

## Title:

Improving diagnosis of non-malarial fevers in Senegal: *Borrelia* and the contribution of tick-borne bacteria

## Authors:

Zoë C. Levine<sup>1,2,3</sup>, Aita Sene<sup>4</sup>, Winnie Mkandawire<sup>1,5</sup>, Awa B. Deme<sup>4</sup>, Tolla Ndiaye<sup>4</sup>, Mouhamad Sy<sup>4</sup>, Amy Gaye<sup>4</sup>, Younouss Diedhiou<sup>4</sup>, Amadou M. Mbaye<sup>4</sup>, Ibrahima Ndiaye<sup>4</sup>, Jules Gomis<sup>4</sup>, Médoune Ndiop<sup>6</sup>, Doudou Sene<sup>6</sup>, Marietou Faye Paye<sup>1</sup>, Bronwyn MacInnis<sup>1,7</sup>, Stephen F. Schaffner<sup>1,7,8</sup>, Daniel J. Park<sup>1</sup>, Aida S. Badiane<sup>4</sup>, Andres Colubri<sup>1,5</sup>, Mouhamadou Ndiaye<sup>4</sup>, Ngayo Sy<sup>9</sup>, \*Pardis C. Sabeti<sup>1,7,8,10</sup>, \*Daouda Ndiaye<sup>4</sup>, \*Katherine J. Sidle<sup>1,11</sup>

## Affiliations

<sup>1</sup>Broad Institute of Harvard and MIT, Cambridge, MA, USA.

<sup>2</sup>Harvard Graduate Program in Biological and Biomedical Science, Boston, MA, USA.

<sup>3</sup>Harvard/MIT MD-PhD Program, Boston, MA, 02115, USA.

<sup>4</sup>Centre International de recherche, de formation en Génomique Appliquée et de Surveillance Sanitaire (CIGASS), Dakar, Senegal.

<sup>5</sup>University of Massachusetts Medical School, Worcester, MA, USA.

<sup>6</sup>Programme National de Lutte contre le Paludisme (PNLP), Ministère de la Santé, Dakar Fann, Senegal.

<sup>7</sup>Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA.

<sup>8</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA.

<sup>9</sup>Service de Lutte Anti Parasitaire, Thies, Senegal.

<sup>10</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA.

<sup>11</sup>Department of Molecular Microbiology and Immunology, Brown University, Providence, RI, USA.

\*Corresponding authors

## 1 Abstract

2  
3 The worldwide decline in malaria incidence is revealing the extensive burden of non-  
4 malarial febrile illness (NMF), which remains poorly understood and difficult to diagnose. To  
5 characterize NMF in Senegal, we collected venous blood and clinical metadata from febrile  
6 patients and healthy controls in a low malaria burden area. Using 16S and unbiased  
7 sequencing, we detected viral, bacterial, or eukaryotic pathogens in 29% of NMF cases.  
8 Bacteria were the most common, with relapsing fever *Borrelia* and spotted fever *Rickettsia*  
9 found in 15% and 3.7% of cases, respectively. Four viral pathogens were found in a total of 7  
10 febrile cases (3.5%). Sequencing also detected undiagnosed *Plasmodium*, including one  
11 putative *P. ovale* infection. We developed a logistic regression model to distinguish *Borrelia*  
12 from NMFs with similar presentation based on symptoms and vital signs. These results  
13 highlight the challenge and importance of improved diagnostics, especially for *Borrelia*, to  
14 support diagnosis and surveillance.

## 15 **Main text**

16

### 17 **Introduction**

18 Febrile illness is a significant cause of morbidity and mortality in West Africa. While  
19 malaria remains the most common single pathogen causing febrile illness, its incidence has  
20 decreased sharply in the last two decades; in Senegal, for example, control measures  
21 decreased malaria incidence from 122 per 1000 in 2006 to 59 per 1000 in 2021<sup>1,2</sup>. As the  
22 malaria burden has decreased, the importance of non-malarial febrile illness (NMFI) has  
23 become more apparent. Unlike malaria, however, many of the pathogens causing NMFI are not  
24 the target of robust surveillance programs, and rapid diagnostic tests (RDTs), the backbone of  
25 both clinical care and surveillance, are not available<sup>3</sup>.

26 A broad range of pathogens across kingdoms can cause NMFI, but surveillance studies  
27 often focus on one or a few pathogens and therefore are limited in their ability to detect all  
28 causes of disease. Causal pathogens include arboviruses, such as Dengue virus and  
29 Chikungunya virus, which cause both sporadic febrile illness and outbreaks in Senegal<sup>4,5</sup>.  
30 Bacterial pathogens are a common cause of severe illness requiring hospitalization, including  
31 several vaccine-preventable illnesses such as *Streptococcus pneumoniae* and *Neisseria*  
32 *meningitidis*<sup>6</sup> and bacterial zoonoses are increasingly recognized as a common cause of  
33 ambulatory febrile illness, including tick and louse-borne relapsing fevers, spotted fever, Q  
34 fever, Leptospirosis and Brucellosis<sup>6-8</sup>. While fungal infections appear less common, *Candida*  
35 and *Cryptococcus* have been reported across West Africa, with cryptococcal disease being  
36 particularly common in HIV-positive patients<sup>6</sup>. Non-malarial parasitic diseases including  
37 trypanosomiasis and leishmaniasis are also common<sup>6</sup>.

38 In the absence of comprehensive surveillance efforts that capture multiple pathogen  
39 types, appropriate public health interventions are hindered by our limited understanding of the  
40 full landscape of common causes of NMFI in the region. Unbiased sequencing, also known as  
41 metagenomic sequencing (mNGS), is a powerful tool for detection of microbial nucleic acids in  
42 clinical samples without *a priori* knowledge of a pathogen<sup>9</sup> and is increasingly used for  
43 surveillance in regions at high risk of emerging and reemerging disease<sup>10-13</sup>. However, the low  
44 abundance of pathogen nucleic acids relative to host and commensal microbial material  
45 requires deep sequencing and can lead to false conclusions<sup>14,15</sup>. Unbiased approaches are  
46 complemented by more targeted strategies, such as 16S sequencing for bacterial pathogens<sup>16</sup>,  
47 that specifically amplify pathogen nucleic acids reducing required sequencing depth and  
48 simplifying interpretation. This technique has been applied to detect bacterial bloodstream  
49 infections<sup>17,18</sup> and tick-borne bacterial illness<sup>19</sup>.

50 This study aimed to determine and characterize the major causes of NMFI in Thiès,  
51 Senegal, a peri-urban community with overall low malaria incidence where malaria transmission  
52 dynamics have been deeply characterized<sup>20</sup>. We collected plasma samples from a cohort of  
53 acutely febrile patients presenting to the Service de Lutte Anti Parasitaire (SLAP) clinic and  
54 healthy controls from the surrounding neighborhoods during the dry and rainy seasons of 2018  
55 and 2019. We aimed to provide insights on clinical presentations of NMFI that can guide  
56 providers at the point of care and genomic characterization to inform design of new detection  
57 tools for clinical diagnosis and public health surveillance.

58

## 59 Results

### 60 **Broad characterization of NMFI reveals viral, bacterial, and eukaryotic pathogens**

61 We first comprehensively characterized the pathogens in plasma samples collected from  
62 acutely febrile patients suspected of malaria and healthy controls across the dry (febrile: n =  
63 100, healthy: n = 54) and rainy (febrile: n = 104, healthy: n = 50) seasons in 2019 (Fig. 1A,  
64 Supplementary Fig. 1B). All febrile patients received a malaria RDT; 39.4% (41/104) of patients  
65 tested positive for malaria in the wet season and no cases were detected in the dry season.  
66 Febrile cases were roughly equally split between children and adults (50.4%  $\geq$  18yrs). Age  
67 distribution was similar ( $p = 0.98$ , Mann-Whitney, two-sided) across case and control groups,  
68 but there were more male febrile cases (Male: n = 110, Female: n = 91) and more female  
69 controls (Male: n = 47, Female: n = 57,  $p = 0.15$  Fisher exact two-sided, Supplementary Fig. 1A).

70 We performed broad detection of viral, bacterial, and eukaryotic pathogens by  
71 sequencing. To detect viral pathogens, we performed RNA-mNGS. We also evaluated the  
72 ability of RNA-mNGS to detect non-viral pathogens, including malaria and fungi. To detect  
73 bacterial infections, we sequenced the v1-2 region of the 16S rRNA gene, which permitted us to  
74 classify the bacterial taxa present in samples with high bacterial load [see methods]. We  
75 detected at least one pathogen in 24% (39/163) of RDT-negative acutely febrile patients and co-  
76 infections in 7% (3/41) of RDT-positive acutely febrile patients (Fig. 1B). The most common  
77 febrile pathogen was *Borrelia*, which was found across the dry (n = 8) and rainy (n = 10)  
78 seasons and in both RDT-negative (n = 17) and RDT-positive (n = 1) patients. We also detected  
79 bacterial infections with *Rickettsia*, *Arcobacter*, *Actinomyces* and *Brevibacterium* (Fig. 1D).  
80 Bacterial pathogens found as co-infections included three *Borrelia/Rickettsia* co-infections, two  
81 *Borrelia/Plasmodium* co-infections, and three *Plasmodium/Rickettsia* co-infections (Fig. 1E).

82 We identified four known vertebrate viruses, Dengue virus (DENV, n = 2), Hepatitis B  
83 virus (HBV, n = 2), Parvovirus B19 (n = 2), and Human immunodeficiency virus 1 (HIV-1, n = 1)  
84 as well as a human virus not currently believed to cause disease, Pegivirus C (GBV-C, n = 4)  
85 (Fig. 1C). Four febrile patients and one healthy control exhibited a high proportion of fungal RNA  
86 reads in the plasma, as compared to healthy controls ( $>$  99th percentile reads/million raw reads  
87 (rpm) for healthy controls), but these reads did not map to any specific fungal pathogen  
88 (Supplementary Fig. 8A). We did not detect any vertebrate viruses in healthy controls, but we  
89 did detect *Borrelia* (n = 1/35 samples sequenced) and *Rickettsia* (n = 1/35).

