Mammal Review

THE Mammal SOCIETY

Mammalian biogeography and the Ebola virus in Africa

Journal:	*Mammal Review*
Manuscript ID	MAMMAL-15-57.R2
Manuscript Type:	Review
Keywords :	zoonotic disease, tropical rainforests, Africa, chorotypes, favourability models
Subject Areas (select one):	Distribution
Mammalian Orders (select all that apply):	Chiroptera, Primates, Rodentia, Artiodactyla, Carnivora, Erinaceomorpha

SCHOLARONE™
Manuscripts

1　*Mammal Review Friday, 19 February 16*

2

3　**Mammalian biogeography and the Ebola virus in Africa**

4

5

6

ABSTRACT

1.   Ebola virus is responsible for the fatal Ebola virus disease (EVD).

2.   Identifying the distribution area of the Ebola virus is crucial to understand risk factors conditioning the emergence of new EVD cases. Existing distribution models have underrepresented the potential contribution that hosts and vulnerable species make in sustaining the virus presence.

3.   In this paper, we map favourable areas for Ebola virus in Africa according to environmental and zoogeographic descriptors, independent of human-to-human transmissions. We combine two different biogeographic approaches: analysis of mammalian distribution types (chorotypes), and distribution modelling of the Ebola virus.

4.   We first obtain a model defining the distribution of environmentally favourable areas for the presence of Ebola virus. Based on a review of mammal taxa affected or suspected of exposure to the Ebola virus, we model favourable areas again, this time according to mammalian chorotypes. We then build a combined model in which both the environment and mammalian distributions explain the favourable areas for Ebola virus in the wild.

5.   We demonstrate that mammalian biogeography contributes to explain the distribution of Ebola virus in Africa, although vegetation may also underscore clear limits to the presence of the virus. Our model suggests that the Ebola virus distribution may be even more widespread than previously suspected, given that additional favourable areas are found throughout the coastal areas of West and Central Africa, stretching from Cameroon to Guinea, and extend further East into the East African Lakes region.

31   6.   Our findings show that the most favourable area for the Ebola virus is significantly

32   associated with the presence of the virus in animals. Such core areas are surrounded

33   by regions of intermediate favourability in which human infections of unknown

34   source were found. This difference in association between human and animals and

35   the virus may offer further insights on how EVD can spread.

36

37   **INTRODUCTION**

38   EVD is a zoonosis caused by filoviruses of the genus *Ebolavirus* (hereafter Ebola virus),

39   of which 4 species (*Zaire ebolavirus*, *Sudan ebolavirus*, *Taï Forest ebolavirus* and *Budibugyo*

40   *ebolavirus*) are known in Africa (Kuhn et al. 2010). These viruses cause often-fatal

41   haemorrhagic fever upon infection in humans (Kuhn et al. 2011). Ebola virus transmission

42   from wildlife has mostly been linked to people handling and butchering wild animals for

43   bushmeat (Leroy et al. 2004a). Because bushmeat is important for human nutrition in Sub-

44   Saharan Africa (Fa et al. 2015), it is fundamental to understand how host factors, together

45   with ecological conditions and human behaviour contributes to Ebola virus outbreaks

46   (Groseth et al. 2007). Recent biogeographical analyses have highlighted the importance of

47   potential hosts in explaining the spatial assemblage of human infectious diseases worldwide

48   (Murray et al. 2015). Biogeography has contributed broadly to questions of infectious disease

49   ecology, management and surveillance (e.g. Cliff & Hagget 1995, Guernier et al. 2004, Smith

50   & Guégan 2010). In this study, we build on this and propose a way to integrate virological,

51   zoogeographical and environmental information through a combination of biogeographical

52   approaches. Using these approaches we define the areas where Ebola virus may find suitable

53   conditions to occur in the wild.

54

55   Despite methodological advances and increasing data availability, a limited

56   understanding of the animals potentially implicated in the zoonosis has hampered mapping

57   the extent of Ebola virus. Although likely hosts for the Ebola virus have been highlighted by

58   some authors (Peterson et al. 2007), existing models describing the virus distribution have

59   either not considered the contribution of hosts in sustaining the virus presence (Peterson et al.

60   2004), or have assumed that only a small number of species, suspected to be the reservoirs for

61   the virus, are meaningful in the biogeography of the virus (Pigott et al. 2014). However, the

62   ecology of the Ebola virus is complex and widely unresolved (Groseth et al. 2007). Thus,

63   imposing restrictions to the selection of animal species considered in a distribution model

64   might underrepresent the zoological substrate that could be determining the virus distribution.

