

1 **Epidemiological, clinical and genomic insights into the ongoing diphtheria outbreak in**
2 **Yemen**

3

4 Edgar Badell^{1,2}, Abdulilah Alharazi³, Alexis Criscuolo⁴, The NCPHL diphtheria outbreak
5 working group[#], Noemie Lefrancq⁵, Valerie Bouchez¹, Julien Guglielmini⁴, Melanie
6 Hennart^{1,6}, Annick Carmi-Leroy^{1,2}, Nora Zidane¹, Marine Pascal-Perrigault¹, Manon
7 Lebreton¹, Helena Martini⁷, Henrik Salje^{5,8}, Julie Toubiana^{1,2,9}, Fekri Dureab¹⁰, Ghulam
8 Dhabaan^{11,*} and Sylvain Brisse^{1,2,*}

9 [#]The NCPHL diphtheria outbreak working group: Abdulaziz A. Rawah , Mohammed A.

10 Aldawla, Ekram M. Al-Awdi, Nabila M. Al-Moalmy, Khaled A. Almoayed, Huda Z. Al-Shami, Ali
11 A. Al-Somainy

12 **Affiliations:**

13 1 Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France

14 2 Institut Pasteur, National Reference Center for Corynebacteria of the diphtheriae complex,
15 Paris, France

16 3 National Centre of the Public Health Laboratories (NCPHL), Sana'a, Yemen

17 4 Hub de Bioinformatique et Biostatistique ☒ Département Biologie Computationnelle,
18 Institut Pasteur, USR 3756 CNRS, Paris, France

19 5 Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, UMR2000, CNRS,
20 Paris, France

21 6 Sorbonne Université, Collège doctoral, F-75005 Paris, France

22 7 Department of Microbiology, National Reference Centre for toxigenic corynebacteria,
23 Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel (VUB), Laarbeeklaan 101, 1090

24 Brussels, Belgium

25 8 Department of Genetics, University of Cambridge, Cambridge, UK

26 9 Université de Paris, General Paediatrics and Paediatric Infectious Diseases Department,
27 Hôpital Necker-Enfants malades, APHP, Paris, France

28 10 Institute for Research in International Assistance, Akkon Hochschule, Berlin, Germany, 2-
29 Heidelberg Institute of Global Health, Heidelberg, Germany

30 11 Mount Sinai Hospital, University Health Network, University of Toronto, Toronto, Canada

31 ***Correspondence:**

32 Sylvain Brisse. Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens, Paris,
33 France. E-mail: sylvain.brisse@pasteur.fr; Tel: +33 1 45 68 83 34;
34 Ghulam Dhabaan: Mount Sinai Hospital, University Health Network, University of Toronto,
35 Toronto, Canada. Ghulam.Dhabaan@utoronto.ca ; Tel: +1 647 949 8012
36 **Keywords:** diphtheria, outbreak, epidemiology, phylogenetics, antimicrobial resistance,
37 genomic sequencing, evolutionary rate, Yemen
38 **Running Title:** Diphtheria in Yemen

39 **Abstract**

40 **Background.** An outbreak of diphtheria, declared in Yemen in October 2017, is still ongoing.

41 **Methods.** Probable cases were recorded through an electronic diseases early warning
42 system. Microbiological culture, genomic sequencing, antimicrobial susceptibility and toxin
43 production testing were performed.

44 **Findings.** The Yemen diphtheria outbreak developed in three epidemic waves, which
45 affected nearly all governorates (provinces) of Yemen, with 5701 probable cases and 330
46 deaths (October 2017 - April 2020). The median age of patients was 12 years (range, 0.17-
47 80). Virtually all outbreak isolates (40 of 43 tested ones) produced the diphtheria toxin. We
48 observed low level of antimicrobial resistance to penicillin. We identified six separate
49 *Corynebacterium diphtheriae* phylogenetic sublineages, three of which are genetically
50 related to isolates from Saudi Arabia and Somalia. The predominant sublineage was resistant
51 to trimethoprim and was associated with unique genomic features, more frequent neck
52 swelling ($p=0.002$) and a younger age of patients ($p=0.06$). Its evolutionary rate was
53 estimated at 1.67×10^{-6} substitutions per site year⁻¹, placing its most recent common
54 ancestor in 2015, and indicating silent circulation of *C. diphtheriae* in Yemen earlier than
55 outbreak declaration.

56 **Interpretation.** We disclose clinical, epidemiological and microbiological characteristics of
57 one of the largest contemporary diphtheria outbreaks and demonstrate clinically relevant
58 heterogeneity of *C. diphtheriae* isolates, underlining the need for laboratory capacity and
59 real-time microbiological analyses to inform prevention, treatment and control of
60 diphtheria.

61 **Funding.** This work was supported by institutional funding from the National Centre of the
62 Public Health Laboratories (Sanaa, Yemen) and Institut Pasteur (Paris, France) and by the
63 French Government Investissement d'Avenir Program.

64 Introduction

65 Diphtheria is a severe infection that typically affects the upper respiratory tract, potentially
66 leading to pseudomembrane formation, neck swelling and suffocation [1, 2]. In addition,
67 diphtheria toxin production by some *C. diphtheriae* strains can cause damage to the heart
68 and other organs. Non-toxigenic strains can also cause invasive infections [3]. Before large
69 scale vaccination, which mainly targets the toxin and is highly effective in preventing the
70 disease, diphtheria was a major cause of death in children [2]. Large outbreaks of diphtheria
71 are often observed following disruption to vaccination programs, as occurred in the Soviet
72 Union in the 1990s [4] and more recently in the Rohingya refugee population in Bangladesh
73 [5] and in Venezuela [6]. Diphtheria is still observed occasionally in high-income countries, in
74 particular among migrants [7, 8].

75 In Yemen, where civil war has been raging since March 2015, a large outbreak of diphtheria
76 was recognized in October 2017 [9, 10][11]. While a vaccination campaign targeting 300,000
77 children began in late November 2018, with plans to scale up to 3 million children and young
78 adults in December, the ongoing conflict has complicated vaccination catch-up efforts [12,
79 13]. Before this outbreak, diphtheria was considered endemic in Yemen, with an average of
80 50 suspected cases reported annually since 2000 [14]. The last documented outbreak of
81 diphtheria in this country occurred in 1981-82, with a total of 149 cases with no death.
82 Yemen is currently also affected by other epidemic diseases including cholera [15] and
83 COVID-19 [16, 17]. As of May 2020, the diphtheria outbreak was still ongoing [18]. The early
84 epidemiology of the outbreak was described [9, 10], but virtually no data have been reported
85 on the clinical and microbiological characteristics of the outbreak.