### 91 **RNA-mNGS identifies viral pathogens known to circulate in Senegal**

92 We next considered the genetic diversity of detected viral pathogens to determine their  
93 relationship to other circulating strains in West Africa. Phylogenetic analysis of complete DENV,  
94 HBV and Parvovirus B-19 genomes indicated that in all three cases, the two isolates were  
95 different genotypes, and thus not closely related (Supplementary Table 1). Notably, there was a  
96 DENV outbreak across Senegal, including Thiès, in 2018. Whole genome phylogenetic analysis  
97 revealed the closest relatives to the 2019 DENV3 isolate from this study were DENV3 genomes  
98 from patients presenting to the SLAP clinic during that outbreak<sup>4</sup> (Supplementary Fig. 2A).  
99 Conversely, the DENV1 genome in this study was more closely related to other isolates from  
100 West Africa than to the DENV1 isolates from the 2018 outbreak (Supplementary Fig. 2A). The  
101 HIV-1 isolate genome could not be assembled and genotyped due to low read count (mean

102 coverage 0.2), but reads mapped across the reference genome including in the *gag*, *pol*, and  
103 *env* genes.

104 While GBV-C is known to infect asymptomatic individuals across the world, we only  
105 detected GBV-C in febrile patients. The prevalence of GBV-C has not been well studied in  
106 Senegal; we found infection in only 1.3% (4/307) of plasma samples sequenced, lower than the  
107 4-11% reported in blood-donor surveys from Sub-Saharan Africa<sup>21,22</sup>. All four isolates were  
108 clustered with previously sequenced human isolates from Sierra Leone, Uganda and Cameroon  
109 and belonged to Genotype I (Supplementary Fig. 2B), the most commonly circulating GBV-C  
110 genotype in West Africa<sup>23</sup>.

111 In order to identify divergent viral species missed at the read level, we performed a  
112 translated nucleic acid search of *de novo* contigs which identified RNA-dependent RNA  
113 Polymerase (RdRp) sequences for two candidate novel viruses, one in the *Narnaviridae* family  
114 and one in the *Reoviridae* family. We detected a *Reoviridae* RdRp sequence across 18 febrile  
115 patients, 12 healthy controls, and 4 non-template controls, suggesting a likely contaminant  
116 (Supplementary Fig. 3). However, given that *Reoviruses* have been isolated from ill patients, we  
117 further investigated these sequences<sup>24</sup>. The *Reovirus* RdRp sequences from this study  
118 clustered together, but were distant from mammalian *Orthoreoviruses*. Given the distance from  
119 mammalian *Orthoreovirus* species and presence in non-template controls, these sequences  
120 were likely from a contaminated reagent rather than true human infections. We also identified a  
121 *Narnaviridae* RdRp sequence across 5 febrile patients and 1 healthy control (Supplementary  
122 Fig. 4), none of whom had any other identifiable pathogen. Because viruses in the *Narnaviridae*  
123 family, which infect plants and fungi, have not been reported as human pathogens and we  
124 detected this novel *Narnavirus* species in both febrile cases and a healthy control, the role of  
125 this virus in disease, if any, is unclear.

126

### 127 ***Unbiased RNA sequencing detects Plasmodium cases missed by RDT***

128 If RNA-mNGS is to be employed as a tool for diagnosis of febrile illness in malaria  
129 endemic regions, we also need to understand the ability of RNA-mNGS to detect malaria as  
130 compared to clinical diagnostics. We know that RDTs have limited sensitivity, detect only certain  
131 *Plasmodium* species, and are susceptible to false negatives in genetically diverse parasites, so  
132 we evaluated our RNA-mNGS reads for evidence of *Plasmodium* infections.

133 RNA-mNGS detected abundant *Plasmodium* nucleic acids (>550 rpm) in 6.8% (11/160)  
134 of RDT negative acutely febrile patients, suggesting these were false negatives. Among RDT  
135 positive patients, there was a wide range in the abundance of *Plasmodium* reads (mean  
136  $1.7 \times 10^4$  rpm, range  $1.5 \times 10^2$  -  $2.6 \times 10^5$  rpm) (Supplementary Fig. 5A). Overall, 63% of RDT  
137 positive patients (26/41) and 77% of smear positive patients (17/22) were also positive by  
138 unbiased RNA sequencing (>550 rpm) (Supplementary Fig. 5C). In patients with detectable  
139 parasitemia by blood smear examination, *Plasmodium* reads did not correlate strongly with  
140 parasite density (Pearson's R = 0.49, Supplementary Fig. 5B).

141 In order to determine why 11 samples were negative by RDT but positive by RNA-  
142 mNGS, we first determined whether any samples belonged to species other than *Plasmodium*  
143 *falciparum*, the target species of the RDTs (Bioline Malaria Ag P.f., Abbot). As previous work  
144 has demonstrated Kraken2 is inaccurate for species-level classification of parasites<sup>25</sup>, we used  
145 DIAMOND to identify one sample with a contig (771 bp) that matched perfectly (100%

146 nucleotide identity, 100% coverage) to *Plasmodium ovale* cytochrome oxidase subunit 1  
147 (Accession: KP050416.1), suggesting a likely *P. ovale* infection. Additionally, we assessed  
148 expression levels of histidine-rich protein 2 (HRP-2), the target antigen of the RDTs used in this  
149 study. Deletions in *pfhrp2* gene have been detected in Senegal and demonstrated to cause  
150 decreased antigen expression and consequently false negative RDT results<sup>26,27</sup>. The number of  
151 RNA-mNGS rpm aligned to *pfhrp2* was significantly different between patients who were RDT  
152 positive and mNGS positive (RDT+/mNGS+) and patients who were RDT negative but  
153 *Plasmodium* RNA-mNGS positive (RDT-/mNGS+) (Supplementary Fig. 5D, Mann Whitney two-  
154 sided,  $p = 0.038$ ). However, given the overlap in expression levels between the two groups and  
155 the low frequency of *pfhrp2* deletion in Senegal, previously estimated at 2.4%<sup>27</sup>, *pfhrp2* deletion  
156 likely explains only some of the false negative RDTs.

157

### 158 **16S sequencing confirms a high burden of Relapsing Fever *Borrelia* and Spotted Fever** 159 ***Rickettsia***

160 As bacteria were the most common cause of NMFI in our 2019 cohort, we extended our  
161 16S sequencing to detect bacterial infections across the complete cohort from 2018-2019, and  
162 further investigated the species causing disease. Among all high bacterial load samples that  
163 underwent 16S sequencing [see methods], we detected *Borrelia* in 15.5% (33/213) and  
164 *Rickettsia* in 5.1% (9/213) of febrile patients. These arthropod-borne bacterial pathogens are  
165 known to circulate in Senegal and cause febrile illness (Fig. 1D, Supplementary Fig. 6C)<sup>8</sup>. 16S  
166 sequencing also identified other bacterial genera containing known pathogens, including  
167 *Bacteroides*, *Actinomyces*, *Brevibacterium*, *Arcobacter* and *Veillonella* across both years  
168 (Supplementary Fig. 6C). However, given the limited taxonomic resolution of v1-2 16S  
169 sequencing, we were unable to determine whether these sequences represented pathogenic  
170 species or harmless commensals.

171 Phylogenetic clustering of *Borrelia* v1-2 16S sequences showed that all *Borrelia*  
172 sequences from this study were similar to each other and fell within the Relapsing Fever (RF)  
173 group (Supplementary Fig. 7A). RF *Borrelia* circulate worldwide and are known to cause febrile  
174 illness in Senegal<sup>28</sup>. All the *Rickettsial* v1-2 16S rRNA gene sequences clustered within the  
175 Spotted Fever group (Supplementary Fig. 7B). Spotted fever *Rickettsia*, including *R. felis*<sup>29</sup> and  
176 *R. africae*<sup>8</sup> have been previously detected in febrile Senegalese patients. We did not detect any  
177 Typhus group *Rickettsia* in our cohort, though *R. typhi* has been reported in an immigrant  
178 traveling from Senegal<sup>30</sup>.

179 A subset of 6 *Borrelia* isolates were typed by amplicon sequencing of the 16S-23S  
180 intergenic spacer (IGS)<sup>31</sup>. All typed isolates were *Borrelia crociduræ* (Fig. 2C). Isolates from this  
181 study were more similar to previously characterized isolates from humans and *Ornithodoros*  
182 *sonrai* in Southern Senegal (unpublished) and *Ornithodoros sonrai* ticks in Mali<sup>32</sup> than isolates  
183 from *Ornithodoros erraticus* in Tunisia<sup>33</sup> and Morocco<sup>34</sup> (Fig. 2C).

184

### 185 **16S sequencing enables detection of low-titer infections**

186 We assessed the ability of RNA-mNGS to detect bacterial infections, based on samples  
187 that underwent both sequencing methods. *Borrelia* RNA-mNGS reads were detected in 68%  
188 (13/19) of 16S positive samples. Conversely, *Rickettsia* RNA-mNGS reads were not detected in  
189 any *Rickettsia* positive samples. Acute RF *Borrelia* infection is known to cause high titers of

190 bacteria in the blood during febrile episodes, while *Rickettsia* is an obligate intraerythrocytic  
191 pathogen and therefore has low titers in the plasma. To assess the extent to which plasma titers  
192 impacted the sensitivity of RNA-mNGS to tick-borne bacterial pathogens, we quantified bacterial  
193 abundance by the percent of total v1-2 16S sequences from a given sample classified as  
194 *Borrelia* or *Rickettsia*, respectively<sup>19</sup>. We observed *Borrelia* abundances ranging from 5-98%,  
195 while *Rickettsia* abundances were lower, ranging from 5% to 16%. RNA-mNGS only detected  
196 *Borrelia* in samples with abundance >20% (Supplementary Fig. 7F). Taken together, these data  
197 suggest that RNA-mNGS can detect high titer bacterial pathogens but will miss less abundant  
198 species.