65   In fact, the role of particular bat species as true reservoirs of Ebola virus is still under

66   discussion, and it is almost certain that there is a significant virus spillover among mammal

67   species not suspected to be the natural hosts (Leroy et al. 2004a, 2004b, Groseth et al. 2007,

68   Lahm et al. 2007, Olival et al. 2014).

69

70   Murray et al. (2015) suggested that mammalian biodiversity, as a whole, could be the

71   strongest predictor explaining similarities between pathogeographic regions of the world. In

72   this study, we analysed the spatial distribution of Ebola virus in Africa, independent of

73   human-to-human transmissions, under the hypothesis that it is influenced by how mammal

74   species are distributed throughout the region. Thus, instead of taking sides with an uncertain

75   selection of probable natural hosts and victims, we tested the explanatory potential of the

76   biogeographic patterns of mammals (involving species co-occurrence and thus potential

77   interactions) in Africa. In order to test this hypothesis, our prediction was that a distribution

78   model of Ebola virus, based on variables defining the existing types of mammalian

79   distributions in Africa, should better describe the virus occurrences recorded in wildlife than

80   a model based on environmental descriptors alone. In this paper, we examined all known

81   literature regarding events of Ebola virus emergence, either EVD outbreaks or recorded

82   presence of the virus in animals (Appendix S1 in Supporting Information). We then

83   combined two different biogeographic approaches: the analysis of distribution types for

84   mammals (chorotypes), and distribution modelling of Ebola virus, to develop a map of

85   favourable areas for the virus according to both environmental conditions and mammal

86   distributions in Africa.

87

88   **METHODS**

89   **Identifying spatial links between Ebola virus and wildlife**

90   Our model to define the favourable areas for Ebola virus was derived from occurrence

91   data of any recorded presence of the virus in wildlife, disregarding whether it was detected in

92   victims (i.e. organisms experiencing EVD symptoms, including human index cases) or in

93   potential hosts (i.e. species without symptoms, in which the virus is harboured). From the

94   literature, we found a total of 91 geo-referenced events (Appendix S1) in which the Ebola

95   virus was transmitted to humans from wildlife or was present in other mammal species.

96   These events were recognized either via the: 1) detection of viral antibodies or viral nucleic

97   acid (n = 58 events, including 22 in animal carcasses), 2) observed abnormal increase of

98   mortality in animal populations, associated with EVD outbreaks (n = 51 events), or 3)

99   identification of index case of EVD in humans (n = 40 events) (note that the total number of

100   occurrence records is 91 because recognition methods overlap). We found published evidence

101   from cases of serological and/or PCR positivity of EVD in animals, or of animal EVD-linked

102   mortality, in 28 mammal species: 10 primates, 3 rodents, 1 shrew, 8 bats, 1 carnivore and 5

103   ungulates (Fig 1; Appendix S1).

104

105   In 45 of the 58 laboratory-confirmed events of Ebola virus presence in wildlife, the virus

106   was either isolated or was detected by polymerase chain reaction (PCR). In 11 out of the

107   other 12 events, the virus was detected using enzyme-linked immunosorbent assay (ELISA)

108   serology. The high sensitivity of this method, recommended for laboratory diagnosis of EVD

109   by the WHO (2014), was demonstrated by Ksiazek et al. (1992, 1999). The remaining case

110   was serologically identified by both immunofluorescence antibody test and western blot

111   analysis (Hayman et al. 2010).

112

113   A large number of incidents of abnormally high animal mortality attributed to Ebola

114   virus were reported since November 1994 (Formenty et al. 1999). These data are of high

115   significance for the development of a model of Ebola virus distribution in wildlife; the

116   detection of unusual animal mortality can uncover the propagation of Ebola virus (WHO

117   2003a, 2003b, Rouquet et al. 2005). Indeed, increased animal mortality preceded the first

118   human cases in most EVD outbreaks that occurred in Gabon and the Republic of Congo

119   between 2001 and 2003 (Leroy et al. 2004a). Of the 51 mortality events considered here, 19

120   cases were temporal and spatially linked to recorded EVD outbreaks in humans, and/or were

121   confirmed by PCR. The remaining 32 cases represented extensions of disease from confirmed

122   events (Leroy et al. 2004a, Rouquet et al. 2005, Caillud et al. 2006, Lahm et al. 2007).