86 Besides vaccination, a critical component of diphtheria management is serotherapy.
87 Unfortunately, diphtheria antitoxin is not readily available in Yemen, as in most world
88 countries [19]. Antimicrobial treatment is also indicated for patient care and to avoid
89 transmission, with penicillin and erythromycin being recommended as first-line therapeutic
90 agents [20][2].

91 Here we report on the clinical, epidemiological and microbiological features of the ongoing
92 Yemen diphtheria outbreak.

93 **Results**

94 **Epidemiology and clinical features of the Yemen diphtheria outbreak, 2017-2020**

95 Large numbers of probable cases were recorded in 2017 (n=560), 2018 (n=2606), 2019
96 (n=2004) and in 2020 (up to week 17: n=649). Altogether, 5701 cases including 330 deaths
97 were reported up to April 26th, 2020, corresponding to a case fatality rate of 5.8%.

98 Between 2017 and 2020 there were three major waves of cases (**Figure 1A**). The first wave
99 occurred from October 2017 until June 2018 with a peak on week 6 of 2018. The second
100 wave was flatter and lasted from June 2018 until June 2019. The third wave started in June
101 2019 and reached its peak in October 2019, with more than 80 cases/week by that time.

102 The outbreak affected nearly all governorates of Yemen (**Figure 1, Table S3**). However, some
103 governorates were more affected and the two first epidemic waves had distinctive
104 geographic signatures (**Figure 1D,E**). The outbreak started in October 2017 in Ibb, where 288
105 cases (52% of the total 2017 cases) were registered; while Al Hodeidah recorded 60 cases
106 (11%) in this year. In 2018, 2606 cases were registered in 21 governorates, with 64% of the
107 cases registered in 5 governorates: Hajja (395 cases, 15%), Al Hodeidah (338, 13%), Sana'a
108 (339, 13%), Ibb (324, 12%) and the capital city of Sana'a (276, 11%). Last, the distribution of
109 reported cases in 2019 shows that 60% of cases occurred in four governorates: Al Hodeida
110 (371 cases, 18%), Hajjah (327 cases, 16%), Sadah (297 cases, 15%) and Sana'a (226 cases,
111 11%).

112
113 Clinical data available for 888 patients (February 2018 - November 2019) showed that
114 patients were more likely to be female (55%), and their median age was 12 (range, 0-80)
115 years (**Figure 2A**). Of the 750 patients with available data, 434 (58%) reported a history of
116 previous vaccination against diphtheria. The vaccinated proportion of patients did not
117 change between 2018 and 2019 (57% vs. 59%, p=0.49). The median age of vaccinated
118 patients was lower (10 years, range 0.17-45) than that of non-vaccinated patients (13 years,
119 range 0.17-80; p<0.001). The age distribution of cases was available for the first two
120 epidemic waves. We found that the median age of infection was 12 years old; however, the
121 attack rate was highest in those 5-15 years in age as compared to other groups. While the
122 attack rate was largely constant between the first two years in those over 5 years old, there
123 was a marked reduction in children 0-4 years between the two waves (**Figure 2B; Figure S1**).

124 Almost all patients (n=818/841, 97%) had fever. Pseudomembrane was present in 787/839
125 (94%), signs of laryngitis in 548/784 (31%), neck swelling in 456/769 (59%) and tonsillitis in
126 821/836 (98%) patients with available data. Difficulty to swallow and to breathe were
127 observed in 710 (86%) and 399 (51%) patients, respectively. No difference in disease
128 expression was observed between vaccinated and non-vaccinated patients. Most patients
129 (n=634/790, 80%) were reported to be under antibiotic therapy at the time of throat
130 sampling.

131

132 Microbiological confirmation of cases was tested for 888 samples referred to the NCPHL
133 from 2018 to 2020. Of the 836 cases with available results, 340 (41%) were microbiologically
134 confirmed by a positive *C. diphtheriae* culture. The temporal and geographic distributions of
135 tested and confirmed cases were similar (**Figure 1B-C**).

136

137 **Phylogenetic diversity of *C. diphtheriae* from Yemen and their relationships with global** 138 **isolates**

139 A random selection of 98 *C. diphtheriae* positive cultures were transferred to Institut Pasteur
140 (Paris, France) for external confirmation and sequencing. Two cultures (YEM0065 = YE-
141 NCPHL-647L) and YEM0070 = YE-NCPHL-647S) with distinct colony morphologies had been
142 isolated from a single sample. We confirmed by qPCR [21] that 47 cultures carried genetic
143 material of *C. diphtheriae*, 43 of which could be isolated in pure culture.

144 Detection of the diphtheria toxin *tox* gene was positive in 40 out of 43 isolates; 3 isolates
145 were *tox*-negative by qPCR (**Table S1**), which was confirmed by genomic analyses. The Elek
146 *in-vitro* test for diphtheria toxin production demonstrated that 39 *tox*-positive isolates
147 produced the toxin (YEM0035 was Elek negative).

148 Phylogenomic analyses revealed the existence of six sublineages (labeled A to F, **Figure 3**).
149 One of them, called sublineage A, comprised 30 (70%) isolates (**Table S1**). The six
150 sublineages were unrelated to each other, as evidenced by their scattered positions in a
151 global phylogenetic tree of *C. diphtheriae* (**Figure S2**), demonstrating multiple occurrences of
152 diphtheria re-emergence or introductions into Yemen.

153 Sublineages A and B were of biovar Gravis, whereas the others were biovar Mitis. The
154 detection of the *spuA* gene, coding for a putative alpha-1,6-glycosidase [22] associated with
155 the glycogen positive phenotype of biovar Gravis, was consistent with these biochemical

156 assignments (**Figure 3A**) and showed that Yemen isolates belonged to two deep lineages of
157 *C. diphtheriae* marked by the presence or absence of this gene (**Figure S2**) as recently
158 reported [23].

159 Genotyping of the *C. diphtheriae* Yemen isolates by 7-gene multilocus sequence typing
160 (MLST, **Table S1**) revealed that sublineage A isolates belonged to ST384. The single
161 previously reported ST384 isolate (DIFT045; <https://pubmlst.org/cdiphtheriae/>) was isolated
162 in Belgium in 2016, and was collected in a patient who had close contact with a person
163 having traveled back from Saudi Arabia [24]. This observation hinted to a possible
164 epidemiological link between the main Yemen outbreak lineage and Saudi Arabia. To
165 reinforce this hypothesis, the genome of DIFT045 was sequenced, confirming its very close
166 genetic relatedness with sublineage A (**Figure 3**). Based on the global phylogenetic tree
167 (**Figure S2**), sublineage A was the Yemenite sublineage closest to the ex-Soviet Union 1990s
168 outbreak strain [25]. However, the large genetic distance between sublineage A isolates and
169 strain NCTC 13129, the reference for the ST8 ex-Soviet Union outbreak, rules out a direct
170 epidemiological link between the two outbreaks.

