n=1) (**Figure 4**). The 16S-01 genotype was conserved among the *L. levettii*, *L.*
613 *mtsangambouensis*, and one of the *L. bandrabouensis* isolates, whereas the 16S-02 genotype
614 was conserved among the remaining three *L. bandrabouensis* isolates (**Table S8**). The *L.*
615 *bandrabouensis* isolate (Sapro03) that contained the 16S-01 genotype differed from the other
616 three *L. bandrabouensis* isolates by >47,000 SNPs (**Figure 4B**). The genome assembly details for
617 all eight isolates are provided in **Table S9**.

618 **Figure 4: Whole genome dendrogram of 802 known pathogenic, intermediate, and**
619 **saprophytic *Leptospira* spp. isolates, including six saprophytic and two pathogenic isolates**
620 **obtained in this study from a single soil sample from site 16.** A) Pairwise genomic distance
621 dendrogram contextualizing the relationship among all known saprophytic (S1 and S2),
622 intermediate (P2), and pathogenic (P1) *Leptospira* spp. reveals the pathogenic isolates LGVF01
623 and LGVF02 represent two genotypes of the same previously undescribed pathogenic species,
624 whereas the saprophytes belong to three known *Leptospira* spp. B) Detailed view of the S1

625 clade that contains four saprophytic isolates obtained during this study. This maximum
626 likelihood phylogeny reveals significant diversity among isolates that fall within the *L.*
627 *bandrabouensis* clade. C) Detailed view of the P1 clade that contains two pathogenic isolates
628 obtained during this study. This maximum likelihood phylogeny reveals two genotypes of a
629 novel pathogenic *Leptospira* spp. isolated from soil in Puerto Rico.

630 **Hamster infection and MAT response**

631 MAT reactivity was detected when sera from hamsters infected with the new
632 pathogenic isolate LGVF02 were tested against the challenge isolate (LGVF02) at a 1:1600
633 dilution. However, no experimentally infected hamsters showed clinical signs of infection or
634 weight loss after inoculation (**Table S10**) and at three weeks post-infection all hamster kidneys
635 were culture negative. In addition, at three weeks post-infection all hamsters were
636 seronegative for the 18 serovars, representative of 15 serogroups, listed in **Table S5**.

637 **Discussion**

638 This study provides new insights into the persistence of pathogenic leptospire in the
639 environment in Puerto Rico and raises important questions regarding host/environmental
640 adaptation, genomic exchange, the evolution or acquisition of pathogenicity in the genus
641 *Leptospira*, and the disease potential of environmental leptospire. Our results also
642 demonstrate the need for updated diagnostics to detect and identify these novel and divergent
643 types.

644 Our survey for pathogenic *Leptospira* spp. in the environment in Puerto Rico revealed
645 that diverse and novel pathogenic types are widespread in soil and water across the island, and

646 more common in soil than water. Thirty-two previously undescribed *lipL32* alleles were
647 identified, which suggests a vast amount of uncharacterized diversity within the pathogenic
648 group of this genus present in the environment in Puerto Rico. However, we only identified a
649 single sequence from one water sample that shared a *lipL32* genotype with *Leptospira* spp.
650 from clade 1; species from clade 1 are commonly found in human leptospirosis infections. This
651 might suggest that clade 1 types are not adapted for survival in soil and water. Indeed, host
652 adaptation among these commonly identified pathogenic types has been suggested (7), which
653 could explain the limited detection of these pathogenic types in soil and water in this study.
654 This is in line with the described transmission dynamics of human leptospirosis in which
655 infecting bacteria are maintained in the kidneys and renal tubules of animal reservoir hosts and
656 excreted through urine whereby transient contamination of soil and water can occur, leading to
657 human infections. In contrast, we identified 33 less common or novel *lipL32* alleles in these
658 environmental samples whose disease potential is unknown. Although the pathogenicity of the
659 novel types is yet to be determined, the presence of the *lipL32* gene suggests pathogenic
660 potential in humans and animals. Of course it is also possible that the opposite is true and that
661 the presence of the *lipL32* gene in these strains is not indicative of pathogenicity but perhaps an
662 artifact of genetic exchange and recombination and/or serves some other purpose in these
663 complex environments. Either way, the abundance and persistence of these types in soil and
664 water suggests adaptation for survival, and potentially proliferation, in these environments.

665 In support of this hypothesis, we found evidence for *lipL32* positive *Leptospira* spp.
666 persisting in water-soaked soil near the edge of the water for more than one year at an urban
667 site in San Juan, Puerto Rico. Two of the five pathogenic genetic clades found in soil at this site

668 were resampled again at two months and 14 months after the initial sampling date. Although
669 survival and proliferation in this environment seems the most likely explanation, possibly via
670 formation of biofilms (74, 75), it is also possible that this site is continually contaminated from
671 some other source. This same urban site was described in a 2015 study wherein the research
672 team evaluated *Leptospira* seroprevalence in humans and carriage in rodents (76). It was
673 determined in that study that *Leptospira* exposure was high in humans in this urban community
674 (seroprevalence of 27.2%) but the species/serotypes identified in humans and the species
675 identified and/or recovered from rodents in that study were not a match to those found in our
676 study but, rather, were the commonly identified human and animal pathogens *L. borgpetersenii*
677 and *L. interrogans*. However, it is important to point out that seroprevalence testing for this
678 other study employed a MAT panel containing common human and animal pathogenic types
679 only (27 serovars represented), so if exposure from these novel soil dwelling types had
680 occurred they may not have been detected with this method.

681 Our discovery of so many novel genotypes at this urban site (n=8) and in Puerto Rico as
682 a whole (n=32) highlights the need to address the limited scope of leptospirosis diagnostics
683 (e.g., a MAT panel that does not include local strains), as it is possible that these novel
684 pathogenic types also are associated with human and animal disease in Puerto Rico but remain
685 undetected in the clinic. Indeed, it is thought that many cases of leptospirosis go undetected or
686 are misdiagnosed as other febrile illnesses due to challenges with implementation and/or
687 limitations of the diagnostic tools themselves (77, 78). If it were shown that these types
688 contribute to human and animal disease and proliferation in the environment is experimentally
689 confirmed, it would contradict the current dogma that pathogenic leptospires require an animal

690 reservoir host for maintenance and proliferation and, thus, suggest a much more complicated
691 scenario of leptospirosis transmission that includes infection by soil dwelling leptospires.

692 From soil sampled at this urban site in San Juan we isolated a novel pathogenic
693 *Leptospira* spp. (represented by strain LGVF02), which was confirmed to express LipL32 *in vitro*.
694 Sera from infected Golden Syrian hamsters were positive by MAT when using the challenge
695 strain, although no signs of acute disease were observed. Interestingly, sera from challenged
696 hamsters were MAT negative when tested against a panel of 18 commonly tested strains
697 representing 15 serogroups; as such, we suspect that this isolate represents an entirely novel
698 serogroup and, thus, would not be detected using a MAT panel of common isolates.
699 Importantly, the absence of disease symptoms in challenged hamsters does not negate its
700 classification as a pathogen; other pathogenic *Leptospira* spp., including *L. borgpetersenii*, *L.*
701 *mayottensis*, and *L. tipperaryensis*, have also failed to cause acute disease in Golden Syrian
702 hamsters (79, 80). LGVF02 was also subjected to LipL32 gene silencing via CRISPRi as part of
703 another study (56) and the results of that study (specifically, the detection of the LipL32 and
704 LipL41 proteins) further support the pathogenic classification of this novel leptospire. The
705 species description of LGVF02 will be treated in a separate publication in which we will
706 investigate whether it represents an entirely new serogroup. Either way, the ability of strain
707 LGVF02 to evade detection with MAT based on common isolates highlights the need to include
708 regionally relevant isolates when conducting MAT in the clinic to ensure detection of these
709 lesser known types and better understand their potential role in human and animal disease.

710 Our attempts to sequence two separate *secY* amplicons (*secY*-203bp and *secY*-549bp)
711 from *lipL32* positive samples suggested that SNPs or other genetic mutations were present in

712 the PCR priming sites. We suspect this because *secY* PCR amplification was only achieved in 32
713 of 86 (37%) *lipL32* positive environmental samples from Puerto Rico. Further evidence of this
714 possibility is supported by other investigators; during the preliminary genome analysis of a
715 recently acquired *L. kirschneri* isolate it was discovered that the reverse priming site for the
716 *secY*-549bp PCR was missing entirely (Camila Hamond, NVSL, personal communication).
717 Accordingly, this *L. kirschneri* isolate and originating sample failed to amplify with *secY*-549bp
718 PCR. The poor performance of both *secY* PCRs with these environmental samples that displayed
719 diverse *lipL32* genotypes, paired with the MAT evasion of the LGVF02 isolate described herein,
720 again highlights the need for updated diagnostics that can account for this novel diversity.
721 However, this goal requires the acquisition of genomes to first characterize this diversity in
722 order that improved diagnostic targets can be developed.

