Host population structure and rare dispersal events drive leptospirosis transmission patterns among *Rattus norvegicus* in Boston, Massachusetts, US

Short Title: Brown rat population dynamics drive leptospirosis transmission in Boston, US

Nathan E. Stone[1], Camila Hamond[2], Joel Clegg[1], Ryelan F. McDonough[1], Reanna M. Bourgeois[1], Rebecca Ballard[1], Natalie B. Thornton[1], Marianece Nuttall[1], Hannah Hertzel[3], Tammy Anderson[2], Ryann N. Whealy[1,4], Skylar Timm[1,4], Alexander K. Roberts[1,4], Verónica Barragán[5], Wanda Phipatanakul[6], Jessica H. Leibler[7], Hayley Benson[3], Aubrey Specht[3], Ruairi White[3], Karen LeCount[2], Tara N. Furstenau[4], Renee L. Galloway[8], Nichola J. Hill[9], Joseph D. Madison[9,10], Viacheslav Y. Fofanov[1,4], Talima Pearson[1], Jason W. Sahl[1], Joseph D. Busch[1], Zachary Weiner[8], Jarlath E. Nally[11], David M. Wagner[1¶*], and Marieke H. Rosenbaum[3¶*]

[1]The Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, US

[2]National Veterinary Services Laboratories, APHIS, United States Department of Agriculture, Ames, IA, US

[3]Department of Infectious Disease and Global Health, Cummings School of Veterinary Medicine, Tufts University, Grafton, MA, US

[4]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, US

[5] Universidad San Francisco de Quito, Colegio de Ciencias Biologicas y Ambientales, Quito, EC

22 [6]Division of Allergy & Immunology, Boston Children's Hospital, Boston, MA, US

23 [7]Department of Environmental Health, Boston University School of Public Health, Boston, MA US

24 [8]Bacterial Special Pathogens Branch, Centers for Disease Control and Prevention, Atlanta, GA, US

25 [9]University of Massachusetts Boston, Biology Department, Boston, MA, US

26 [10]Center for Conservation Genomics, Smithsonian's National Zoo & Conservation Biology Institute,

27 Washington, DC, US

28 [11]Infectious Bacterial Diseases Research Unit, ARS, United States Department of Agriculture, Ames,

29 IA, US

30 ¶These authors contributed equally to this work.

31 * Corresponding authors

32 Marieke.Rosenbaum@tufts.edu

33 Dave.Wagner@nau.edu

34

35 **Keywords: Leptospirosis, urban setting, whole genome, *Rattus norvegicus*, population genetics**

36

37 **Formatted for PLoS Biology**

38 **No maximum word count**

39 **Figures (4)**

40 **Tables (1)**

41 **Supplementary Figures (7)**

42 **Supplementary Tables (9)**

43 **Supplementary Files (1)**

Author Email Addresses

Nathan E. Stone: Nathan.Stone@nau.edu

Camila Hamond: Camila.Hamond@usda.gov

Joel Clegg: jrc665@nau.edu

Ryelan F. McDonough: Ryelan.McDonough@nau.edu

Reanna M. Bourgeois: reannamb21@gmail.com

Rebecca Ballard: Rebecca.Ballard@nau.edu

Natalie B. Thornton: nbt39@nau.edu

Marianece Nuttall mkn75@nau.edu

Hannah Hertzel: hrhertzel@gmail.com

Tammy Anderson: tammy.m.anderson@usda.gov

Ryann N. Whealy: rnw42@nau.edu

Skylar Timm: sab832@nau.edu

Alexander K. Roberts: akr348@nau.edu

Verónica Barragán: vbarragan@usfq.edu.ec

Wanda Phipatanakul: wanda.phipatanakul@childrens.harvard.edu

Jessica H. Leibler: jleibler@bu.edu

Hayley Benson: haylbenson@gmail.com

Aubrey Specht: aubrey.g.specht@gmail.com

Ruairi White: ruairi.g.white@gmail.com

Karen LeCount: karen.j.lecount@usda.gov

Tara N. Furstenau: Tara.Furstenau@nau.edu

Renee L. Galloway: zul0@cdc.gov

Nichola J. Hill: Nichola.Hill@umb.edu

Joseph D. Madison: josephdmadison7@gmail.com

Viacheslav Y. Fofanov: viacheslav.fofanov@nau.edu

Talima Pearson: talima.pearson@nau.edu

Jason W. Sahl: Jason.Sahl@nau.edu

Joseph D. Busch: Joseph.Busch@nau.edu

Zachary Weiner: xxd7@cdc.gov

Jarlath E. Nally: jarlath.nally@usda.gov

David M. Wagner: Dave.Wagner@nau.edu

Marieke H. Rosenbaum: Marieke.Rosenbaum@tufts.edu

Abstract

Leptospirosis (caused by pathogenic bacteria in the genus *Leptospira*) is prevalent worldwide but more common in tropical and subtropical regions. Transmission can occur following direct exposure to infected urine from reservoir hosts, such as rats, or a urine-contaminated environment, which then can serve as an infection source for additional rats and other mammals, including humans. The brown rat, *Rattus norvegicus*, is an important reservoir of leptospirosis in urban settings. We investigated leptospirosis among brown rats in Boston, Massachusetts and hypothesized that rat dispersal in this urban setting influences the movement, persistence, and diversity of *Leptospira*. We analyzed DNA from 328 rat kidney samples collected from 17 sites in Boston over a seven-year period (2016-2022); 59 rats representing 12 of 17 sites were positive for *Leptospira*. We used 21 neutral microsatellite loci to genotype 311 rats and utilized the resulting data to investigate genetic connectivity among sampling sites. We generated whole genome sequences for 28 *Leptospira* isolates obtained from frozen and fresh tissue from some of the 59 *Leptospira*-positive rat kidneys. When isolates were not obtained, we attempted *Leptospira* genomic DNA capture and enrichment, which yielded 14 additional *Leptospira* genomes from rats. We also generated an enriched *Leptospira* genome from a 2018 human case in Boston. We found evidence of high genetic structure and limited dispersal among rat populations that is likely influenced by major roads and/or other unknown dispersal barriers, resulting in distinct rat population groups within the city; at certain sites these groups persisted for multiple years. We identified multiple distinct phylogenetic clades of *L. interrogans* among rats, with specific clades tightly linked to distinct rat populations. This pattern suggests *L. interrogans* persists in local rat populations and movement of leptospirosis in this urban rat community is driven by rat dispersal. Finally, our genomic analyses of the 2018 human leptospirosis case in Boston suggests a link to rats as the source. These findings will be useful for guiding rat control and human leptospirosis mitigation efforts in this and other urban settings.

Introduction

Rats (*Rattus norvegicus and R. rattus*) are highly invasive species globally, persisting on all continents (except Antarctica) and in most habitats (1, 2). They cause devastating impacts to natural ecosystems (3), and are considered a serious public health threat (4, 5). *Rattus norvegicus* (the brown rat) is ubiquitous in urban environments (6, 7) and has been documented to carry/transmit (either directly or indirectly) an abundance of bacterial and viral pathogens in urban ecosystems, including *Leptospira* spp., *Bartonella* spp., *Borrelia* spp., *Yersinia pestis*, *Streptobacillus moiliformis*, antibiotic-resistant *Staphylococcus* spp., *Shigella* spp., *Campylobacter* spp., *Escherichia coli*, *Rickettsia* spp., *Salmonella* spp., influenza A virus, and Seoul hantavirus (8-11).

Multiple genetic studies utilizing various methods have assessed *R. norvegicus* population structure, movements, and dispersal patterns in urban settings in both temperate and tropical climates (2, 12-17), consolidated in a recent review (6). Despite this somewhat limited body of research, findings are generally uniform across all cities assessed (Vancouver, Canada; New York City, US; New Orleans, US; Baltimore, US; Salvador, Brazil; and Hauts-de-Seine, France): 1) individual *R. norvegicus* utilize small home ranges, display site fidelity, and display philopatry to their birth site; and 2) dispersal events are more common in adult rats (male or female), but rare overall, and often inhibited by major roads and rivers, which act as dispersal barriers.

Leptospirosis is the most globally widespread bacterial zoonoses resulting in over 1 million human infections annually and nearly 60,000 deaths (18,19). Brown rats are asymptomatic, chronic carriers of pathogenic *Leptospira* spp. (20-23) and the most important known source of human leptospirosis infections (24). Leptospires colonize and replicate in the proximal renal tubules of rats (20-22), with transmission occurring after leptospires are shed through urine, either via direct contact with the urine or indirectly via environmental or fomite contamination (24). Globally, >30% of *R. norvegicus* are estimated to be infected with *Leptospira* spp. (25).

126 Understanding how rat movements relate to *Leptospira* transmission in urban settings is of
127 vital importance to guide the rational design of disease mitigation strategies. A recent study of rat
128 population structure and pathogen density in an urban setting (Vancouver) suggested *L. interrogans*
129 prevalence was associated with rat population connectivity (16), wherein related rats sampled from
130 neighboring blocks shared the same pathogen status (*i.e.*, infected/not infected). Our study builds
131 upon this finding by pairing fine-scale population genetic analyses of rats in another urban city
132 (Boston, US) with fine-scale genomic analyses of the *Leptospira* they carry and transmit.
133 Historically, the acquisition of *Leptospira* genomes from reservoir (*e.g.*, rats) and incidental
134 (*e.g.*, humans) hosts has relied on the ability to obtain cultured isolates, which is inherently
135 challenging (26). Obtaining isolates from human leptospirosis cases is further inhibited by the
136 administration of antibiotics used to treat the disease (27), after which isolation of live leptospires is
137 almost impossible. As such, genomic analyses of infecting strains are lacking, and these data are
138 critical for improved understanding of leptospirosis ecology and epidemiology. We implemented two
139 new culture-independent techniques for the genetic/genomic characterization of infecting *Leptospira*
140 spp. in rats and humans in Boston: DNA capture and enrichment (28) and amplicon sequencing
141 (AmpSeq; described herein). Our goal was to gain a deeper understanding of the nuanced *Leptospira*
142 transmission network among urban *R. norvegicus* and potential spillover into humans. Our findings
143 are useful for guiding rat control and human leptospirosis mitigation efforts in urban settings.
144

145 # Results

146 ## *lipL32* qPCR in rodent samples

147 In total, 59 *R. norvegicus* kidney samples and one fresh urine sample (derived from an animal
148 that also had a positive kidney) were *lipL32* PCR positive, revealing an overall *Leptospira* infection
149 rate of 18.0% (59 of 328 animals) in this *R. norvegicus* dataset (**S1 Table**). *Leptospira*-positive

150 kidneys included 37 frozen samples (17.6%; 37/210) yielding an average Ct value of 28.77 ± 5.58,
151 and 23 fresh samples from 22 rats (18.6%; 22/118) yielding an average Ct value of 27.1 ± 6.06. Four
152 bycatch mice and one squirrel were *lipL32* PCR negative (**S2 Table**). Fifty-six of the infected *R.*
153 *norvegicus* were detected among 20/33 collections representing 12/17 sampling sites (**Fig 1 and S1**
154 **Table**) and positive rats were detected in all sampling years [2016=3/42 (7.1%), 2017=6/50 (12.0%),
155 2018=16/68 (23.5%), 2019=9/42 (21.4%), 2020=3/8 (37.5%), 2021=14/70 (20.0%), 2022=8/48
156 (16.7%)]. Precise locations were unknown for infected *R. norvegicus* collected in 2020 (samples B1,
157 B2, and B4; **S1 and S2 Tables**). Infected rats were more commonly adults (58/59) than juveniles
158 (1/59) (X^2 =60.68, p <0.001), and a higher proportion of infected rats were collected in fall (0.277)
159 and winter (0.283) compared to spring (0.196) and summer (0.123); a chi-square test indicated a non-
160 random association between *Leptospira* infection and season (X^2 =10.33, p <0.05). We also detected
161 a significant association between *Leptospira* infection and site (X^2 =26.58, p <0.05); however, this
162 association is likely driven by capture of a single rat from site 4 that was infected, and 3/5 infected
163 rats captured at site 15 (60%); all other sites ranged from 0-33.3% positivity. No other significant
164 associations were observed between other sampling variables (sex, sampling year, genetic group) and
165 the detection of *Leptospira*.
166
167 **Fig 1. *Rattus norvegicus* population assignment.** STRUCTURE population assignment of 311 *R.*
168 *norvegicus* with collections color coded according to genetic group (S1 Table); likely dispersal
169 barriers and major open spaces are indicated with black bars and green shading, respectively. A)
170 Inferred migrants are represented by one-sided arrows; shared genetic groups among collections from
171 sites 13, 14, and 16, and separately sites 9 and 10, are indicated by thin lines; and *lipL32* PCR
172 positivity at 12 of 17 sites is indicated with asterisks. Two migrants (sites 2 and 6) were both *lipL32*
173 PCR positive and are denoted by asterisks on one-sided arrows. B) Each individual *R. norvegicus*

174 (n=311) is assigned a probability of membership ($Q \geq 0.75$) to one of 12 genetic groups (S3 Table).
175 Collection ID is indicated on the X-axis and probability (Q) on the Y-axis; each thin vertical line
176 represents the Q value for an individual rat. The map in panel A was created using ArcGIS software
177 by Esri. ArcGIS and Arc-Map are the intellectual property of Esri and are used herein under license.
178 Copyright Esri. All rights reserved. For more information about Esri software, please visit
179 www.esri.com. Basemap: Light Gray Canvas Base v2
180 https://www.arcgis.com/home/item.html?id=291da5eab3a0412593b66d384379f89f.
181

182 # Culture

183 Twenty-eight isolates were obtained from *lipL32* PCR positive kidneys, including 15 from 37
184 frozen samples and 13 from 22 fresh samples. Nearly half of the collections (15/33) are represented
185 by these isolates as well as all sampling years (2016=1, 2017=2, 2018=6, 2019=3, 2020=3, 2021=7,
186 2022=6; **S2 Table**). The frozen samples were originally cultured in both T80/T40/LH and HAN
187 media, which were incubated at 29 and 37°C, respectively. However, it became evident that they
188 were heavily contaminated, and we observed that the contaminants grew better in HAN compared to
189 T80/T40/LH. Therefore, use of T80/T40/LH at 29°C was prioritized. Fresh kidneys were positive by
190 culture in both HAN and T80/40/LH media.
191

192 # Serotyping of *Leptospira* isolates

193 All isolates had high agglutination titers with reference antiserum specific for serogroup
194 Icterohaemorrhagiae (**S2 Table**). Isolates were completely seronegative (at 1:100 or above) (29) for
195 any other serogroup tested, except for three samples: MAR_57 had a titer of 1:102,400 against
196 Icterohaemorrhagiae and 1:100 against Pyrogenes; Cad103 had a titer of 1:51,200 against
197 Icterohaemorrhagiae, 1:200 against Djasiman, and 1:100 against Hardjo; and R304 had a titer of

198 1:52,400 against Icterohaemorrhagiae and 1:100 against Pyrogenes. We consider all three of these
199 isolates to be serogroup Icterohaemorrhagiae as it is not uncommon to see cross reactions with other
200 serogroups (30).
201

202 Evaluation of virulence

203 Intraperitoneal inoculation of all hamsters with 10^8 leptospires derived from a frozen *R.*
204 *norvegicus* kidney (strain R47) resulted in acute disease requiring all hamsters to be euthanized
205 four days post-infection. Liver and kidney samples from each infected hamster were positive by both
206 culture and *lipL32* qPCR; Ct values for positive liver and kidney samples were on average
207 19.04 ± 0.63 and 23.64 ± 0.75, respectively.
208

209 Rat MSAT results

210 MSAT marker validation

211 Twenty-one microsatellite (MSAT) markers were suitable for use in this population genetic
212 analyses of *R. norvegicus*. Only one marker (D12Rat76) deviated from Hardy-Weinberg equilibrium
213 in one collection (8f) at the Bonferroni corrected a =0.000595 (a =0.05/84 tests) and no signatures of
214 linkage disequilibrium were detected. Null alleles, large allelic dropout, or scoring errors due to
215 stuttering also were not detected, although we did detect excessive homozygosity in two markers
216 (D12Rat76 and D5Rat83) for collection 2c and one marker in four other collections (D5Rat33 in 5b,
217 D10Mit5 in 6g1, D7Rat13 in 12a, and D1Cebr3 in 12ab). No single marker displayed excessive
218 homozygosity in more than one collection. To check for allele scoring errors, we replicated PCRs for
219 all 21 loci across 94 *R. norvegicus* DNAs (30.2% of the final dataset) and detected an allele scoring
220 error rate of 0.23% (9 errors from 3,948 possible allele calls).
221

Population genetic analyses: determination of genetic groups and population assignment using STRUCTURE

The delta-*K* method estimated 2 and 12 genetic groups as the most likely number of populations sampled in the 311 individual rats from 33 collections/17 sites (**S1A Fig**). The *K* of 12 is more consistent with what is known about the philopatric nature of *R. norvegicus* (6, 12) and physical distances among our collection sites. Variance among the L(K) values is very low at *K* =12 but rapidly increases beyond this number (**S1B Fig**); this phenomenon is discussed in the STRUCTURE manual and can be expected to occur after the real *K* is reached (31). Furthermore, the higher delta-*K* value at *K* =2 could simply be that *K* =1 is consistently a very poor solution and the model improved dramatically between *K* =1 and 2. The 12 genetic groups were each assigned an arbitrary color in STRUCTURE (**Fig 1**) and that color was retained across all analyses, figures, and tables. Thus, we refer to the genetic groups by the following color designations: Blue, Brown, Evergreen, Green, Light Blue, Orange, Pink, Purple, Red, Steel, Tan, Yellow, and admixed (indicated by the color White) (**Fig 1 and S1 Table**). To aid in visualizing patterns across space and time, collections were color coded according to the most dominant genetic group (≥50%) present in that collection (**defined in S1 Table and displayed in Figs 1 and 2**). In total, 247 rats were assigned to one of these genetic groups at $Q \geq 0.75$ (**see Materials and Methods**), whereas 64 rats were assigned to the admixed group (**S3 Table**). Due to repeated sampling over multiple years at sites 1, 5, 6, 8, 9, 10, 12, and 17, we detected stable rat populations at 7/8 sites for up to five years: site 1 (Blue group: 2016-2018), site 5 (Light Blue group: 2016-2017), site 6 (Tan group: 2018-2022 and Yellow group: 2021-2022), site 8 (Red group: 2017-2018 and Steel group: 2021-2022), site 9 (Brown group: 2021-2022), site 12 (Green group: 2016-2017), and site 17 (Pink group: 2018-2019); as well as population turnover/invasion followed by persistence at two sites: site 6 (Tan group in 2018 to Yellow and Tan