123

124   Only one of the 40 index cases of haemorrhagic fever syndrome considered in humans

125   was not confirmed to be EVD through laboratory testing of clinical samples. This episode

126   took place in Olloba (Republic of the Congo) in May 2002, and then sample collection was

127   not possible (WHO 2002; Rouquet et al. 2005). However, a previous EVD index case in

128   Olloba had been confirmed just five months before (Leroy et al. 2004a).

129

Environmental model of Ebola virus

130

131   Individual data points were represented as 5-km buffers around the original virus

132   occurrence locations, and we overlapped these buffers to a 1°×1°-resolution grid covering the

133   whole of Africa (n = 2,547 grid cells). With this buffer approach, we aimed to provide the

134   occurrence-allocation in grid cells with some flexibility. In central Africa, 5 km is less than

135   0.05 times the side-length of a 1°×1° grid cell; given that the Ebola virus is hosted by animals,

136   being less than 5-km far from a cell border is virtually equivalent to being located at the

137   limits of two cells. The spatial resolution employed prevented autocorrelation that could exist

138   as a consequence of spatial dependence among very close (< 1°) observations (Legendre &

139   Legendre 1998). The final number of grid cells for detected presence of Ebola virus was 40,

140   of which 33 corresponded to Zaire ebolavirus (Fig 1). From these, an environmental model of

141   Ebola virus (i.e the distribution of environmentally favourable areas for the presence of Ebola

142   virus) was made by employing the Favourability Function (Real et al. 2006, Acevedo & Real

143   2012):

144

$$F = \frac{P}{1-P} \Big/ \left(\frac{n_1}{n_0} + \frac{P}{1-P} \right) \qquad (1),$$

146

147   where F is environmental favourability (0-1), P is probability of occurrence, and n_1 and n_0 are

148   presence and absence numbers, respectively (absences were considered to be those squares

149   not included in the presences subset). P was calculated for the entire African continent

150   through a forward-stepwise logistic regression, according to 24 predictor variables describing

151   types of ecosystems, abiotic factors and anthropogenic pressures on wildlife (see variable

152   descriptions and sources in Appendix S2 in Supporting Information). Land-cover variables

153   were computed as cover percentages in every grid, and the rest of the variables were

154   estimated by averaged grid-values. The logistic regression was based on the 40 Ebola virus

155   occurrence squares (= n_1), considering the four Ebola virus species together. The cases of

156   serological or PCR positivity of EVD, of animal mortality related with EVD, and of index

157   cases of EVD in humans were not differentiated in the occurrence data employed for model

158   building ─events of abnormal animal mortality not directly confirmed in laboratory only

159   contributed to 5 presences, all of which were adjacent to the main core of EVD outbreaks in

160   Gabon and Republic of Congo. Models based on the Favourability Function distinguish those

161   localities with environmental conditions that favour the species' existence from those with

162   detrimental characteristics for its presence, irrespective of the species' proportion of

163   occurrences within the study area. This property is essential for our objectives, as it enables

164   direct comparison between models when several species are involved in the analytical design,

165   and allows for model combinations through fuzzy logic (Barbosa & Real 2012). Fuzzy logic

166   provides metrics for fuzzy sets, i.e. classes of objects with a continuum of membership

167   degrees; these sets are characterized by a membership function that assigns to each object a

168   real number in the interval [0, 1] (Zadeh 1965). In biogeography, environmental

169   favourability, chorotypes, biotic boundaries and species richness can be identified as fuzzy

170   sets (Estrada et al. 2008, Olivero et al. 2011, 2013, Romero et al. in press), thus fuzzy logic

171   can be easily applied to these concepts (see below).

172

173   Adding a spatial descriptor to this variable set allowed us to consider autocorrelation

174   resulting from the spatial structure of the virus distribution (Sokal & Oden 1978), and so take

175   into account the impact of dispersal barriers, geological history and biotic interactions as

176   potential predictors (Fa et al. 2014). To this end, we followed the "trend surface approach"

177   (Legendre & Legendre 1998). Thus, a series of spatial variables resulting from average X and

178   average Y combinations for every square of the grid were examined through a backward

179   stepwise logistic regression. Then we used as spatial descriptor the lineal combination (logit)

180   of spatial variables resulting from the logistic regression.