723 We generated whole genome sequences for two pathogenic and six saprophytic
724 *Leptospira* isolates obtained from a single soil sample collected from serially sampled site 16.
725 The whole genome dendrogram contextualized the phylogenetic relationships of these isolates
726 to all other known *Leptospira* spp. and revealed that the two pathogenic isolates, LGVF01 and
727 LGVF02, were unique strains (separated by 20,980 SNPs) representing a single, novel pathogenic
728 species. We classified these isolates as pathogenic based upon the presence of the *lipL32* gene
729 in their genomes and the expression of the LipL32 protein, as well as their phylogenetic
730 relationships to other species. The saprophytic isolates represent three known species, of which
731 two (*L. bandrabouensis* and *L. mtsangambouensis*) have previously only been identified from
732 the island of Mayotte in the Indian ocean, and the third (*L. levettii*) has previously been
733 identified only in Mayotte, Malaysia, and New Caledonia. In other words, these three species

734 have not previously been identified in Puerto Rico, the Caribbean, or the Western Hemisphere.
735 In total, four *Leptospira* species were isolated from a single 10g soil sample from site 16 in
736 Puerto Rico, and the *lipL32* analysis from this same site identified four additional phylogenetic
737 clades that were not represented by the isolates obtained from this sample (**Figure 3A**),
738 suggesting the presence of even more pathogenic species in soil from this site. Finally, the SNP
739 diversity identified among isolates within the same species from a single soil sample is
740 considerable (20,980 between LGVF01 and LGVF02 and >47,000 SNPs differentiating two
741 subgroups of *L. bandrabouensis*). Together, these findings reveal a tremendous amount of
742 unrecognized *Leptospira* spp. diversity within the environment in Puerto Rico, and very likely
743 many other geographic locations, that has yet to be characterized.

744 Soil may serve as a long-term reservoir for contamination of waterways with pathogenic
745 *Leptospira* spp. in Puerto Rico. At four of the five sites where pathogenic leptospires were
746 detected in both soil and water, the *Leptospira* clade(s) identified in water were a subset of
747 those identified in soil and a higher proportion of soil samples were positive when compared to
748 water samples at these sites. These observations suggest soil may serve as a reservoir for
749 pathogenic *Leptospira* that leads to the contamination of waterways in Puerto Rico, which
750 could occur when the soil is disturbed by animal or human activity, or during weather events
751 such as heavy rains and floods. This is in line with the dynamics of *Leptospira* environmental
752 contamination reviewed elsewhere (28). However, our findings also provide support for the
753 hypothesis that some pathogenic *Leptospira* spp. are maintained in the soil long term, rather
754 than just transiently being present due to urine contamination from an infected reservoir host.
755 If this is the case, the potential for genetic exchange among pathogenic isolates that could lead

756 to the emergence of novel outbreak strains is something that needs to be explored further. A
757 possible example of this may be the documented outbreak between 1999-2003 in Thailand that
758 was associated with a novel type of *L. interrogans* that emerged, caused illness in humans for
759 several years, and then inexplicably seemed to fade out (81). It is possible this type of outbreak
760 is due to the exchange of genetic material among transiently present (via urine contamination)
761 and soil dwelling pathogenic leptospires that result in increased pathogenic potential of certain
762 strains in these complex environments. A comprehensive comparative genomics study
763 conducted by Xu *et al.*, 2016 (7) revealed that horizontal gene transfer events resulting in the
764 acquisition of novel genes in *Leptospira* are fairly common.

765 Overall, this study provides evidence for the existence of divergent lineages of soil
766 dwelling pathogenic *Leptospira* spp. in Puerto Rico; the disease potential of these novel types in
767 humans and animals is unknown. The genetic diversity within this genus found in the
768 environment in Puerto Rico is vast and, thus, there is a need to characterize this diversity to
769 account for it in the diagnosis of leptospirosis. Many cases of leptospirosis are thought to go
770 undetected, or are misdiagnosed as other febrile illnesses, and it is possible that these novel
771 types play a role in such cases. A follow up study in Puerto Rico to investigate the possibility
772 that these novel types are found in human and animal samples would be highly valuable.
773 Furthermore, environmental surveys aimed at the discovery of these novel types in all areas
774 where leptospirosis is a public health concern will be of critical importance towards improving
775 diagnostics and public health outcomes.

776 **Acknowledgements**

777 We thank Ryelan McDonough and Amber Jones for laboratory assistance preparing
778 samples for whole genome sequencing and Karen LeCount for performing MAT. USDA is an
779 equal opportunity provider and employer. Mention of trade names or commercial products in
780 this publication is solely for the purpose of providing specific information and does not imply
781 recommendation or endorsement by the U.S. Department of Agriculture or the Centers for
782 Disease Control and Prevention. This work was funded in part by the United States Department
783 of Health and Human Services Centers for Disease Control and Prevention (Award#
784 6NU1ROT000006-01-04). The Brazilian agency FAPESP also financially supported this work;
785 LGVF is funded with a fellowship from FAPESP (2017/06731-8 and 2019/20302-8).

786 References

- 787 1. Vincent AT, Schiettekatte O, Goarant C, Neela VK, Bernet E, Thibeaux R, et al. Revisiting the
788 taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. *PLoS*
789 *Negl Trop Dis*. 2019;13(5):e0007270.
- 790 2. Levett PN. Leptospirosis. *Clin Microbiol Rev*. 2001;14(2):296-326.
- 791 3. Adler B, de la Pena Moctezuma A. *Leptospira* and leptospirosis. *Vet Microbiol*. 2010;140(3-
792 4):287-96.
- 793 4. Stoddard RA, Gee JE, Wilkins PP, McCaustland K, Hoffmaster AR. Detection of pathogenic
794 *Leptospira* spp. through TaqMan polymerase chain reaction targeting the LipL32 gene. *Diagn Microbiol*
795 *Infect Dis*. 2009;64(3):247-55.
- 796 5. Ko AI, Goarant C, Picardeau M. *Leptospira*: the dawn of the molecular genetics era for an
797 emerging zoonotic pathogen. *Nat Rev Microbiol*. 2009;7(10):736-47.
- 798 6. Picardeau M, Bulach DM, Bouchier C, Zuerner RL, Zidane N, Wilson PJ, et al. Genome sequence
799 of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the
800 pathogenesis of leptospirosis. *PLoS One*. 2008;3(2):e1607.
- 801 7. Xu Y, Zhu Y, Wang Y, Chang YF, Zhang Y, Jiang X, et al. Whole genome sequencing revealed host
802 adaptation-focused genomic plasticity of pathogenic *Leptospira*. *Sci Rep*. 2016;6:20020.
- 803 8. Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, Lovett MA, et al. Leptospirosis: a zoonotic
804 disease of global importance. *Lancet Infect Dis*. 2003;3(12):757-71.
- 805 9. Picardeau M. Diagnosis and epidemiology of leptospirosis. *Med Mal Infect*. 2013;43(1):1-9.
- 806 10. Putz EJ, Nally JE. Investigating the Immunological and Biological Equilibrium of Reservoir Hosts
807 and Pathogenic *Leptospira*: Balancing the Solution to an Acute Problem? *Front Microbiol*. 2020;11:2005.
- 808 11. Monahan AM, Callanan JJ, Nally JE. Proteomic analysis of *Leptospira interrogans* shed in urine of
809 chronically infected hosts. *Infect Immun*. 2008;76(11):4952-8.
- 810 12. Haake DA, Levett PN. Leptospirosis in humans. *Curr Top Microbiol Immunol*. 2015;387:65-97.