245 groups from 2020-2021) and site 8 (Red group from 2017-2018 to Steel group from 2021-2022) (**Fig**
246 **1**).
247
248 **Fig 2. *Leptospira interrogans* serogroup Icterohaemorrhagiae whole genome phylogeny**. A
249 maximum likelihood phylogeny of 28 *Leptospira interrogans* serogroup Icterohaemorrhagiae isolate
250 genomes from rats, two publicly available *L. interrogans* serogroup Icterohaemorrhagiae serovar
251 Copenhageni complete genomes (GenBank accession# GCA_002073495.2 and GCA_010978155.1),
252 and 13 enriched genomes (12 rats and 1 human) based upon a concatenated SNP alignment of 127
253 positions out of a core genome size of 2,123,028 nucleotides (nts). The phylogeny was rooted with
254 reference genome *L. interrogans* serogroup Icterohaemorrhagiae serovar Copenhageni strain Fiocruz
255 L1-130 (GenBank accession# GCA_000007685.1). Four major clades were observed (1-4 in white
256 text) and bootstrap values are indicated at specific nodes. The color of the genome ID corresponds to
257 the rat genetic groups in **Fig 1**; black text indicates that no genetic group was assigned. One inferred
258 migrant (indicated with a black arrow) from site 6c was trapped at site 2c and carried *Leptospira*
259 *interrogans* identical to its source collection (see R335-Migrant in clade 1).
260
261 Using Q values ≥0.75, we detected potential *R. norvegicus* migrations as indicated by the
262 trapping of a rat assigning to a distinct genetic group among a collection of rats belonging to a
263 different genetic group; any highly admixed individuals were not treated as migrants due to the
264 potential for mis-assignment. In total, we detected eleven such migration events, seven of which were
265 adult rats (5 females, 2 males) and four were juveniles (1 female, 3 males). Rat movement distances
266 ranged from ~0.8-5.3 km. Migrations were suspected from collection 17c or 17d to collection 14d (1
267 juvenile female, Q =0.93); collection 6c to collection 2c (2 juveniles males, Q =0.91 and 0.95, and 1
268 adult male, Q =0.76); collection 13d, 14d, or 16d (all belonging to the Evergreen genetic group) to

collection 15d (1 adult female, Q =0.81); collection 5a or 5b to collection 12a (1 adult female, Q =0.92); collection 9g1, 9g2, or 10c (Brown genetic group) to collection 6g1 (1 adult male, Q =0.76); collection 15d to collection 10d (1 adult female, Q =0.97); collection 12a or 12b to collection 5a (1 juvenile male, Q =0.82); collection 8b or 8c to collection 1b (1 adult female, Q =0.83); and collection 13d, 14d, or 16d (Evergreen genetic group) to collection 7d (1 adult female, Q =0.87) (**Fig 1 and S3 Table**). Two migrants (R335 from collection 2c and Cad89 from 6g1) were *lipL32* positive (**Fig 1 and S3 Table**). Some suspected migrations could also result from home range overlap among neighboring collections.

Population genetic analyses: genetic structure

Principal Coordinates Analysis (PCoA) generally supported the 12 *R. norvegicus* genetic groups identified by STRUCTURE and indicated two genetic groups (Blue and Steel) may be more genetically isolated from the other ten groups (**S2 Fig**). The collections comprising these two genetic groups are geographically separated from each other and the other genetic groups by major roadways (**Figs 1 and 3).** This PCoA analysis of standardized genetic distances mirrors the geographic distribution of these genetic groups (**Fig 1)**, suggesting a relationship between *R. norvegicus* genetic and geographic distance in this study system.

Fig 3. *Rattus norvegicus* trap locations. Map indicating 64 trap locations (white circles) at 17 sites (numbers), likely dispersal barriers (black bars), and major open spaces (green shading). Circle size corresponds to the number of total rat samples collected from each trap location. This map was created using ArcGIS software by Esri. ArcGIS and Arc-Map are the intellectual property of Esri and are used herein under license. Copyright Esri. All rights reserved. For more information about Esri

software, please visit www.esri.com. Basemap: Light Gray Canvas Base v2 https://www.arcgis.com/home/item.html?id=291da5eab3a0412593b66d384379f89f.

Fairly high estimates of global F_{ST} were observed in the FSTAT analysis, with θ =0.184 (95% CI =0.155, 0.214). Of 66 pairwise F_{ST} estimates among 12 *R. norvegicus* genetic groups, all were significant at a =0.05 and only 16/66 were not significant at the Bonferroni corrected a =0.000758. That said, pairwise F_{ST} values were high among the 66 pairwise comparisons, ranging from 0.0826-0.2922 (**S4 Table**). All significant p-values (16/16) were associated with the Purple, Red, Orange, and Brown genetic groups, a result in agreement with the pairwise analysis of similarities (ANOSIM) (see below). Likewise, an Analysis of molecular variance (AMOVA) generated a moderate yet significant PhiPT value of 0.312 (p <0.001), revealing that ~31% of the molecular variance in this dataset was present among *R. norvegicus* genetic groups (*i.e.,* intersite variation), whereas 69% was present within them (*i.e.,* intrasite variation).

Population genetic analyses: isolation by distance and fine-scale genetic structure

To explore signatures of isolation by distance, we used the RELATE function in PRIMER v5.2.9 and the standardized genetic distances and geographic locations (latitude and longitude) of the 247 individual *R. norvegicus* forming the 12 genetic groups (**see Materials and Methods**). The Spearman's rank correlation coefficient (Rho) was calculated as 0.232 (p <0.001), identifying a statistically significant relationship between *R. norvegicus* genetic distance and geographic distance. The global ANOSIM R statistic was 0.921 (p <0.0001), documenting that some of the 12 genetic groups formed distinct geographic groups. Most of the pairwise ANOSIM comparisons among the 12 genetic groups (61 of 66) were significant even with a very conservative Bonferroni corrected a =0.000758, indicating that most of the 12 genetic groups formed distinct geographic groups. Only

316 5/66 pairwise comparisons were not significant at *a* =0.000758 (Brown-Orange, Orange-Purple,
317 Purple-Red, Pink-Purple, and Purple-Evergreen), suggesting the Purple group has a genetic
318 connection with three other groups (Red, Pink, and Evergreen).

319 Consistent with the signal of *R. norvegicus* genetic isolation by distance across all sites,
320 microsatellite data revealed clear evidence of fine-scale spatial genetic structure within collection
321 sites. First, the spatial autocorrelation analysis yielded correlation values significantly greater than
322 zero in the first six distance classes (0-100 m out to 0-600 m), suggesting that, on average, relatives
323 are generally living in close proximity (**S3 Fig**). This is supported by the significantly high value of
324 average relatedness of individuals within collections, at 0.313 (95% CI = 0.268, 0.357), which
325 indicates relatedness ranging from half (0.25) to full (0.5) siblings (**S4 Table**). Therefore, dispersing
326 *R. norvegicus* would need to travel >600 m, on average, to move beyond sites occupied by relatives.
327 Finally, F_{IS} values were only positive in three genetic groups [Blue (0.11), Yellow (0.078), and Red
328 (0.006)], suggesting inbreeding within collections at sites 1 and 2 (Blue), and perhaps site 6 (Yellow)
329 (**Fig 1 and S4 Table**).

330

331 *Leptospira* genomics and genotyping

332 In total, 52 of 59 *lipL32* PCR positive *R. norvegicus* and one of two positive human samples
333 were assigned to *L. interrogans* (86.9%) using the genotyping methods described herein (**S2 Table**).
334 Whole genome analyses revealed all isolates derived from frozen or fresh kidneys (n=28) and all
335 successful enrichments (n=15) were *L. interrogans* (**S2 Table**) and most like other serogroup
336 Icterohaemorrhagiae serovar Copenhageni/Icterohaemorrhagiae genomes (**S4 Fig**). Enriched
337 genomes for samples R47 and R316, which were included as genomic analysis controls, were
338 identical to their isolate genome counterparts (**Fig 2**). Phylogenomic analysis of all 28 isolates and 12
339 enrichments [the latter selected due to average breadth and depth of coverage values >63% and >18x,

340 respectively (**S5 Table**)], revealed four distinct clades that largely associate with the genetic groups
341 of the *R. norvegicus* from which they were derived (**Fig 2**). For example, 14 of 15 *L. interrogans*
342 genomes assigning to clade 1 were obtained from rats belonging to the Yellow, Tan, and Green
343 genetic groups, all occurring in or near Boston Common Garden, and the Tan and Yellow groups co-
344 occur at site 6. However, the Tan group, the most prominent in Boston Common Garden in 2018,
345 was replaced or modified as indicated by the emergence of the Yellow group in 2021 and 2022.
346 Likewise, clade 3 contains mostly *L. interrogans* genomes from the Red and Steel genetic groups (6
347 of 7). Both genetic groups contain rats from the same site (8) but different collections and years (8b
348 and 8c =Red, 8f and 8fg =Steel), perhaps indicating the removal of one rat genetic group (Red) and
349 replacement with another (Steel). Similarly, clade 2 contains mostly *L. interrogans* genomes from
350 rats belonging to the Evergreen genetic group (3 of 5); this clade also contains the *L. interrogans*
351 genome derived from a human leptospirosis case in 2018 (MA-2). Finally, clade 4 contains only *L.*
352 *interrogans* genomes from rats belonging to the Pink genetic group, comprised of the most
353 geographically isolated collections (17c and 17d) (**Fig 1**). Our phylogenetic placement of two lower
354 coverage enriched genomes (R289 and R292; **see S5 Table**) placed them into clade 1 among other *L.*
355 *interrogans* genomes derived from rats from collection 6c and belonging to the Tan genetic group
356 (**S5 Fig**).

357 Our higher resolution [3,280,516 nucleotides (nts)] phylogenomic comparison of *L.*
358 *interrogans* clade 2 revealed the human derived genome (MA-2) from a 2018 leptospirosis case
359 differed by only two single nucleotide polymorphisms (SNPs) from three *L. interrogans* genomes
360 obtained from rats collected in 2019, 2021, and 2022, respectively, when comparing a core genome
361 of >3.2 mb (**Fig 4**). The four *L. interrogans* genomes comprising this clade were from rats belonging
362 to the Evergreen genetic group or admixed group and were collected at sites 9 (collections 9f and
363 9g1) and 16 (collection 16d) (**Fig 1 and S1 Table**).

364

365 **Fig 4. *Leptospira interrogans* serogroup Icterohaemorrhagiae clade 2 whole genome phylogeny.**
366 Maximum likelihood phylogeny of four *R. norvegicus* and one human derived *L. interrogans*
367 genomes from clade 2 (**see Fig 2**) based upon the concatenated SNP alignment of 33 positions out of
368 a core genome of 3,280,516 nucleotides. The color of the genome ID corresponds to the rat genetic
369 group in **Fig 1**; black text indicates that no genetic group was assigned.

370

371 Our phylogenomic analysis of 24 whole genome assemblies representing all 18 P1 clade
372 pathogenic *Leptospira* spp. placed the 24 assemblies into 18 unique species clades, as expected (**S6A**
373 **Fig**). All nine isolate genomic DNAs subjected to AmpSeq analysis were accurately placed into their
374 respective species clades (**S6B Fig**), which were conserved in the AmpSeq phylogeny despite the
375 large reduction in size of the SNP alignment used to infer it (WGS =53,132 vs. AmpSeq =375 nts)
376 and the core genome included in the analyses (WGS =167,849 vs. AmpSeq =1,195 nts) (**S6 Fig**). We
377 note the proportion of variable sites (variable sites/core genome) represented in both the whole
378 genome and AmpSeq phylogenies are essentially equal (WGS =0.317 vs. AmpSeq =0.314), which
379 likely accounts for the similar branch lengths separating the species clades. The average number of
380 reads that mapped for each isolate was 44,994 ± 12,348 and the number of AmpSeq loci that
381 amplified for each isolate ranged from 22 for *L. kmetyi* to 39 each for *L. interrogans* and *L.*
382 *santarosai* (**S6 Table**).

383 Of the 20 *R. norvegicus* and one human sample subjected to AmpSeq analysis, 13 rats yielded
384 adequate PCR amplification and sequencing coverage (≥7 loci with >10x reads) of *Leptospira* loci to
385 confidently assign a species classification of *L. interrogans* (**S6 Table**). Nine displayed enough
386 coverage (≥25 loci with >10x reads) to assign within species phylogenetic placement to the *L.*
387 *interrogans* serogroup Icterohaemorrhagiae serovar Copenhageni/Icterohaemorrhagiae clade (**S7**

388 **Fig**). The remaining seven rats and one human sample were not assigned a species classification due
389 to inadequate PCR amplification and/or sequencing coverage (≤2 loci with >10x reads); two rats
390 (R66 and R108) had no mapped loci with >10x reads, two (R324 and R379) had one mapped locus,
391 and four samples (Cad1, MAR_36, MAR_5 and MA-1) had two mapped loci (**S6 Table**). Regardless,
392 the few loci that did amplify for samples R324, R379, Cad1, MAR_36, MAR_5, and MA-1 all
393 mapped to *L. interrogans* references except for MAR_5 and Cad1, which had one and two loci,
394 respectively that mapped to *L. borgpetersenii* (**data not shown**). The *L. interrogans* serovar
395 Copenhageni strain Fiocruz L1-130 gDNA control amplified robustly at all 42 loci (1,444,384
396 assigned reads in total) and was accurately phylogenetically placed among other *L. interrogans*
397 serovar Copenhageni strains (**S7 Fig**).
398

399 # Discussion

400 Our population genetic analyses of *R. norvegicus* in Boston revealed the existence of distinct
401 rat populations for periods of up to five years, most likely due to low dispersal rates among rat
402 populations and inbreeding. Our analyses revealed significant relatedness among individuals within a
403 genetic group and, furthermore, that related individuals are generally living within close proximity to
404 each other (**S3 Fig**), an observation consistent with other studies of *R. norvegicus* home ranges,
405 dispersal, and site fidelity in urban settings (6, 7, 14, 17, 32-35). In an analysis of gene flow among
406 *R. norvegicus* in two US, one Canadian, and one Brazilian city, Combs et al., found that rats were
407 highly related to each other within 500 m (12). Our findings in Boston were similar, as rats would
408 need to travel, on average, >600 m to encounter an unrelated individual. Although dispersal events
409 were observed in this study (**Fig 1 and S1 and S3 Tables**), they were rare [only 11/311 rats (3.5%)
410 were presumptive migrants], and this low rate is consistent with the observed patterns of strong
411 genetic structure and stability of genetic groups over time (**Fig 1**).

412 The 11 instances of inferred migrations included seven trapped adults (5 females, 2 males)
413 and four juveniles (1 female, 3 males) (**S1 and S3 Tables**). Rat migration distances ranged from
414 ~0.8-5.3 km in Boston, but because of the methodology we used to assign rats to specific collections
415 for population genetic analyses (**see Materials and Methods**), short-distance migrations (< 0.5 km)
416 would only be detected among adjacent sites separated by major roads (*e.g.,* sites 8 and 9 in **Fig 3**).
417 Although short-distance migrations are thought to be more common (6, 7, 36), long-distance
418 migrations of up to 11.5 km have been documented in brown rats within urban settings (14) and
419 could be anthropogenically facilitated (6, 37). We detected a juvenile female that migrated over 5.3
420 km from site 17 (Pink group) to site 14 (Evergreen group) (**Fig 1**). Since both sites are connected by
421 a large road and subway line, and adjacent to the same large river, it is possible human mediated
422 transport on a car, truck, train, or water vessel facilitated migration. We also detected migration of
423 three males (2 juveniles, 1 adult) from site 6 (Boston Common Garden) to site 2 in 2017 (**Fig 1 and**
424 **S1 and S3 Tables**). These migration events were <1 km and all three rats (R335, R336, and R337)
425 were collected on the same day (30-Nov-2018) at the same trap location (Government Center-1) (**S1**
426 **and S2 Tables**). We are unsure if this finding represents a group migration or if it is the result of a
427 common dispersal throughway between sites 6 and 2, or some other unknown dispersal mechanism.
428 To our knowledge, group migrations have not been described previously in the literature. In general,
429 our observed rate of inferred migrations (3.5%) was similar to other studies of *R. norvegicus*
430 population genetics in urban ecosystems; Gardner-Santana et al., reported a migration rate of 4.7% in
431 Baltimore, US (14), and Kajdacsi et al., reported 6.8% in Salvador, Brazil (15).
432 Major roads were likely the largest contributor to low rates of migration of *R. norvegicus*
433 among neighboring sites in Boston (**Figs 1 and 3**). We observed several examples of major roads
434 acting (and smaller roads failing to act) as dispersal barriers among three intensely sampled sites (6,
435 8, and 9). We compared the dispersal dynamics within Boston Common Garden (site 6: sampled

436 2016-2022) that spans ~0.83 km and is bisected by a 3-lane road with those between Melnea Cass
437 Blvd SE (site 8: sampled 2017-2022) and Melnea Cass Blvd N (site 9: sampled 2021-2022), which
438 are directly adjacent to each other but separated by an 8-lane road (**Figs 1 and 3**). The 3-lane road
439 bisecting the northeastern and southwestern portions of Boston Common Garden did not act as a
440 dispersal barrier for *R. norvegicus* as both the Tan and Yellow groups were detected ubiquitously
441 throughout the park. In contrast, the 8-lane road separating sites 8 and 9 prevented admixing among
442 them and migration between them (**Fig 1**), despite the sites being separated by only ~0.25 km.
443 Roadways have frequently been described as significant barriers to rat dispersal (6, 7, 17, 35, 38, 39)
444 and larger roadways deter dispersal more than smaller ones (40). We also observed a pattern of
445 genetic connectivity among sites separated by >1 km (see Evergreen group; sites 13, 14, and 16 in
446 **Fig 1**) but connected by a large park complex containing green spaces, open water, and a network of
447 river underpasses (**Figs 1 and 3**). Taken together, these results suggest different dispersal dynamics
448 in commercial/residential areas (shorter movements) vs. park settings (longer movements) in this
449 urban ecosystem, likely associated with road size and access to dispersal throughways. These
450 observations are in concert with previous findings of *R. norvegicus* movement patterns in urban
451 ecosystems (6, 12, 13, 41).