171 Sublineage E isolates were closely related to two isolates collected from migrants from
172 Somalia [26], and sublineage C isolates were closely related to isolates from siblings with a
173 Somalian origin [7]. Other Yemen sublineages were not phylogenetically closely related to
174 isolates from previously sequenced *C. diphtheriae* isolates (**Figure S2**).

175 We defined the genomic diversity of Yemen and DIFT045 isolates by clustering their protein-
176 coding genes (**Figure S3**). A total of 3619 gene families were identified, 1790 of which were
177 shared by all isolates (core genome). Genomes had 2277 genes on average (range 2234-
178 2427), and each of the six sublineages possessed 54 to 140 genes not observed in the other
179 sublineages, underlying the potential for large phenotypic variation among them.

180

181 **Microevolution of sublineage A**

182 We subdivided sublineage A into two branches, with branch A.1 being further subdivided
183 into A.1.1 and A.1.2 (**Figure 3**). The latter comprised the Belgium isolate DIFT045 and
184 YEM0063. The epidemiologically predominant branch A.1 was highly homogeneous
185 genetically, with a maximum of 86 genome-wide SNPs among its members (observed
186 between YEM0063 and YEM0089), indicating that A.1 isolates share a recent common
187 ancestor. The amount of nucleotide substitutions observed in isolates of this sublineage was

188 significantly related to their isolation time (**Figure S4**), demonstrating measurable
189 evolutionary divergence (**Figure 3B**), with a substitution rate estimated at 1.67×10^{-6}
190 substitutions per site per year (95%CI = 3.98×10^{-7} - 2.99×10^{-6}). The last common ancestor
191 of A.1.1 was estimated to have existed in March 2015 (Nov. 2003 - Sep. 2016). These results
192 indicate that this subgroup of Yemenite diphtheria isolates circulated and diversified silently
193 within the country several years before the diphtheria outbreak declaration in 2017.

194

195 **Antimicrobial susceptibility of Yemen *C. diphtheriae* isolates**

196 Antimicrobial susceptibility data were heterogeneous among Yemen sublineages (**Figure 3**,
197 **Table S2**). Importantly, full susceptibility or only low levels of penicillin resistance were
198 observed for all Yemen diphtheria isolates. While eight isolates (18.6%) were penicillin
199 resistant (**Table S2**), the two most resistant isolates had a minimal inhibitory concentration
200 (MIC) of 0.25 mg/L, which is only twice the clinical breakpoint of resistance (>0.125 mg/L).
201 Further, the MIC50 and MIC90 of penicillin were 0.094 and 0.19 mg/L, respectively (**Table**
202 **S2**). Amoxicillin and erythromycin were also active against all isolates, as were many other
203 tested molecules. Resistance to ciprofloxacin (but not to moxifloxacin) was observed in two
204 isolates, which had a D93G change in the quinolone-resistance determining region of their
205 *gyrA* gene. Resistance to trimethoprim was found in 32 (74.4%) isolates, including most
206 members of sublineage A (**Figure 3**), but none was resistant to the combination of this agent
207 with sulfamethoxazole (**Table S2**). Four isolates possessed homologs of the *sul1* gene, and
208 decreased phenotypic susceptibility to sulfonamide was observed in two of these. Resistance
209 to tetracycline was observed in 11 isolates, which all belonged to minor sublineages whereas
210 sublineage A was susceptible (**Figure 3A**); five of these isolates possessed known tetracycline
211 resistance genes encoding the *tet33* efflux pump or the *tetO* ribosomal protection protein.
212 Finally, the two isolates of sublineage B possessed an *aadA2* gene coding for an ANT(3")
213 aminoglycoside modifying enzyme targeting spectinomycin and streptomycin. Accordingly,
214 these two isolates were susceptible to gentamicin and kanamycin.

215

216 **Association of sublineage A with clinical signs and genomic features potentially linked to** 217 **virulence**

218 Clinical features of *C. diphtheriae* infections caused by isolates from sublineage A were
219 distinctive, with patients tending to be younger ($p=0.06$) and with more frequent swollen

220 neck ($p=0.002$; **Table 1**). We screened the genomic sequences for the presence of putative
221 pathogenicity-associated genes. Sublineage A isolates possessed genes *iusABC* coding for an
222 ABC-type iron uptake system, the *chtAB* genes that are homologous to *htaAB* genes for
223 hemin binding, a pathogenicity island of *C. diphtheriae* called PICD-11 [27] comprising a
224 putative collagen binding protein suggested to be associated with cutaneous or invasive
225 infections [28], and the *spaABC-strA* and *spaDEF-strBC* gene clusters coding for two distinct
226 types of fimbriae (**Figure 3**). In contrast to sublineage A, most other Yemen isolates did not
227 harbor the above gene clusters. In turn, their genomes comprised genes *sidBA-ddpABC* for
228 putative siderophore biosynthesis, *irp1ABCD* for a putative ABC-type iron transporter [29]
229 and the *irp2ABCDEFGHI* and *irp2JKLMN* operons. Genes coding for SpaH fimbriae were
230 absent from all Yemen genomes. Broader putative virulence genes variation was observed at
231 the global scale, with a marked dichotomy between the two major lineages defined by gene
232 *spuA* (**Figure S2**).

233 We further searched for sublineage A.1-enriched genes by analyzing the pangenome of
234 Yemen isolates together with the sublineage A-related strain NCTC13129. 77 genes were
235 enriched in sublineage A.1 (present in >95% A.1 isolates, and absent in >85% of other
236 isolates), among which six were uniquely found in all members of this sublineage (**Table S4**).
237 Among these, one gene was annotated as a 5-formyltetrahydrofolate cyclo-ligase gene, the
238 homolog of which was previously shown to increase the resistance level to antifolates in
239 *Mycobacterium smegmatis* and *Escherichia coli* [30]. Further analyses showed that this gene
240 is in fact found in all *Corynebacterium* genomes, but the sublineage A.1 version is shorter
241 (417 versus a median of 585 bp among all *Corynebacterium* species). Of note, the size of this
242 gene does not vary at all in *C. diphtheriae* except for isolates of the Yemen outbreak
243 sublineage A (585 bp). A possible causal link of this genetic truncation with resistance to
244 trimethoprim, which was observed mostly in sublineage A (but not solely A.1; **Figure 3**), will
245 be interesting to investigate in future studies.