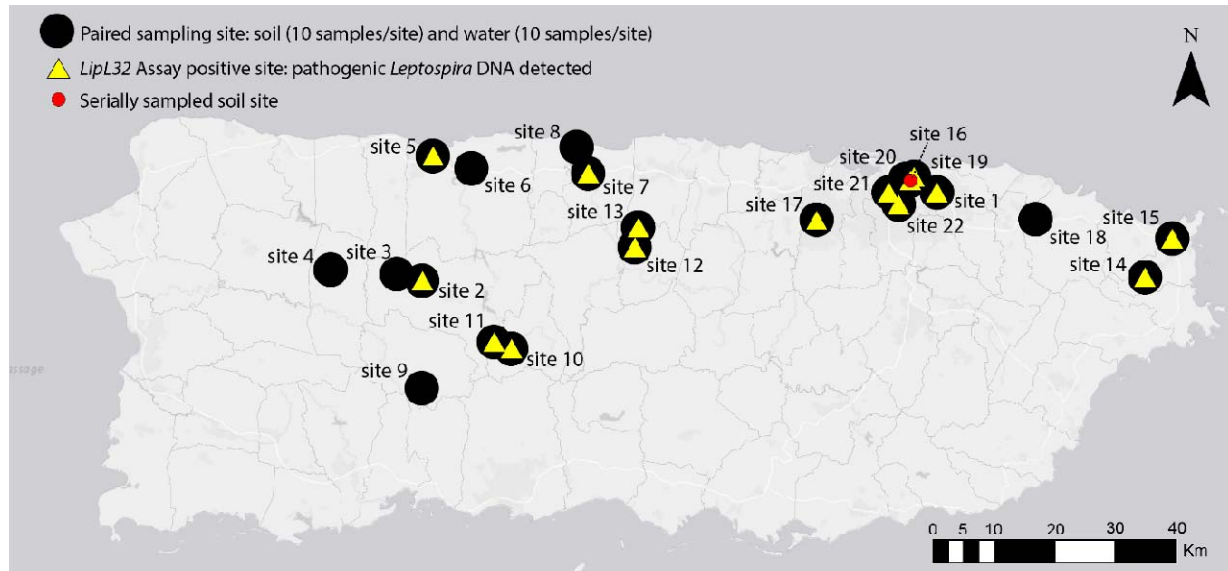
- 811 13. Leonard FC, Quinn PJ, Ellis WA, O'Farrell K. Duration of urinary excretion of leptospires by cattle
812 naturally or experimentally infected with *Leptospira interrogans* serovar hardjo. Vet Rec.
813 1992;131(19):435-9.
- 814 14. Rocha BR, Narduche L, Oliveira CS, Martins G, Lilenbaum W. Molecular demonstration of
815 intermittent shedding of *Leptospira* in cattle and sheep and its implications on control. Cienc Rural.
816 2017;47(8).
- 817 15. Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, et al. Global Morbidity
818 and Mortality of Leptospirosis: A Systematic Review. PLoS Negl Trop Dis. 2015;9(9):e0003898.
- 819 16. Goarant C. Leptospirosis: risk factors and management challenges in developing countries. Res
820 Rep Trop Med. 2016;7:49-62.
- 821 17. Marinova-Petkova A, Guendel I, Strysko JP, Ekpo LL, Galloway R, Yoder J, et al. First Reported
822 Human Cases of Leptospirosis in the United States Virgin Islands in the Aftermath of Hurricanes Irma and
823 Maria, September-November 2017. Open Forum Infect Dis. 2019;6(7):ofz261.
- 824 18. Ramos LJS, Lazaro PM, Barbosa JS, Cardona CP. Spatial and Epidemiological Analysis of Human
825 Leptospirosis in Puerto Rico, 1996 to 2014. Geofocus-Rev Int Cie. 2018(21):227-51.
- 826 19. Peters A, Vokaty A, Portch R, Gebre Y. Leptospirosis in the Caribbean: a literature review. Rev
827 Panam Salud Publica. 2017;41:e166.
- 828 20. Mwachui MA, Crump L, Hartskeerl R, Zinsstag J, Hattendorf J. Environmental and Behavioural
829 Determinants of Leptospirosis Transmission: A Systematic Review. PLoS Negl Trop Dis.
830 2015;9(9):e0003843.
- 831 21. Herrmann-Storck C, Brioudes A, Quirin R, Deloumeaux J, Lamaury I, Nicolas M, et al.
832 Retrospective review of leptospirosis in Guadeloupe, French West Indies 1994-2001. West Indian Med J.
833 2005;54(1):42-6.
- 834 22. Pappachan MJ, Mathew S, Aravindan KP, Khader A, Bharghavan PV, Kareem MM, et al. Risk
835 factors for mortality in patients with leptospirosis during an epidemic in northern Kerala. Natl Med J
836 India. 2004;17(5):240-2.
- 837 23. Epstein PR. Climate change and human health. N Engl J Med. 2005;353(14):1433-6.
- 838 24. Sanders EJ, Rigau-Perez JG, Smits HL, Deseda CC, Vorndam VA, Aye T, et al. Increase of
839 leptospirosis in dengue-negative patients after a hurricane in Puerto Rico in 1996 [correction of 1966].
840 Am J Trop Med Hyg. 1999;61(3):399-404.
- 841 25. Concepcion-Acevedo J, Patel A, Luna-Pinto C, Pena RG, Cuevas Ruiz RI, Arbolay HR, et al. Initial
842 Public Health Laboratory Response After Hurricane Maria - Puerto Rico, 2017. MMWR Morb Mortal
843 Wkly Rep. 2018;67(11):333-6.
- 844 26. Nedelman M. Suspected leptospirosis cases increasing in Puerto Rico after hurricane Maria.
845 <https://www.cnn.com/2017/10/24/health/leptospirosis-puerto-rico/index.html>. 2017.
- 846 27. Casanovas-Massana A, Pedra GG, Wunder EA, Jr., Diggle PJ, Begon M, Ko AI. Quantification of
847 *Leptospira interrogans* Survival in Soil and Water Microcosms. Appl Environ Microbiol. 2018;84(13).
- 848 28. Bierque E, Thibeaux R, Girault D, Soupe-Gilbert ME, Goarant C. A systematic review of
849 *Leptospira* in water and soil environments. PLoS One. 2020;15(1):e0227055.
- 850 29. Tarantola A, Goarant C. Leptospirosis in French Historical Medical Literature: Weil's Disease or
851 Kelsch's Disease? Am J Trop Med Hyg. 2018;99(6):1366-8.
- 852 30. Miller E, Barragan V, Chiriboga J, Weddell C, Luna L, Jimenez DJ, et al. *Leptospira* in river and soil
853 in a highly endemic area of Ecuador. BMC Microbiol. 2021;21(1):17.
- 854 31. Munoz-Zanzi C, Mason MR, Encina C, Astroza A, Romero A. *Leptospira* contamination in
855 household and environmental water in rural communities in southern Chile. Int J Environ Res Public
856 Health. 2014;11(7):6666-80.

- 857 32. Flores B, Escobar K, Muzquiz JL, Sheleby-Elias J, Mora B, Roque E, et al. Detection of Pathogenic
858 Leptospire in Water and Soil in Areas Endemic to Leptospirosis in Nicaragua. *Trop Med Infect Dis.*
859 2020;5(3).
- 860 33. Barragan V, Olivas S, Keim P, Pearson T. Critical Knowledge Gaps in Our Understanding of
861 Environmental Cycling and Transmission of *Leptospira* spp. *Appl Environ Microbiol.* 2017;83(19).
- 862 34. Hall CM, Jaramillo S, Jimenez R, Stone NE, Centner H, Busch JD, et al. *Burkholderia pseudomallei*,
863 the causative agent of melioidosis, is rare but ecologically established and widely dispersed in the
864 environment in Puerto Rico. *PLoS Negl Trop Dis.* 2019;13(9):e0007727.
- 865 35. Stone NE, Sidak-Loftis LC, Sahl JW, Vazquez AJ, Wiggins KB, Gillece JD, et al. More than 50% of
866 *Clostridium difficile* Isolates from Pet Dogs in Flagstaff, USA, Carry Toxigenic Genotypes. *PloS one.*
867 2016;11(10):e0164504.
- 868 36. Ahmed A, Engelberts MF, Boer KR, Ahmed N, Hartskeerl RA. Development and validation of a
869 real-time PCR for detection of pathogenic *Leptospira* species in clinical materials. *PLoS One.*
870 2009;4(9):e7093.
- 871 37. Ahmed N, Devi SM, Valverde Mde L, Vijayachari P, Machang'u RS, Ellis WA, et al. Multilocus
872 sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species.
873 *Ann Clin Microbiol Antimicrob.* 2006;5:28.
- 874 38. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for
875 Windows 95/98/NT. *Nucleic Acids Symposium Series.* 1999(41):95-8.
- 876 39. Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, et al. Detection of Low-
877 Level Mixed-Population Drug Resistance in *Mycobacterium tuberculosis* Using High Fidelity Amplicon
878 Sequencing. *Plos One.* 2015;10(5).
- 879 40. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible,
880 interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.*
881 2019;37(8):852-7.
- 882 41. M. M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads.
883 *EMBNetjournal.* 2011;17(1):10-2.
- 884 42. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution
885 sample inference from Illumina amplicon data. *Nat Methods.* 2016;13(7):581-3.
- 886 43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol*
887 *Biol.* 1990;215(3):403-10.
- 888 44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture
889 and applications. *BMC Bioinformatics.* 2009;10:421.
- 890 45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
891 *Nucleic Acids Res.* 2004;32(5):1792-7.
- 892 46. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics
893 Analysis version 6.0. *Molecular biology and evolution.* 2007;30(12):2725-9.
- 894 47. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-
895 BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389-
896 402.
- 897 48. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic
898 algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution.*
899 2015;32(1):268-74.
- 900 49. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model
901 selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587-9.
- 902 50. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast
903 Bootstrap Approximation. *Molecular biology and evolution.* 2018;35(2):518-22.