199 We compared the sensitivity of our 16S sequencing to qPCR and Giemsa stained blood  
200 smear examination, the current gold standard for clinical diagnosis of RF *Borrelia*<sup>35</sup>. We  
201 examined smears from a subset of RF *Borrelia* patients. 28.6% (4/14) of examined smears were  
202 positive; all 4 positive patients had a high RF *Borrelia* load ( $\geq 90\%$ , Mann-Whitney two-sided  $p$   
203 = 0.002 compared to 16S positive/smear negative) (Supplementary Fig. 6B, D). Although blood  
204 smear examination is the most widely used diagnostic, qPCR diagnostics are known to have  
205 higher sensitivity<sup>35,36</sup>. Using a pan-*Borrelia* qPCR assay previously used for detection of *Borrelia*  
206 in patient blood samples<sup>37</sup> 79% (27/34) of *Borrelia* samples were confirmed by qPCR, while 6  
207 samples that were negative by 16S sequencing were positive by qPCR. Bacterial load  
208 measured by qPCR and 16S abundance correlated well (Pearson's  $R = 0.72$ , Supplementary  
209 Fig. 6G). While 16S sequencing was done only on a subset of samples, pan-*Borrelia* qPCR was  
210 performed on all febrile cases and healthy controls in 2019 and all febrile cases in 2018. An  
211 additional 5 cases of *Borrelia* were identified in febrile patients by qPCR in samples that did not  
212 undergo 16S sequencing (Supplementary Fig. 7E).

213

### 214 ***Borrelia* infection presents similarly to other febrile illness but can be distinguished by** 215 **key features**

216 Given the limited sensitivity of available smear-based diagnostics compared to molecular  
217 methods, we sought to assess the clinical syndrome associated with qPCR-confirmed *Borrelia*  
218 infections, compared to RDT-confirmed malaria and non-*Borrelia* NMFI ("other febrile"), to guide  
219 differential diagnosis at the point of care. *Borrelia* infections occurred in a consistent proportion  
220 of the study population in the two years (Fisher exact  $p = 0.61$ ) and in both seasons (Fisher  
221 exact  $p = 0.24$ ). *Borrelia* infections occurred across all ages (t-test compared to all other febrile  
222  $p = 0.45$ ) (Fig. 2B) and at similar rates in male and female patients (Fisher exact  $p = 0.87$ ), while  
223 male patients were more likely to test positive for malaria (Fisher exact  $p = 1.6e-7$ ).

224 A high proportion of *Borrelia* positive patients reported generalized symptoms that were  
225 common across all febrile patients, including headache (100%), body aches (59%), and  
226 dizziness (46%), as reported in previous studies of RF *Borrelia* (Fig. 3A)<sup>36</sup>. *Borrelia* infection  
227 was distinguished by vomiting (49%), vomiting was observed more often in *Borrelia*-positive  
228 patients (stat = 2.8,  $p = 0.0026$ , Fisher Exact two-sided). *Borrelia*-positive patients were less  
229 likely than other febrile patients to report sore throat (7.7%, stat = 0.21,  $p = 0.0039$ , Fisher Exact  
230 two-sided). Notably, RF *Borrelia* can invade the central nervous system in severe cases; while  
231 we did not observe neurological symptoms such as seizures or loss of consciousness, one  
232 *Borrelia* positive patient reported eye pain, a potential symptom of neuroinvasive infection not  
233 observed in any other febrile patients<sup>36,38</sup>. A high proportion of *Borrelia*-positive patients reported

234 contact with a febrile person (38%), contact with rats (57%), and prior travel (32%), but these  
235 exposures were common across malaria and other febrile patients and were not associated with  
236 *Borrelia* infection in particular (Fig. 3B).

237 *Borrelia* patients showed significant differences in temperature and blood glucose. On  
238 average, *Borrelia* positive patients had a higher fever (mean = 38.6°C, SD = 1.1°C) than other  
239 febrile patients (mean = 38.0°C, SD = 1.0°C, Mann-Whitney two-sided  $p = 0.0011$ ). *Borrelia*  
240 patients (mean = 1.07 g/L, sd = 0.19) were also hypoglycemic compared to other febrile patients  
241 (Other febrile: mean = 1.22 g/L, sd = 3.97, Mann-Whitney two-sided  $p = 1.204e-02$ ) (Fig. 3C).

242 We assessed whether the addition of a complete blood cell count with differential could  
243 help distinguish RF *Borrelia* from other febrile diseases. *Borrelia* positive patients exhibited  
244 abnormal blood counts, including lymphopenia (mean  $1.5 \times 10^9$  cells/L, std 1.04),  
245 granulocytosis (mean  $7.47 \times 10^9$  cells/L, std 4.23), and thrombocytopenia ( $148 \times 10^9$  cells/L  
246 std 66) (Fig. 3D). However, this hematologic response was not significantly different from the  
247 response observed in other febrile patients. On a subset of patients, we measured 23 cytokines  
248 and chemokines to further distinguish *Borrelia* from other infections (Fig. 3E). *Borrelia* positive  
249 patients were distinct from viral infections in elevated CRP (Mann-Whitney two-sided,  $p =$   
250 0.030), IL-10 ( $p = 0.00038$ ), and MIP-1Beta ( $p = 0.0049$ ) and decreased IP-10 ( $p = 0.00070$ ),  
251 IFN-alpha ( $p = 0.0022$ ), and MCP-1 ( $p = 0.00038$ ) (Fig. 3E).

252 Given the highly overlapping presentation of disease, we developed a weighted logistic  
253 regression model to distinguish *Borrelia* infection from other NMFI based on clinical symptoms,  
254 vital signs, and key demographic information. We evaluated the performance of the model with  
255 bootstrapping and found the model was able to predict *Borrelia* infection with high performance  
256 (recall: 0.861, 95% CI [0.837 - 0.910], precision: 0.792, 95% CI [0.696 - 0.892], and F1-score:  
257 0.823, 95% CI [0.766 - 0.900]) (Fig. 4A). The model identified features that were significant in  
258 univariate analysis, including vomiting, temperature, blood glucose, and sore throat.  
259 Interestingly, common symptoms that were observed across all febrile patients, such as  
260 headache, body ache, and dizziness were not useful predictors on their own but were  
261 associated with increased risk of *Borrelia* when evaluated in the context of other symptoms and  
262 exposures. To test whether additional laboratory tests might further increase performance of the  
263 model, we incorporated complete blood count with differential (CBC) values. The addition of  
264 CBC values moderately increased model performance (Supplementary Fig. 9), but the  
265 difference was not significant and the test dataset was small ( $n = 163$ ). When blood biomarkers  
266 were incorporated, decreased platelet count was an important predictor of *Borrelia*.

## 267 268 **Discussion**

269 In this study, we employed untargeted RNA-mNGS and targeted 16S sequencing to  
270 understand the causes of acute febrile illness at an ambulatory clinic in Thiès, Senegal. As the  
271 first unbiased investigation of causes of fever in Senegal, our findings highlight the importance  
272 of looking across many pathogen types simultaneously to understand their contributions—both  
273 individually and as co-infections—to disease, and enable comparison of the strengths of  
274 genomic tools as well as clinical and epidemiological data to support disease characterization.

275 Arthropod-borne bacterial pathogens, *Borrelia* and *Rickettsia*, are the major identifiable  
276 causes of NMFI, but remain underdiagnosed. The frequency of these pathogens is broadly  
277 consistent with previous findings in targeted studies from other regions of Senegal<sup>8,39</sup>. *Borrelia*

278 spp. associated with human and zoonotic infection have been identified in ticks across West  
279 Africa<sup>28,36,40,41</sup> but human surveillance has not been done.

280 Our unbiased approaches and broad sampling illuminated co-infections and missed  
281 diagnoses that may be clinically relevant. Both *Borrelia* and *Rickettsia* occurred frequently as  
282 co-infections in our study. *Rickettsia* was detected more often as a co-infection with  
283 *Plasmodium* (n = 3) or *Borrelia* (n = 2) than as a stand-alone infection (n = 3).  
284 *Borrelia/Plasmodium* co-infections have been previously detected in Senegal<sup>42</sup> and *in vitro*  
285 evidence suggests co-infection increases the risk of severe malarial illness<sup>43</sup>. *Plasmodium/R.*  
286 *felis* co-infections have been previously identified in Senegal<sup>44</sup> and *Anopheles gambiae*, the  
287 primary malaria vector in Senegal, may be able to transmit *Rickettsia*<sup>45</sup>. Whether *Rickettsia* is  
288 contributing to pathogenesis or simply co-transmitting in these cases remains an open question.  
289 More research is needed to understand the interaction between the parasite and *Rickettsia*  
290 species. Given the common occurrence of co-infections and potential for negative outcomes  
291 without proper treatment of both pathogens, it is important that both surveillance and diagnostic  
292 approaches do not stop at detection of the first pathogen.

293 Despite using multiple methods and searching across kingdoms, we did not find any  
294 pathogen in over 70% of NMFI cases in this study. Although our enrollment criteria focused on  
295 suspicion of malaria, many of these patients reported sore throat and difficulty breathing (2019:  
296 35%, 40/113) or abdominal pain (4.4%, 5/113). Given that only plasma samples were collected  
297 and sequenced, our approach may have missed infections limited to other body compartments,  
298 including the respiratory tract and digestive tract. Our results also demonstrate the difficulty of  
299 using RNA-mNGS to detect DNA pathogens and pathogens with different cellular compositions.  
300 Several of the bacterial infections in this cohort, particularly those with low abundance in the  
301 blood, were detected by 16S but missed by RNA-mNGS. Similarly, we were not able to resolve  
302 putative fungal hits to a suitable level of taxonomic resolution; to capture fungal pathogens,  
303 extraction methods should be optimized for lysis of fungal cells and more targeted methods  
304 such as ITS sequencing could be used. We observed wide variations in the number of  
305 *Plasmodium* reads among RDT positive patients and weak correlation with parasite density,  
306 suggesting that factors beyond parasite burden affect the number of reads recovered. There are  
307 also still many challenges in interpreting the RNA-mNGS data. In particular, RNA-mNGS is  
308 highly susceptible to contamination, such as the novel *Reoviridae* RdRp sequences we detected  
309 across febrile cases, healthy controls and non-template controls. Negative controls should be  
310 included alongside samples throughout processing to measure the “contaminome,” which likely  
311 differs between sample types and labs<sup>15</sup>.