181

182   In order to control Type-I errors caused by the large number of variables employed in the

183   process, we used Benjamini & Hochberg's (1995) False Discovery Rate (FDR). This control

184   was performed before the stepwise variable selection, so that only significant variables under

185   an FDR of $q < 0.05$ were accepted in a multivariate environmental model. To minimise

186   multicollinearity, we also avoided Pearson correlation values higher than 0.8 between

187   predictor variables.

188

189   **Zoogeographic model of Ebola virus**

190   We defined a zoogeographic model of Ebola virus as the distribution of favourable areas

191   for the presence of Ebola virus according to mammalian chorotypes. A "chorotype" is a

192   distribution pattern followed by one or several species, which can be operatively recognized

193   within an area (Baroni-Urbani et al. 1978, Real et al. 2008); chorotypes, or types of

194   distribution, represent the shared geographical, ecological and evolutionary context of several

195   species (Real et al. 2008). The zoogeographic model of Ebola virus was produced in three

196   steps: (1) we generated a list of mammal species to be considered; (2) we then defined the

197   mammalian chorotypes with potential for explaining the Ebola virus range in Africa; and (3)

198   we finally built a model sustained on a set of predictor variables based on the chorotypes.

199

200   ***Selection of mammal species to be considered***

201   Two different criteria were used to select the list of mammal species based on the a-

202   priori analysis of probable (*sensu lato*) links between mammals and Ebola virus: (1)

203   taxonomic proximity; and (2) biogeographic coincidence. Our aim here was to describe

204   mammalian distribution patterns with enough potential to explain the distribution of Ebola

205   virus; this does not mean that the species selected are proposed to be Ebola virus reservoirs or

206   susceptible taxa.

207

208   Taxonomically related species might be susceptible to similar pathogens (Plyusnin &

209   Morzunov 2001). Using the 28 mammal species recorded to be linked to the Ebola virus (see

210   Appendix S1), we then generated a preliminary list of 216 species (Appendix S3), by

211   considering congenerics: *Mus, Praomys, Sylvisorex, Tadarida*; and whole suprageneric

212   groups if they were represented by more than 5% of their species (a lower proportion was

213   considered to be anecdotic) among the probable host and susceptible species: Pteropodidae,

214   Hystricidae, Hominidae, Cercopithecidae, Viverridae and Suidae families, and forest tribes of

215   the family Bovidae. We also included pouched rats (*Cricetomys*) and grasscutters

216   (*Thryonomys*) in our preliminary list, despite there is no current evidence of relationship with

217   the Ebola virus. We considered that, because members of these rodent genera are so

218   numerically important in the bushmeat trade in central and West Africa (Alexander et al.

219   2014), their inclusion in the analysis was warranted. Thus, our taxonomic criterion was

220   defined so that it discarded groups very far from any suspect hosts (e.g. proboscideans,

221   equids or felines), whilst it considered all those taxa including species whose membership of

222   the Ebola virus host system cannot be categorically denied.

223

224   We then preselected, from our preliminary list of 216 species, those whose biogeography

225   was potentially able to explain the distribution of Ebola virus. To this end, those species

226   geographically coinciding with or nested within the Ebola virus range were chosen. Given

227   that the area covered by the Ebola virus is poorly known, we estimated the geographical

228   coincidence between the virus and mammalian distributions according to models defining the

229   spatial structure of their respective ranges. Thus, for every mammal species, and for Ebola

230   virus, we built spatial models that described their distributions according to purely spatial

231   variables (i.e. combinations of average X and average Y). These models were generated

232   employing the Favourability Function (see above). For mammals, presences/absences were

233   derived from polygon shapefiles for each species available from the IUCN website (IUCN

234   2012, compiled or modified in 2008) projected to a 1º×1º-grid system, which is the maximum

235   spatial resolution at which "extent-of-occurrence" range maps (such as those provided by

236   IUCN) are suitable for analysis (Hurlbert & Jetz 2007). The spatial model of Ebola virus was

237   based on the same presence/absence data set considered for the environmental model. The

238   degree of geographic coincidence was quantified through the fuzzy overlap (or fuzzy

239   similarity) index, whilst the degree of nesting was measured using the fuzzy inclusion index

240   (Dubois & Prade 1980; Olivero et al. 2011); both measures ranged from 0 to 1. The

241   mathematic formulation of the fuzzy overlap (i.e. intersection divided by union) is equivalent

242   to Jaccard's similarity index (Olivero et al. 2011), whose theoretical distribution is randomly