246 Discussion

247 The outbreak of diphtheria in Yemen was first notified in October 2017 and its early
248 epidemiology was described [9, 11, 14]. Here, we provide an update on the epidemiological
249 situation, a clinical description of the characteristics of recorded suspected cases, and
250 detailed information on microbiological features of *C. diphtheriae* outbreak isolates.

251 Based on currently reported probable cases, the Yemen diphtheria outbreak is one of the
252 largest *C. diphtheriae* contemporaneous outbreaks. We identified three discernible waves
253 with distinctive geographic signatures. We note that reporting is very likely to be an
254 underestimate in the context of the Yemen health care disruption, and to be uneven
255 geographically given the heterogeneity in the continuation of health care facility operations.
256 Nevertheless, the eDEWS system [31] proved invaluable in the surveillance of diphtheria in
257 the conflict context.

258 Clinical characteristics were typical of diphtheria in terms of laryngeal manifestations,
259 pseudomembrane presence and neck swelling [32]. Yemenite patients were mostly young,
260 which is also typical, with the exception of the ex-Soviet Union event [4]. Vaccination is
261 highly effective against the clinical expression of toxigenic *C. diphtheriae* infections, and
262 countries with high population coverage have virtually eliminated endemic diphtheria [2]. In
263 Yemen, the vaccination coverage of the third dose of pentavalent vaccine in 2017 and 2018
264 (Figure S5) was <80% in several governorates, which is considered as insufficient to reach
265 population immunity [33]. We found that more than 50% of patients were reported to be
266 previously vaccinated. We did not have access to the exact vaccination status, but we can
267 assume that this observation reflects an incomplete vaccination without booster. This might
268 also explain the median age > 5 years of patients, comparable to other outbreaks in
269 populations where vaccination was incomplete [34][35]. From November 2017 to March
270 2018, a mass vaccination campaign targeted nearly 2.7 million children aged 6 weeks to 15
271 years in 11 governorates [36]. Here we found that from the first to the second wave, there
272 was a reduction in infection risk in 0-4 years old, consistent with some success from the
273 mass immunization program. In 2019, diphtheria vaccination was further conducted in 186
274 districts of the 12 Northern governorates [37], with 1.2 million children 6 weeks to 5 years of
275 age vaccinated with the Penta vaccine and 2.2 million children 5-15 years with the Td
276 vaccine. Vaccination against diphtheria in selected districts of Southern governorates and

277 Sa'adah has started in July 2020. However, despite efforts to control the disease, cases were
278 still being reported in 2020, and future studies will be needed to evaluate their effect.

279 This work represents a unique genomic analysis of a large contemporaneous outbreak of
280 diphtheria. A prominent observation was that at least six phylogenetic sublineages,
281 ancestrally unrelated within the global diversity of *C. diphtheriae*, contributed to the
282 resurgence of diphtheria in Yemen. This important observation shows that strains with
283 distinctive genomic and phenotypic characteristics can coexist within a single diphtheria
284 outbreak. The multiplicity of strains reflects a silent diversity reservoir of *C. diphtheriae*
285 despite vaccination, which is designed to protect against disease expression rather than
286 colonization and transmission. Antimicrobial resistance remains rare in *C. diphtheriae* [38],
287 and our results show that the main recommended agents (penicillin G, aminopenicillin and
288 erythromycin) are active against Yemen *C. diphtheriae* despite a low-level penicillin
289 resistance in some isolates. However, antimicrobial susceptibility profiles were
290 heterogeneous, with a strong phylogenetic sublineage effect, and were largely concordant
291 with genomic data, implying a possible future role for molecular testing of antimicrobial
292 resistance in *C. diphtheriae*.

293 Our work also uncovers a large heterogeneity in virulence-associated genomic features
294 among circulating isolates. The diphtheria toxin gene was observed in most isolates, which
295 were all confirmed to produce the toxin *in vitro* except one, which was non-toxigenic toxin
296 gene bearing (NTTB). However, three isolates were *tox*-negative, indicating that such isolates
297 can cause diphtheria-like respiratory symptoms. Currently, with the important exception of
298 diphtheria toxin, the links between virulence genes of *C. diphtheriae* strains and clinical
299 expression are unknown [39]. Whereas two Yemen sublineages are of biovar Gravis, four
300 were of biovar Mitis. Biovar Gravis isolates have been considered more virulent than biovar
301 Mitis isolates [40, 41]. Interestingly, our work reveals that the distribution of virulence-
302 associated genes is strongly contrasted between two major lineages of *C. diphtheriae*, which
303 were defined here by the presence of the Gravis-specific gene *spuA*. Our genomic data
304 suggest that inter-lineage heterogeneity in iron acquisition, adhesion and colonization
305 capacities may exist and could underlie epidemiological or clinical differences among them.
306 We found that patients infected with sublineage A, of biovar Gravis, tended to be younger
307 and with more frequent neck swelling. However, as neck swelling is more commonly
308 observed in young patients, we cannot disentangle sublineage and age effects here. The

309 underlying mechanisms of lymph node manifestations in diphtheria, described as nonspecific
310 acute lymphadenitis [1], are unknown. Future studies should investigate the impact of the
311 genomic diversity of *C. diphtheriae* on the pathogenicity and clinical expression of
312 diphtheria.

313 Genomic sequencing of bacterial pathogens is a powerful approach to define relationships
314 across time, sources and geography [42][15]. The *C. diphtheriae* genotypes observed in
315 Yemen were distinct from strains from the large ex-Soviet Union 1990s outbreak, whereas
316 no data exists on the strains involved in other recent outbreaks in Venezuela, Haiti and
317 Bangladesh. Migration and trade from the horn of Africa into Yemen are major drivers of
318 pathogenic strain spread, as observed for cholera [15]. The nearly complete lack of
319 information on circulating *C. diphtheriae* from neighboring countries has restricted our
320 ability to define the dynamics of geographic spread of *C. diphtheriae* in this world region.
321 However, a few *C. diphtheriae* genomes were linked epidemiologically and genetically to the
322 neighboring countries Somalia and Saudi Arabia.

323 We estimated the evolutionary rate of *C. diphtheriae* and found it is similar to other human
324 pathogens sampled over short time scales, such as *Staphylococcus aureus* or *Streptococcus*
325 *pyogenes* [43]. Our genomic data analyses thus imply that the diversification of the major
326 Yemen outbreak sublineage largely predates the detection of the outbreak by the disease
327 surveillance system. These observations concur with the multiplicity of sublineages that a
328 reservoir of *C. diphtheriae* diversity existed in Yemen before the outbreak recognition. We
329 note that our rate and age estimates have large confidence intervals, which is explained by
330 the small number of available genomes.