- 904 51. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and
905 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst*
906 *Biol.* 2010;59(3):307-21.
- 907 52. Team R. RStudio: Integrated Development for R. PBC, Boston, MA: RStudio; 2020.
- 908 53. Hornsby RL, Alt DP, Nally JE. Isolation and propagation of leptospire at 37 degrees C directly
909 from the mammalian host. *Sci Rep-Uk.* 2020;10(1).
- 910 54. Chakraborty A, Miyahara S, Villanueva SY, Saito M, Gloriani NG, Yoshida S. A novel combination
911 of selective agents for isolation of *Leptospira* species. *Microbiol Immunol.* 2011;55(7):494-501.
- 912 55. Zuerner RL. Laboratory maintenance of pathogenic *Leptospira*. *Curr Protoc Microbiol.*
913 2005;Chapter 12:Unit 12E 1.
- 914 56. Fernandes LGV, Hornsby RL, Nascimento A, Nally JE. Genetic manipulation of pathogenic
915 *Leptospira*: CRISPR interference (CRISPRi)-mediated gene silencing and rapid mutant recovery at 37
916 degrees C. *Sci Rep.* 2021;11(1):1768.
- 917 57. Kozarewa I, Turner DJ. 96-plex molecular barcoding for the Illumina Genome Analyzer. *Methods*
918 *Mol Biol.* 2011;733:279-98.
- 919 58. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids*
920 *Res.* 2012;40(Database issue):D48-53.
- 921 59. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new
922 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.*
923 2012;19(5):455-77.
- 924 60. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094-
925 100.
- 926 61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
927 format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
- 928 62. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast
929 genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17(1):132.
- 930 63. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0:
931 fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17(3):261-72.
- 932 64. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in
933 diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods-Uk.*
934 2016;8(1):12-24.
- 935 65. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068-9.
- 936 66. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment
937 and comparison. *Nucleic Acids Res.* 2002;30(11):2478-83.
- 938 67. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid
939 method for the identification of SNPs in WGS datasets that supports flexible input and output formats.
940 *Microb Genom.* 2016;2(8):e000074.
- 941 68. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-
942 TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular*
943 *biology and evolution.* 2020;37(5):1530-4.
- 944 69. Nally JE, Hornsby RL, Alt DP, Bayles D, Wilson-Welder JH, Palmquist DE, et al. Isolation and
945 characterization of pathogenic leptospire associated with cattle. *Vet Microbiol.* 2018;218:25-30.
- 946 70. Cole JR, Jr., Sulzer CR, Pursell AR. Improved microtechnique for the leptospiral microscopic
947 agglutination test. *Appl Microbiol.* 1973;25(6):976-80.
- 948 71. Ellis WA, Montgomery J, Cassells JA. Dihydrostreptomycin treatment of bovine carriers of
949 *Leptospira interrogans* serovar hardjo. *Res Vet Sci.* 1985;39(3):292-5.
- 950 72. Picaudeau M. Virulence of the zoonotic agent of leptospirosis: still terra incognita? *Nat Rev*
951 *Microbiol.* 2017;15(5):297-307.

- 952 73. Baker GC, Smith JJ, Cowan DA. Review and re-analysis of domain-specific 16S primers. J
953 Microbiol Methods. 2003;55(3):541-55.
- 954 74. Thibeaux R, Soupe-Gilbert ME, Kainiu M, Girault D, Bierque E, Fernandes J, et al. The zoonotic
955 pathogen *Leptospira interrogans* mitigates environmental stress through cyclic-di-GMP-controlled
956 biofilm production. NPJ Biofilms Microbiomes. 2020;6(1):24.
- 957 75. Ristow P, Bourhy P, Kerneis S, Schmitt C, Prevost MC, Lilenbaum W, et al. Biofilm formation by
958 saprophytic and pathogenic leptospires. Microbiol-Sgm. 2008;154:1309-17.
- 959 76. Briskin EA, Casanovas-Massana A, Ryff KR, Morales-Estrada S, Hamond C, Perez-Rodriguez NM,
960 et al. Seroprevalence, Risk Factors, and Rodent Reservoirs of Leptospirosis in an Urban Community of
961 Puerto Rico, 2015. J Infect Dis. 2019;220(9):1489-97.
- 962 77. Izurieta R, Galwankar S, Clem A. Leptospirosis: The "mysterious" mimic. J Emerg Trauma Shock.
963 2008;1(1):21-33.
- 964 78. Musso D, La Scola B. Laboratory diagnosis of leptospirosis: a challenge. J Microbiol Immunol
965 Infect. 2013;46(4):245-52.
- 966 79. Zuerner RL, Alt DP, Palmer MV. Development of chronic and acute golden Syrian hamster
967 infection models with *Leptospira borgpetersenii* serovar Hardjo. Vet Pathol. 2012;49(2):403-11.
- 968 80. Cordonin C, Turpin M, Bascands JL, Dellagi K, Mavingui P, Tortosa P, et al. Three *Leptospira*
969 Strains From Western Indian Ocean Wildlife Show Highly Distinct Virulence Phenotypes Through
970 Hamster Experimental Infection. Frontiers in Microbiology. 2019;10.
- 971 81. Thaipadungpanit J, Wuthiekanun V, Chierakul W, Smythe LD, Petkanchanapong W, Limpaboon
972 R, et al. A dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis
973 in Thailand. PLoS Negl Trop Dis. 2007;1(1):e56.
- 974