452 *Rattus norvegicus* population turnover occurred at two sites (6 and 8), followed by robust
453 expansion of the new/invading population (**Fig 1**). In Boston Common Garden we observed a large
454 population in 2018 (collection 6c: Tan group) that was mostly displaced by the emergence of the
455 Yellow group in 2021 (collection 6f). Both genetic groups remained (6g1 and 6g2) at the conclusion
456 of sampling in 2022 although the Yellow group was most often detected in 2021 and 2022 (**Fig 1B**).
457 We also observed genetic signatures of admixture among the Tan and Yellow genetic groups
458 (individuals assigning to both groups or neither) in the later collections (6f and 6g2 in **Fig 1B**), which
459 could suggest these genetic groups are in the process of merging. Complete population turnover

460 occurred at site 8 between October 2018 (Red group) and June 2021 (Steel group) (**Fig 1**). A rat
461 abatement campaign took place at site 8 but did not begin until 5-Nov-2021, suggesting population
462 turnover from the Red to Steel group occurred before that campaign began. Furthermore, the Steel
463 group persisted throughout the campaign (collections 8fg and 8g in **Fig 1**). It is unknown what
464 occurred at site 8 between October 2018 and June 2021 that led to the population change, but our
465 results suggest that rat abatement is unlikely to lead to population turnover unless complete
466 eradication can be achieved, as populations can rapidly rebound to pre-intervention sizes once
467 interventions cease (42-45). However, if an intervention resulted in removal of dominant individuals
468 followed by immigration of other individuals into this open niche, population turnover could occur
469 (7, 42). We suspect an open niche once recolonized would result in the rapid expansion of a new
470 genetic group due to rapid urban rat reproduction rates (up to five litters per year) (7) paired with the
471 tendency of *R. norvegicus* to breed with individuals from their same genetic group, a result of
472 dispersal barriers, site fidelity, and small home ranges (6, 12).

473 *Leptospira interrogans* persists locally in *R. norvegicus* populations in Boston. We detected
474 pathogenic *Leptospira* DNA in 18.0% of rats representing 20/33 collections and 12/17 sampling sites
475 (**Fig 1 and S1 Table**) with 98.3% of infected rats classified as adults. To put these number into
476 context, several other studies that examined *Leptospira* carriage in *R. norvegicus* in urban ecosystems
477 reported 7-82% in the highly endemic tropical climate city of Salvador, Brazil (46-48), 11.1% in the
478 temperate city of Vancouver, Canada (49), and 35.7% in New Orleans, US (50). Furthermore, a 2019
479 review of *Leptospira* carriage in rats globally reported an overall carriage rate in *R. norvegicus* of
480 30.3% (25), with higher carriage rates associated with tropical climates. Of the studies that assessed
481 the relationship between *Leptospira* positivity and age or sexual maturity, all reported significantly
482 higher proportions of positive adults than juveniles (47-49). Using a suite of molecular methods, we
483 determined the *Leptospira* species circulating in *R. norvegicus* in Boston was *L. interrogans* and

484 most similar to other serogroup Icterohaemorrhagiae genomes (**S4 and S7 Figs**). The 28 cultured
485 isolates were confirmed to belong to serogroup Icterohaemorrhagiae via the microscopic
486 agglutination test (MAT) (**S2 Table**), whereas the samples subjected to enrichment (n=14) and
487 AmpSeq (n=9) could only be inferred based upon phylogenetic placement (**Fig 2 and S4, S5, and S7**
488 **Figs**). Detection of only *L. interrogans* serogroup Icterohaemorrhagiae in this rat dataset was not
489 surprising. Indeed, the exclusive detection of *L. interrogans* serogroup Icterohaemorrhagiae in *R.*
490 *norvegicus* from the US was documented in a recent review (25). In that same article, it was also
491 reported that the most frequently identified *Leptospira* in rats globally (in 36 of 43 countries) belong
492 to serogroup Icterohaemorrhagiae. To that point, in the broad phylogenetic framework of *L.*
493 *interrogans,* the genomes from Boston *R. norvegicus* formed a monophyletic clade with four other *L.*
494 *interrogans* serogroup Icterohaemorrhagiae genomes from Saint Kitts (SK1), Brazil (203 and L1-
495 130), and Malaysia (898) (**S4 Fig**); this analysis included diverse *L. interrogans* representing >9
496 serovars. Importantly, monophyletic groupings like this can be teased apart with clade-focused whole
497 genome analysis, which was demonstrated by Santos et al., in a comprehensive study of 67 *L.*
498 *interrogans* serogroup Icterohaemorrhagiae isolates originating from >15 countries across the globe
499 (51). At that scale, they describe the detection of 1,072 SNPs and 258 indels among the 67 isolates,
500 and phylogenetic clusters that broadly associated with geographic locations. Likewise, by using
501 clade-focused whole genome analysis we identified >100 SNPs and four distinct clades among *L.*
502 *interrogans* serogroup Icterohaemorrhagiae genomes from *R. norvegicus* in Boston (**Fig 2 and S5**
503 **Fig**).

504 Distinct clades of *L. interrogans* correspond to specific geographic locations and rat
505 populations in Boston and persist locally through time despite population turnover (**Fig 2**). Rats from
506 sites 6 and 12 exclusively carried *L. interrogans* from clade 1 (**Fig 2 and S5 Fig**), the virulence of
507 which was validated in the hamster model of leptospirosis, despite being separated by >1.3 km and

508 belonging to three different rat genetic groups (Tan, Yellow, and Green); at site 6 the rat population
509 shifted from the Tan to Yellow group between 2018 and 2021 (**Fig 1**). Furthermore, collections from
510 sites 6 and 12 spanned seven years (2016-2022) (**S1 Table**). This finding suggests connectivity
511 among these sites and/or that stable rat populations at these sites maintain endemic clades of *L.*
512 *interrogans* over time. A greenbelt containing a public walkway connects these two sites almost
513 directly, and although eight roads bisect this greenbelt, none exceed 4-lanes and six are only 2-lanes,
514 suggesting movement between these sites is possible (40). It is also important to consider that
515 although these two sites contain three distinct genetic groups, the Green group from site 12 was
516 sampled in 2016 and 2017, whereas the Tan and Yellow groups from site 6 were sampled in 2018-
517 2022. It is therefore possible that the Green group no longer existed beyond 2017 and that if rats were
518 trapped at site 12 in 2018-2022, they would resemble the Tan and Yellow genetic groups, or –
519 alternatively – that the Green group was present at site 6 prior to 2018. A single rat was trapped at
520 site 6 in 2016 (Cad18) and although it was not assigned to a genetic group due to Q <0.75 (*i.e.,*
521 admixed), the most probable assignment was to the Green group with Q =0.358 (**S3 Table**). Under
522 either scenario, the exclusive detection of clade 1 *Leptospira* at sites 6 and 12 over a period of seven
523 years suggests a connection between them. Likewise, *L. interrogans* from clade 2 was exclusively
524 found in rats from sites 9 and 16 (**Fig 2 and S5 Fig**). These sites were separated by nearly 2.5 km and
525 the samples were collected over different years; the site 16 rats were trapped in 2019 and the site 9
526 rats in 2021-2022. As with clade 1, we see two distant locations sharing a single *L. interrogans*
527 genotype, suggesting a connection between them, but in this case, there is no obvious dispersal
528 throughway linking them. That said, we see evidence of long-distance migrations throughout this
529 study system (**Fig 1**), so it is plausible that a *Leptospira*-positive migrating rat from site 16 to site 9
530 (during or post 2019) led to the establishment of the clade 2 genotype at site 9 that we observed in
531 2021-2022. Population turnover of the rats with persistence of the same *Leptospira* genotype was

532 also observed at site 8 (**Fig 2 and S5 Fig**) where the rats displayed a genetic change from Red to
533 Steel between 2018 and 2021, yet the *L. interrogans* clade 3 genotype remained the same. Finally, *L.*
534 *interrogans* clade 4 contained only rats from site 17 collected in 2018-2019. Site 17 is the most
535 geographically distant site in this study being >4.5 km away from its nearest geographic neighbor,
536 site 16 (**Figs 1 and 3**), so it is not surprising to see an exclusive clade of *L. interrogans* among rats
537 from site 17. These results provide a clearer picture of *Leptospira* transmission among rat
538 populations and provide compelling evidence for local persistence of distinct *L. interrogans*
539 serogroup Icterohaemorrhagiae genotypes among *R. norvegicus* in Boston.

540 *Leptospira* genotypes can be dispersed between populations by migrating rats**.** We detected
541 the migration of a *Leptospira*-positive adult rat (R335) from site 6 to site 2 in 2018 (**Fig 1**) and the
542 genetic signature of the *Leptospira* it carried (clade 1) further supported this assignment (**Fig 2**).
543 Importantly, this event clearly illustrates a mechanism by which *Leptospira* genotypes could be
544 transferred between rat populations. A *Leptospira*-positive migrant (Cad89) that moved from site 9 to
545 site 6 (**S3 Table and Fig 1**) was also detected. Unfortunately, a genome was not obtained from this
546 sample so fine-scale genomic assignment was not possible. However, AmpSeq assigned it to the *L.*
547 *interrogans* serogroup Icterohaemorrhagiae clade (**S7 Fig**). Regardless, it illustrates movement of
548 *Leptospira* via rat migrations between sites 9 and 6 and it is therefore reasonable to assume that
549 migration of a *Leptospira*-positive rat could have occurred in the opposite direction (site 6 to 9),
550 ultimately resulting in a rat trapped at site 9 carrying a *L. interrogans* clade 1 genotype. We also
551 observed single rat populations that carried diverse *L. interrogans* that assigned to multiple locations
552 in the phylogeny (*e.g.*, 15d, 9g1, 8f) (**Fig 2**), suggestive of movement of *L. interrogans* between rat
553 populations. For example, a single rat trapped at site 9 (Cad91) contained *L. interrogans* from clade 1
554 (**Fig 2**). Although this rat was not assigned to a genetic group the most probable assignment was to
555 the Brown group (Q =0.589) (**S3 Table**). We see evidence of rat mediated *L. interrogans* movement

556 at site 8 where three *L. interrogans* genotypes were recovered (**Fig 2**). Because the clade 3 genotype
557 was the most common at site 8 and maintained over three years despite the turnover of the rat
558 population from the Red to Steel group, we suggest the clade 3 *L. interrogans* is established at that
559 site, whereas the other two genotypes (n=1 each) were the result of *Leptospira*-positive rat migrations
560 into site 8. Finally, we found clade 3 *L. interrogans* in a rat from site 13 that clearly assigned to the
561 Evergreen genetic group (**Fig 1**); clade 3 genotypes were otherwise only found in rats from site 8
562 (**Fig 2**). Taken together, these data provide strong support for *R. norvegicus* migrations being a
563 mechanism by which *L. interrogans* serogroup Icterohaemorrhagiae genotypes are transferred among
564 rat populations in Boston, an important consideration when designing rodent control interventions or
565 construction, as both can drive long-distance migrations and intraspecific antagonism (42) and
566 human interventions may increase rat population pathogen prevalence (52).

567 The mechanism(s) by which identical *Leptospira* genomes becomes established in rat
568 populations at these sites is unclear. *Rattus norvegicus* are well known chronic and asymptomatic
569 carriers of *L. interrogans* serogroup Icterohaemorrhagiae (20-22), and the transmission cycle via
570 carriage of *Leptospira* in renal tubules, excretion through urine, environmental contamination,
571 ultimately leading to direct or indirect exposure to other rats as well as other susceptible mammals,
572 including humans, is well established (24). During chronic carriage in the rat, one might expect to
573 observe the accumulation of mutations over time corresponding to an inherent *L. interrogans*
574 mutation rate (to our knowledge, this is unknown). Indeed, we do observe this pattern across certain
575 rats (R269, R289, R292 in clade 1; R365, R366 in clade 2; R336-19 in clade 4) (**Fig 2 and S5 Fig**),
576 presumably the result of long-term chronic infection where *L. interrogans* was in a replicative state.
577 More often, however, we observed identical genomes from rats spanning up to a seven-year period,
578 suggesting a dormant or very low replication state of *L. interrogans* in this urban ecosystem. A
579 possible mechanism for this could be biofilm formation (53) in the renal tubules of the rats (54).

580 However, laboratory analysis of *R. norvegicus* infected with *L. interrogans* serovar Copenhageni
581 revealed they shed leptospires continuously (10^5 -10^7 leptospires per mL) in their urine for the
582 duration of the experiment (159 days) (22), suggesting that in this laboratory setting there was no
583 such dormant state. Alternatively, *Leptospira* biofilm formation could take place in the environment
584 (55, 56). It is important to note that because we are only analyzing the shared genomic space among
585 all the rat derived *L. interrogans* genomes (core genome >2.1 million nts) there could be mutations or
586 acquired/lost pan genomic content outside of the core genome that we did not detect. That said, if
587 replication were occurring at a constant rate, we would still expect to observe mutations
588 accumulating in the core-genome. Additionally, when we increased our resolution for clade 2 to >3.2
589 million nts (**Fig 4**), we observed the same identical genomes shared among three rats (MAR-47,
590 Cad86, R367). Furthermore, our genomic analysis of clade 2 also included a human derived *L.*
591 *interrogans* genome (MA-2 in **Figs 2 and 4**). This genome was enriched from urine DNA obtained
592 from a human leptospirosis case in 2018 and only differs by two SNPS (core genome >3.2 million
593 nts) from *L. interrogans* genomes obtained from three rats at sites 9 and 16 in 2019 and 2022 (**Fig 4**).
594 Due to the clonality of these *L. interrogans* genomes derived from two host species [human
595 (incidental) and rat (reservoir)] and collected over a five-year period (2018-2022) paired with the
596 knowledge that rats are recognized globally as carriers of serogroup Icterohaemorrhagiae [while
597 other hosts (*e.g.*, dogs) are incidental carriers], we suggest the most likely explanation for the human
598 infection is exposure to a fomite or the environment that was contaminated by rats from sites 9 or 16
599 or another *R. norvegicus* population carrying the same *L. interrogans* clade 2 genotype in the city.
600 That said, any attempts to link this human case to a specific site should be interpreted with caution
601 given that *L. interrogans* dispersal between sites does occur (**Figs 1 and 2**).
602 This is the first study to pair fine-scale population genetic analyses of *R. norvegicus* in an
603 urban ecosystem with genomic analysis of the *L. interrogans* they harbor. This analysis was possible

604 due to improvements in culturing techniques (described herein) enabling high rates of *Leptospira*
605 recovery (47.5%) paired with advances in culture independent molecular capabilities (28) that
606 together enabled the acquisition of 40 *L. interrogans* genomes from rats (out of 59 *lipL32* PCR
607 positives) and one from a human. Importantly, these genomes were derived not just from fresh
608 samples, but also frozen and archived samples (tissues and DNA), thereby greatly expanding the
609 types of samples that could be included in these genomic analyses. A particularly salient outcome
610 was the acquisition of a human derived *L. interrogans* genome from trace amounts of archived DNA
611 (22.4 ng total). Although not much is known about the origin of this urine sample beyond it
612 originating from a human leptospirosis case in Boston in 2018 due to patient privacy rights, our
613 analysis placed it among rat derived genomes collected exclusively at sites 9 and 16 (**Fig 4**),
614 suggesting the source of infection was localized to a relatively small geographic area within Boston
615 (**Fig 1**) that was accessible to rats from both sites.