331 Although we provide a glimpse into the microbiological characteristics of diphtheria in
332 Yemen, the available samples of *C. diphtheriae* sent for biological confirmation currently do
333 not provide a complete picture of the outbreak. Samples referred to NCPHL were distributed
334 across a large time period but governorates of the Southern part of the country were largely
335 under-represented, as 91% of confirmed samples came from Northern governorates (**Table**
336 **S3**). This limitation reflects the current situation of the country, divided by conflict into South
337 and North parts.

338 In summary, this work sheds light on epidemiological and clinical aspects of the current
339 Yemen diphtheria outbreak. We demonstrate high phylogenetic, genomic and phenotypic
340 variation of *C. diphtheriae* within a single outbreak, suggesting that the characterization of

341 isolates at the level of individual patients or local chains of transmission is relevant for
342 patient management during diphtheria outbreaks. This underlines the need to reinforce
343 laboratory capacity in the context of diphtheria outbreaks. Concerted international efforts
344 should be developed aiming at enhancing surveillance and defining in real time the
345 characteristics of *C. diphtheriae* clinical isolates. Diphtheria is a largely forgotten disease, and
346 contemporaneous research into its pathophysiology is scarce. Further, in face of increasing
347 disruption of vaccination campaigns and the shortage of diphtheria antitoxin supply, studies
348 into the determinants of local persistence of *C. diphtheriae* and its spread at regional or
349 global scales are needed in order to better control the reemergence of this dreadful
350 pathogen.

351 **Methods**

352 **Definitions and surveillance data**

353 Probable cases were defined based on clinical examination: any person with illness
354 characterized by an adherent membrane on the tonsils, pharynx and/or nose and any one of
355 the following: laryngitis, pharyngitis or tonsillitis [31]. Confirmed cases were defined as a
356 probable case from which a *C. diphtheriae* isolate was cultivated in the NCPHL laboratory.
357 Probable cases were based on the health facility surveillance system and were compiled
358 from electronic diseases early warning system (eDEWS) releases of weekly reporting of
359 diphtheria cases [31]. For the vaccination coverage, we used the 2017 and 2018 annual
360 reports by the national Expanded Programme on Immunization (EPI), obtained from the
361 Ministry of Public Health and Population (MoPHP) of Yemen. The attack rate per age group
362 was calculated by dividing the number of cases of each age group by the size of the
363 population within each age group (source: [https://www.census.gov/data-](https://www.census.gov/data-tools/demo/idb/informationGateway.php)
364 [tools/demo/idb/informationGateway.php](https://www.census.gov/data-tools/demo/idb/informationGateway.php)).

365

366 **Biological confirmation, clinical data and external microbiological and genomic analyses**

367 A total of 888 throat swab samples were received by the NCPHL and analyzed using local
368 microbiological procedures (Supplementary data). Demographic data, prior vaccination,
369 signs and symptoms and current antibiotic of patients were collected for all samples. A
370 random selection of 98 samples available on July 2019 was sent to Institut Pasteur for
371 external analyses (Supplementary data).

372

373 **Data availability**

374 The whole genome sequencing data generated in this study were deposited in the European
375 Nucleotide Archive (ENA) database and are accessible through the study PRJEB34206
376 (<https://www.ebi.ac.uk/ena/data/view/PRJEB34206>). Raw read data are available under the
377 accessions ERR4332853-96. Contig sequences are available under the accessions
378 CAJDXH01000000-CAJDYY01000000.

379 **Ethical statement**

380 Data were collected as part of routine case management under an emergency response
381 mandate from the government of Yemen. The scientific committee of the NCPHL, Sana'a
382 authorized the usage of the clinical data included in this study.

383 **Author contributions**

384 A.H., the NCPHL diphtheria outbreak working group (A.A.R, M.A.A., E.M.A.-A., N.M.A.-M.,
385 K.A.A, H.Z. A.-S., A.A.A.-S.), A.C-L., E.B., M.P.P. and M.L. performed the microbiological
386 cultures of the isolates and their biochemical and molecular characterizations. F.D., G.D.,
387 N.L. and H.S. collated and analyzed the epidemiological data. A.H., G.D. and J.T. collated and
388 analyzed the clinical data. A.C., V.B., M.H., J.G., N.Z. and S.B. analyzed the genomic data.
389 H.M. provided the DIFT045 strain. S.B. and G. D. initially designed, and subsequently
390 coordinated, the project. S.B. designed the microbiological aspects of the study and wrote
391 the initial version of the manuscript. All authors provided input to the manuscript and
392 reviewed the final version.

393 **Conflicts of interests**

394 The authors declare no competing interests.

395 **Acknowledgments**

396 We thank Dr Taha Almotuakel (Yemen Ministry of Health) for continuous support, and
397 Isabelle Cailleau (Institut Pasteur) and the Medecins Sans Frontieres Organization for
398 logistical support. The NCPHL also acknowledges the World Health Organization for
399 providing some reagents used in this study. This work used the computational and storage
400 services (TARS cluster) provided by the IT department at Institut Pasteur.

401 **Role of the funding source**

402 This work was supported financially by institutional funding from the NCPHL-Sana'a and
403 Institut Pasteur. The French National Reference Center for Corynebacteria of the diphtheriae
404 complex received support from Institut Pasteur and Public Health France (Santé publique
405 France, Saint Maurice, France). This work was also supported financially by the French
406 Government's Investissement d'Avenir program Laboratoire d'Excellence "Integrative
407 Biology of Emerging Infectious Diseases" (ANR-10-LABX-62-IBEID). The funders had no role
408 in the study design, collection, analysis, interpretation of the data and manuscript writing.
409 Institut Pasteur and the NCPHL supported the decision to submit the paper for publication.

410 **Figure Legends**

411 **Figure 1. Epidemic curve and geographic and temporal origins of the isolates.**

412 A. Weekly numbers of suspected cases (purple), samples (light blue) and confirmed isolates
413 (dark blue) of diphtheria in Yemen from January 2017 to April 2020, with three epidemic
414 waves being highlighted. Sequenced isolates are shown below the epidemic curve, colored
415 according to their sublineages (see Figure 3). B and C. Geographical location of the 888
416 samples of this study (B) and of the 340 confirmed samples (C). D and E. Geographical
417 location of the suspected cases during the first epidemic wave (D) and the second epidemic
418 wave (E). The governorate maps of Yemen was created using a shape file approved by the
419 Humanitarian Country Team in October 2019 ([https://data.humdata.org/dataset/yemen-
420 admin-boundaries](https://data.humdata.org/dataset/yemen-admin-boundaries)).

421

422 **Figure 2. Age profile of suspected cases.**

423 A. Age distribution of suspected cases during the first and second waves. B. Attack rates per
424 age group, per wave. Red shows the first wave, green the second. The mean attack rate
425 during the two waves was 6.6 per 100,000.