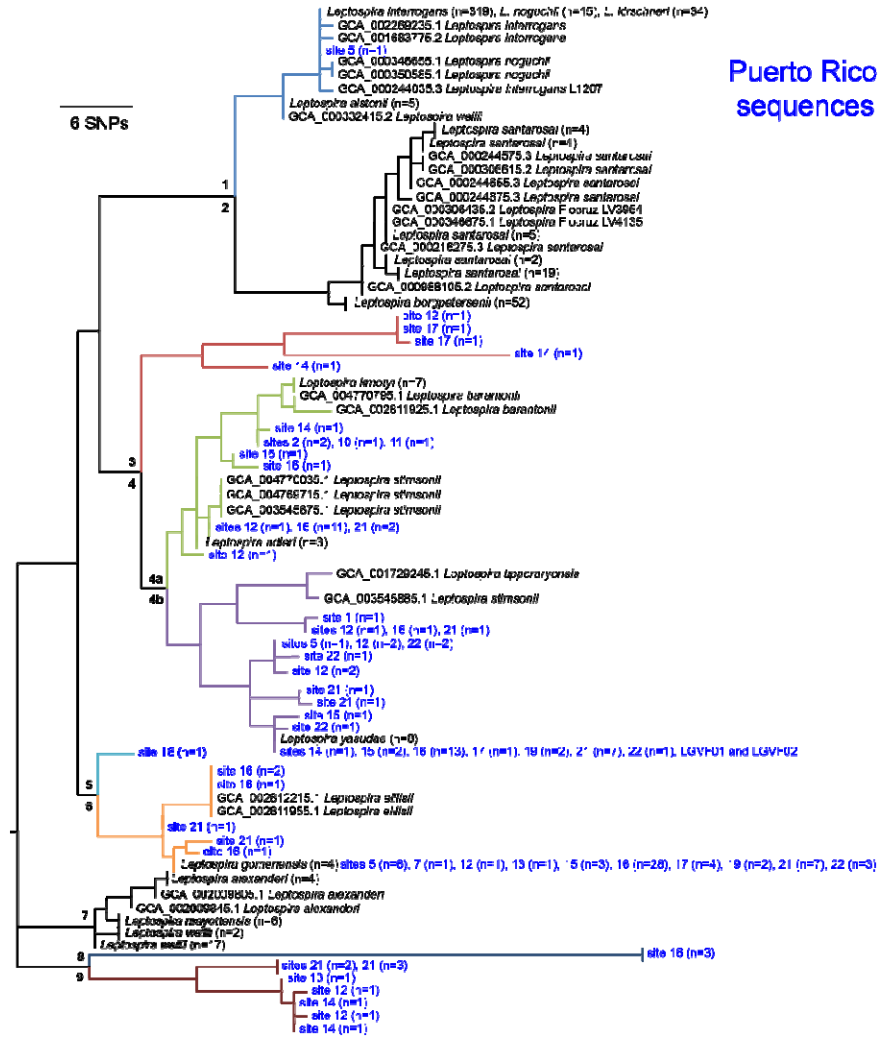
975 **Figures**



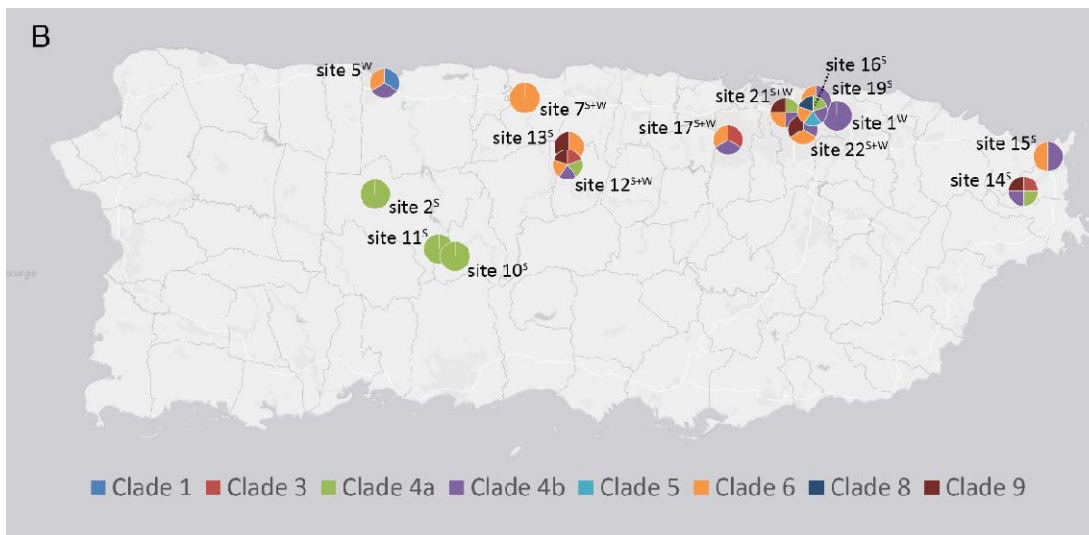
976

977 **Figure 1: Map of Puerto Rico indicating sites where environmental samples were collected for**
978 **the detection of pathogenic *Leptospira* spp. DNA from 2018-2020. Black dots represent paired**
979 **water and soil sites (n=22); 20 samples were collected from each site (10 soil and 10 water).**
980 **Yellow triangles indicate sites where pathogenic *Leptospira* DNA was detected (15/22 sites).**
981 **One site in San Juan (site 16, red circle) was serially sampled three times over a 14-month**
982 **period to assess persistence of pathogenic *Leptospira* in soil.**

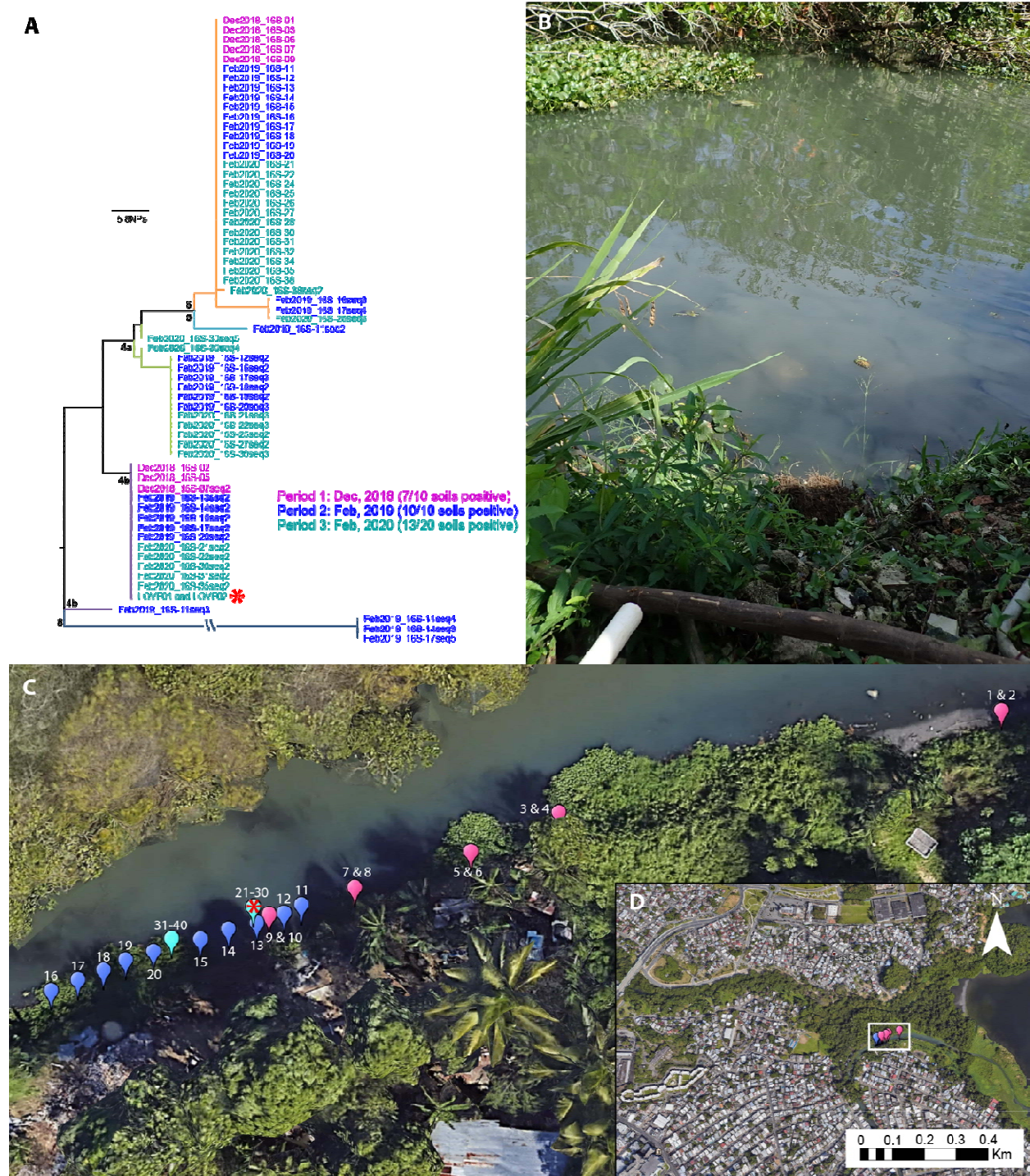
A



B

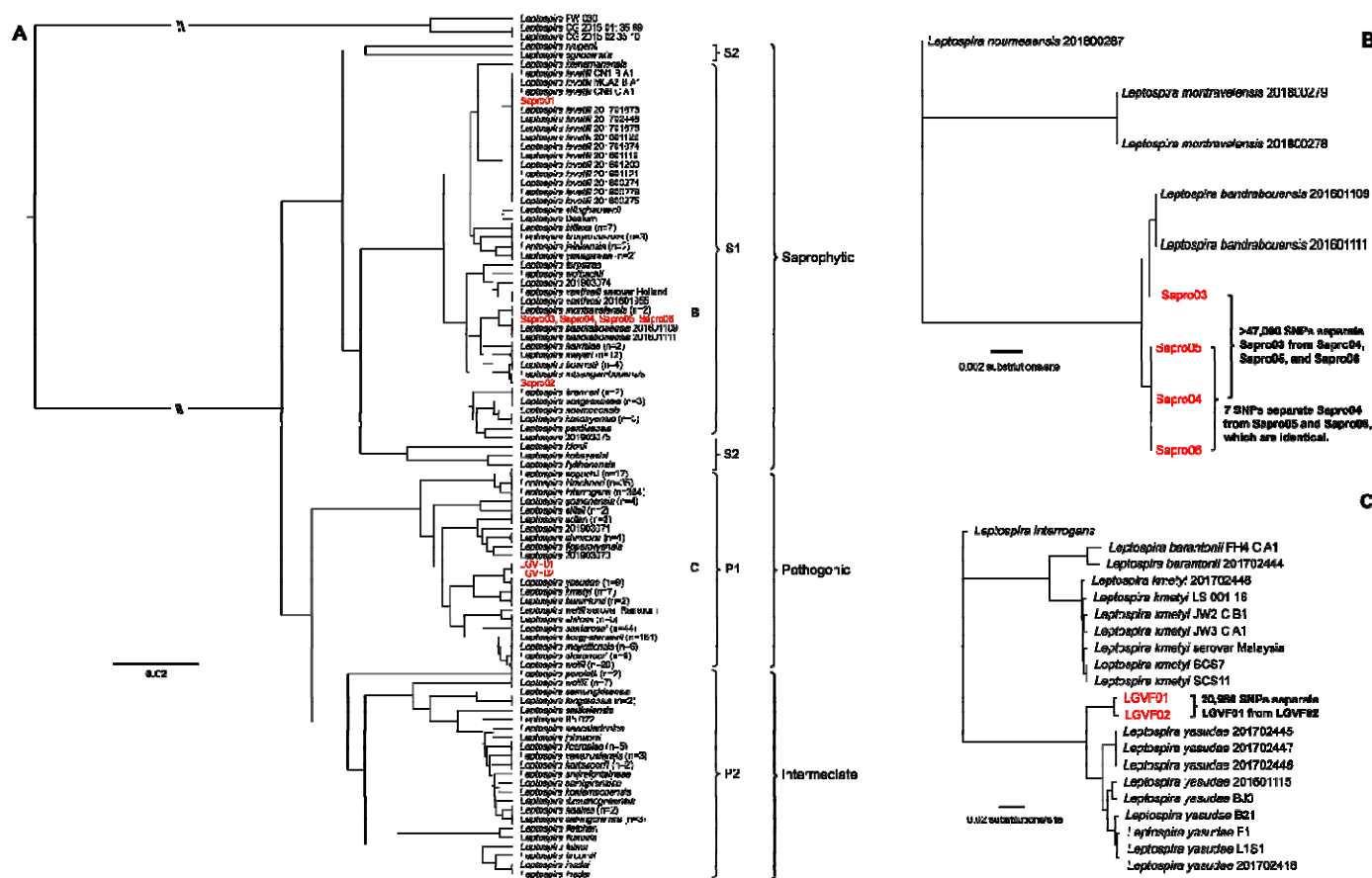


984 **Figure 2: Genetically diverse and previously undescribed lineages of pathogenic *Leptospira***
985 **are widespread in Puerto Rico.** A) Maximum likelihood phylogeny constructed using a 202bp
986 segment of the *lipL32* gene, including sequences obtained from soil and water samples
987 collected in this study (blue text) and those available in public databases (black text). Nine
988 major genetic clades are represented, seven of which were present in Puerto Rico, including
989 four not previously described. Clades are labelled 1 through 9 on the tree and color coded. B)
990 Sites where each major clade was collected, with colors within the pie charts corresponding to
991 those in the phylogeny; superscript indicates if soil (S) and/or water (W) samples were positive
992 at each site. Fifteen positive sites are included on this map.