243   distributed around 0.33 (Baroni-Urbani 1980). Thus, we considered a geographic coincidence

244   to be significant when the fuzzy overlap was > 0.33 (i.e. when both Ebola virus and the

245   mammal had more than one third of their ranges in common). There is no theoretical

246   reference for estimating significance of fuzzy inclusion values hence, we considered that a

247   mammal distribution was nested within that of Ebola virus when at least two thirds of the

248   mammal range was included within the Ebola virus range (i.e. fuzzy inclusion > 0.66). A

249   final list of 96 species was ultimately considered for entry in the zoogeographic model of

250   Ebola virus (Appendix S3), 89 of which fulfilled the two selection criteria. Another 7 species

251   were considered because there is published record of probable infection by Ebola virus.

252

253

254   ***Defining and mapping mammalian chorotypes***

255   The distributions of the 96 above-selected species were classified hierarchically

256   according to the Baroni-Urbani & Buser (1976) similarity index, using the UPGMA

257   agglomerative algorithm (Sneath & Sokal 1973). All clusters in the resulting classification

258   dendrogram were assessed for statistical significance with the method in Olivero et al.

259   (2011), using the RMacoqui 1.0 software (http://rmacoqui.r-forge.r-project.org/). The number

260   of resulting chorotypes was not predefined; all groups of distributions that were significantly

261   clustered were considered chorotypes.

262

263   The resulting chorotypes were mapped following the accumulated favourability

264   approach as in Fa et al. (2014). The accumulated favourability is considered a fuzzy-logic

265   method for estimating species richness (Estrada et al. 2008). For all species forming part of

266   the same chorotype, we built an environmental model using the method shown above for

267   Ebola virus; then, in every grid cell of the study area, we added the favourability values

268   defined by these environmental models. So, a cell showing high-accumulated favourability

269   for the species of a chorotype is defined to have favourable conditions for the presence of a

270   large number of these species. Mapping chorotypes this way allowed for further downscaling

271   to a higher spatial resolution (see below).

272

273   ***Building the zoogeographic model of Ebola virus***

274   To build the zoogeographic model of Ebola virus, we followed the same procedure as for

275   the environmental model. In this case, the model was built using every chorotype as a

276   predictor variable. The accumulated favourability for the species of a chorotype was

277   employed as a variable.

278

279   Although every chorotype was based on a finite cluster of species (i.e. the chorotypical

280   cluster), the distribution of every species has a certain degree of membership in all detected

281   chorotypes (Olivero et al. 2011). We proffered a biogeographically-justified list of mammal

282   species whose link with Ebola virus is worth investigating, based on a species membership

283   degree > 0.5 in at least one of the chorotypes entered in the zoogeographic model of Ebola

284   virus. The degree of membership of every species in each chorotype was calculated as the

285   average of the Baroni-Urbani & Buser similarities between a species and all distributions in

286   the chorotypical cluster. The theoretical distribution of this similarity index is randomly

287   distributed around 0.5 (Baroni-Urbani & Buser 1976); we considered membership to be

288   significant above this value.

289

290   **Environmental/zoogeographic model assessment and combination**

291   Both the environmental and zoogeographic models were assessed using calibration

292   (Hosmer & Lemeshow 2000), i.e. testing whether favourability values reflected the existing

293   observations of Ebola virus presence in wildlife. We also compared the performance of both

294   models in relation to goodness of fit using -2×log-likelihood; discrimination capacity using

295   the Area Under the receiver-operating-characteristic Curve (AUC) (Lobo et al. 2008); and

296   classification capacity using sensitivity, specificity, correct classification rate (CCR), Cohen's

297   Kappa (Fielding & Bell 1997), and under- and overprediction rates (Barbosa et al. 2013).

298   Classification measures were based on the 0.5 favourability threshold because probability is

299   equal to the overall prevalence at this level (Acevedo & Real 2012). In a calibrated model,

300   Hosmer-Lemeshow index should be non-significant; for higher goodness of fit, -2×log-

301   likelihood should be lower; for better discrimination, AUC should be higher; for better

302   classification, sensitivity, specificity, CCR and Kappa should be higher whereas under-

303   prediction and over-prediction should be lower.