312 We find that combinations of clinical signs and symptoms can increase suspicion for  
313 *Borrelia* and support targeting clinical care. Due to the lack of available point-of-care tests, many  
314 clinics in LMICs rely on malaria RDTs as the primary diagnostic for febrile illness and may give  
315 blanket antibiotic treatment to RDT-negative patients<sup>46</sup>. Concerns have been raised about this  
316 practice<sup>47</sup>, including driving antibiotic resistance and exposing patients to unnecessary risks or  
317 side effects. For example, tetracycline antibiotic treatment—the recommended therapeutic for  
318 *Borrelia* and *Rickettsial* infections—poses the risk of Jarisch-Herxheimer reaction (JHR), a  
319 severe inflammatory response to spirochete lysis that often requires close patient monitoring<sup>48</sup>.  
320 We identify a number of clinical and immunological features that could bolster support for a  
321 *Borrelia* diagnosis. Our weighted logistic regression model indicates that when larger cohorts



322 and diverse data types are aggregated, *Borrelia* can be differentiated from other NMFI with  
323 reasonable accuracy. The addition of laboratory measures, such as CBC, did not substantially  
324 improve differential diagnosis. Further, data from a subset of our cohort supports the hypothesis  
325 that a chemokine panel including TRAIL, IP-10 and CRP, previously employed for detection of  
326 bacterial infections in hospitalized children<sup>49</sup>, could apply more broadly and help distinguish  
327 bacterial infections such as *Borrelia* from other causes of febrile illness in the ambulatory  
328 setting. However, none of these approaches are currently actionable as point-of-care tests and  
329 their lower precision would still necessitate confirmatory diagnostic testing.

330 The challenge of quickly and accurately differentiating *Borrelia* infections underscores  
331 the need for improved diagnostics. The diagnostic gold standard for *Borrelia* remains  
332 microscopic examination of blood smears, which has low sensitivity in our study and others<sup>35</sup>.  
333 Further, smear cannot distinguish species, and is highly dependent on a trained practitioner.  
334 Our study showed that molecular methods, including 16S sequencing and mNGS, can detect  
335 *Borrelia* infections, but are not practical in the clinical setting. Increased availability of qPCR  
336 assays and improved point of care nucleic acid diagnostics could enable detection and  
337 treatment of *Borrelia* across differently resourced clinical settings and improve our  
338 understanding of the geographic range and diversity of this pathogens

339 An increasingly heterogeneous pathogen landscape will require the integration of  
340 targeted diagnostics and broad surveillance to effectively combat acute febrile infections moving  
341 forward. Decreased malaria transmission coupled with changes in the range and incidence of  
342 other pathogens—influenced in part by global climatic shifts—is broadening the array of  
343 pathogens accounting for the majority of febrile disease<sup>50</sup>. Fully understanding this landscape  
344 will require surveillance systems utilizing unbiased approaches that have been validated for  
345 viral, bacterial and eukaryotic pathogens and diverse sample types. In tandem, for *Borrelia* and  
346 other pathogens known to frequently cause NMFI, the development and availability of diagnostic  
347 tests that are cheap, rapid and sensitive will be key to enable appropriate clinical treatment at  
348 the point of care and support deeper investigations of the pathophysiology of disease.

349

## 350 **Methods**

### 351 **Sample collection**

352 Febrile cases were selected from patients presenting to the SLAP outpatient clinic in  
353 Thies, Senegal during the collection period under local IRB (SEN15/46) and Harvard IRB  
354 (IRB19-0023). Informed consent was obtained for all enrollees; for minors under 18 years of age  
355 or individuals unable to provide their own consent, the consent of a parent or legal guardian was  
356 obtained. The study team explained to potential participants that their participation was strictly  
357 voluntary and that they could withdraw from the study at any time without any penalties or  
358 consequences and translations were made for potential participants (or their parent / legal  
359 guardian) that do not understand or cannot read the language in which the consent form was  
360 produced.

361 Patients who met the following inclusion criteria and gave consent were enrolled: (1)  
362 Febrile symptoms within the 3 days up to and including the day of presentation, (2) Age 2-75  
363 years, and (3) Ambulatory with no signs of severe malarial (glucose < 2.2mM, hemoglobin < 5  
364 gms/dL). A subset of DENV positive samples from 2018 were previously sequenced and  
365 published<sup>4</sup> and were excluded from this study other than being used as a viral outgroup in the

366 chemokine/cytokine analyses. At the time of enrollment, a structured interview including  
367 personal information, demographics, and self-reported symptoms was completed, vital signs  
368 were measured and blood was drawn by a trained practitioner. *Plasmodium falciparum* RDT  
369 (Bioline Malaria Ag P.f., Abbott), thick blood smear, and thin blood smear was performed and  
370 blood glucose (HemoCue 201/301) and hemoglobin (HemoCue Glucose 201) were measured.

371

### 372 **Sample processing**

373 Blood samples were stored on ice upon collection (maximum 8 hours). In the lab, blood  
374 samples were split into two aliquots. 200 ul of the first aliquot was used for an automated blood  
375 count with differential (Mindray BC 20s) and 200uL was used for a blood spot (Whatman filter  
376 paper) The remaining aliquot was centrifuged at 2300 rpm for 10 min to separate plasma, buffy  
377 coat and red blood cells. 140 uL of plasma was transferred to a tube containing 760uL of  
378 inactivation buffer (AVL, Qiagen). The remaining plasma was stored without inactivation. Red  
379 blood cells, buffy coat, plasma, and inactivated plasma were flash frozen and stored at -80°C  
380 until further processing. Total nucleic acids were extracted with Qiagen QIAmp Viral Mini RNA  
381 according to the manufacturer's protocol. Total nucleic acids were split into two aliquots and  
382 treated with Lucigen RNase I or Ambion TURBO DNase to obtain purified DNA and RNA,  
383 respectively.

384

### 385 **Unbiased RNA sequencing**

#### 386 *Library preparation and sequencing*

387 Unbiased RNA sequencing libraries were prepared as described previously<sup>51</sup>. Briefly,  
388 cDNA synthesis was performed from DNase-treated RNA using random hexamers (Illumina)  
389 and SSIV. Libraries were prepared from cDNA using Nextera XT (Illumina) and UD index  
390 primers (IDT/Illumina) with the following modifications to account for low cDNA input: the volume  
391 of ATM was decreased from 1uL to 0.5uL per 10uL reaction and PCR cycles were increased to  
392 17 cycles. Libraries were purified using AMPure XP (Agilent), quantified using KAPA  
393 Biosystems Universal Library Quantification, and pooled equally for 75bp paired end  
394 sequencing on NovaSeq SP to obtain at least 2 million reads per sample.

395

#### 396 *Taxonomic classification of RNA-mNGS reads*

397 Unbiased RNA sequencing reads were processed using viral-ngs<sup>52</sup>. Briefly, reads were  
398 demultiplexed, adapter sequences were trimmed, and sliding-window quality filtering was  
399 performed. Human reads were filtered out using BMTAGGER and BLASTN. Cleaned reads  
400 were uploaded to NCBI SRA under BioProject PRJNA662334 (Accession numbers:  
401 SRR24622550-SRR24622641, SRR24995052-SRR24995258). Samples with insufficient reads  
402 (<2 million) were removed from further analysis (n = 1, SHC1064, Accession: SRR24995248).

403

404 Taxonomic classification of human-depleted reads was performed using Kraken2<sup>53</sup> with  
405 the JHU PlusPF database (JHU, downloaded 12-13-2022). Kraken2 results were thresholded to  
406 consider only viral genera with at least 5 reads/million raw reads. Results were filtered to  
407 remove viral taxa detected in non-template controls (at least one non-template control) and  
408 healthy patients (at least 2 healthy patients). Further, viral genera with non-vertebrate hosts  
409 were filtered out.

410 To verify Kraken2 classifications, we performed a protein-sequence similarity search  
411 using DIAMOND-blastx<sup>54</sup>. Cleaned, de-duplicated reads were used for *de novo* contig  
412 assembly with SPAdes and DIAMOND-blastx was run on all *de novo* contigs > 100 bp long with  
413 the complete nr database (downloaded December 2022). The least common ancestor of the top  
414 e-value hits was identified using a custom script (lakras/bio-helper-  
415 scripts/blast/retrieve\_top\_blast\_hits\_LCA\_for\_each\_sequence.pl). A classification was  
416 considered verified by DIAMOND if the cumulative length of contigs with a high identity match  
417 (mean identity of top e-value hits  $\geq 95\%$ ) to the expected viral genus was at least 1kb. After  
418 DIAMOND-blastx verification, raw reads were aligned to the NCBI Virus RefSeq and assessed  
419 for evenness and depth of coverage.

420 To assess for presence of divergent viral taxa that are not well represented in the  
421 database and may be missed by Kraken2, DIAMOND-blastx results were filtered for viral hits  
422 with low identity (20-80% amino acid identity). Results were filtered for viral families with a per-  
423 sample cumulative *de novo* contig length  $\geq 500$  bp. *Narnaviridae* and *Reoviridae* sequences  
424 were aligned with all available RdRp sequences for each family in nr using MAFFT and trimmed  
425 with trimAl. Maximum likelihood phylogenetic trees were generated in IQ-TREE<sup>55</sup> with a GTR-  
426 gamma substitution model and bootstrapping (n = 1000) and visualized in FigTree v1.4.4<sup>56</sup>.

#### 427 *Viral genome assembly and phylogenetic analysis*

428 For virus-positive samples, reference-guided *de novo* assembly was performed using  
429 viral-ngs (assembly with SPAdes, reference-assisted improvements with MUMMER and  
430 MUSCLE/MAFFT and gap sealing with Gap2Seq). The NCBI Virus RefSeq as well as any  
431 >80% complete genomes available from Senegal were provided as references. For samples  
432 with near-complete or complete (>80%) genomes, genotype was determined with  
433 GenomeDetective (DENV, HBV) or by phylogenetic tree (Parvovirus B-19, GBV-C).  
434 Phylogenetic trees were generated with all >80% complete references from Africa downloaded  
435 from NCBI viruses (DENV1, DENV3, HBV), or all global >80% complete genomes available in  
436 NCBI virus (Parvovirus B-19, GBV-C). Sample sequences and reference sequences were  
437 aligned using MAFFT v7<sup>57</sup> and alignments were trimmed with trimAl v1.4.rev15<sup>58</sup> to remove any  
438 bases with coverage in <80% of sequences. Maximum likelihood phylogenetic trees were  
439 generated in IQ-TREE<sup>55</sup> with a GTR-gamma substitution model and bootstrapping (n = 1000).  
440 Visualizations were generated with FigTree v1.4.4<sup>56</sup>.