993
 994 **Figure 3: Site 16 in Caño Martín Peña, San Juan, Puerto Rico where sampling was conducted**
 995 **at three separate timepoints to assess environmental persistence of pathogenic *Leptospira* in**
 996 **soil near the edge of the water. A) Maximum likelihood phylogeny of all *lipL32* sequences**

997 obtained from this soil site over the course of three sampling periods (color coded by date of
998 sampling period); two of five clades detected at this site were identified at all three timepoints,
999 suggesting environmental persistence. Clade colors correspond to Figure 2. Two isolates
1000 representing a novel pathogenic *Leptospira* species were obtained from soil sample 27 (red
1001 asterisk). B) Photo of exact location where samples 21-30 were collected and the isolates were
1002 obtained. C) Transect where 40 soil samples were collected over three sampling periods.
1003 Balloons are color coded to indicate sampling period as in the phylogeny (panel A) and the red
1004 asterisk denotes the location where the *Leptospira* isolates originated. D) Location of site 16 in
1005 Caño Martín Peña.
1006



1007

1008 **Figure 4: Whole genome dendrogram of 802 known pathogenic, intermediate, and**
 1009 **saprophytic *Leptospira* spp. isolates, including six saprophytic and two pathogenic isolates**

1010 **obtained in this study from a single soil sample from site 16. A) Pairwise genomic distance**

1011 **dendrogram contextualizing the relationship among all known saprophytic (S1 and S2),**

1012 **intermediate (P2), and pathogenic (P1) *Leptospira* spp. reveals that the pathogenic isolates**

1013 **LGVF01 and LGVF02 represent two genotypes of the same previously undescribed pathogenic**

1014 **species, whereas the saprophytes belong to three known *Leptospira* spp. B) Detailed view of**

1015 **the S1 clade that contains four saprophytic isolates obtained during this study. This maximum**

1016 **likelihood phylogeny reveals significant diversity among isolates that fall within the *L.***

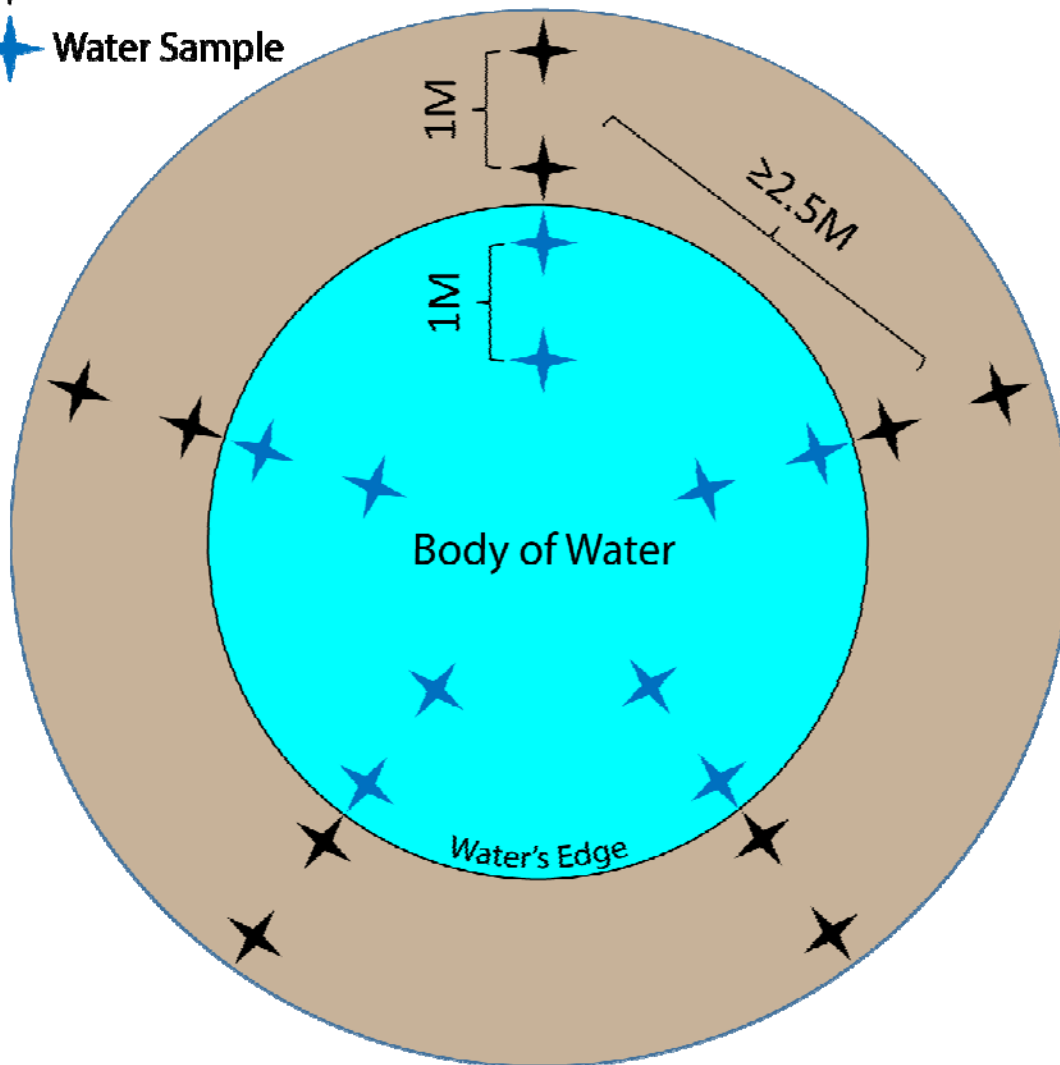
1017 ***bandrabouensis* clade. C) Detailed view of the P1 clade that contains two pathogenic isolates**

1018 obtained during this study. This maximum likelihood phylogeny reveals two genotypes of a
1019 novel pathogenic *Leptospira* spp. isolated from soil in Puerto Rico.