616 We have previously applied AmpSeq to assess the genetic diversity of *lipL32* sequences from
617 pathogenic leptospires [original design described by Stone et al., (56)]. AmpSeq is based upon
618 traditional PCR methods, Illumina short-read sequencing (both technologies that are accessible to
619 researchers and clinicians across the world) and requires <10 µL of a *Leptospira*-positive clinical
620 DNA sample. Furthermore, it is relatively inexpensive on a per-sample basis due to the ability to
621 simultaneously sequence hundreds of samples on a single Illumina run. For this study, we
622 implemented and assessed an expanded *Leptospira* AmpSeq panel that is capable of providing
623 *Leptospira* species identification among all 69 currently described and validated *Leptospira* spp.
624 from the P1, P2, S1, and S2 clades (57) (**S1 File**), as well as sub-species phylogenetic placement
625 within *L. interrogans* (*i.e.*, serogroup associated clades) (**S7 Fig**). Although we did not extensively
626 assess sub-species placement among other P1 *Leptospira* spp., we expect a similar outcome for *L.*
627 *kirschneri*, *L. noguchii*, *L. alexanderi*, *L. santarosai*, *L. weilii*, *L. borgpetersenii*, and *L. mayottensis*

628 due to the high number of discriminating loci (<35) expected to amplify in each species (**S1 File**).
629 Furthermore, our phylogenetic analysis of P1 clade pathogenic *Leptospira* spp. via sequences
630 generated with AmpSeq suggests sub-species placement (**S6B Fig**), even with as few as 22
631 amplicons (**S6 Table**). Importantly, the AmpSeq method can identify mixtures of *Leptospira* spp.
632 within a single sample (56), which is an under-characterized aspect of *Leptospira* ecology. Molecular
633 advances such as these will greatly facilitate an increased understanding of mixed infections and their
634 potential role in leptospirosis infection and disease ecology. Thus, we suggest it could serve as an
635 applicable tool for species and sub-species identification of circulating and clinically relevant
636 leptospires in regions throughout the world.
637

638 # Conclusion

639 In this study, we paired population genetic analyses of *R. norvegicus* in an urban setting with
640 genomic analysis of the infecting leptospires they transmit to gain a deeper understanding of the
641 nuanced transmission patterns of *Leptospira* among rats and spillover into humans. We found that
642 rats in Boston displayed strong signals of population structure, likely associated with behavioral
643 norms and dispersal barriers. We also found the *Leptospira* carried by rats in Boston was *L.*
644 *interrogans* and most similar to other serogroup Icterohaemorrhagiae isolates, and that there was
645 fine-scale genetic structure within that phylogenetic group that associated with distinct rat
646 populations. We found evidence that long-distance dispersal events lead to the movement of unique
647 *L. interrogans* genotypes throughout the city, but these events were rare. Our analysis identified
648 highly clonal genomes of *L. interrogans* that persisted within certain populations of rats for up to
649 seven years. Finally, our genomic analysis of *L. interrogans* obtained from a human leptospirosis
650 case using DNA capture and enrichment strongly suggests a link to rats as the source and provides a
651 likely geographic origin within Boston where this transmission event might have occurred, although

652 we cannot rule out the possibility that the same genotype of *L. interrogans* was also carried by rats in
653 an unsampled location in Boston.
654

655 Materials and Methods

656 Study location and rodent trapping

657 The City of Boston, Massachusetts (125.4 sq. km), in the eastern US has an estimated
658 population of ~650,000 people (US Census Bureau 2020 Decennial Census) with a temperate climate
659 (cold winters and hot, humid summers). *Rattus norvegicus* collections between July 2016 and
660 November 2020 were conducted in collaboration with the City of Boston's Inspectional Services
661 Department (COB-ISD), as described previously (10, 11).
662 For rats trapped between 2021-2022, 6-10 traps were placed at each site (see site definitions
663 below) at least 15 meters apart in locations with evidence of recent rat activity (*i.e.*, near burrows or
664 along runs with evidence of fresh rat droppings/urine). Rats were trapped using a commercial lethal
665 snap trap (T-Rex, Bell Labs, Windsor, WI, US) fixed to a remote sensing device (RATMO®, Arctic
666 Systems, Copenhagen, DK) with zip-ties. Traps were baited with meat jerky (Slim-Jim original
667 flavor, Conagra Brand, Chicago, IL, US) and placed inside a weighted, locked commercial bait
668 station (Protecta EVO Express Bait Station, Bell Labs, Windsor, WI, US). When rats were trapped,
669 the remote sensing device alerted the study team in real time for prompt retrieval and reset of the
670 trap. Traps were placed in the evening, attended from dusk until dawn, and deactivated and shut
671 down the following morning. Trapping was conducted in accordance with protocols reviewed and
672 approved by the Cummings School of Veterinary Medicine IACUC (G2020-126).
673 Rat carcasses were stored in a cooler with icepacks until necropsies were performed within
674 six hours of trapping. Rats were weighed, sexed, dipped in 70% ethanol, and placed in dorsal
675 recumbency. Sterile tissue scissors and forceps were used to reflect the abdominal skin and a new set

676 of sterile instruments were used to open the abdomen and aseptically harvest the kidneys. For
677 samples collected in 2021-2022, one fresh kidney was placed into 50 mL conical tubes containing 10
678 mL liquid HAN medium (26) and shipped overnight at ambient temperature to the National Centers
679 for Animal Health (NCAH) in Ames, IA. Archived frozen kidneys (2016-2020) were stored at -80°C
680 until they were shipped overnight from Tufts University to NCAH on dry ice.
681

682 *lipL32* qPCR detection of *Leptospira* in rodent samples

683 We screened for the presence of *Leptospira* DNA in DNA extracted from 210 frozen kidneys,
684 118 fresh kidneys, and one fresh urine sample (kidney was also evaluated from this individual) from
685 328 *R. norvegicus* rats collected in Boston from 2016-2022; four kidneys from bycatch mice and one
686 kidney from a gray squirrel were also included. Kidneys were transferred to a 150 mm petri dish, the
687 capsule (fibrous outer layer of the kidney) removed, and then placed in a 710 mL Whirl-Pak bag
688 (World Bioproducts, LLC, Madison, Wisconsin, US) containing 10 mL of liquid HAN, and
689 immediately macerated manually. DNA was extracted using 500 µL of kidney macerate or 200 µL of
690 urine, and the Maxwell RSC Purefood Purification Pathogen kit (Promega Corporation, Madison,
691 Wisconsin, US). We screened all samples for the presence of pathogenic *Leptospira* spp. using a
692 TaqMan® (Thermo Fischer Scientific, Waltham, MA, US) qPCR assay specific for the *lipL32* gene
693 as previously described, using primer set LipL32-47Fd and LipL32-301Rd with LipL32-189P probe
694 (58-60). To control for qPCR inhibitors, we added TaqMan® exogenous Internal Positive Control to
695 the master mix to confirm DNA amplification, detect false negatives, and qualitatively detect
696 presence of amplification inhibitory substances. All samples were assayed in triplicate and
697 considered positive when duplicate or triplicates were positive with Ct values <40 (58, 60). Using
698 chi-square tests of independence, we examined associations between *Leptospira* positive rats and six
699 sampling variables: sex, maturity (juvenile vs. adult), sampling season [spring (19-Mar to 20-Jun),

700 summer (20-Jun to 22-Sept), fall (22-Sept to 21-Dec), and winter (21-Dec to 19-Mar)], year, site, and
701 rat genetic group (assigned herein).
702

703 Culture

704 Remaining tissue from frozen and fresh kidneys positive by *lipL32* qPCR was used for
705 culture. Macerate from frozen kidneys (200 µL) was inoculated into 5 mL of semisolid T80/40/LH
706 media (61) and incubated at 29°C. Two 10-fold serial dilutions were made from the fresh kidney
707 macerate into liquid HAN media in 3% CO_2 (26) and T80/40/LH media (61) and incubated at 37°C
708 and 29°C, respectively. Cultures were maintained for up to six months and checked periodically by
709 dark-field microscopy for the presence of leptospires.
710

711 Serotyping of isolates

712 *Leptospira* isolates were serotyped by the MAT with a panel of polyclonal rabbit reference
713 antisera representing 13 serogroups: Australis, Autumnalis, Ballum, Bataviae, Canicola,
714 Grippotyphosa, Hebdomadis, Icterohaemorrhagiae, Mini, Pomona, Pyrogenes, Sejroe, and Tarassovi.
715 The serogroup of each isolate was assigned according to the reference antiserum yielding the highest
716 agglutination titer.
717

718 Evaluation of virulence

719 One rodent isolate (designated strain R47) was propagated in liquid HAN medium at 37°C in
720 3% CO_2 and evaluated for virulence by intraperitoneal injection of 10^8 leptospires in a final volume
721 of 1 mL into one group of four female golden Syrian hamsters (*Mesocricetus auratus*). Negative
722 control hamsters were inoculated with media alone. Hamsters were monitored daily for clinical signs

723 of disease, including weight loss, lethargy, bloody discharge from the nose or urogenital tract, and
724 sudden death; and were euthanized when moribund. At time of euthanasia, liver and kidney tissue
725 were harvested for culture and *lipL32* qPCR. These experiments were conducted in accordance with
726 protocols as reviewed and approved by the Animal Care & Use Committee at the NADC and as
727 approved by USDA institutional guidelines.
728

729 *Rattus norvegicus* collection definition

730 *Rattus norvegicus* were trapped over seven years (2016-2022) at 64 trap locations (**S7 Table**)
731 within 17 geographic sites in Boston (**Fig 3**). Site numbers were assigned from east to west and
732 defined as trapping grids separated by ≤0.5 km. Exceptions to this were trapping locations ≤0.5 km
733 apart but separated by major roadways, which are known dispersal barriers for *R. norvegicus* and *R.*
734 *rattus* (6); these were considered separate sites. A large city park complex (~0.83 km northeast to
735 southwest) comprised of two smaller parks [Boston Common (northeast) and Boston Public Garden
736 (southwest)] was considered a single site and is referred to hereafter as Boston Common Garden. We
737 grouped individual rats into 33 collections based upon site ID and sampling date, whereby samples
738 collected >4 months apart were considered unique collections. These criteria were selected based
739 upon the ~3 months needed to reach sexual maturity for *R. norvegicus* (62) and previous work
740 describing genetic relatedness among *R. norvegicus* in urban settings with distances of up to 0.5 km
741 (12). Thus, nomenclature for collections is as follows: site is indicated by numerals ranging from 1-
742 17 ordered from east to west followed by letters corresponding to year of collection where a=2016,
743 b=2017, c=2018, d=2019, e=2020, f=2021, and g=2022. If multiple collections were obtained within
744 a year, a second numeral follows the letter designation (*e.g.,* 9g2 indicates the second collection at
745 site 9 in 2022). These definitions resulted in 33 collections from 17 sites (**S1 Table**).

746 Three sites were sampled intensely between August 2021 and August 2022: Boston Common
747 Garden (site 6), Melnea Cass Blvd SE (site 8), and Melnea Cass Blvd N (site 9) (**Fig 3**). Boston
748 Common Garden is a heavily landscaped urban park system with primarily mowed grass ground
749 cover and concrete foot paths. The southwest and northeast portions are bisected by a 3-lane road,
750 and the southwest portion is the site of a historical municipal burial ground. The two Melnea Cass
751 Blvd sites are heavily developed urban areas of mixed commercial and residential use and separated
752 by an 8-lane road. These three sites were identified in consultation with the COB-ISD as having
753 active rat populations and burrow systems. They were selected for intensive sampling to facilitate a
754 deeper understanding of the relationship between rat population genetics and leptospirosis
755 transmission in parks versus mixed commercial/residential environments, and to assess how a major
756 road may impact dispersal among neighboring rat populations, which we hypothesize will impact the
757 movement of pathogenic *Leptospira*. Furthermore, we used these sites to assess how genetic
758 signatures change over time across stable rat populations (sites 6 and 9) and the effects of a rat
759 control intervention (baiting and trapping) at site 8 that began 5 November 2021.
760

761 *Rattus norvegicus* microsatellite development and genotyping

762 Marker selection, optimization, and multiplexing

763 Twenty-one MSAT markers were utilized for population genetic analyses of *R. norvegicus*
764 samples (**Table 1**). All markers have been previously published (14, 15, 17, 32, 63) and were
765 selected for use in this study based upon: 1) geographic regions where previously utilized (*i.e.*,
766 eastern US cities as a proxy for presumed genetic compatibility with our Boston study site), 2)
767 number of described alleles in other study systems (prioritizing markers with multiple alleles), and 3)
768 successful use among diverse populations of *R. norvegicus* as inferred by inclusion in multiple
769 studies from diverse western hemisphere countries (*i.e.,* US and Brazil). The 21 markers were

770 optimized using a panel of six *R. norvegicus* DNA samples from different sites in Boston, one
771 Sprague Dawley rat gDNA control (Zyagen, San Diego, CA, US), and a non-template control
772 (molecular grade water).

773 We generated multi-locus genotypes for 328 rat kidney DNAs using these 21 MSAT markers
774 amplified in nine PCR reactions (six multiplexes and three singles, **Table 1**). All PCRs were carried
775 out in 10 µL volumes containing the following reagents (given in final concentrations): 10-20 ng of
776 DNA template, 1x PCR buffer, 2.5 mM MgCl2, 0.2 mM dNTPs, 0.8 U Platinum Taq polymerase
777 (Invitrogen, Carlsbad, CA, US), and 0.1-0.5 µM of each primer pair (see **Table 1**). Thermocycling
778 used the following conditions: 10 minutes at 95°C to release the Platinum Taq® antibody, followed
779 by 38 cycles of 60 s at 94°C, 30 s at the annealing temperature (Ta) and 30 s at 72°C. The Ta,
780 dilution, multiplex, and pooling schemes for each marker are provided in **Table 1**. Diluted PCR
781 products were electrophoresed on an ABI 3130xl sequencer with LIZ®-1200 size standard (Applied
782 Biosystems, Foster City, CA, US) and analyzed using GeneMarker v2.6.7 (SoftGenetics, State
783 College, PA, US).

784 **Table 1. *Rattus norvegicus* MSAT marker primers and conditions.**

Marker	T_a	Mix	Primer pair conc. (µM)	Post-PCR dilution	Dye	*A*	Allele sizes (bp)	Citation and primer sequence (5'-3')
D12Rat76	60	1	0.5	1/50	NED	10	86, 88, 90, 92, 94, 96, 98, 100, 102, 104	[1] F-TGCCTTTTAAAATGATGTGCA R-ATTGGCAATGCACTCATGTG
D1Wox23	60	1	0.3	1/50	VIC	10	186, 190, 194, 198, 202, 206, 210, 214, 218, 222	[2] F-TCTGACCCATACTTGTACTTTGC R-AATTTCTGCCTCTTTTTCTCAG
D7Rat97	60	1	0.3	1/50	6FAM	11	172, 174, 176, 178, 180, 184, 186, 188, 192, 196, 198	[3] F-CAAGTTTTCCTCTGCCCAAG R-GCTGTCATTCCACTGGGTTT
D11Mgh5	60	2	0.4	1/75	6FAM	9	216, 226, 228, 230, 232, 234, 236, 238, 248	[1] F-CAGCTCTAATTCCAGAAAGGTTT R-GAATCGATTGACAGATGTCTGTG
D3Cebr3	60	2	0.3	1/75	PET	5	163, 165, 167, 177, 181	[4] F-CAGGGAATGCAGAAGATACAG R-GTGGCTTTAGGACTCTGGAG
D6Cebr1	60	2	0.4	1/75	NED	7	220, 224, 226, 230, 232, 234, 236	[4] F-TGGTTTGGTTGGGGAGAA R-GTGCTGTCAGGGAAAGATGTA
D4Wox7	55	3	0.1	1/50	VIC	5	130, 134, 138, 142, 146	[2] F-GATAGCATAAAATCCCTAGAGGTT R-TCGATTTATCTGAAACCATCAC
D6Wox2	55	3	0.3	1/50	6FAM	10	96, 100, 104, 108, 116, 120, 124, 128, 132, 136	[2] F-CCAGTCCATACTTATCCATCTG R-CATTTAGATAGGTGATAGATTCAG

D1Cebr9	55	3	0.3	1/50	6FAM	6	261, 263, 265, 267, 269, 271	[4] F-GGATTTGGCTCCCTTTAAG R-CAGTAACTCTGGTTCATGTACTCC
D7Rat13	64	4	0.4	1/25	NED	12	128, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164	[5] F-GACTTCTGCTACACGCCACA R-CAGCCCTAGAAGGAAATGCA
D5Rat83	64	4	0.4	1/25	VIC	11	164, 168, 170, 178, 184, 186, 190, 192, 194, 196, 200	[5] F-ACTTGGAAACAGGGAGATGG R-GGGTCTTCAGGATGGCAATGT
D12Wox1	64	4*	0.4	1/50	6FAM	8	392, 396, 400, 404, 408, 412, 416, 420	[2] F-GACATTAAGGGGTCTTCCTAAG R-TATCTTTGCAACGCTGAGG
D10Rat20	64	5	0.4	1/25	PET	6	109, 113, 115, 117, 119, 123	[1] F-AGTGATTGCCATACCTGCCT R-GAAATGGCCAGGATAAACCA
D20Rat46	64	5	0.4	1/25	6FAM	13	140, 142, 144, 148, 154, 156, 160, 162, 164, 168, 172, 174, 180	[1] F-AAGTACTGAGTGGGCTGCGT R-GGCAAAACACCAATGCCTAT
D10Mit5	64	5	0.2	1/25	NED	9	135, 137, 139, 141, 143, 145, 147, 149, 151	[1] F-TGCTGGGTGAACCAGAGAG R-CTGCCCTCCAAACCACC
D9Rat13	64	5	0.4	1/25	VIC	5	101, 109, 119, 121, 123	[5] F-CCCATCTTTACACCTCCCAA R-GGAAAGGAAACTGGAGGGTC
D19Wox11	66	6	0.3	1/15	NED	5	212, 216, 220, 224, 228	[2] F-CTACCCACCCATCTATTCATCC R-GTTTCCAGCACCCATGTCC
D1Cebr3	66	6	0.3	1/15	VIC	6	79, 81, 83, 105, 107, 109	[4] F-CTTGGGAGCTGGGAGTGT R-GAAGGCTGAGGTATGAAGACTG
D16Rat81	66	6	0.4	1/15	PET	7	148, 150, 152, 154, 156, 158, 160	[1] F-GAGCCTTAGCACAGTGGCTT R-GGCCCACATGTGCATGTATA
D5Rat33	66	6*	0.4	1/15	6FAM	15	102, 106, 110, 112, 116, 118, 120, 122, 124, 126, 130, 132, 134, 138, 140	[3] F-TGGAGAAAAGAAGAACCTCCA R-GTGCCCTCAGACTGAACTC
D1Cebr4	66	6*	0.4	1/30	VIC	8	278, 284, 286, 288, 290, 292, 294, 296	[4] F-GACCTCCTGCCCTTCACTG R-TGAAAAATGAATTGCTTGTG

[a][1] (64), [2] (65), [3] (66), [4] (67), [5] (63).
[b]Additional single PCRs are pooled into a corresponding mix (*i.e.*, multiplex PCR) prior to capillary electrophoresis. $\boldsymbol{T_a}$ = PCR annealing temperature. ***A*** = number of observed alleles in Boston, MA *Rattus norvegicus* (2016-2019 and 2021-2022).