#### 442 *Plasmodium RNA-mNGS analysis*

443 In order to quantify *Plasmodium* rpm, the number of reads classified as *Plasmodium* at  
444 the genus level by Kraken2 was divided by the raw read count in millions. The cut-off for  
445 considering a sample positive for *Plasmodium* by mNGS was determined by calculating the 99th  
446 percentile for healthy control samples, 550 *Plasmodium* rpm. Cleaned deduplicated reads were  
447 aligned to the *P. falciparum* *pfrp2* (PlasmoDB PF3D7\_0831800) with NovoAlign.

#### 449 *Fungal RNA-mNGS analysis*

450 In order to detect possible fungal infections, we calculated the rpm classified as Fungi by  
451 Kraken2<sup>53</sup> at the kingdom level and determined the 99th percentile for healthy controls, 3399  
452 rpm. We identified 5 samples (4 febrile, 1 control) with high fungal reads. We attempted

454 taxonomic classification of these reads with Kraken2 at the genus level and saw hits across  
455 multiple genera for many samples (Supplementary Fig. 8B). To try to improve fungal  
456 classification, we performed a nucleic acid search of all *de novo* contigs from these 5 samples  
457 with megablast against nt (Supplementary Fig. 8C) and a translated nucleic acid search with  
458 DIAMOND-blastx<sup>54</sup> (as described above). For both searches, we found the least common  
459 ancestor of the top hits (as described above), and filtered hits in the kingdom Fungi (taxid: 4751)  
460 with >90% identity and >30% coverage.

461  
462 **qPCR**  
463 Total bacterial load was quantified from extracted plasma DNA with primers targeting the  
464 V1-2 region (see Supplementary Table 2 for primer sequences) with Power SYBR Green PCR  
465 MasterMix (Thermo Fisher) under the following conditions: 300nM each F/R primer, 95°C hold  
466 for 10 min, 40 cycles of 95°C for 15 sec, 50°C for 1 min, 75°C for 30s. Pan-*Borrelia* qPCR was  
467 performed with Power SYBR Green PCR MasterMix (Thermo Fisher) under the following  
468 conditions: 400 nM each F/R primer, 95°C hold for 10 min, 40 cycles of 95°C for 15 sec, 60°C  
469 for 1 min.

470  
471 **16S sequencing**  
472 *Library preparation and sequencing*  
473 Samples with high bacterial load (V1-2 qPCR CT < 31.5) were selected for 16S  
474 sequencing (2019: febrile n = 129, healthy n = 35, 2018: febrile n = 84). Libraries were  
475 generated by amplification with tailed universal primers Tail-V1-2F and Tail-V1-2R  
476 (Supplementary Table 2) targeting variable regions 1-2. Amplification was performed on 5ul of  
477 RNase-treated DNA using Q5 high fidelity polymerase (NEB) with forward and reverse primers  
478 at 100nM each with the following cycling conditions: 98°C for 30 sec; 35 cycles of 95°C for 15  
479 sec and 63°C for 2 min, 4°C hold. PCR products were cleaned using Ampure XP (Agilent) with  
480 0.7X ratio to remove primer dimers. Adapters and barcodes (BroadDuplex Seq) were added  
481 with a second PCR reaction using Q5 polymerase and the following cycling conditions: 98°C for  
482 30 sec; 18 cycles of 95°C for 15 sec, 60°C for 15S, 72°C for 30S, followed by 72°C for 5min,  
483 4°C hold. A subset of libraries were visualized with hsDNA BioAnalyzer (Agilent) and all libraries  
484 were quantified with KAPA Biosystems Universal Library Quantification Kit before pooling  
485 equally across samples and 250 bp paired-end sequencing on Illumina MiSeq v2 with 40%  
486 PhiX, given the low diversity of the single-amplicon library.

487  
488 *Analysis*  
489 Sequencing reads were demultiplexed using viral-ngs and demultiplexed FASTQ files  
490 were imported into qiime2<sup>59</sup> for further analysis. Briefly, after sliding window quality filtering and  
491 adapter trimming with cut-adapt (minimum-length 20), paired reads were joined with vsearch  
492 (tuncqual 15, minlen 35, minovlen 10, maxdiffs 3), and ASVs were generated with Deblur (trim-  
493 length 280). ASVs were taxonomically classified by blastn search of ASV sequences against the  
494 NCBI 16S rRNA db (downloaded June 2023). The least common ancestor of the top e-value  
495 hits was determined using a custom script (lakras/bio-helper-  
496 scripts/blast/retrieve\_top\_blast\_hits\_LCA\_for\_each\_sequence.pl). Abundance was calculated  
497 by dividing the number of ASVs in each genus by the total number of ASVs in each sample.

498 Abundance data was filtered to show only taxa accounting for greater than 5% of ASVs. Taxa  
499 detected in non-template controls (n >= 1) and healthy controls (n >= 2) were filtered out.

500 For *Borrelia* and *Rickettsia* positive samples, sample ASVs were aligned with MAFFT<sup>57</sup>  
501 (Geneious) with curated high-quality 16S sequencing from the NCBI 16S rRNA BioProject  
502 (BioProject: 33175) from each genus respectively. Alignments were trimmed with trimAl to  
503 remove positions with gaps in >90% of sequences and a maximum-likelihood phylogenetic tree  
504 was generated using IQ-TREE<sup>55</sup> with bootstrapping (n = 1000) and annotated with FigTree  
505 v1.4.4<sup>56</sup>.

506

### 507 **IGS Sequencing**

508 The intergenic spacer (IGS) region was amplified using a nested-PCR. First, the region  
509 was amplified from 5uL of extracted DNA with primers targeting the IGS region (IGS-outer F/R,  
510 see Supplementary Table 2) using Q5 polymerase and the following cycling conditions: 98°C for  
511 30 sec; 35 cycles of 94°C for 30 sec, 66°C for 30S, 74°C for 60s, followed by 74°C for 2 min,  
512 10°C hold. After a 0.75X AMPure XP cleanup to remove excess primer, a nested PCR was  
513 performed (IGS-inner F/R, see Supplementary Table 2) using Q5 polymerase and the following  
514 cycling conditions: 98°C for 30 sec; 30 cycles of 94°C for 30 sec, 67°C for 30S, 74°C for 60s,  
515 followed by 74°C for 2 min, 10°C hold. Amplified product was cleaned with AMPure XP (0.75X),  
516 visualized with BioAnalyzer TapeStation and bi-directional Sanger sequencing was performed  
517 (Azenta). Paired Sanger sequencing traces were analyzed in Geneious to generate a  
518 consensus sequenced and aligned with all available reference sequences for *B. crocidurae*, *B.*  
519 *duttonii*, and *B. recurrentis* using MAFFT<sup>57</sup>. A maximum-likelihood phylogenetic tree was  
520 generated using IQ-TREE<sup>55</sup> and annotated with FigTree v1.4.4<sup>56</sup>.

521

### 522 **Blood smear examination**

523 Thick and thin blood smears were fixed with methanol and stained with 3mL 10%  
524 Giemsa for 10 minutes. Stained smears were examined under a brightfield microscope for  
525 evidence of *Borrelia*. Slides were first scanned at low (40X) magnification and then at high  
526 (100X) magnification with oil immersion to quantify organisms per field (representative image,  
527 Supplementary Fig. 6B).

528

### 529 **Serology**

530 We assessed levels of common cytokines and chemokines using the Luminex platform  
531 on plasma from a subset of patients and controls. Specifically, we selected 10 healthy controls  
532 and 30 cases representing 10 each with a confirmed diagnosis of; malaria (based on malaria  
533 RDT administered at enrollment), *Borrelia* (based on pan-*Borrelia* qPCR<sup>37</sup>), or viral infection.  
534 Samples with a viral infection were all confirmed to have Dengue virus and are a subset of  
535 those cases previously described elsewhere<sup>4</sup>. Cases were selected randomly within those  
536 meeting each criteria. All samples were collected in 2018. We used the Inflammation 20-Plex  
537 Human ProcartaPlex™ Panel (Invitrogen), according to the manufacturer's instructions. We  
538 additionally performed a 2-plex assay for Human IP-10 and TRAIL (using IP-10 and TRAIL  
539 ProcartaPlex™ simplex kits from Invitrogen); two markers previously found to differentiate  
540 bacterial infections in hospitalized children<sup>49</sup>, as these were not included on the predefined  
541 panel. All samples were run in duplicate on a Luminex MAGPIX® instrument in Senegal. We

542 confirmed that standards showed the expected values and calculated the average across  
543 duplicates, which we report in the text.

544

### 545 **Multivariate model**

546 We developed a weighted logistic regression model that considered 46 candidate  
547 predictors based on clinical symptoms, vital signs, and key demographic information to predict  
548 *Borrelia* infection. Before modeling, the dataset underwent the following preprocessing steps:  
549 imputation of missing values with the Multivariate Imputation by Chained Equations (MICE)  
550 algorithm from the fancyimpute Python library<sup>60</sup>, one hot-encoding of categorical features, and  
551 standardization of continuous features using the StandardScaler from the scikit-learn library.  
552 Given that normal heart rate varies significantly with age, heart rates were binned and treated  
553 as a categorical variable. Heart rate categories were high, low, or normal based on the Pediatric  
554 Advanced Life Support guidelines as follows: for age 3-5, normal range 80-120, age 6-11  
555 normal range 75-118, age 12+ normal range 60-100<sup>61</sup>.

556 The sample size was relatively small, with 526 patients and only 38 q-PCR confirmed  
557 cases of *Borrelia* infection. Given this inherent class imbalance in the dataset, a combination of  
558 techniques was employed to address this challenge. The Synthetic Minority Oversampling  
559 Technique (SMOTE) was applied to generate synthetic samples for the minority class (q-PCR  
560 confirmed *Borrelia* case)<sup>62</sup>. This technique mitigates class imbalance by interpolating between  
561 existing samples and creating synthetic instances of the minority class. During model training,  
562 class weights were also applied to adjust the influence of each class. The class weights were  
563 inversely proportional to the class frequencies, with higher weights assigned to the minority  
564 class.