1020 **Supporting information**

✦ Soil Sample

✦ Water Sample



1021

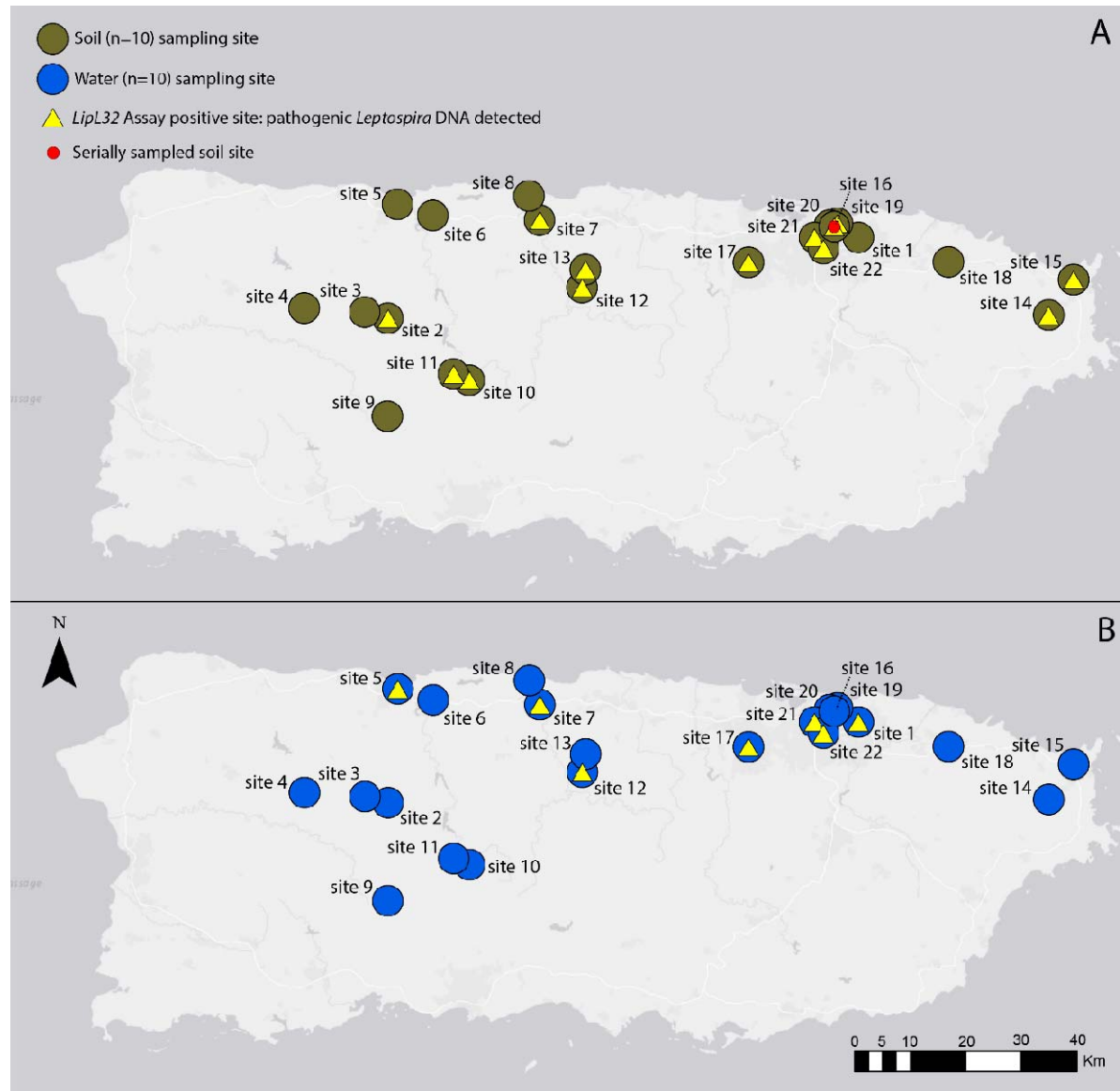
1022 **Figure S1: Sampling design for paired soil and water sites. At each site, five transects were**

1023 **sampled (two soil and two water samples per transect) for a total of 20 samples per site.**

1024 Black and blue stars represent soil and water samples, respectively. Transects were a minimum

1025 of 2.5 meters apart. This sampling design was also applied in a linear fashion for the sampling of

1026 rivers and streams.



1027

1028 **Figure S2: Map of Puerto Rico indicating where soil (panel A) and water (panel B) samples**

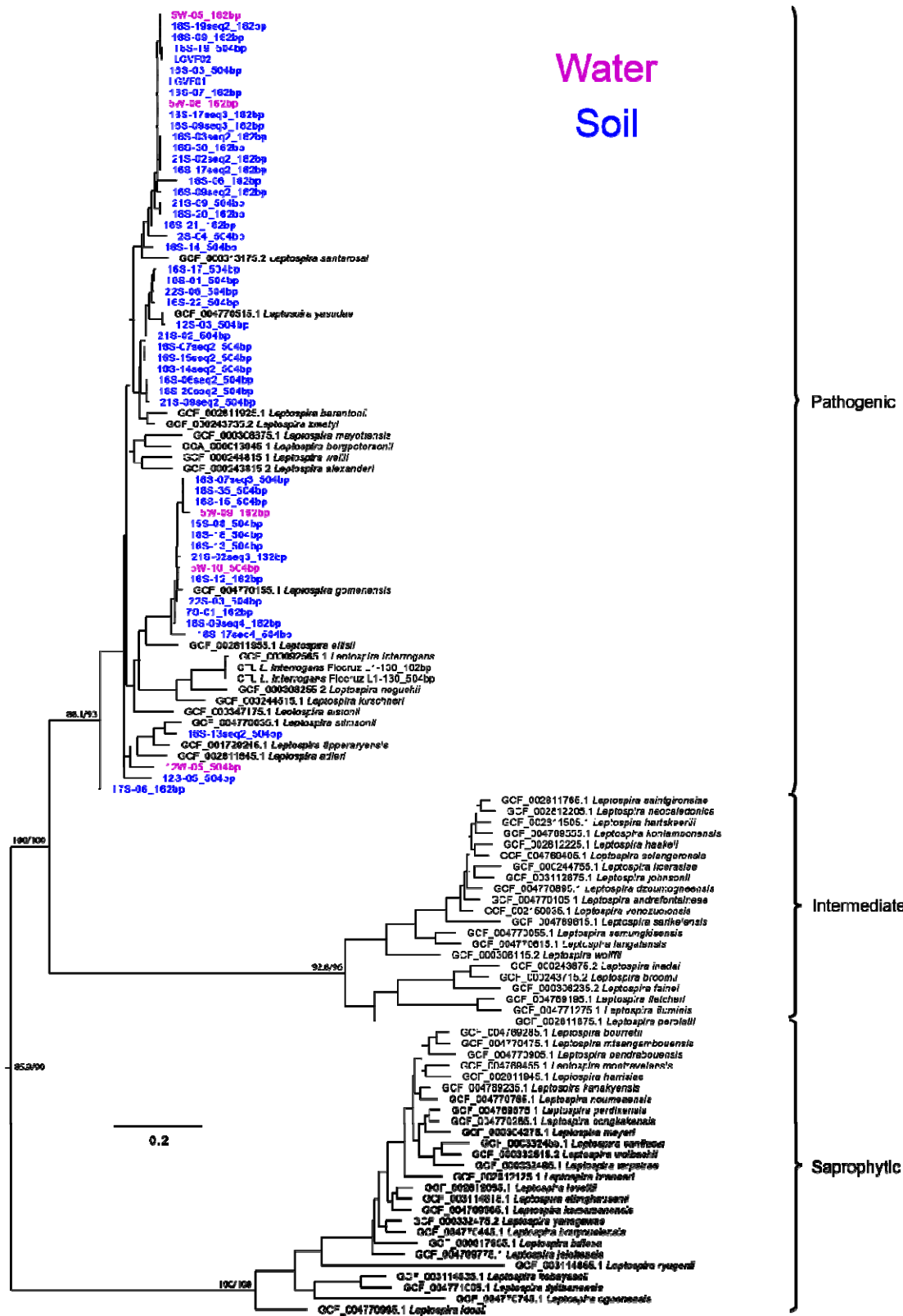
1029 **were collected for the detection of pathogenic *Leptospira* spp. (2018-2020).** The yellow

1030 triangles indicate sites where pathogenic *Leptospira* DNA was detected. Eight soil sites were

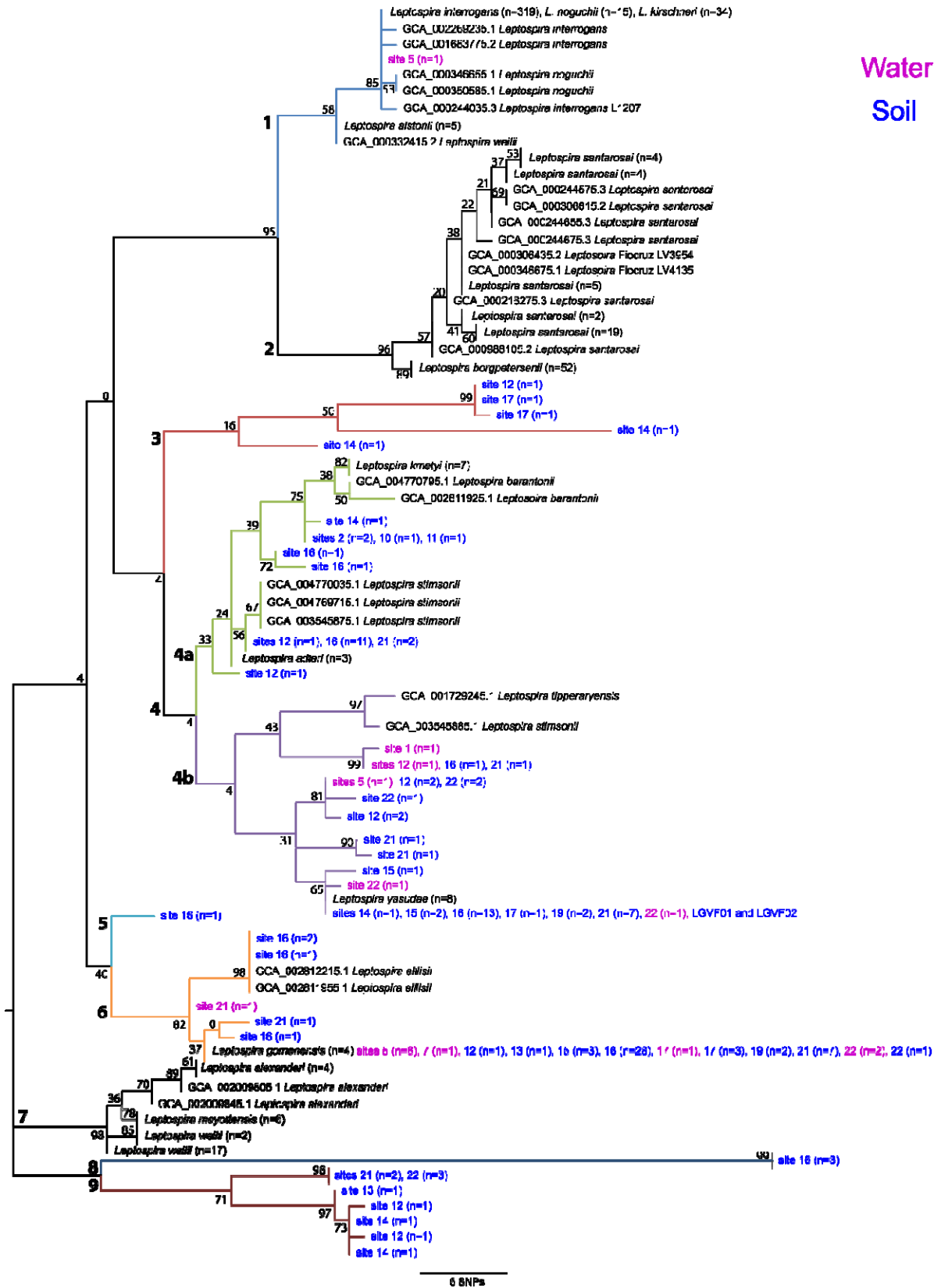
1031 positive for pathogenic *Leptospira* DNA, compared to two water sites; both soil and water were