MSAT sample data filters

Our final MSAT dataset included only rats with complete genotypes at ≥19 markers; one individual sample failed to meet this criterion and was removed. We also removed three individuals with evidence of DNA cross contamination as indicated by the presence of >2 alleles at multiple markers. Finally, location provenance was incomplete for 13 individuals (including all eight rats collected during 2020), which were also removed. This resulted in a final MSAT dataset of 311 *R.*

796 *norvegicus* representing 33 collections from 17 sites collected between 2016-2019 and 2021-2022
797 (**S1 Table**).
798

799 ## MSAT marker validation analyses

800 We ran a series of tests on the 21 MSAT markers to validate neutrality and utility in
801 population genetic analyses of *R. norvegicus* from Boston using all 311 individuals organized into 33
802 collections. Validations included a query for identical multi-locus genotypes in GenAlEx (68) and a
803 test for linkage disequilibrium in FSTAT v2.9.4 based on the log-likelihood ratio G statistic (69). We
804 used MICRO-CHECKER (70) to test for null alleles, large allelic dropouts, and scoring errors due to
805 PCR stuttering using the following settings: maximum expected allele size of 420 bp, a 95%
806 confidence interval (CI), and 1000 repetitions. We performed a test of Hardy-Weinberg equilibrium
807 for each marker, also in GenAlEx, but using only collections with ≥20 rats (*i.e.*, 5b, 6c, 6g2, and 8f,
808 **see S1 Table**). Finally, we ran PCR replicates on a subset of 94 *R. norvegicus* DNAs (30.2%) from
809 the final dataset to check for allele scoring errors.

810

811 ## Population genetic analyses: determination of genetic groups and population
812 ## assignment using STRUCTURE

813 We determined the most likely number of genetic groups and conducted population
814 assignment analyses using STRUCTURE v2.3.4 analysis (71) on the MSAT genotypes obtained from
815 the 311 *R. norvegicus* DNAs. We set Bayesian parameters to assume genetic admixture and
816 correlated allele frequencies in the study system, and to ignore collection sites. These parameters
817 were run at *K* values from 1–15, with 25,000 burn-in iterations followed by 100,000 run iterations.
818 Bayesian analysis was repeated ten times for each *K*. Resulting log(Prob K) values were analyzed

819 with the delta-*K* method (72) to estimate the most likely number of genetic populations. We assigned
820 rats to a specific genetic group and identified potential migrants if their probability of membership
821 (*Q*) to a specific genetic group was ≥0.75. Individual rats not meeting these criteria were grouped
822 broadly into an "admixed" group, regardless of collection site (**S3 Table**). Using these results, we
823 sorted all 311 rats according to their predicted genetic groups [12 distinct genetic groups (n=247) and
824 one admixed group (n=64)] for additional population genetic analyses [~24% of individuals (64
825 admixed and 11 potential migrants, **see Results**) were no longer grouped according to their initial
826 collection site].
827

828 **Population genetic analyses: estimates of genetic structure at two scales**

829 We estimated overall genetic structure among Boston *R. norvegicus* using PCoA and three
830 different estimators of F_{ST}. For PCoA, we used standardized genetic distances among all pairs of
831 individuals in GenAlEx (68). We then conducted *a posteriori* F_{ST} analyses using only individuals
832 from the 12 genetic groups identified in STRUCTURE (n=247) but also removed juveniles (weight
833 <140g) (73) and migrants, which could artificially increase or decrease, respectively, F_{ST} estimates
834 among the 12 predicted genetic groups. Collection 11b was the single exception because only
835 juveniles were trapped and they appeared to form a distinct genetic group (Orange, **see Results**).
836 These filters resulted in a final dataset for this analysis of 119 individual rats among the 12 genetic
837 groups; all genetic groups were represented by ≥4 individuals (up to 19; **S3 Table**). We generated
838 F_{ST} estimates of theta (θ) in FSTAT version 2.9.4 using Weir and Cockerham's method (74) to
839 estimate population structure. Pairwise population F_{ST} values were also estimated in FSTAT and *p*-
840 values for all comparisons were generated using a =0.05. We performed AMOVA on this same
841 reduced set (n=119) using standardized genetic distances in GenAlEx with 999 permutations to
842 estimate PhiPT (an F_{ST} analog). The third F_{ST} estimator was a one-way global ANOSIM in PRIMER

843 v5.2.9 (PRIMER-e, Albany, AK, NZ), conducted on a reduced set of 247 individual rats from the 12
844 distinct genetic groups (the 64 individuals forming the admixed group were removed but migrants
845 and juveniles were included; **S3 Table**); the R statistic was generated using 10,000 permutations.
846 Afterwards, a post-hoc test was conducted to determine the genetic groups contributing to R, wherein
847 pairwise comparisons among all genetic groups were conducted. The global alpha was divided by 66
848 (number of pairwise comparisons) to calculate a conservative Bonferroni corrected a =0.000758.
849

850 ## Population genetic analyses: isolation by distance

851 Because *R. norvegicus* relatives often live in close proximity, we performed a fine-scale
852 analysis of spatial genetic structure (SGS) on all 311 rats using spatial autocorrelation in GenAlEx
853 with options set to SinglePOP, even distance classes (n=20), and 999 permutations. Distance class
854 intervals were increased in 100 m increments (0-100 m, 0-200 m, etc.) up to a maximum of 0-2,000
855 m. We assessed isolation by distance using the RELATE function in PRIMER v5.2.9 and
856 standardized genetic distances and geographic locations (latitude and longitude) of the 247 rats
857 forming the 12 distinct genetic groups (**S3 Table**). We generated Spearman's rank correlation
858 coefficient (Rho) using 10,000 permutations. Inbreeding coefficient F_{IS} was estimated for each
859 genetic group and global relatedness was estimated among all groups using the reduced set of 119
860 rats (no juveniles or migrants) in FSTAT (**S3 Table**).
861

862 # *Leptospira* genomics and genotyping

863 ## Whole genome sequencing of isolates

864 Whole genome sequencing of DNA extracted from 5 mL cultures of each isolate was
865 performed as previously described (75, 76). Briefly, each isolate was propagated in HAN liquid
866 media at 29°C, which was harvested by centrifugation at 10,000 x *g* for 15 min and DNA extracted

867 using the Maxwell RSC Purefood Purification Pathogen kit. For Illumina library preparation, the
868 concentration of reconstituted genomic DNA was determined with a Qubit dsDNA HS assay using a
869 Qubit® 3.0 fluorometer (Invitrogen, Carlsbad, CA, US). Whole-genome sequences were obtained
870 (MiSeq Desktop Sequencer, 2x250 v2 paired-end chemistry and the Nextera XT DNA Library
871 Preparation Kit, Illumina, San Diego, California, US) per manufacturer's instructions.
872

873 **Capture and enrichment of *Leptospira* DNA from rat and human samples**

874 We attempted *Leptospira* genome capture and enrichment from 22 PCR positive rat samples
875 collected between 2017-2019, 20 culture negative and two culture positive (R316 and R47); the latter
876 served as genome analysis controls. We also attempted enrichment on two separate human samples
877 from Boston: one serum from 2017 (MA-1) and one concentrated urine from 2018 (MA-2). These
878 sample DNAs were provided by the CDC and had *lipL32* PCR Ct values of 38.03 and 31.6,
879 respectively. DNA from these 24 samples were subjected to pan pathogenic *Leptospira* DNA capture
880 and enrichment as previously described (28). Briefly, sample DNAs were diluted separately to ~4
881 ng/µL in a volume of 55 µL, sonicated to an average size of 285 bp using a Q800R2 sonicator
882 (QSonica, Newtown, CT, US), and then single indexed libraries were prepared using Agilent Sure-
883 Select methodology (XT-HS kit, Agilent, Santa Clara, CA, US). The libraries (some pooled) were
884 subjected to one or two rounds of DNA capture and enrichment (**S5 Table**) and sequenced on an
885 Illumina MiSeq instrument using either a MiSeq v3 600 cycle (2x300 reads) or v2 300 cycle kit
886 (2x150 reads). Three samples (R309, R334-19, and R379) failed to generate successful sequencing
887 libraries and were excluded from DNA capture and enrichment and Illumina sequencing and six
888 others failed to enrich *Leptospira* DNA and were excluded from sequencing (R66, R67, R108, R318,
889 R324 and MA-1) (**S2 and S5 Tables**). These unsuccessful attempts were associated with low
890 abundance samples [PCR Ct values >38 (n=2)], enzymatic inhibition [failed library preparation

891 (n=3)], or perhaps contained divergent or non-pathogenic species of *Leptospira* or some other
892 genetically similar species not compatible with our pan pathogenic *Leptospira* DNA capture and
893 enrichment system (n=4 including all three *Leptospira*-positive rats from collection 5b). Pooling
894 schemes, rounds of DNA capture and enrichment, and Illumina sequencing kits used for each sample
895 are provided in **S5 Table**.
896

897 **Read classifications, read mapping, SNP calling, and phylogenomics**

898 To estimate the percentage of *Leptospira* reads in enriched sequences and for taxonomic
899 classification of enrichments and isolates, reads were mapped against the standard Kraken database
900 with Kraken v2.1.2 (77). To determine breadth of coverage for enriched genomes, reads were aligned
901 against *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (GenBank accession#
902 GCA_000007685.1) with minimap2 v2.22 (78) and the per base depth of coverage was calculated
903 with Samtools v1.6 (79). Two *Leptospira* phylogenies were built. The first included a diverse set of
904 *L. interrogans* (representing >9 serovars) for the purpose of contextualizing Boston rat and human
905 derived genomic diversity within *L. interrogans*. Single nucleotide polymorphisms were identified
906 among 28 genomes from cultured isolates (GenBank accession numbers provided in phylogeny), 21
907 publicly available *L. interrogans* genomes, and 13 enriched genomes (12 rats and 1 human) by
908 aligning reads against reference genome *L. interrogans* serovar Copenhageni strain Fiocruz L1-130
909 using minimap2 v2.22 and calling SNPs from the BAM file with GATK v4.2.2 (80) using a depth of
910 coverage ≥3x and a read proportion of 0.9. SNPs that fell within duplicated regions, based on a
911 reference self-alignment with NUCmer v3.1 (81), were filtered from downstream analyses. All
912 methods were wrapped by NASP v1.2.1 (82). A maximum likelihood phylogeny was inferred on the
913 concatenated SNP alignments using IQ-TREE v2.2.0.3 with default parameters (83), the "-fast"
914 option, and the integrated ModelFinder method (84); the phylogeny was midpoint rooted. Second, to

915 provide increased genomic resolution within *L. interrogans* and assess fine scale relatedness among
916 the Boston genomes we replaced the 21 diverse *L. interrogans* genomes with two complete, publicly
917 available *L. interrogans* serovar Copenhageni genomes (GenBank accession# GCA_002073495.2
918 and GCA_010978155.1). A maximum likelihood phylogeny was inferred on the concatenated SNP
919 alignments using IQ-TREE v2.2.0.3 as described above, except with 1000 bootstrap replicates; the
920 phylogeny was rooted with *L. interrogans* serovar Copenhageni strain Fiocruz L1-130.
921

922 *Leptospira interrogans* reference phylogeny and WG-FAST placement of enriched 923 samples

924 Two enriched samples (R289 and R292) had lower breadth and depth of coverage than the
925 other enrichments (**S5 Table**) and were inserted into the phylogeny with *WG-FAST* v1.2 (85) using
926 default settings. Reads were simulated from genome assemblies with ART vMountRainier (86) using
927 the command "-p -na -ss MSv3 -l 250 -f 75 -m 300 -s 30". Based on the position of the enriched
928 reads in the complete *L. interrogans* phylogeny (**see Results**), simulated reads were aligned against
929 *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 with minimap2 v2.24 (78) and SNPs were
930 called with GATK v4.2.6.1 (80). A maximum likelihood phylogeny was inferred on the concatenated
931 SNPs with RAxML-NG v. 1.1.0 (87) and rooted with *L. interrogans* serovar Copenhageni strain
932 Fiocruz L1-130.
933

934 Human and rat phylogenomic comparison

935 To provide additional resolution among *Leptospira* genomes derived from four *R. norvegicus*
936 and one human that grouped together within a single *L. interrogans* serovar Copenhageni clade
937 (designated as clade 2, **see Results**), we conducted a phylogenomic analysis using NASP v1.2.1 and

938 IQ-TREE v2.2.0.3 as described above, except *L. interrogans* serovar Copenhageni strain SK1
939 (GenBank accession# GCA_010978155.1) was used as the reference and root.
940

941 # AmpSeq methods

942 We performed highly multiplexed amplicon sequencing (AmpSeq) on two subsets of rat
943 samples that were PCR positive for *Leptospira* but not included (or were unsuccessful) in the above
944 steps for culturing or genome enrichment. The first subset consisted of rats from 2016 and 2020-2022
945 that were PCR positive but culture negative (n=11); we included one culture positive rat (Cad99) to
946 serve as an analysis control. The final subset for AmpSeq contained any 2017-2019 positive samples
947 that failed DNA capture and enrichment (n=8 rats and n=1 human) (**S2 Table**).
948 In total, 42 primer sets were designed and validated for the *Leptospira* AmpSeq system.
949 Thirteen primer sets were either modified from existing primer sets (60, 88-98) targeting genes
950 commonly used for molecular typing or detection of *Leptospira* (*lipl32*, 16S, *secY*, *Loa22*, *OmpL1*,
951 *flaB*, LIC_11108, LFB1, *ligB*), or redesigned from MLST loci (*glmU*, *lipL41*, *icdA*, *sucA*) (99-101) to
952 be compatible with AmpSeq (153-415 bp in length) (**S8 Table**). We also designed 29 new primer
953 sets targeting loci determined, via comparative genomic analyses, to be SNP-rich and highly
954 discriminatory among 778 genome assemblies from 64 pathogenic and non-pathogenic *Leptospira*
955 spp. (**S9 Table**). Assemblies were downloaded from the NCBI RefSeq database (102, 103) using
956 ncbi-genome-download v0.3.1 https://github.com/kblin/ncbi-genome-download/tree/0.3.1 and
957 aligned to reference genome *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 using
958 NUCmer (104), and SNPs were identified using NASP v1.2.0 (82). The SNPs were analyzed using
959 VaST (105) to identify a set of targets maximizing discriminatory power among the assembled
960 genomes. To avoid paralogous sequence variants and loci with highly variable alleles, we only
961 analyzed SNPs located within a genus-wide core genome (106). To maximize amplification of these

962 targets across *Leptospira* species, primers were designed with ambiguous bases for variants that
963 occurred in ≥5% of genome assemblies.

964 To generate a conservative estimate of species coverage and discrimination for this AmpSeq
965 system, we used *in silico* PCR to predict amplification for each of the 42 loci across 69 *Leptospira*
966 spp. genome assemblies listed on the International Leptospirosis Society's resource webpage
967 (accessed 15-Jan-2024; https://leptosociety.org/resources/); these genomes represent all validated
968 pathogenic (P1 and P2) and saprophytic (S1 and S2) *Leptospira* spp. (57, 107). Genomes were
969 downloaded with the ncbi-genome-download tool (https://github.com/kblin/ncbi-genome-download)
970 from the RefSeq database (102) and renamed based on annotation information in the RefSeq record.
971 PCR primers were screened against the 69 *Leptospira* genomes with the -search_pcr2 method in
972 USEARCH v11.0.667 (108) with maximum number of SNP mismatches allowed in each primer set
973 to two, using parameters: "-strand both -maxdiffs 2 -maxamp 500". Predicted genome hits were
974 compiled for each primer set. This *in silico* analysis predicted amplification in all 69 *Leptospira*
975 species. The number of expected loci for each species varied, ranging from a single locus to all 42
976 loci. The pathogenic P1 clade was well represented with predicted amplification of 18-42 loci in all
977 20 species (**S1 File**).