426

427 **Figure 3. Phylogenetic analysis of Yemen isolates with their virulence and resistance
428 characteristics.**

429 A. Phylogenetic tree of the Yemen isolates, with the distinction of 6 sublineages (A to F;
430 sublineage A was subdivided into A.1 [A.1.1, A.1.2] and A.2). On the right of the tree, the
431 black squares denote a resistance phenotype, while a black circle denote the presence of the
432 *tox*-gene or other virulence-associated genes. Open squares denote an intermediate
433 resistance phenotype. Scale bar, 0.001 nucleotide substitutions per site. B. Detailed
434 phylogenetic tree of the sublineage A. Nodes' ages and 95%CI are shown at the nodes. Scale
435 bar, 0.000005 nucleotide substitutions per site (*i.e.*, 1 substitution every 200,000
436 nucleotides, or approximately 10 substitutions per genome).

437 References

- 438 1. **Hadfield TL, McEvoy P, Polotsky Y, Tzinerling VA, Yakovlev AA.** The pathology of diphtheria. *J*
439 *Infect Dis* 2000;181 Suppl 1:S116-120.
- 440 2. **Sharma NC, Efstratiou A, Mokrousov I, Mutreja A, Das B, et al.** Diphtheria. *Nat Rev Dis Primer*
441 2019;5:81.
- 442 3. **Patey O, Bimet F, Riegel P, Halioua B, Emond JP, et al.** Clinical and molecular study of
443 *Corynebacterium diphtheriae* systemic infections in France. Coryne Study Group. *J Clin Microbiol*
444 1997;35:441–445.
- 445 4. **Dittmann S, Wharton M, Vitek C, Ciotti M, Galazka A, et al.** Successful control of epidemic
446 diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. *J*
447 *Infect Dis* 2000;181 Suppl 1:S10-22.
- 448 5. **Rahman MR, Islam K.** Massive diphtheria outbreak among Rohingya refugees: lessons learnt. *J*
449 *Travel Med*;26. Epub ahead of print 1 January 2019. DOI: 10.1093/jtm/tay122.
- 450 6. **Paniz-Mondolfi AE, Tami A, Grillet ME, Márquez M, Hernández-Villena J, et al.** Resurgence of
451 Vaccine-Preventable Diseases in Venezuela as a Regional Public Health Threat in the Americas.
452 *Emerg Infect Dis* 2019;25:625–632.
- 453 7. **Meinel DM, Kuehl R, Zbinden R, Boskova V, Garzoni C, et al.** Outbreak investigation for
454 toxigenic *Corynebacterium diphtheriae* wound infections in refugees from Northeast Africa and
455 Syria in Switzerland and Germany by whole genome sequencing. *Clin Microbiol Infect Off Publ*
456 *Eur Soc Clin Microbiol Infect Dis* 2016;22:1003.e1-1003.e8.
- 457 8. **Scheifer C, Rolland-Debord C, Badell E, Reibel F, Aubry A, et al.** Re-emergence of
458 *Corynebacterium diphtheriae*. *Med Mal Infect* 2019;49:463–466.
- 459 9. **Dureab F, Müller O, Jahn A.** Resurgence of diphtheria in Yemen due to population movement. *J*
460 *Travel Med*;25. Epub ahead of print 1 January 2018. DOI: 10.1093/jtm/tay094.
- 461 10. **Dureab F, Al-Sakkaf M, Ismail O, Kuunibe N, Krisam J, et al.** Diphtheria outbreak in Yemen: the
462 impact of conflict on a fragile health system. *Confl Health* 2019;13:19.
- 463 11. **WHO.** Diphtheria – Yemen. *Disease Outbreak News*. [https://www.who.int/csr/don/22-](https://www.who.int/csr/don/22-december-2017-diphtheria-yemen/en/)
464 [december-2017-diphtheria-yemen/en/](https://www.who.int/csr/don/22-december-2017-diphtheria-yemen/en/) (2017).
- 465 12. **OCHA.** Yemen: Diphtheria Outbreak - Nov 2017. [https://reliefweb.int/disaster/ep-2017-000175-](https://reliefweb.int/disaster/ep-2017-000175-yem)
466 [yem](https://reliefweb.int/disaster/ep-2017-000175-yem) (2017).
- 467 13. **Meyer D.** Yemen-- First cholera, now diphtheria.
468 [https://www.outbreakobservatory.org/outbreakthursday-1/11/30/2017/yemen-first-cholera-](https://www.outbreakobservatory.org/outbreakthursday-1/11/30/2017/yemen-first-cholera-now-diphtheria)
469 [now-diphtheria](https://www.outbreakobservatory.org/outbreakthursday-1/11/30/2017/yemen-first-cholera-now-diphtheria) (2017).
- 470 14. **WHO-EMRO.** Weekly Epidemiological Monitor, Volume 10, Issue 47 (19 November 2017).
471 [https://reliefweb.int/report/yemen/who-emro-weekly-epidemiological-monitor-volume-10-](https://reliefweb.int/report/yemen/who-emro-weekly-epidemiological-monitor-volume-10-issue-47-19-november-2017)
472 [issue-47-19-november-2017](https://reliefweb.int/report/yemen/who-emro-weekly-epidemiological-monitor-volume-10-issue-47-19-november-2017) (2017).
- 473 15. **Weill F-X, Domman D, Njamkepo E, Almesbahi AA, Naji M, et al.** Genomic insights into the
474 2016-2017 cholera epidemic in Yemen. *Nature* 2019;565:230–233.