565 A logistic regression model was chosen for its interpretability and ability to estimate the  
566 probabilities of class membership. After a univariate screening for statistically significant  
567 predictors using chi-squared test ( $p$ -value  $< 0.05$ ), several Feature Selection techniques were  
568 applied to construct the final set of predictors in the model. Before training the model, feature  
569 importance was assessed using a combination of three different feature selection methods;  
570 mutual information classification<sup>63</sup>, Recursive Feature Elimination<sup>64</sup>, and Lasso L1  
571 regularization<sup>65,66</sup> to ensure consistent elimination of redundant variables. The selected features  
572 were used in the final weighted logistic regression model to analyze the performance of *Borrelia*  
573 prediction. The regularization parameter from Lasso was tuned using cross-validation of  
574 different alpha values and the best alpha value was used to train the final model and evaluate  
575 its performance on the test set.

576 Model performance was evaluated with bootstrapping and 5-fold cross validation to  
577 quantify optimism and generalizability. The optimism-corrected metrics include F1-score,  
578 precision, recall, area under the ROC curve (AUC-ROC), and area under the precision-recall  
579 curve (AUC-PR). Odds ratios and their corresponding confidence intervals for all the model  
580 coefficients were calculated to assess the impact of predictor variables on the odds of the  
581 outcome. All predictive analyses were conducted using the Python programming language. The  
582 scikit-learn library was utilized for preprocessing, model training, and evaluation. Statistical  
583 analyses and visualization of odds ratios and confidence intervals were done using the R  
584 statistical package.

585

586 **Author contributions**

587 According to the CRediT format, author contributions are as follows: Conceptualization: A.B.D.,  
588 A.S.B., Z.C.L., N.S., P.C.S., K.J.S., and D.N.; Methodology: A.B.D., A.S.B., Z.C.L., N.S., P.C.S.,  
589 K.J.S., and D.N.; Software: D.J.P. ; Validation: P.C.S., K.J.S., and D.N.; Formal Analysis: Z.C.L.,  
590 K.J.S., W.M., and A.C.; Investigation: Z.C.L, A.S., A.G., T.N., M.S., Y.D., A.M., and I.N.; Data  
591 Curation: J.G., M.Ndiop, D.S.; Writing - Original Draft: Z.C.L. Writing - Review & Editing: K.J.S.,  
592 S.F.S., B.M, and P.C.S. with input from all authors; Visualization: Z.C.L., W.M.; Supervision:  
593 P.C.S., K.J.S., and D.N.; Project Administration: M.F.P., M.Ndiaye.; Funding acquisition: P.C.S.,  
594 D.N., K.J.S., D.J.P.

595

596 **Funding Statement**

597 Support for this project was provided in part by the National Institutes of Allergy and Infectious  
598 Diseases through the H3Africa and Center for Research in Emerging Infectious Disease  
599 programs (U54HG007480, U01HG007480 and U01AI151812 to C.T.H. and P.C.S.), and the  
600 Genomic Center for Infectious Disease (U19AI110818 to P.C.S.). Z.C.L was supported by the  
601 National Institute of General Medical Sciences (T32GM007753 and T32GM144273). K.J.S  
602 received support from the American Society of Tropical Medicine and Hygiene Centennial  
603 Award. P.C.S. is an investigator supported by the Howard Hughes Medical Institute (HHMI). The  
604 content is solely the responsibility of the authors and does not necessarily represent the official  
605 views of the National Institute of General Medical Sciences, National Institutes of Allergy and  
606 Infectious Diseases, or the National Institutes of Health.

607

608 **Data and Code Availability**

609 RNA-mNGS short read sequence data have been deposited in NCBI's SRA and Genbank  
610 databases under BioProject PRJNA662334 Accession numbers SRR24622550-SRR24622641,  
611 SRR24995052-24995258. 16S v1-2 amplicon sequencing data and assembled viral genomes  
612 will be deposited shortly under BioProject PRJNA662334, data currently available upon request.  
613 The Python source code for the multivariate *Borrelia* prediction model is available at  
614 [github.com/colabobio/borrelia-diagnosis-prediction-models](https://github.com/colabobio/borrelia-diagnosis-prediction-models).

615

616 **Acknowledgements**

617 We thank the Programme National de Lutte contre le Paludisme (National Malaria Control  
618 Program) for providing malaria surveillance data, and Fama and Astou for support with data  
619 entry and organization. We thank Lydia Krasilnikova for advice and custom scripts for BLAST  
620 analysis, and Flavia Negrete for support with data uploads and sharing. We thank Jacob  
621 Lemieux and Gordon Adams for providing valuable guidance on *Borrelia* sequencing.

622

623 **Competing Interests**

624 P.C.S. is a co-founder of, shareholder in, and consultant to Sherlock Biosciences, Inc. and  
625 Delve Bio, as well as a Board member of and shareholder in Danaher Corporation.

## References

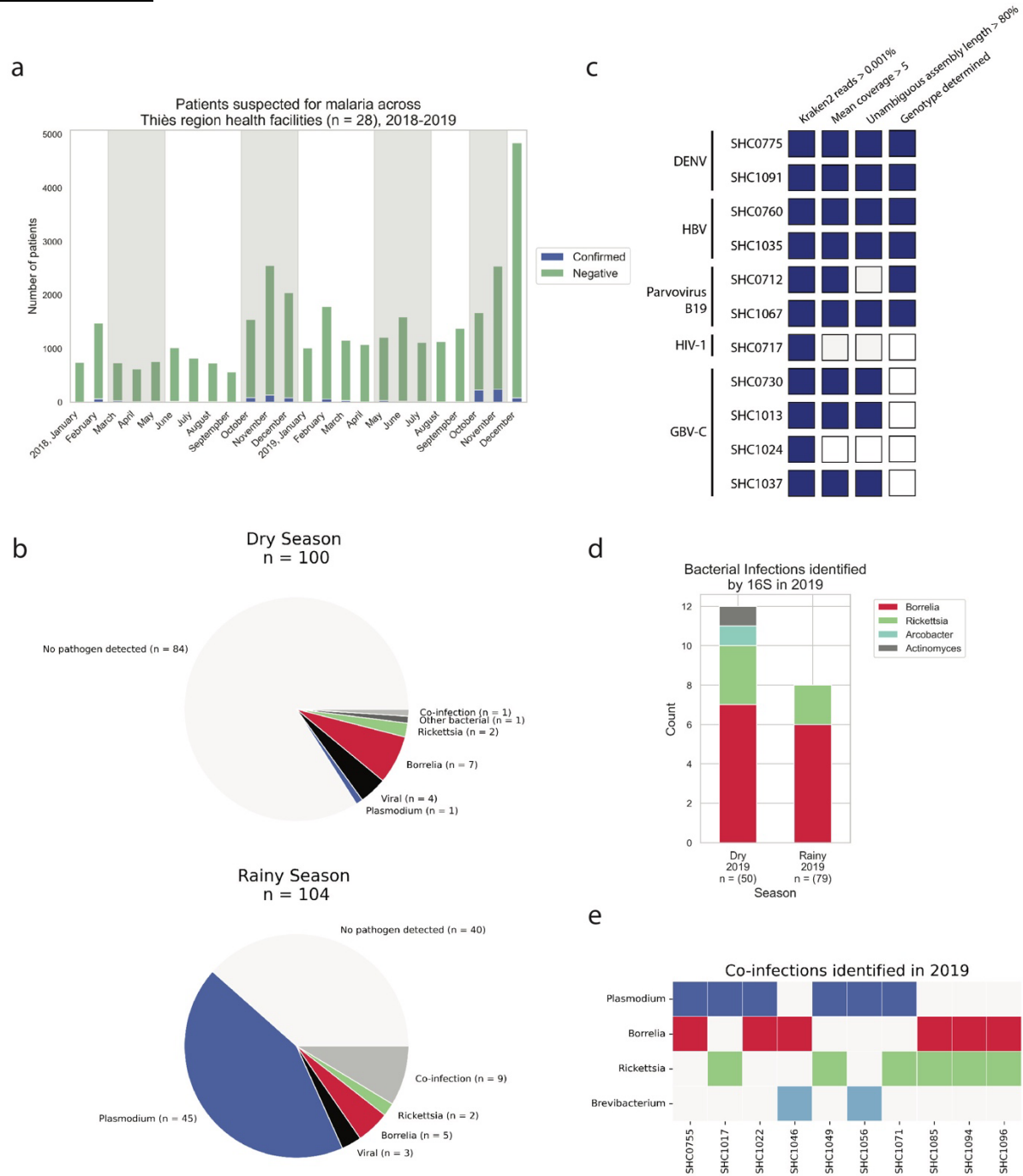
1. Thwing, J. *et al.* Declines in Malaria Burden and All-Cause Child Mortality following Increases in Control Interventions in Senegal, 2005-2010. *Am. J. Trop. Med. Hyg.* **97**, 89–98 (2017).
2. World Bank. Incidence of malaria (per 1,000 population at risk). (2021).
3. WHO. Who informal consultation on fever management in peripheral health care settings a global review of evidence and. (2013).
4. Gaye, A. *et al.* Genomic investigation of a dengue virus outbreak in Thiès, Senegal, in 2018. *Sci. Rep.* **11**, 10321 (2021).
5. Sow, A. *et al.* Chikungunya Outbreak in Kedougou, Southeastern Senegal in 2009-2010. *Open Forum Infect Dis* **5**, ofx259 (2018).
6. Maze, M. J. *et al.* The epidemiology of febrile illness in sub-Saharan Africa: implications for diagnosis and management. *Clin. Microbiol. Infect.* **24**, 808–814 (2018).
7. Elven, J. *et al.* Non-malarial febrile illness: a systematic review of published aetiological studies and case reports from Africa, 1980-2015. *BMC Med.* **18**, 279 (2020).
8. Sokhna, C. *et al.* Point-of-care laboratory of pathogen diagnosis in rural Senegal. *PLoS Negl. Trop. Dis.* **7**, e1999 (2013).
9. Gu, W., Miller, S. & Chiu, C. Y. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol.* **14**, 319–338 (2019).
10. Ramesh, A. *et al.* Metagenomic next-generation sequencing of samples from pediatric febrile illness in Tororo, Uganda. *PLoS One* **14**, e0218318 (2019).
11. Mwakibete, L. *et al.* Metagenomic next-generation sequencing to characterize potential etiologies of non-malarial fever in a cohort living in a high malaria burden area of Uganda. *PLoS Glob Public Health* **3**, e0001675 (2023).
12. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
13. Oguzie, J. U. *et al.* Metagenomic surveillance uncovers diverse and novel viral taxa in febrile patients from Nigeria. *Nat. Commun.* **14**, 4693 (2023).
14. Houldcroft, C. J., Beale, M. A. & Breuer, J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **15**, 183–192 (2017).
15. Chrisman, B. *et al.* The human ‘contaminome’: bacterial, viral, and computational contamination in whole genome sequences from 1000 families. *Sci. Rep.* **12**, 9863 (2022).
16. Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).
17. Decuypere, S. *et al.* Diagnosis of Bacterial Bloodstream Infections: A 16S Metagenomics Approach. *PLoS Negl. Trop. Dis.* **10**, e0004470 (2016).
18. Watanabe, N. *et al.* Detection of pathogenic bacteria in the blood from sepsis patients using 16S rRNA gene amplicon sequencing analysis. *PLoS One* **13**, e0202049 (2018).
19. Kingry, L. *et al.* Targeted Metagenomics for Clinical Detection and Discovery of Bacterial Tick-Borne Pathogens. *J. Clin. Microbiol.* **58**, (2020).
20. Sy, M. *et al.* Plasmodium falciparum genomic surveillance reveals spatial and temporal trends, association of genetic and physical distance, and household clustering. *Sci. Rep.* **12**, 938 (2022).
21. Casteling, A. *et al.* GB virus C prevalence in blood donors and high risk groups for parenterally transmitted agents from Gauteng, South Africa. *J. Med. Virol.* **55**, 103–108 (1998).
22. Compston, L. I. *et al.* Prevalence of persistent and latent viruses in untreated patients infected with HIV-1 from Ghana, West Africa. *J. Med. Virol.* **81**, 1860–1868 (2009).
23. Lauck, M. *et al.* GB virus C coinfections in west African Ebola patients. *J. Virol.* **89**, 2425–2429 (2015).