1032 positive at five sites. Site 16, which was used to assess environmental persistence in soil is

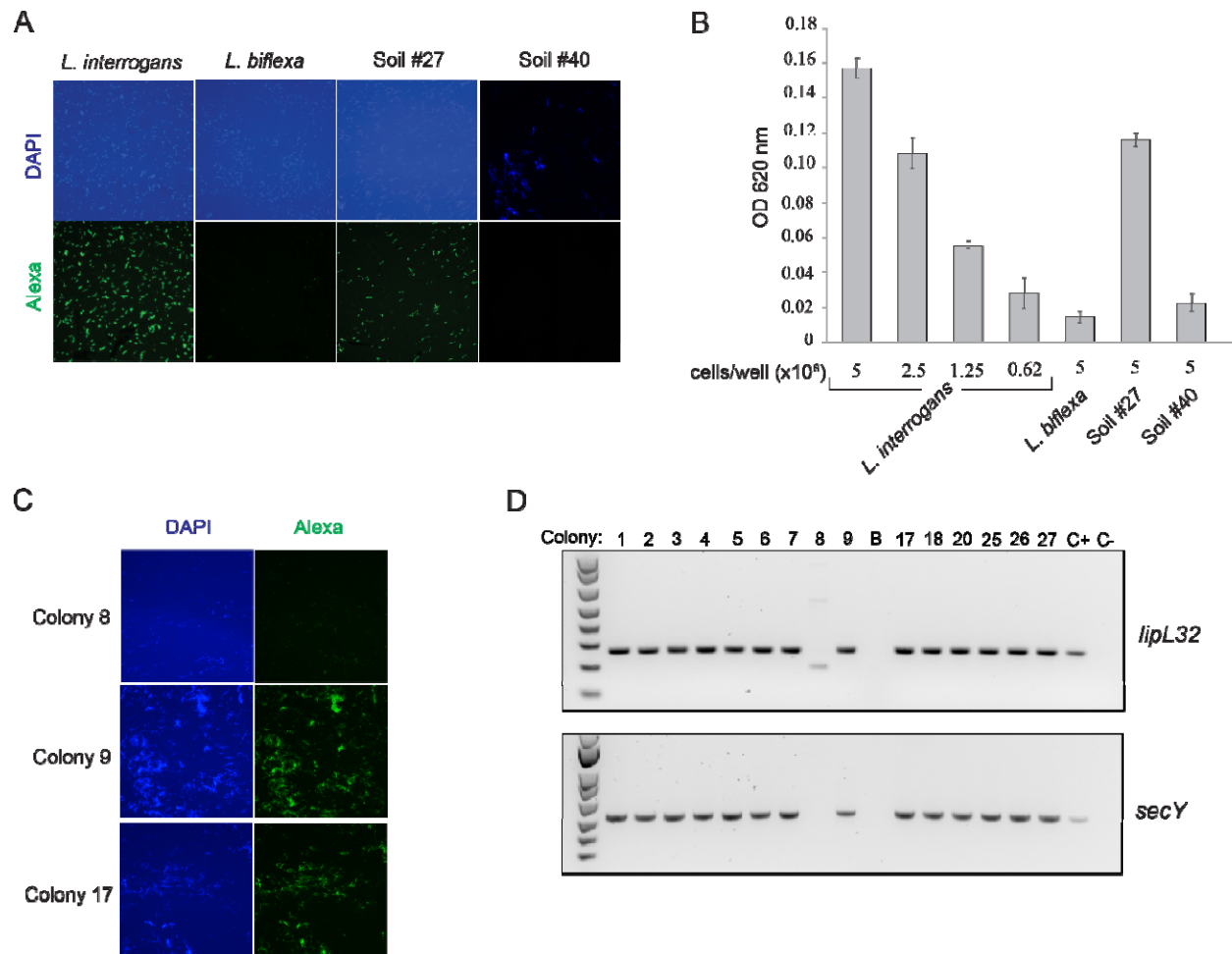
1033 indicated with a red circle.



1035 **Figure S3: Maximum likelihood phylogeny of the *secY* gene reveals that pathogenic *Leptospira***
1036 **was identified in Puerto Rico soil and water and provides support for the phylogenetic**
1037 **relationships observed with *lipL32* sequences.** Sequences generated from soil and water
1038 samples collected in Puerto Rico are in blue or pink text, the novel pathogenic Puerto Rico
1039 isolates from soil are also included; reference sequences for pathogenic, intermediate, and
1040 saprophytic *Leptospira* are in black. Bootstrap/aLRT support values are indicated on branch
1041 nodes.



1043 **Figure S4: Maximum likelihood phylogeny built using a 202bp region of the *lipL32* gene.**
1044 Sequences generated from soil (blue text) and water (pink text) samples collected in Puerto
1045 Rico are displayed. Reference sequences representing all other known pathogenic *Leptospira*
1046 spp. are in black. Nine major clades are represented of which seven were identified in Puerto
1047 Rico and four of those (3, 5, 8, and 9) have not been described previously. Identical genotypes
1048 were present in soil and water from Puerto Rico at two sites (17 and 21), with both occurrences
1049 representing a single genotype from clade 6 (in orange). Bootstrap values are indicated on each
1050 branch. Major clades are numbered 1 through 9 and color-coding matches Figure 2.
1051



1052

1053 **Figure S5: Identification and isolation of pathogenic *Leptospira* spp. from Puerto Rico soil at**

1054 **serially sampled site 16. *LipL32* PCR positive (Soil #27) and negative (Soil #40) samples were**

1055 **subjected to FAT (Panel A) and ELISA (Panel B) along with positive and negative controls (*L.***

1056 ***interrogans* and *L. biflexa*, respectively) to assess expression of the LipL32 pathogenicity**

1057 **protein. Individual colonies from Soil #27 were subjected to additional FAT testing (Panel C) and**

1058 **confirmatory PCR (Panel D) to verify that the obtained isolates were pathogenic.**

1059

1060 **Supplemental File 1: Table S1 Results summary.** Summary of detection and sequencing results
1061 for environmental samples from Puerto Rico, 2018-2020.

1062 **Supplemental File 1: Table S2 Sampling data.** Sample and site ID, location, elevation, and pH.

1063 **Supplemental File 1: Table S3 Accession numbers.** GenBank accession numbers for all
1064 sequences used to build the *lipL32*, *secY*, and whole genome phylogenies.

1065 **Supplemental File 1: Table S4 Statistical tests.** Statistical tests used to assess associations
1066 between the detection of pathogenic *Leptospira* spp. and environmental or sampling variables.

1067 **Supplemental File 1: Table S5 MAT panel.** Antigens used in the microscopic agglutination test
1068 (MAT) that were unreactive when tested against sera from hamsters infected with the novel
1069 pathogenic *Leptospira* spp. isolated from Puerto Rico soil.

1070 **Supplemental File 1: Table S6 *lipL32* sequences.** All *lipL32* sequences generated during this
1071 study using Sanger and AmpSeq methods.

1072 **Supplemental File 1: Table S7: *secY* sequences.** All *secY* sequences generated during this study
1073 for two *secY* amplicons (203 and 549bp).

1074 **Supplemental File 1: Table S8: 16S sequences.** Near full length 16S sequences (1330bp)
1075 generated for six saprophytic isolates.

1076 **Supplemental File 1: Table S9: Genome assemblies.** Genome assembly details including
1077 assembly size, number of contigs, and NCBI assembly accession numbers.

1078 **Supplemental File 1: Table S10: Hamster infection.** The weight of each hamster was recorded
1079 daily, beginning two days preinfection.