978 AmpSeq was performed on two sample sets. The first set was used to validate the method and
979 consisted of nine genomic DNA extracts obtained from isolates from seven P1 clade *Leptospira* spp.
980 [*L. interrogans* (n=2), *L. kirschneri*, *L. santarosai*, *L. noguchii* (n=2), *L. borgpetersenii*, *L.*
981 *alexanderi*, and *L. kmetyi*]. The second set contained unknown *Leptospira* spp. and consisted of DNA
982 from 20 *Leptospira* PCR positive *R. norvegicus* kidney samples and one human serum sample (MA-
983 1) (**S2 Table**) along with *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (ATCC,
984 Manassas, Virginia US, catalog# BAA-1198D-5), which served as a PCR and analysis control, and
985 also molecular grade water as a non-template negative control. The 42 loci were arranged into four

986 multiplexes and PCRs were carried out in 10 µL volumes containing the following reagents (given in
987 final concentrations): 10-20 ng of DNA template, 1x PCR buffer, 2.5 mM MgCl2, 0.2 mM dNTPs,
988 0.8 U Platinum Taq polymerase (Invitrogen, Carlsbad, CA, US), and 0.1-0.4 µM of each primer pair
989 (**see S8 Table**). Thermocycling utilized the following conditions: 10 minutes at 95°C to release the
990 Platinum Taq® antibody, followed by 38 cycles of 60 s at 94°C, 30 s at 55°C and 30 s at 72°C. PCRs
991 were pooled together in equal amounts in a final volume of 25 µL (6.25 µL of each multiplex) and
992 sequence libraries were prepared for all positive samples as previously described (56, 109). Uniquely
993 indexed sample libraries were pooled together in equimolar amounts and sequenced in parallel on an
994 Illumina NextSeq instrument using a 600 cycle (2x300) P1 kit (Illumina, San Diego, CA, US).
995

996 # AmpSeq analysis

997 ## Read mapping and classification

998 Illumina sequencing reads for each individual sample were analyzed using the Amplicon
999 Sequencing Analysis Pipeline (ASAP; https://github.com/TGenNorth/ASAP), wherein reads from
1000 each locus are mapped against a reference database using Bowtie2 and binned according to alignment
1001 score (110); the number of paired end reads assigning to each reference are then counted. To build a
1002 comprehensive reference database of *Leptospira* species to be used for read mapping and
1003 classification, we implemented two strategies. First, we downloaded reference sequences from
1004 GenBank for the thirteen previously published loci and removed redundant sequences. Second, for
1005 the 29 new loci we bioinformatically extracted each locus from the reference set of 69 *Leptospira*
1006 spp. genome assemblies described above. We first amplified each locus separately (using the same
1007 PCR conditions above for the AmpSeq multiplexes) using gDNA from *L. interrogans* serovar
1008 Copenhageni strain Fiocruz L1-130. PCR products were visualized on a 2% agarose gel to ensure
1009 robust amplification and treated with ExoSAP-IT (Affymetrix, Santa Clara, CA, US) using 1 µL of

1010 ExoSAP-IT and 7 µL of PCR product under the following conditions: 37°C for 15 minutes followed
1011 by 80°C for 15 minutes. Treated products were then diluted (based on amplicon intensity) and
1012 sequenced using forward primer UT1F-acccaactgaatggagc and reverse primer UT2R-
1013 acgcacttgacttgtcttc (universal sequences incorporated into each unique primer pair, **see S8 Table**) in a
1014 BigDye® Terminator v3.1 Ready Reaction Mix (Applied Biosystems, Foster City, CA, US). We
1015 used 10 µL volumes for sequencing reactions containing the following reagents (given in final
1016 concentrations): 5x Sequencing Buffer, 1 µL BigDye® Terminator v3.1 Ready Reaction Mix, 1 µM
1017 primer, and 5 µL diluted PCR product. Thermocycling conditions were: 96°C for 20 seconds,
1018 followed by 30 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Sanger
1019 sequences were assembled, trimmed, and converted to fasta format using SeqMan Pro (DNASTAR
1020 Lasergene, Madison, WI, USA), which were used to extract each locus (if present) from the 69
1021 *Leptospira* spp. genome assemblies using BLASTN (111) and a custom python script
1022 (https://gist.github.com/jasonsahl/2a232947a3578283f54c).
1023

1024 **Whole genome versus AmpSeq phylogenetic resolution**

1025 We compared phylogenetic resolution and tree topology among two maximum likelihood
1026 phylogenies generated, respectively, from whole genome reads and AmpSeq reads. Using 24 genome
1027 assemblies representing 18 P1 clade pathogenic *Leptospira* spp. (GenBank accession numbers
1028 provided in the phylogeny), we conducted a whole genome phylogenetic analysis using NASP v1.2.1
1029 and IQ-TREE v2.2.0.3 as described above with 1000 bootstrap replicates and *L. interrogans* serovar
1030 Copenhageni strain L1-130 as the reference. We then conducted an identical analysis but also
1031 included the AmpSeq generated reads from the nine *Leptospira* isolate genomic DNAs in our
1032 validation set. Inclusion of AmpSeq generated reads reduced the core genome to only nucleotide

1033 positions 1) included in the 42 loci, and 2) shared among the genome (n=24) and AmpSeq (n=9)
1034 reads. Both phylogenies were midpoint rooted.
1035

1036 **Species id and within species genotyping**

1037 Rat and human DNAs that amplified at ≥7 loci (out of 42) with >10x coverage were assigned
1038 a *Leptospira* species classification. To facilitate genotyping within species for samples with high
1039 breadth of coverage (*i.e.,* 25-42 loci amplified with >10x coverage), SNPs were identified from
1040 AmpSeq reads and a reference set of 22 publicly available *L. interrogans* complete genome
1041 assemblies (GenBank accession numbers provided in the phylogeny). AmpSeq reads contain
1042 Illumina adaptors, universal tail sequences, and locus specific primers, all of which must be removed
1043 prior to SNP discovery if precise phylogenetic placement is the analysis goal; adaptors and universal
1044 tails were removed in ASAP and locus specific primers were trimmed using cutPrimers
1045 (https://github.com/aakechin/cutPrimers) (112). Trimmed AmpSeq reads and genome assemblies
1046 were aligned against reference genome *Leptospira kirschneri* serovar Grippotyphosa strain
1047 RedPanda1 (GenBank accession# GCA_027563495.1) using NASP v1.2.1 (82). A maximum
1048 likelihood phylogeny was then inferred on the concatenated SNP alignments using IQ-TREE v2.2.0.3
1049 with the "-fast" option, default parameters (83), and the integrated ModelFinder method (84); the
1050 phylogeny was rooted with *Leptospira kirschneri* serovar Grippotyphosa strain RedPanda1.
1051

1052 # Acknowledgments

1053 The authors extend their gratitude to the City of Boston's Inspectional Services Department and the
1054 Boston Public Health Commission for their assistance. The authors thank the *Mycobacteria* &
1055 *Brucella* Section from the NVSL for their technical assistance in genome sequencing. USDA is an

equal opportunity provider and employer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Funding Disclosure

This research was supported in part by an appointment to the Animal and Plant Health Inspection Service (APHIS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA). ORISE is managed by ORAU under DOE contract number DE-SC0014664. This project was also funded in part by the National Institute of Allergy and Infectious Diseases in the National Institutes of Health (NIH) award number R01AI172924 (Pearson), the National Institute of Environmental Health Sciences under award number K08ES035460 (Rosenbaum), the National Center for Advancing Translational Sciences, NIH under award number KL2TR002545 (Rosenbaum, Hertzel, Benson, Specht), and K-24 AI 106822 (Phipatanakul). Work conducted by JDM was completed while supported by the NSF Postdoctoral Research Fellowships in Biology Program, Award No. 1907311 (Madison).

Competing Interests

The authors declare they have no competing interests.

Author Contributions

Nathan E. Stone: Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing original draft, Writing review and editing

Camila Hamond: Investigation, Methodology, Validation, Writing original draft, Writing review and editing

Joel Clegg: Investigation, Validation

Ryelan F. McDonough: Investigation, Validation

Reanna M. Bourgeois: Investigation, Validation

Rebecca Ballard: Investigation, Validation

Natalie B. Thornton: Investigation

Marianece Nuttall: Investigation

Hannah Hertzel: Investigation

Tammy Anderson: Investigation, Validation, Writing review and editing

Ryann N. Whealy: Investigation, Validation

Skylar Timm: Investigation

Alexander K. Roberts: Investigation

Verónica Barragán: Investigation

Wanda Phipatanakul: Funding acquisition, Investigation

Jessica H. Leibler: Investigation

Hayley Benson: Investigation

Aubrey Specht: Investigation

Ruairi White: Methodology, Investigation, Resources

Karen LeCount: Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing

Tara N. Furstenau: Investigation, Methodology, Writing original draft

Renee L. Galloway: Investigation, Methodology

Nichola J. Hill: Investigation, Methodology

Joseph D. Madison: Funding acquisition, Investigation, Methodology, Writing reviewing and editing

Viacheslav Y. Fofanov: Investigation, Methodology

Talima Pearson: Funding acquisition, Project administration

Jason W. Sahl: Investigation, Methodology, Supervision, Writing review and editing

Joseph D. Busch: Project administration, Supervision, Writing original draft, Writing review and editing

Zachary Weiner: Project administration

Jarlath E. Nally: Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing original draft, Writing review and editing

David M. Wagner: Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing original draft, Writing review and editing

Marieke Rosenbaum: Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing original draft, Writing review and editing

References

1. Lund M. Commensal Rodents. Rodent Pests and Their Control, 2nd Edition. 2015:19-32.
2. Desvars-Larrive A, Pascal M, Gasqui P, Cosson JF, Benoit E, Lattard V, et al. Population genetics, community of parasites, and resistance to rodenticides in an urban brown rat (Rattus norvegicus) population. PLoS One. 2017;12(9):e0184015.
3. Harper GA, Bunbury N. Invasive rats on tropical islands: Their population biology and impacts on native species. Glob Ecol Conserv. 2015;3:607-27.
4. Kosoy M, Khlyap L, Cosson JF, Morand S. Aboriginal and Invasive Rats of Genus *Rattus* as Hosts of Infectious Agents. Vector-Borne Zoonot. 2015;15(1):3-12.

1131 5. Meerburg BG, Singleton GR, Kijlstra A. Rodent-borne diseases and their risks for public health. Crit
1132 Rev Microbiol. 2009;35(3):221-70.
1133 6. Byers KA, Lee MJ, Patrick DM, Himsworth CG. Rats About Town: A Systematic Review of Rat
1134 Movement in Urban Ecosystems. Frontiers in Ecology and Evolution. 2019;7.
1135 7. Feng AYT, Himsworth CG. The secret life of the city rat: a review of the ecology of urban Norway and
1136 black rats (Rattus norvegicus and Rattus rattus). Urban Ecosyst. 2014;17(1):149-62.
1137 8. Sabour S, Azimi T, Nasser A, Hadi N, Mohsenzadeh A, Shariati A. A global overview of the most
1138 important zoonotic bacteria pathogens transmitted from Rattus norvegicus to humans in urban environments.
1139 Infect Med (Beijing). 2022;1(3):192-207.
1140 9. Himsworth CG, Parsons KL, Jardine C, Patrick DM. Rats, cities, people, and pathogens: a systematic
1141 review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers.
1142 Vector Borne Zoonotic Dis. 2013;13(6):349-59.
1143 10. Gerbig GR, Piontkivska H, Smith TC, White R, Mukherjee J, Benson H, et al. Genetic characterization
1144 of Staphylococcus aureus isolated from Norway rats in Boston, Massachusetts. Vet Med Sci. 2023;9(1):272-81.
1145 11. Cummings CO, Hill NJ, Puryear WB, Rogers B, Mukherjee J, Leibler JH, et al. Evidence of Influenza
1146 A in Wild Norway Rats (Rattus norvegicus) in Boston, Massachusetts. Front Ecol Evol. 2019;7.
1147 12. Combs M, Byers KA, Ghersi BM, Blum MJ, Caccone A, Costa F, et al. Urban rat races: spatial
1148 population genomics of brown rats (Rattus norvegicus) compared across multiple cities. Proc Biol Sci.
1149 2018;285(1880).
1150 13. Combs M, Puckett EE, Richardson J, Mims D, Munshi-South J. Spatial population genomics of the
1151 brown rat (Rattus norvegicus) in New York City. Mol Ecol. 2018;27(1):83-98.
1152 14. Gardner-Santana LC, Norris DE, Fornadel CM, Hinson ER, Klein SL, Glass GE. Commensal ecology,
1153 urban landscapes, and their influence on the genetic characteristics of city-dwelling Norway rats (Rattus
1154 norvegicus). Mol Ecol. 2009;18(13):2766-78.
1155 15. Kajdacsi B, Costa F, Hyseni C, Porter F, Brown J, Rodrigues G, et al. Urban population genetics of
1156 slum-dwelling rats (Rattus norvegicus) in Salvador, Brazil. Mol Ecol. 2013;22(20):5056-70.
1157 16. Byers KA, Booker TR, Combs M, Himsworth CG, Munshi-South J, Patrick DM, et al. Using genetic
1158 relatedness to understand heterogeneous distributions of urban rat-associated pathogens. Evol Appl.
1159 2021;14(1):198-209.
1160 17. Richardson JL, Burak MK, Hernandez C, Shirvell JM, Mariani C, Carvalho-Pereira TSA, et al. Using
1161 fine-scale spatial genetics of Norway rats to improve control efforts and reduce leptospirosis risk in urban slum
1162 environments. Evol Appl. 2017;10(4):323-37.
1163 18. Levett PN. Leptospirosis. Clin Microbiol Rev. 2001;14(2):296-326.
1164 19. Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, et al. Global Morbidity
1165 and Mortality of Leptospirosis: A Systematic Review. PLoS Negl Trop Dis. 2015;9(9):e0003898.
1166 20. Sterling CR, Thiermann AB. Urban rats as chronic carriers of leptospirosis: an ultrastructural
1167 investigation. Vet Pathol. 1981;18(5):628-37.
1168 21. Thiermann AB. The Norway Rat as a Selective Chronic Carrier of Leptospira-Icterohaemorrhagiae. J
1169 Wildlife Dis. 1981;17(1):39-43.
1170 22. Monahan AM, Callanan JJ, Nally JE. Proteomic analysis of *Leptospira interrogans* shed in urine of
1171 chronically infected hosts. Infect Immun. 2008;76(11):4952-8.
1172 23. Bonilla-Santiago R, Nally JE. Rat model of chronic leptospirosis. Curr Protoc Microbiol. 2011;Chapter
1173 12:Unit 12E 3.
1174 24. Haake DA, Levett PN. Leptospirosis in humans. Curr Top Microbiol Immunol. 2015;387:65-97.
1175 25. Boey K, Shiokawa K, Rajeev S. Leptospira infection in rats: A literature review of global prevalence
1176 and distribution. PLoS Negl Trop Dis. 2019;13(8):e0007499.
1177 26. Hornsby RL, Alt DP, Nally JE. Isolation and propagation of leptospires at 37 degrees C directly from
1178 the mammalian host. Sci Rep-Uk. 2020;10(1).
1179 27. Musso D, La Scola B. Laboratory diagnosis of leptospirosis: a challenge. J Microbiol Immunol Infect.
1180 2013;46(4):245-52.
1181 28. Stone NE, McDonough RF, Hamond C, LeCount K, Busch JD, Dirsmith KL, et al. DNA Capture and
1182 Enrichment: A Culture-Independent Approach for Characterizing the Genomic Diversity of Pathogenic
1183 Leptospira Species. Microorganisms. 2023;11(5):1282.

1184 29. Cole JR, Jr., Sulzer CR, Pursell AR. Improved microtechnique for the leptospiral microscopic
1185 agglutination test. Appl Microbiol. 1973;25(6):976-80.
1186 30. Chirathaworn C, Inwattana R, Poovorawan Y, Suwancharoen D. Interpretation of microscopic
1187 agglutination test for leptospirosis diagnosis and seroprevalence. Asian Pac J Trop Biomed. 2014;4(Suppl
1188 1):S162-4.
1189 31. Pritchard J. Documentation for STRUCTURE software: Version 2.3 2003. Available from:
1190 https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf.
1191 32. Costa F, Richardson JL, Dion K, Mariani C, Pertile AC, Burak MK, et al. Multiple Paternity in the
1192 Norway Rat, Rattus norvegicus, from Urban Slums in Salvador, Brazil. J Hered. 2016;107(2):181-6.
1193 33. Glass GE, Klein SL, Norris DE, Gardner LC. Multiple Paternity in Urban Norway Rats: Extended
1194 Ranging for Mates. Vector Borne Zoonotic Dis. 2016;16(5):342-8.
1195 34. Davis DE. The characteristics of rat populations. Q Rev Biol. 1953;28(4):373-401.
1196 35. Davis DE, Emlen JT, Stokes AW. Studies on Home Range in the Brown Rat. J Mammal.
1197 1948;29(3):207-25.
1198 36. Easterbrook JD, Shields T, Klein SL, Glass GE. Norway rat population in Baltimore, Maryland, 2004.
1199 Vector Borne Zoonotic Dis. 2005;5(3):296-9.
1200 37. Berthier K, Garba M, Leblois R, Navascués M, Tatard C, Gauthier P, et al. Black rat invasion of inland
1201 Sahel: insights from interviews and population genetics in south-western Niger. Biol J Linn Soc.
1202 2016;119(4):748-65.
1203 38. Traweger D, Slotta-Bachmayr L. Introducing GIS-modelling into the management of a brown rat
1204 (Rattus norvegicus Berk.) (Mamm. Rodentia Muridae) population in an urban habitat. J Pest Sci. 2005;78(1):17-
1205 24.
1206 39. Worth CB. Field and Laboratory Observations on Roof Rats, Rattus-Rattus (Linnaeus), in Florida. J
1207 Mammal. 1950;31(3):293-304.
1208 40. Petrie GF, Todd RE. Observations in Upper Egypt on the range of excursion of the house rodents: R.
1209 rattus and Acomys cahirinus. Rep Notes Pub Hlth Lab Cairo. 1923;5:14-8.
1210 41. Glass GE. Comparative ecology and social interactions of Norway rat (Rattus norvegicus) populations
1211 in Baltimore, Maryland. Lawrence, Kan. :: Museum of Natural History, the University of Kansas; 1989.
1212 42. Davis DE, Christian JJ. Population Consequences of a Sustained Yield Program for Norway Rats.
1213 Ecology. 1958;39(2):217-22.
1214 43. Margulis HL. Rat Fields, Neighborhood Sanitation, and Rat Complaints in Newark, New-Jersey. Geogr
1215 Rev. 1977;67(2):221-31.
1216 44. Vadell MV, Cavia R, Suárez OV. Abundance, age structure and reproductive patterns of Rattus
1217 norvegicus and Mus musculus in two areas of the city of Buenos Aires. Int J Pest Manage. 2010;56(4):327-36.
1218 45. Wundram IJ, Ruback RB. Urban Rats - Symbol, Symptom and Symbiosis. Hum Organ. 1986;45(3):212-
1219 9.
1220 46. Costa F, Ribeiro GS, Felzemburgh RD, Santos N, Reis RB, Santos AC, et al. Influence of household rat
1221 infestation on leptospira transmission in the urban slum environment. PLoS Negl Trop Dis. 2014;8(12):e3338.
1222 47. de Faria MT, Calderwood MS, Athanazio DA, McBride AJ, Hartskeerl RA, Pereira MM, et al. Carriage
1223 of Leptospira interrogans among domestic rats from an urban setting highly endemic for leptospirosis in Brazil.
1224 Acta Trop. 2008;108(1):1-5.
1225 48. Costa F, Porter FH, Rodrigues G, Farias H, de Faria MT, Wunder EA, et al. Infections by Leptospira
1226 interrogans, Seoul virus, and Bartonella spp. among Norway rats (Rattus norvegicus) from the urban slum
1227 environment in Brazil. Vector Borne Zoonotic Dis. 2014;14(1):33-40.
1228 49. Himsworth CG, Bidulka J, Parsons KL, Feng AY, Tang P, Jardine CM, et al. Ecology of Leptospira
1229 interrogans in Norway rats (Rattus norvegicus) in an inner-city neighborhood of Vancouver, Canada. PLoS Negl
1230 Trop Dis. 2013;7(6):e2270.
1231 50. Peterson AC, Ghersi BM, Riegel C, Wunder EA, Jr., Childs JE, Blum MJ. Amplification of pathogenic
1232 Leptospira infection with greater abundance and co-occurrence of rodent hosts across a counter-urbanizing
1233 landscape. Mol Ecol. 2021;30(9):2145-61.
1234 51. Santos LA, Adhikarla H, Yan X, Wang Z, Fouts DE, Vinetz JM, et al. Genomic Comparison Among
1235 Global Isolates of L. interrogans Serovars Copenhageni and Icterohaemorrhagiae Identified Natural Genetic
1236 Variation Caused by an Indel. Front Cell Infect Microbiol. 2018;8:193.