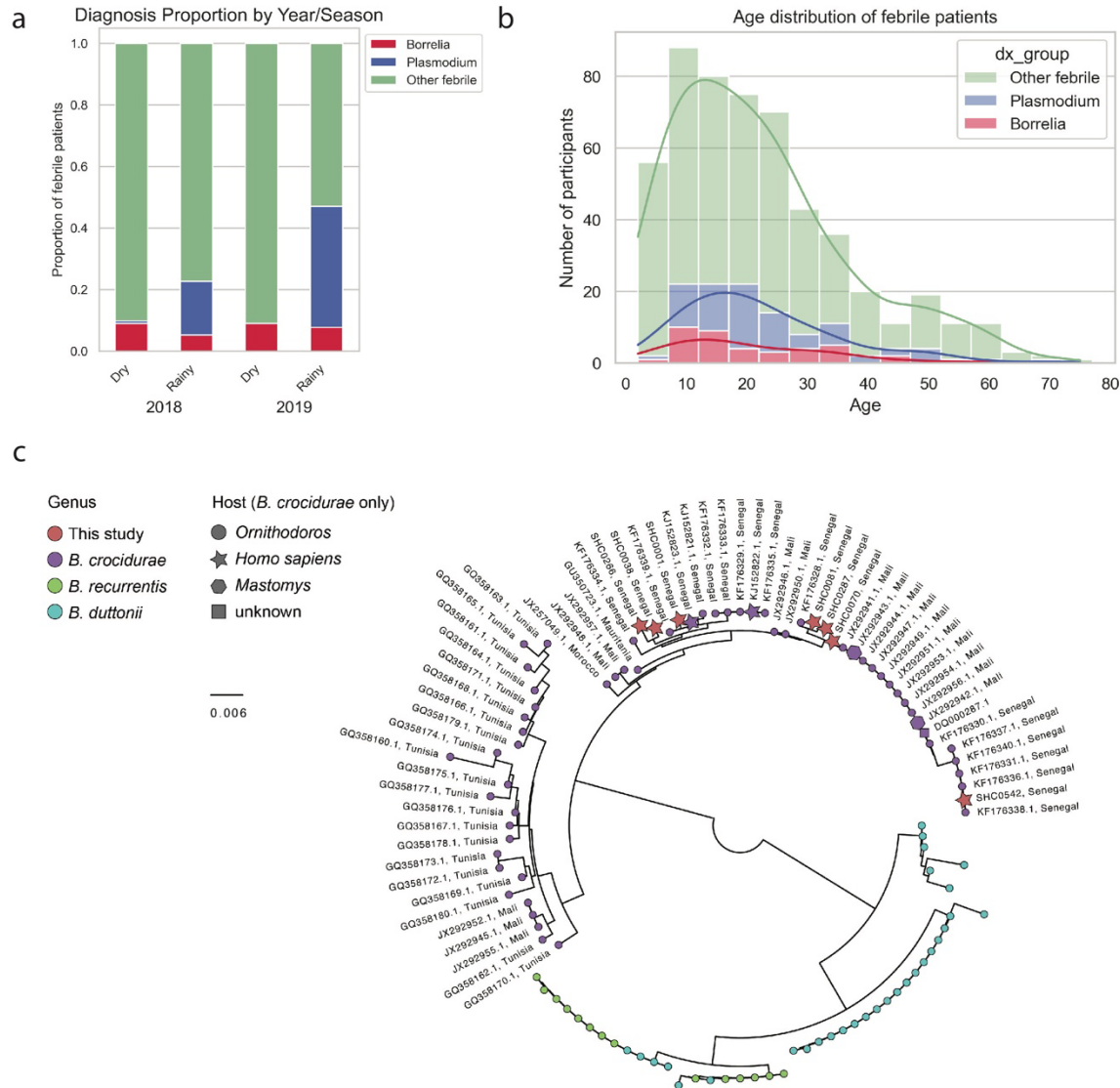
24. Ouattara, L. A. *et al.* Novel human reovirus isolated from children with acute necrotizing encephalopathy. *Emerg. Infect. Dis.* **17**, 1436–1444 (2011).
25. Frickmann, H. *et al.* Metagenomic Sequencing for the Diagnosis of Plasmodium spp. with Different Levels of Parasitemia in EDTA Blood of Malaria Patients-A Proof-of-Principle Assessment. *Int. J. Mol. Sci.* **23**, (2022).
26. Deme, A. B. *et al.* Analysis of pfhrp2 genetic diversity in Senegal and implications for use of rapid diagnostic tests. *Malar. J.* **13**, 34 (2014).
27. Agaba, B. B. *et al.* Systematic review of the status of pfhrp2 and pfhrp3 gene deletion, approaches and methods used for its estimation and reporting in Plasmodium falciparum populations in Africa: review of published studies 2010–2019. *Malar. J.* **18**, 355 (2019).
28. Vial, L. *et al.* Incidence of tick-borne relapsing fever in west Africa: longitudinal study. *Lancet* **368**, 37–43 (2006).
29. Socolovschi, C. *et al.* Rickettsia felis-associated unruptive fever, Senegal. *Emerg. Infect. Dis.* **16**, 1140–1142 (2010).
30. Ángel-Moreno, A., Bolaños, M., Santana, E. & Pérez-Arellano, J. L. Tifus murino importado de Senegal en un inmigrante viajero. *Enfermedades Infecciosas y Microbiología Clínica* **24**, 406–407 (2006).
31. Bunikis, J. *et al.* Typing of Borrelia relapsing fever group strains. *Emerg. Infect. Dis.* **10**, 1661–1664 (2004).
32. Schwan, T. G. *et al.* Endemic foci of the tick-borne relapsing fever spirochete Borrelia crocidurae in Mali, West Africa, and the potential for human infection. *PLoS Negl. Trop. Dis.* **6**, e1924 (2012).
33. Bouattour, A. *et al.* Borrelia crocidurae infection of Ornithodoros erraticus (Lucas, 1849) ticks in Tunisia. *Vector Borne Zoonotic Dis.* **10**, 825–830 (2010).
34. Diatta, G. *et al.* Epidemiology of tick-borne borreliosis in Morocco. *PLoS Negl. Trop. Dis.* **6**, e1810 (2012).
35. Fotso Fotso, A. & Drancourt, M. Laboratory Diagnosis of Tick-Borne African Relapsing Fevers: Latest Developments. *Front Public Health* **3**, 254 (2015).
36. Jakab, Á., Kahlig, P., Kuenzli, E. & Neumayr, A. Tick borne relapsing fever - a systematic review and analysis of the literature. *PLoS Negl. Trop. Dis.* **16**, e0010212 (2022).
37. Kingry, L. C. *et al.* Surveillance for and Discovery of Borrelia Species in US Patients Suspected of Tickborne Illness. *Clin. Infect. Dis.* **66**, 1864–1871 (2018).
38. Lim, L. L. & Rosenbaum, J. T. Borrelia hermsii causing relapsing Fever and uveitis. *Am. J. Ophthalmol.* **142**, 348–349 (2006).
39. Mediannikov, O. *et al.* Borrelia crocidurae infection in acutely febrile patients, Senegal. *Emerg. Infect. Dis.* **20**, 1335–1338 (2014).
40. Trape, J.-F. *et al.* The epidemiology and geographic distribution of relapsing fever borreliosis in West and North Africa, with a review of the Ornithodoros erraticus complex (Acari: Ixodida). *PLoS One* **8**, e78473 (2013).
41. Cutler Sally J., Idris Jibrin M., Ahmed Akeem O. & Elelu Nusirat. Ornithodoros savignyi, the Tick Vector of ‘Candidatus Borrelia kalaharica’ in Nigeria. *J. Clin. Microbiol.* **56**, 10.1128/jcm.00532–18 (2018).
42. Diallo, M. A. *et al.* Plasmodium falciparum malaria co-infection with tick-borne relapsing fever in Dakar. *Malar. J.* **16**, 24 (2017).
43. Normark, J. *et al.* Maladjusted host immune responses induce experimental cerebral malaria-like pathology in a murine Borrelia and Plasmodium co-infection model. *PLoS One* **9**, e103295 (2014).
44. Mediannikov, O. *et al.* Common epidemiology of Rickettsia felis infection and malaria, Africa. *Emerg. Infect. Dis.* **19**, 1775–1783 (2013).
45. Dieme, C. *et al.* Transmission potential of Rickettsia felis infection by Anopheles gambiae mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8088–8093 (2015).