- 475 16. **Dureab F, Al-Awlaqi S, Jahn A.** COVID-19 in Yemen: preparedness measures in a fragile state.
476 *Lancet Public Health* 2020;5:e311.
- 477 17. **Dhabaan GN, Al-Soneidar WA, Al-Hebshi NN.** Challenges to testing COVID-19 in conflict zones:
478 Yemen as an example. *J Glob Health* 2020;10:010375.
- 479 18. **eDEWS.** Electronic Integrated Disease Early Warning and Response System, Volume 08, Issue19,
480 Epi week 19,(04- 10 May,2020).
- 481 19. **Wagner KS, Stickings P, White JM, Neal S, Crowcroft NS, et al.** A review of the international
482 issues surrounding the availability and demand for diphtheria antitoxin for therapeutic use.
483 *Vaccine* 2009;28:14–20.
- 484 20. **Kneen R, Pham NG, Solomon T, Tran TM, Nguyen TT, et al.** Penicillin vs. erythromycin in the
485 treatment of diphtheria. *Clin Infect Dis Off Publ Infect Dis Soc Am* 1998;27:845–850.
- 486 21. **Badell E, Guillot S, Tulliez M, Pascal M, Panunzi LG, et al.** Improved quadruplex real-time PCR
487 assay for the diagnosis of diphtheria. *J Med Microbiol* 2019;68:1455–1465.
- 488 22. **Santos AS, Ramos RT, Silva A, Hirata R, Mattos-Guaraldi AL, et al.** Searching whole genome
489 sequences for biochemical identification features of emerging and reemerging pathogenic
490 *Corynebacterium* species. *Funct Integr Genomics* 2018;18:593–610.
- 491 23. **Hennart M, Panunzi LG, Rodrigues C, Gaday Q, Baines SL, et al.** Population genomics and
492 antimicrobial resistance in *Corynebacterium diphtheriae*. *bioRxiv* 2020;2020.05.19.101030.
- 493 24. **Martini H, Soetens O, Litt D, Fry NK, Detemmerman L, et al.** Diphtheria in Belgium: 2010-2017. *J*
494 *Med Microbiol* 2019;68:1517–1525.
- 495 25. **Cerdeño-Tárraga AM, Efstratiou A, Dover LG, Holden MTG, Pallen M, et al.** The complete
496 genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res*
497 2003;31:6516–6523.
- 498 26. **Berger A, Dangel A, Schober T, Schmidbauer B, Konrad R, et al.** Whole genome sequencing
499 suggests transmission of *Corynebacterium diphtheriae*-caused cutaneous diphtheria in two
500 siblings, Germany, 2018. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull*;24. Epub
501 ahead of print January 2019. DOI: 10.2807/1560-7917.ES.2019.24.2.1800683.
- 502 27. **Trost E, Tauch A.** Comparative genomics and pathogenicity islands of *Corynebacterium*
503 *diphtheriae*, *Corynebacterium ulcerans*, and *Corynebacterium pseudotuberculosis*. In:
504 *Corynebacterium diphtheriae and related toxigenic species: genomics, pathogenicity and*
505 *applications*. Burkovski, A. Ed.; Springer; 2014. pp. 39–65.
- 506 28. **Timms VJ, Nguyen T, Crighton T, Yuen M, Sintchenko V.** Genome-wide comparison of
507 *Corynebacterium diphtheriae* isolates from Australia identifies differences in the Pan-genomes
508 between respiratory and cutaneous strains. *BMC Genomics* 2018;19:869.
- 509 29. **Qian Y, Lee JH, Holmes RK.** Identification of a DtxR-regulated operon that is essential for
510 siderophore-dependent iron uptake in *Corynebacterium diphtheriae*. *J Bacteriol* 2002;184:4846–
511 4856.
- 512 30. **Ogwang S, Nguyen HT, Sherman M, Bajaksouzian S, Jacobs MR, et al.** Bacterial conversion of
513 folinic acid is required for antifolate resistance. *J Biol Chem* 2011;286:15377–15390.

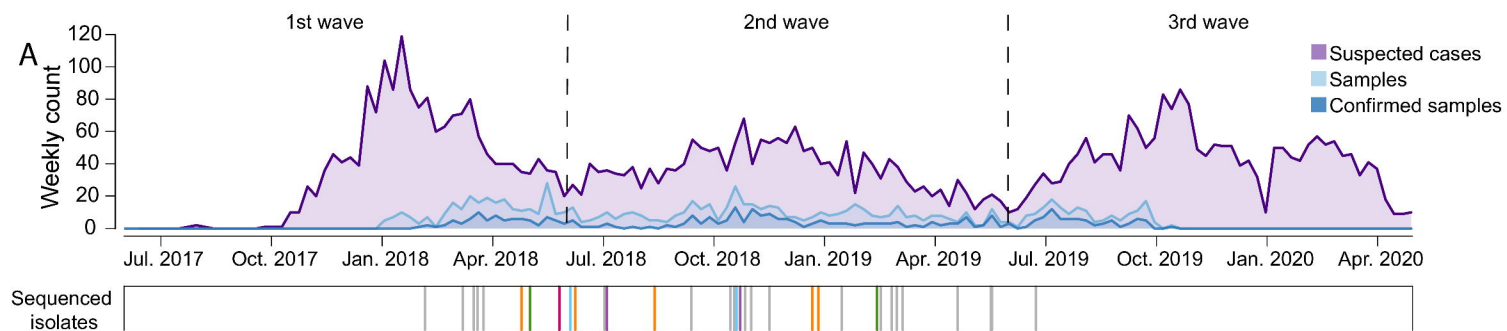
- 514 31. **Dureab F.** *The usefulness of the electronic Disease Early Warning System (eDEWS) in the*
515 *humanitarian crisis of Yemen.* University of Heidelberg. [https://archiv.ub.uni-](https://archiv.ub.uni-heidelberg.de/volltextserver/27986/)
516 [heidelberg.de/volltextserver/27986/](https://archiv.ub.uni-heidelberg.de/volltextserver/27986/) (2019).
- 517 32. **Dash N, Verma S, Jayashree M, Kumar R, Vaidya PC, et al.** Clinico-epidemiological profile and
518 predictors of outcome in children with diphtheria: a study from northern India. *Trop Doct*
519 2019;49:96–101.
- 520 33. **Blumberg LH, Prieto MA, Diaz JV, Blanco MJ, Valle B, et al.** The Preventable Tragedy of
521 Diphtheria in the 21st Century. *International journal of infectious diseases*: IJID: official
522 *publication of the International Society for Infectious Diseases*;71. Epub ahead of print June 2018.
523 DOI: 10.1016/j.ijid.2018.05.002.
- 524 34. **Murhekar M.** Epidemiology of Diphtheria in India, 1996-2016: Implications for Prevention and
525 Control. *Am J Trop Med Hyg* 2017;97:313–318.
- 526 35. **Galazka A.** The changing epidemiology of diphtheria in the vaccine era. *J Infect Dis* 2000;181
527 Suppl 1:S2-9.
- 528 36. **OCHA-services.** Diphtheria vaccination campaign for 2.7 million children concludes in Yemen.
529 [https://reliefweb.int/report/yemen/diphtheria-vaccination-campaign-27-million-children-](https://reliefweb.int/report/yemen/diphtheria-vaccination-campaign-27-million-children-concludes-yemen)
530 [concludes-yemen](https://reliefweb.int/report/yemen/diphtheria-vaccination-campaign-27-million-children-concludes-yemen) (2018).
- 531 37. **WHO-EMRO.** Situation Report, September 2019, Issue NO.9: Yemen Conflict.
532 <https://reliefweb.int/sites/reliefweb.int/files/resources/Yem-Sitrep-Sept-2019.pdf> (2019).
- 533 38. **Zasada AA.** Antimicrobial susceptibility and treatment. In: *Corynebacterium diphtheriae and*
534 *related toxigenic species genomics, pathogenicity and applications.* Burkovski A., editor. New
535 York: Springer; 2014. pp. 239–246.
- 536 39. **Sangal V, Hoskisson PA.** Evolution, epidemiology and diversity of *Corynebacterium diphtheriae*:
537 New perspectives on an old foe. *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis*
538 2016;43:364–370.
- 539 40. **McLeod JW.** THE TYPES MITIS, INTERMEDIUS AND GRAVIS OF CORYNEBACTERIUM DIPHTHERIAE:
540 A Review of Observations during the Past Ten Years. *Bacteriol Rev* 1943;7:1–41.
- 541 41. **Barksdale L.** *Corynebacterium diphtheriae* and its relatives. *Bacteriol Rev* 1970;34:378–422.
- 542 42. **Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, et al.** Whole genome-based
543 population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat Microbiol*
544 2016;2:16185.
- 545 43. **Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, et al.** Genome-scale rates of evolutionary
546 change in bacteria. *Microb Genomics* 2016;2:e000094.