1237 52. Lee MJ, Byers KA, Donovan CM, Bidulka JJ, Stephen C, Patrick DM, et al. Effects of Culling on
1238 Leptospira interrogans Carriage by Rats. Emerg Infect Dis. 2018;24(2):356-60.
1239 53. Thibeaux R, Soupe-Gilbert ME, Kainiu M, Girault D, Bierque E, Fernandes J, et al. The zoonotic
1240 pathogen *Leptospira interrogans* mitigates environmental stress through cyclic-di-GMP-controlled biofilm
1241 production. NPJ Biofilms Microbiomes. 2020;6(1):24.
1242 54. Santos AAN, Ribeiro PDS, da Franca GV, Souza FN, Ramos EAG, Figueira CP, et al. Leptospira
1243 interrogans biofilm formation in Rattus norvegicus (Norway rats) natural reservoirs. PLoS Negl Trop Dis.
1244 2021;15(9):e0009736.
1245 55. Davignon G, Cagliero J, Guentas L, Bierque E, Genthon P, Gunkel-Grillon P, et al. Leptospirosis:
1246 toward a better understanding of the environmental lifestyle of. Front Water. 2023;5.
1247 56. Stone NE, Hall CM, Ortiz M, Hutton SM, Santana-Propper E, Celona KR, et al. Diverse lineages of
1248 pathogenic Leptospira species are widespread in the environment in Puerto Rico, USA. PLoS Negl Trop Dis.
1249 2022;16(5):e0009959.
1250 57. Vincent AT, Schiettekatte O, Goarant C, Neela VK, Bernet E, Thibeaux R, et al. Revisiting the
1251 taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. PLoS Negl
1252 Trop Dis. 2019;13(5):e0007270.
1253 58. Galloway RL, Hoffmaster AR. Optimization of LipL32 PCR assay for increased sensitivity in
1254 diagnosing leptospirosis. Diagn Microbiol Infect Dis. 2015;82(3):199-200.
1255 59. Hamond C, Browne AS, de Wilde LH, Hornsby RL, LeCount K, Anderson T, et al. Assessing rodents
1256 as carriers of pathogenic Leptospira species in the U.S. Virgin Islands and their risk to animal and public health.
1257 Sci Rep. 2022;12(1):1132.
1258 60. Stoddard RA, Gee JE, Wilkins PP, McCaustland K, Hoffmaster AR. Detection of pathogenic *Leptospira*
1259 spp. through TaqMan polymerase chain reaction targeting the LipL32 gene. Diagn Microbiol Infect Dis.
1260 2009;64(3):247-55.
1261 61. Ellis WA, Montgomery J, Cassells JA. Dihydrostreptomycin treatment of bovine carriers of *Leptospira*
1262 *interrogans* serovar hardjo. Res Vet Sci. 1985;39(3):292-5.
1263 62. Modlinska K, Pisula W. The Norway rat, from an obnoxious pest to a laboratory pet. Elife. 2020;9.
1264 63. Lack JB, Hamilton MJ, Braun JK, Mares MA, Van den Bussche RA. Comparative phylogeography of
1265 invasive Rattus rattus and Rattus norvegicus in the US reveals distinct colonization histories and dispersal. Biol
1266 Invasions. 2013;15(5):1067-87.
1267 64. Jacob HJ, Brown DM, Bunker RK, Daly MJ, Dzau VJ, Goodman A, et al. A genetic linkage map of the
1268 laboratory rat, Rattus norvegicus. Nat Genet. 1995;9(1):63-9.
1269 65. Heiberg AC, Leirs H, Siegismund HR. Reproductive success of bromadiolone-resistant rats in absence
1270 of anticoagulant pressure. Pest Manag Sci. 2006;62(9):862-71.
1271 66. Robertson BC, Gemmell NJ. Defining eradication units to control invasive pests. J Appl Ecol.
1272 2004;41(6):1042-8.
1273 67. Giraudeau F, Apiou F, Amarger V, Kaisaki PJ, Bihoreau MT, Lathrop M, et al. Linkage and physical
1274 mapping of rat microsatellites derived from minisatellite loci. Mamm Genome. 1999;10(4):405-9.
1275 68. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching
1276 and research--an update. Bioinformatics. 2012;28(19):2537-9.
1277 69. Goudet J. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2).
1278 2002.
1279 70. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: software for
1280 identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 2004;4(3):535-8.
1281 71. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data.
1282 Genetics. 2000;155(2):945-59.
1283 72. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software
1284 STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611-20.
1285 73. Macdonald DW, Barrett P. Mammals of Europe. Princeton, N.J.: Princeton University Press; 1993. 312
1286 p., 126 p. of plates p.
1287 74. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution.
1288 1984;38(6):1358-70.

1289 75. Hamond C, Dirsmith KL, LeCount K, Soltero FV, Rivera-Garcia S, Camp P, et al. Leptospira
1290 borgpetersenii serovar Hardjo and Leptospira santarosai serogroup Pyrogenes isolated from bovine dairy herds
1291 in Puerto Rico. Front Vet Sci. 2022;9:1025282.
1292 76. Hamond C, LeCount K, Putz EJ, Bayles DO, Camp P, Goris MGA, et al. Bovine Leptospirosis Due to
1293 Persistent Renal Carriage of Leptospira borgpetersenii Serovar Tarassovi. Front Vet Sci. 2022;9:848664.
1294 77. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol.
1295 2019;20(1):257.
1296 78. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100.
1297 79. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
1298 format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
1299 80. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis
1300 Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res.
1301 2010;20(9):1297-303.
1302 81. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence
1303 sets. Curr Protoc Bioinformatics. 2003;Chapter 10:Unit 10 3.
1304 82. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid method
1305 for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb Genom.
1306 2016;2(8):e000074.
1307 83. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm
1308 for estimating maximum-likelihood phylogenies. Molecular biology and evolution. 2015;32(1):268-74.
1309 84. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model
1310 selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9.
1311 85. Sahl JW, Schupp JM, Rasko DA, Colman RE, Foster JT, Keim P. Phylogenetically typing bacterial
1312 strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data.
1313 Genome Med. 2015;7(1):52.
1314 86. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator.
1315 Bioinformatics. 2012;28(4):593-4.
1316 87. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly
1317 tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453-5.
1318 88. Barragan V, Chiriboga J, Miller E, Olivas S, Birdsell D, Hepp C, et al. High Leptospira Diversity in
1319 Animals and Humans Complicates the Search for Common Reservoirs of Human Disease in Rural Ecuador.
1320 PLoS Negl Trop Dis. 2016;10(9):e0004990.
1321 89. Merien F, Portnoi D, Bourhy P, Charavay F, Berlioz-Arthaud A, Baranton G. A rapid and quantitative
1322 method for the detection of Leptospira species in human leptospirosis. FEMS Microbiol Lett. 2005;249(1):139-
1323 47.
1324 90. Matthias MA, Lubar AA, Lanka Acharige SS, Chaiboonma KL, Pilau NN, Marroquin AS, et al. Culture-
1325 Independent Detection and Identification of Leptospira Serovars. Microbiol Spectr. 2022;10(6):e0247522.
1326 91. Backstedt BT, Buyuktanir O, Lindow J, Wunder EA, Jr., Reis MG, Usmani-Brown S, et al. Efficient
1327 Detection of Pathogenic Leptospires Using 16S Ribosomal RNA. PLoS One. 2015;10(6):e0128913.
1328 92. Iwasaki H, Chagan-Yasutan H, Leano PS, Koizumi N, Nakajima C, Taurustiati D, et al. Combined
1329 antibody and DNA detection for early diagnosis of leptospirosis after a disaster. Diagn Microbiol Infect Dis.
1330 2016;84(4):287-91.
1331 93. Benacer D, Zain SNM, Lewis JW, Khalid M, Thong KL. A duplex endpoint PCR assay for rapid
1332 detection and differentiation of Leptospira strains. Rev Soc Bras Med Trop. 2017;50(2):239-42.
1333 94. Cerqueira GM, McBride AJ, Hartskeerl RA, Ahmed N, Dellagostin OA, Eslabao MR, et al.
1334 Bioinformatics describes novel Loci for high resolution discrimination of leptospira isolates. PLoS One.
1335 2010;5(10):e15335.
1336 95. Di Azevedo MIN, Borges A, Kremer F, de Melo J, Carvalho-Costa FA, Lilenbaum W. Genetic Analysis
1337 of LigA, LipL32, Loa22, and OmpL1 Antigens from Leptospira spp. Sejroe Serogroup: How Close We Are To
1338 a Vaccine Against Bovine Leptospirosis? Curr Microbiol. 2023;80(9):310.
1339 96. Ahmed A, Engelberts MF, Boer KR, Ahmed N, Hartskeerl RA. Development and validation of a real-
1340 time PCR for detection of pathogenic *Leptospira* species in clinical materials. PLoS One. 2009;4(9):e7093.

97. Rahelinirina S, Moseley MH, Allan KJ, Ramanohizakandrainy E, Ravaoarinoro S, Rajerison M, et al. Leptospira in livestock in Madagascar: uncultured strains, mixed infections and small mammal-livestock transmission highlight challenges in controlling and diagnosing leptospirosis in the developing world. Parasitology. 2019;146(14):1707-13.
98. Haake DA, Suchard MA, Kelley MM, Dundoo M, Alt DP, Zuerner RL. Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer. J Bacteriol. 2004;186(9):2818-28.
99. Ahmed N, Devi SM, Valverde Mde L, Vijayachari P, Machang'u RS, Ellis WA, et al. Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. Ann Clin Microbiol Antimicrob. 2006;5:28.
100. Boonsilp S, Thaipadungpanit J, Amornchai P, Wuthiekanun V, Bailey MS, Holden MT, et al. A single multilocus sequence typing (MLST) scheme for seven pathogenic Leptospira species. PLoS Negl Trop Dis. 2013;7(1):e1954.
101. Varni V, Ruybal P, Lauthier JJ, Tomasini N, Brihuega B, Koval A, et al. Reassessment of MLST schemes for Leptospira spp. typing worldwide. Infect Genet Evol. 2014;22:216-22.
102. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733-45.
103. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614-24.
104. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12.
105. Furstenau TN, Cocking JH, Sahl JW, Fofanov VY. Variant site strain typer (VaST): efficient strain typing using a minimal number of variant genomic sites. BMC Bioinformatics. 2018;19(1):222.
106. Guglielmini J, Bourhy P, Schiettekatte O, Zinini F, Brisse S, Picardeau M. Genus-wide Leptospira core genome multilocus sequence typing for strain taxonomy and global surveillance. PLoS Negl Trop Dis. 2019;13(4):e0007374.
107. Sykes JE, et al.,. Understanding leptospirosis: application of state-of-the-art molecular typing tools with a One Health lens. AJVR. 2022;In Press.
108. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.
109. Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F, et al. Detection of Low-Level Mixed-Population Drug Resistance in *Mycobacterium tuberculosis* Using High Fidelity Amplicon Sequencing. Plos One. 2015;10(5).
110. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
111. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
112. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. J Comput Biol. 2017;24(11):1138-43.

Supporting Information

S1 Fig. Plots from the delta-*K* method. The hypothetical number of subpopulations (*K*) is plotted on the X-axis against A) delta-*K*, and B) The logarithm probability for each *K* [L(*K*)]. Ten independent runs in STRUCTURE were performed for each *K* value.

1386 **S2 Fig. Principle Coordinates Analysis plot.** Principle Coordinates Analysis plot of standardized
1387 genetic distances illustrating genetic similarity among 311 *Rattus norvegicus* collected in Boston
1388 (2016-2022). Each rat is color coded according to its probability of membership (Q ≥0.75) to one of
1389 12 genetic groups assigned using STRUCTURE (**Table S3 and Fig 1**). The percentage of genetic
1390 variation explained by these two axes is indicated.
1391
1392 **S3 Fig. Spatial autocorrelation analysis graph.** Spatial autocorrelation analysis graph assessing the
1393 correlation between genetic and geographic distance (Y-axis) using twenty even distance classes
1394 corresponding to 100 m intervals (X-axis). The blue line is the correlation value (r), and the red
1395 dashed lines are the 95% confidence intervals (U=upper and L=lower) about the null hypothesis of
1396 no spatial structure as determined by bootstrap resampling. Values of r above the upper 95%
1397 confidence interval are significant.
1398
1399 **S4 Fig. *Leptospira interrogans* whole genome phylogeny.** A maximum likelihood midpoint rooted
1400 phylogeny of 28 *Leptospira interrogans* serogroup Icterohaemorrhagiae isolate genomes and 13
1401 enriched genomes from 40 rats and one human (bold text), together with 22 publicly available *L.*
1402 *interrogans* isolate genomes representing >9 serovars (GenBank accession numbers are provided in
1403 the figure annotations), based upon a concatenated SNP alignment of 27,678 positions out of a core
1404 genome size of 1,186,500 nucleotides (nts).
1405
1406 **S5 Fig. *Leptospira interrogans* serogroup Icterohaemorrhagiae WG-FAST phylogeny** A
1407 maximum likelihood phylogeny of 28 *Leptospira interrogans* serogroup Icterohaemorrhagiae isolate
1408 genomes from rats, together with two publicly available *L. interrogans* serogroup
1409 Icterohaemorrhagiae serovar Copenhageni complete genomes (GenBank accession#

1410 GCA_002073495.2 and GCA_010978155.1), and 13 enriched genomes (12 rats and 1 human), based
1411 upon a concatenated SNP alignment of 127 positions out of a core genome size of 2,123,028
1412 nucleotides (nts). The phylogeny was rooted with reference genome *L. interrogans* serogroup
1413 Icterohaemorrhagiae serovar Copenhageni strain Fiocruz L1-130 (GenBank accession#
1414 GCA_000007685.1). The color of the genome ID corresponds to the rat genetic group in Fig 1; black
1415 text indicates that no genetic group was assigned. Two enriched samples (R289 and R292) had lower
1416 breadth and depth of coverage than the other enrichments (**see S5 Table**) and were inserted into the
1417 rat phylogeny based upon canonical SNPs with WG-FAST.
1418
1419 **S6 Fig. Pathogenic *Leptospira* whole genome and AmpSeq phylogenies.** A) Whole genome
1420 maximum likelihood midpoint rooted phylogeny including 18 P1 clade pathogenic *Leptospira* spp.
1421 B) compared to an AmpSeq phylogeny that also includes AmpSeq data generated from nine
1422 *Leptospira* isolate gDNAs (blue text). Species specific clades were maintained (color coded) and
1423 similar phylogenetic topologies were observed despite a large reduction in the core-genome of the
1424 AmpSeq tree (1,195 nts) compared to the whole genome tree (167,849 nts). Bootstrap values are
1425 indicated at specific nodes. The nine AmpSeq *Leptospira* isolate gDNAs grouped in their respective
1426 species clades.
1427
1428 **S7 Fig. *Leptospira interrogans* AmpSeq phylogeny.** Maximum likelihood phylogeny based upon an
1429 alignment of 4,260 shared nucleotide positions using *Leptospira* AmpSeq data generated from nine
1430 *lipL32* qPCR positive *Rattus norvegicus* (bold text) along with 22 *Leptospira interrogans* isolate
1431 genomes (GenBank accession numbers are provided in the figure annotations). The phylogeny is
1432 rooted with *Leptospira kirschneri* serovar Grippotyphosa strain RedPanda1 (GenBank accession#
1433 GCA_027563495.1).