46. Hopkins, H. *et al.* Impact of introduction of rapid diagnostic tests for malaria on antibiotic prescribing: analysis of observational and randomised studies in public and private healthcare settings. *BMJ* **356**, j1054 (2017).
47. WHO informal consultation on fever management in peripheral health care settings: A global review of evidence and practice. <https://www.who.int/publications-detail-redirect/9789241506489>.
48. Butler, T. The Jarisch-Herxheimer Reaction After Antibiotic Treatment of Spirochetal Infections: A Review of Recent Cases and Our Understanding of Pathogenesis. *Am. J. Trop. Med. Hyg.* **96**, 46–52 (2017).
49. Papan, C. *et al.* A host signature based on TRAIL, IP-10, and CRP for reducing antibiotic overuse in children by differentiating bacterial from viral infections: a prospective, multicentre cohort study. *Clin. Microbiol. Infect.* **28**, 723–730 (2022).
50. Rocklöv, J. & Dubrow, R. Climate change: an enduring challenge for vector-borne disease prevention and control. *Nat. Immunol.* **21**, 479–483 (2020).
51. Matranga, C. B. *et al.* Unbiased Deep Sequencing of RNA Viruses from Clinical Samples. *J. Vis. Exp.* (2016) doi:10.3791/54117.
52. viral-ngs. viral-ngs v2.1.33.16. <https://viral-pipelines.readthedocs.io/en/v2.1.33.16/description.html> (2023).
53. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
54. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
55. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–5 (2016).
56. Rambaut. *FigTree*. (2018).
57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
58. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
59. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
60. Rubinsteyn, A. & Feldman, S. fancyimpute: Version 0.0. 16. *Version v0. 0.16. May* (2016).
61. American Heart Association. *Pediatric Advanced Life Support Provider Manual*. (2016).
62. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *arXiv [cs.AI]* (2011).
63. Hancer, E., Xue, B. & Zhang, M. Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems* **140**, 103–119 (2018).
64. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
65. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).
66. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).

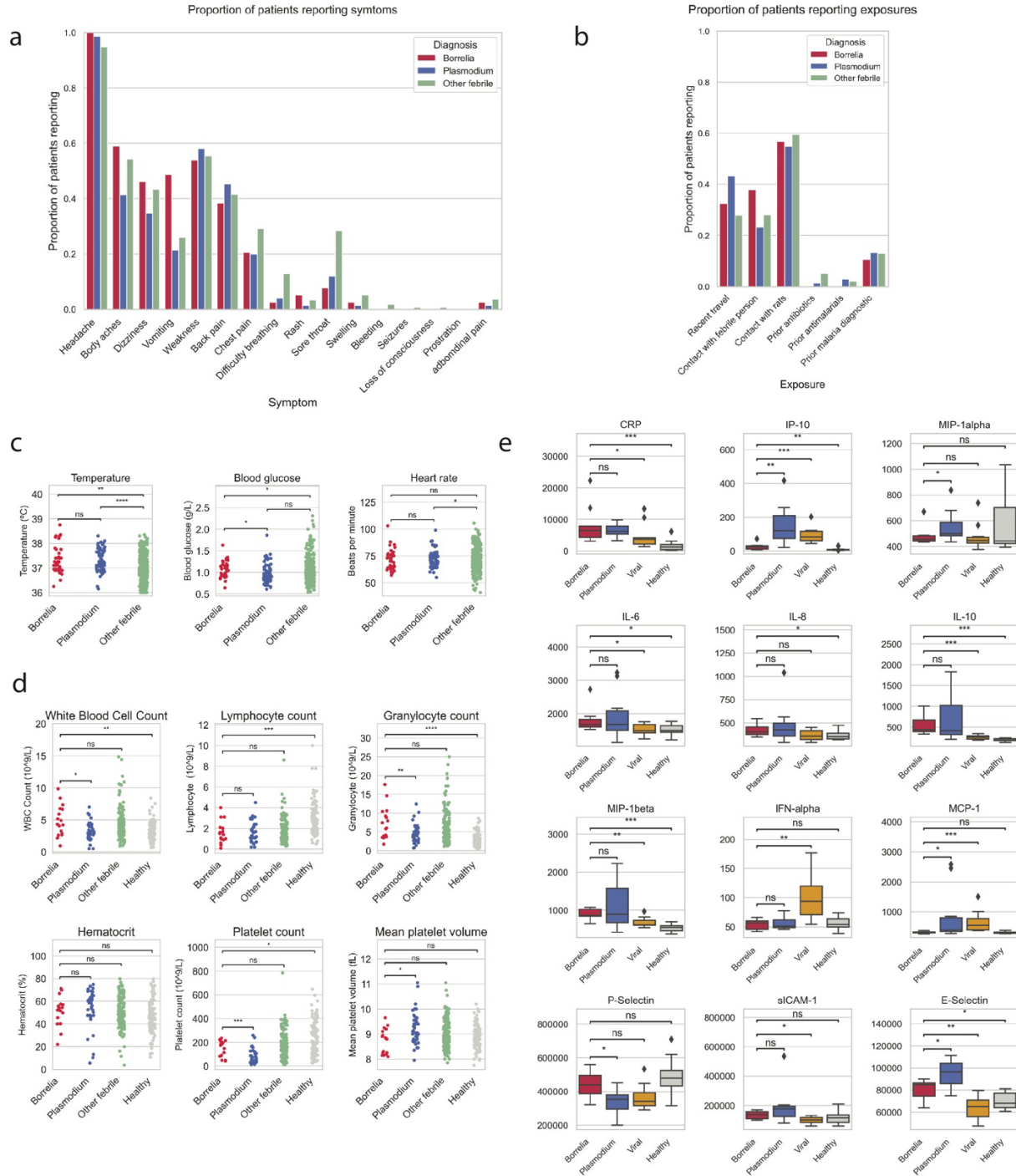
## Main figures:



**Figure 1:** Overview of burden of malarial and NMFI in Thiès, Senegal and NMFI pathogens identified in 2019 **a.** Patients suspected of malaria across 28 health facilities in the Thiès region (data provided by Programme National de lutte contre le Paludisme); gray shading indicates periods of sample collection for this study. **b.** Types of pathogens detected in febrile patients in dry and rainy seasons in 2019. Plasmodium detected by RDT or RNA-mNGS, bacterial infections detected by 16S amplicon sequencing, and viral infections detected by RNA-mNGS. **c.** Viral infections detected by RNA-mNGS. **d.** Bacterial infections detected by 16S sequencing and **e.** Co-infections detected in febrile patients from 2019.

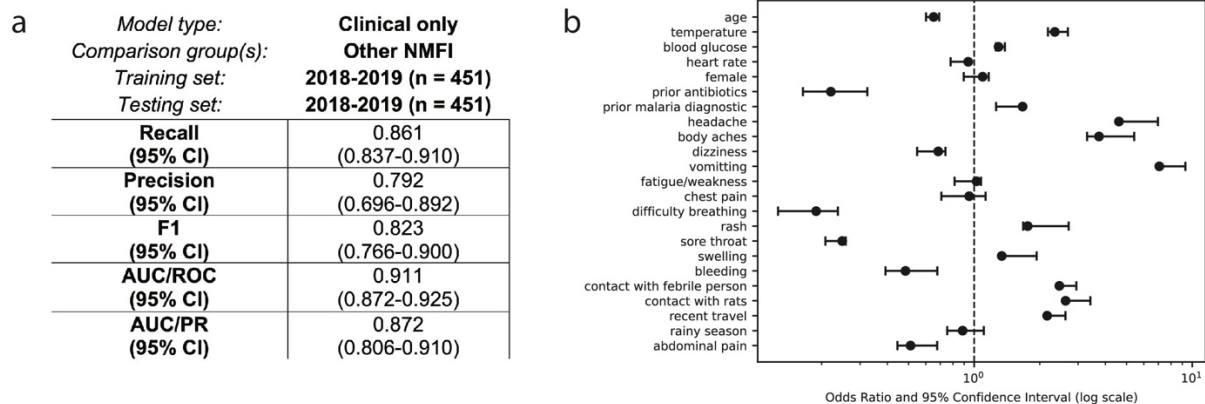


**Figure 2:** Characterization of qPCR-confirmed *Borrelia* and RDT-confirmed *Plasmodium* across both years **a.** *Borrelia*, *Plasmodium*, and non-*Borrelia* NMFI (Other febrile) incidence across dry and rainy seasons in 2018 and 2019. **b.** Age distribution of *Borrelia*, *Plasmodium*, and other febrile patients. **c.** Maximum likelihood phylogenetic tree (midpoint rooted) of IGS sequences from this study (red) in the context of available reference sequences for *B. crocidurae* (purple), *B. duttonii* (cyan), and *B. recurrentis* (green). For *B. crocidurae* isolates, location of collection indicated in label and host species indicated with node shape as available.



**Figure 3:** Clinical characteristics across both years of *Borrelia* (n = 39), *Plasmodium* (n = 75), and other febrile (n = 412) patients including **a.** symptoms reported by patients, **b.** exposures reported by patients, and **c.** vital signs. For 2019 only, **d.** blood cell counts in *Borrelia* (n = 17), *Plasmodium* (n = 41), other febrile (n = 146), and healthy (n = 104) **e.** Serological profile of a subset of patients with *Borrelia* (n = 9), *Plasmodium* (n = 11), viral infection (n = 10) compared to healthy patients (n = 10); cytokines/chemokines with at least one significant difference between groups shown [see Supplementary File 1 for results for all cytokines/chemokines tested].

\*Mann-Whitney-Wilcoxon test two-sided, p-value annotation legend: ns:  $p \leq 1.00e+00$ , \*:  $1.00e-02 < p \leq 5.00e-02$ , \*\*:  $1.00e-03 < p \leq 1.00e-02$ , \*\*\*:  $1.00e-04 < p \leq 1.00e-03$ , \*\*\*\*:  $p \leq 1.00e-04$



**Figure 4:** Results of best performing weighted logistic regression model to distinguish *Borrelia* infection from other NMFI using clinical data, including demographics, symptoms, exposures, and vital signs (trained on 2018-2019 data, tested on the same data with bootstrapping and cross-validation, n = 451). **a.** Model performance metrics and **b.** odds ratios for all features retained in the final model after feature selection.