547

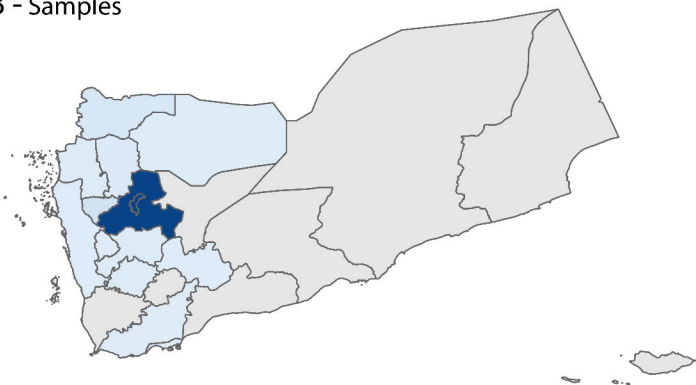
548 **Table 1. Clinical characteristics of patients infected by sublineage A or other sublineages**
549 **circulating in Yemen**

Characteristics, (%) otherwise stated	Sublineage A N=28	Other sublineages N=14	P value*
Age, years, median (range)	8 (2-35)	14 (2-35)	0.06
Male	14 (50)	5 (36)	0.38
History of previous vaccination	11 (44)	7 (64)	0.24
Fever	28 (100)	14 (100)	-
Pseudomembrane	28 (100)	14 (100)	
Neck swelling	22 (84)	4 (33)	0.002

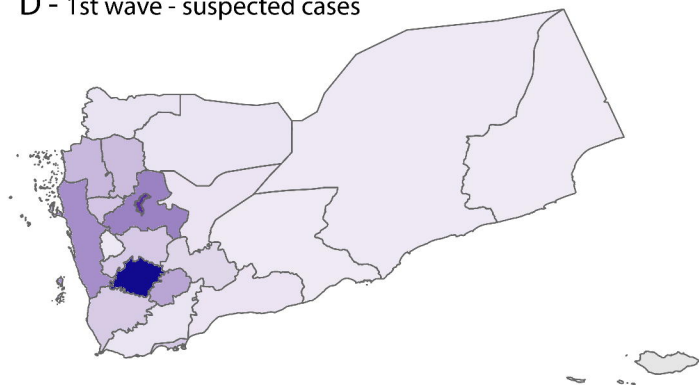
550 * Differences between cases infected with *C. diphtheriae* sublineage A compared to the ones
551 infected with isolates from other sublineages (B, C, D, E or F) were assessed by the Mann-
552 Whitney U test, the Student's t test, the χ^2 method, and Fisher's exact test where
553 appropriate.



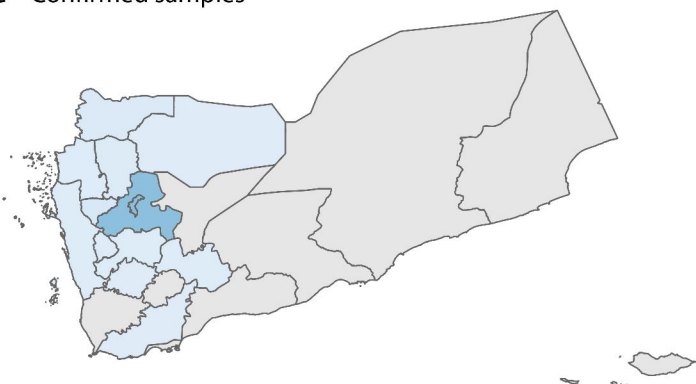
B - Samples



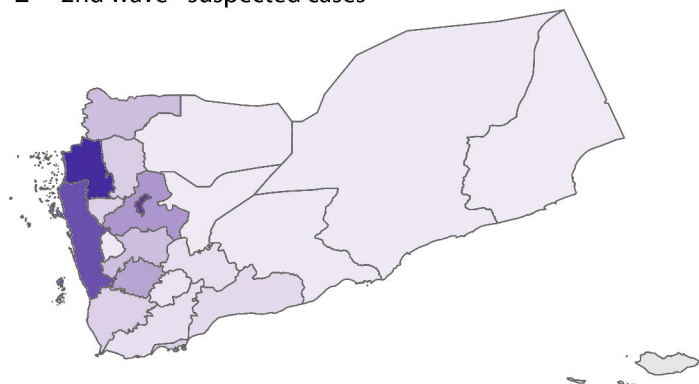
D - 1st wave - suspected cases



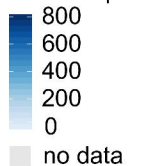
C - Confirmed samples



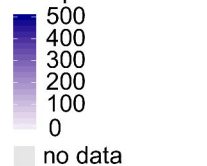
E - 2nd wave - suspected cases



Number of samples

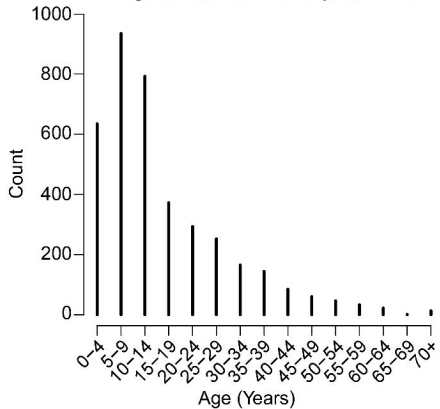


Number of suspected cases



0 100 200 300 400km

A Age distribution of suspected cases



B Changes in attack rates per age group