1434

1435 **S1 Table. *Rattus norvegicus* collections.**

1436 **S2 Table. Analyses summary and results for all rodent and human samples.**

1437 **S3 Table. *Rattus norvegicus* population assignment.**

1438 **S4 Table. *Rattus norvegicus* FSTAT results**

1439 **S5 Table. DNA capture and enrichment setup and results.**

1440 **S6 Table. *Leptospira* AmpSeq results**

1441 **S7 Table. *Rattus norvegicus* trapping locations.**

1442 **S8 Table. *Leptospira* AmpSeq primers and conditions**

1443 **S9 Table. *Leptospira* AmpSeq primer design genomes**

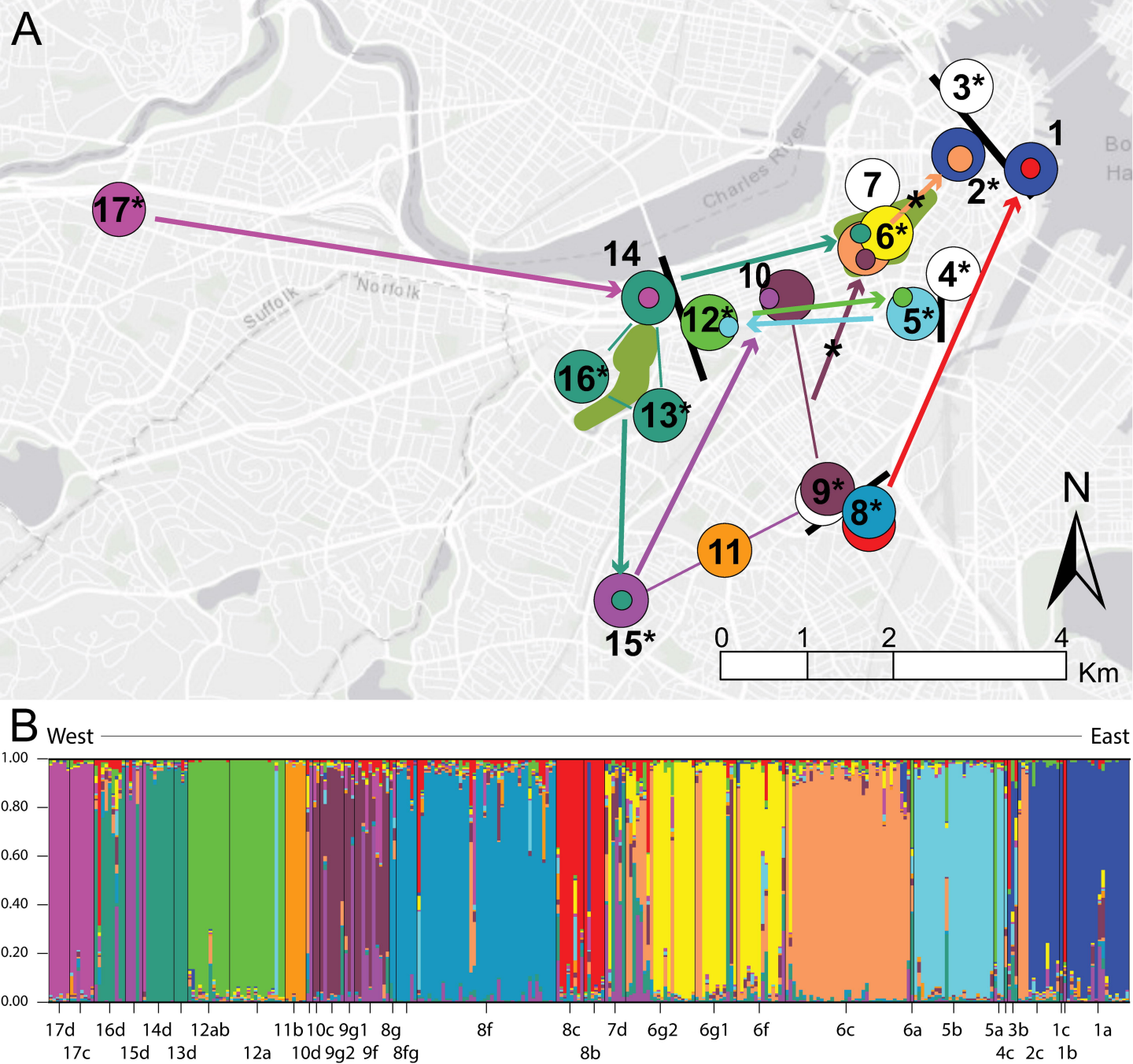
1444 **S1 File: Predicted amplification of 42 AmpSeq loci against 69 *Leptospira* spp.**

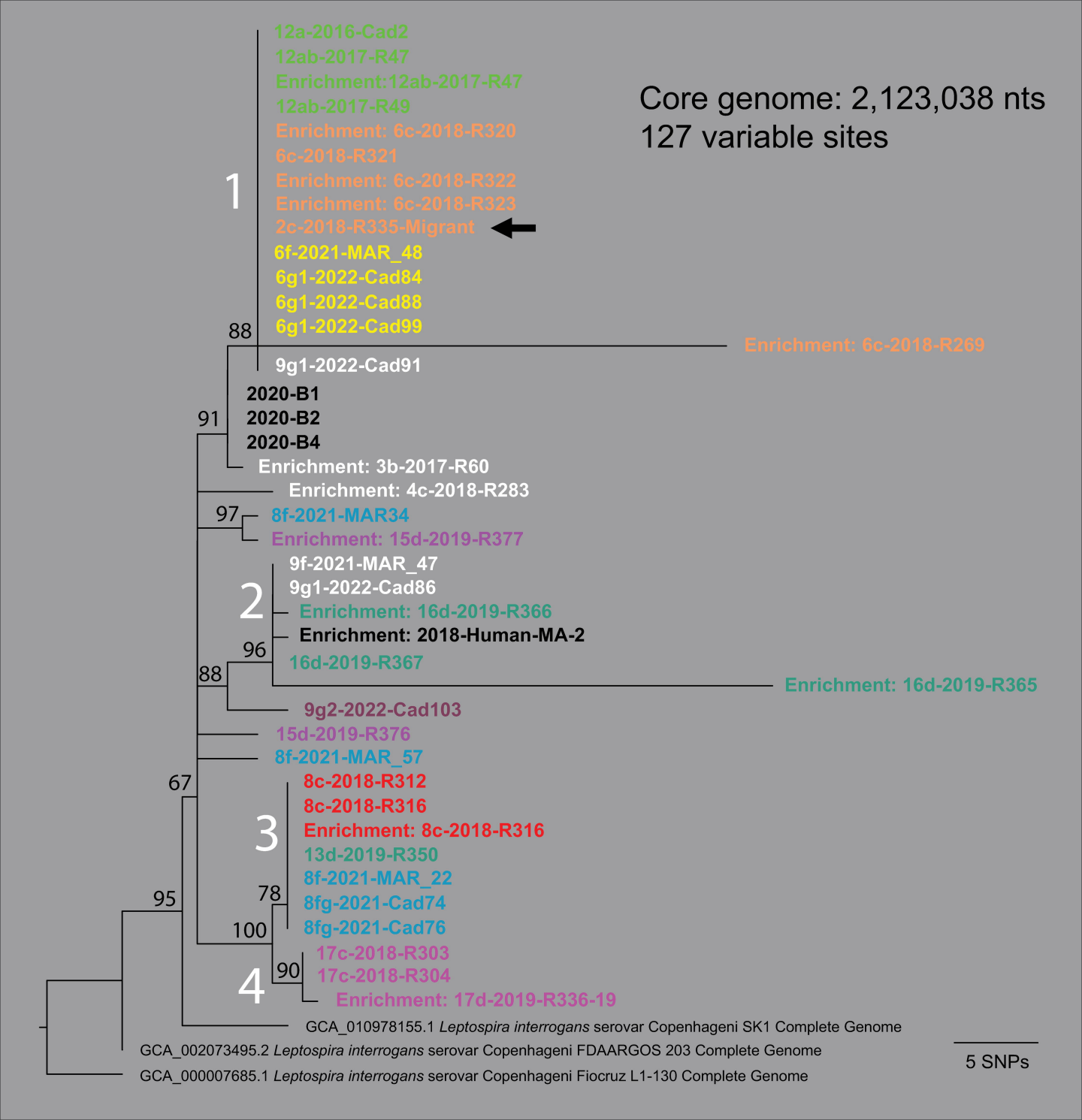
1445

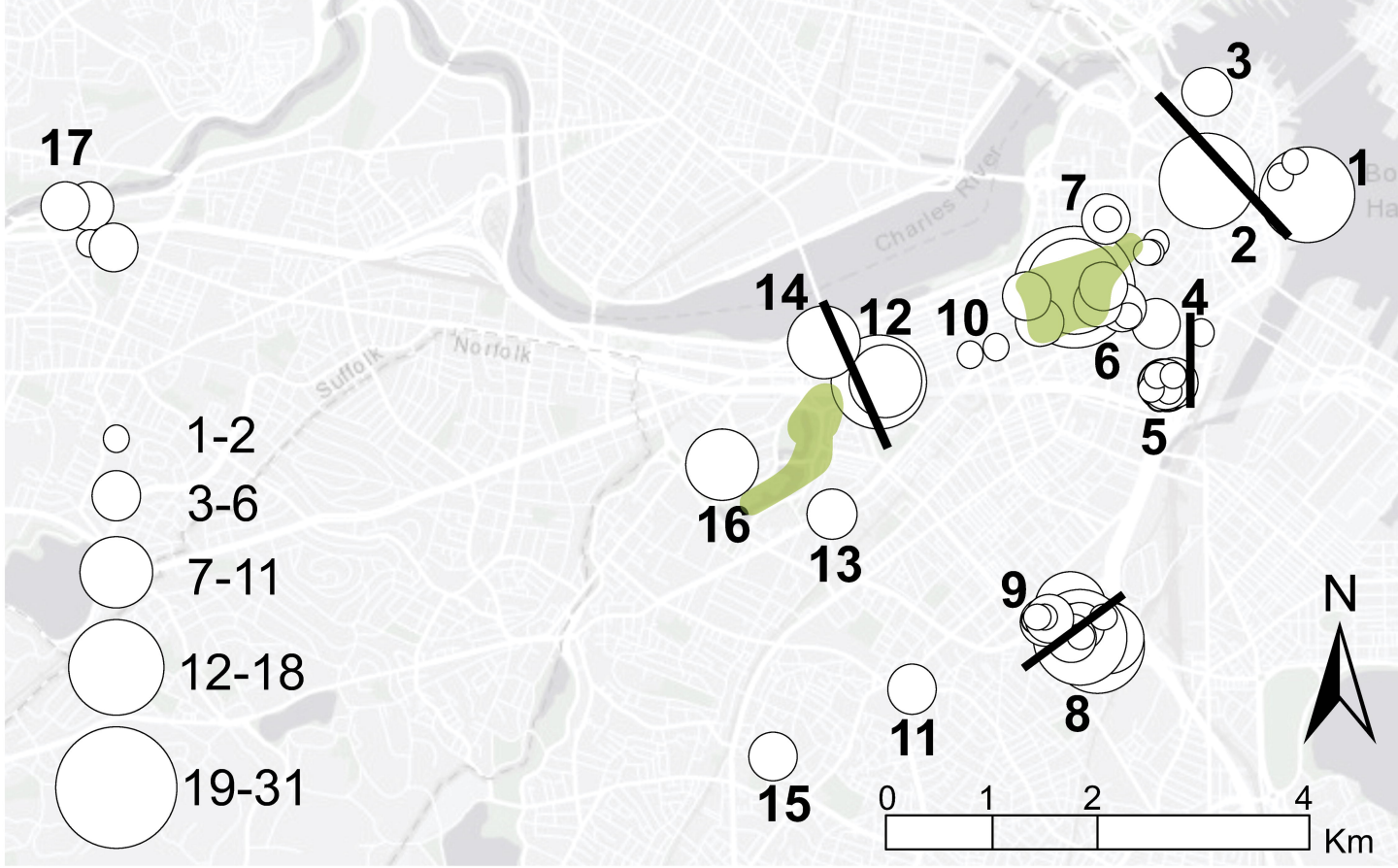
1446 Data Availability Statement

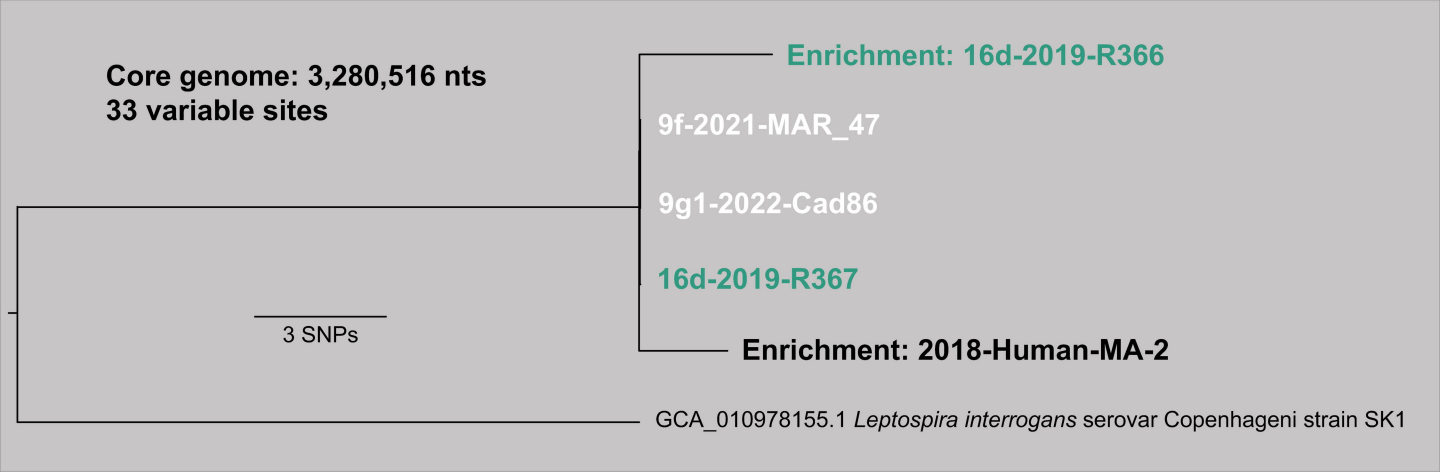
1447 MSAT multi-locus genotypes for 311 *R. norvegicus* are provided in **S2 Table**. Whole genome
1448 sequencing reads have been deposited in GenBank under BioProject PRJNA1080226 with SRA
1449 accession numbers SRR28374870-SRR28374875, SRR28374877-SRR28374878, SRR28374880,
1450 SRR28374882-SRR28374900 for isolate genomes and SRR28374876, SRR28374879,
1451 SRR28411353-SRR28411357, SRR28411359-SRR28411365, SRR28411370 for enriched genomes.
1452 Raw AmpSeq data for up to 42 loci are also provided in GenBank under the same BioProject with
1453 SRA accession numbers SRR28374881, SRR28411338-SRR28411352, SRR28411358,
1454 SRR28411366-SRR28411369 and SRR28765937- SRR28765945.

1455


A
17*
14
12*
10
16*
13*
15*
11
9*
8*
7
6*
5*
4*
3*
2*
1
Charles River
Suffolk
Norfolk
N
0
1
2
4
Km
B
West
East
1.00
0.80
0.60
0.40
0.20
0.00
17d
17c
16d
15d
14d
13d
12ab
12a
11b
10d
10c
9g2
9g1
9f
8g
8fg
8f
8c
8b
7d
6g2
6g1
6f
6c
6a
5b
5a
4c
3b
2c
1c
1b
1a


Core genome: 2,123,038 nts
127 variable sites
1
12a-2016-Cad2
12ab-2017-R47
Enrichment:12ab-2017-R47
12ab-2017-R49
Enrichment: 6c-2018-R320
6c-2018-R321
Enrichment: 6c-2018-R322
Enrichment: 6c-2018-R323
2c-2018-R335-Migrant
6f-2021-MAR_48
6g1-2022-Cad84
6g1-2022-Cad88
6g1-2022-Cad99
88
Enrichment: 6c-2018-R269
9g1-2022-Cad91
2020-B1
91
2020-B2
2020-B4
Enrichment: 3b-2017-R60
Enrichment: 4c-2018-R283
97
8f-2021-MAR34
Enrichment: 15d-2019-R377
2
9f-2021-MAR_47
9g1-2022-Cad86
Enrichment: 16d-2019-R366
Enrichment: 2018-Human-MA-2
96
16d-2019-R367
88
Enrichment: 16d-2019-R365
9g2-2022-Cad103
15d-2019-R376
8f-2021-MAR_57
67
3
8c-2018-R312
8c-2018-R316
Enrichment: 8c-2018-R316
13d-2019-R350
8f-2021-MAR_22
78
95
8fg-2021-Cad74
100
8fg-2021-Cad76
4
17c-2018-R303
90
17c-2018-R304
Enrichment: 17d-2019-R336-19
GCA_010978155.1 Leptospira interrogans serovar Copenhageni SK1 Complete Genome
GCA_002073495.2 Leptospira interrogans serovar Copenhageni FDAARGOS 203 Complete Genome
GCA_000007685.1 Leptospira interrogans serovar Copenhageni Fiocruz L1-130 Complete Genome
5 SNPs


17
3
1
2
7
Charles River
14
12
10
4
6
5
Suffolk
Norfolk
16
13
9
8
11
15
1-2
3-6
7-11
12-18
19-31
N
0
1
2
4
Km


Core genome: 3,280,516 nts
33 variable sites
Enrichment: 16d-2019-R366
9f-2021-MAR_47
9g1-2022-Cad86
16d-2019-R367
Enrichment: 2018-Human-MA-2
3 SNPs
GCA_010978155.1 Leptospira interrogans serovar Copenhageni strain SK1