304

305   The environmental and the zoogeographic models of Ebola virus were combined so that

306   the final model gave favourability values based on the degree to which conditions are both

307   environmentally and zoogeographically favourable for Ebola virus presence. To this end, we

308   used the fuzzy intersection between the environmental model and the zoogeographic model

309   of Ebola virus (Romero et al. in press), by measuring the minimum favourability value for

310   either of the two models in each grid cell.

311

312   We analysed the relative contribution of environment and zoogeography to the combined

313   model. To do this, we mapped where favourability values were derived from either the

314   environmental or the zoogeographic model. We, thus, identified where, and to what extent,

315   one of these models acted as a limiting factor whilst the other model showed a higher

316   favourability for presence of Ebola virus. Finally, we employed the sensitivity index to assess

317   the capacity of the combined model to classify recorded presences of Ebola virus and of the

318   four African Ebola virus species separately.

319

320   We compared the distribution of known EVD outbreaks in humans with our resulting

321   favourability maps for the Ebola virus. For this purpose, we overlapped the locations of index

322   cases in humans with a favourability map divided into three regions depending on their

323   favourability values, according to the thresholds proposed by Muñoz & Real (2006). If the

324   predicted favourability was higher than 0.8, which means that the odds are more than 4:1

325   favourable to Ebola virus, the square was considered as highly favourable. Those areas with a

326   favourability value lower than 0.2 (odds less than 1:4) were considered of low favourability

327   for Ebola virus. The remaining squares were regarded as intermediate favourability areas.

328

329   The potential for contacts between human populations and wildlife could have

330   influenced detection of Ebola virus infections beyond environmental and zoogeographic

331   factors that could favour the presence of Ebola virus. Therefore, we tested whether Ebola

332   virus occurrences poorly explained by the combined model of Ebola virus (i.e. not explained

333   by environmental conditions nor by mammalian biogeography) could be accounted for by the

334   presence of human populations. To do this, we performed a logistic regression of the 40

335   Ebola virus occurrence squares, using favourability values for Ebola virus as the independent

336   variable. The residuals of this regression were then related to rural population density and

337   with distance to roads using linear regression (for variable sources, see Appendix S2).

338

339   To increase the potential uses or our output for management, surveillance or analyses

340   requiring an ecological context for Ebola virus, the combined model was downscaled to

341   $0.1°×0.1°$ resolution squares, by employing the "direct downscaling approach" (Bombi &

342   d'Amen 2012). This method was classified by Bierkens et al. (2000) as "downscaling based

343   on mechanistic models through a deterministic [favourability] function". To do so, we

344   applied the favourability equations involved in the combined model to predictor variables at

345   this resolution. A 10-fold shortening of the grain size (referring to pixel side length) does not

346   severely affect predictions of species distributions (Bombi & d'Amen 2012).

347

348   **RESULTS**

349   **The environmental model of Ebola virus**

350   Three variables (*terra-firme* rain forests, natural vegetation/cropland mosaics and small

351   annual temperature range) were significantly associated with the areas of high environmental

352   favourability for Ebola virus presence ($P < 0.01$) (Figs 1 and 2). Pearson correlation values

353   between these variables were lower than 0.51. Vegetation/cropland mosaics and constant

354 temperatures complemented the predictive power of forests, especially within deforested

355 areas. Crucially, the large swamp forest areas along the lower course of the Congo River

356 ("cuvette congolaise", where EVD outbreaks have so far not been recorded) did not appear as

357 favourable as other *terra firme* central and West African forest areas (Fig 2).

358

359 The zoogeographic model of Ebola virus

360   Our analyses revealed the presence of 16 significant types of distributions, or chorotypes

361 (Fig 3). Four of these chorotypes significantly supported the zoogeographic model of Ebola

362 virus ($P < 0.001$): Rainforest chorotype (RF), West-African Forest chorotype (WAF), North-

363 Western Congolian Forest chorotype (NWCF), and the single-species chorotype of *Mus*

364 *goundae* (MG) (Fig 3). The zoogeographic model of Ebola virus, based on these chorotypes,

365 showed high favourability values within the rain forests of Central and West Africa, with a

366 decline in favourability towards the East, dropping even more dramatically South of the

367 Congo River.

368

369   Having identified the four chorotypes that, according to the zoogeographic model, favour

370 Ebola virus presence, we propose a list of 64 mammal species as a guide for future

371 investigations in the search for Ebola virus hosts and potentially susceptible species (see Fig

372 3; Appendix S3).

373

374 Environmental/zoogeographic model assessment and combination

375   Both the environmental and zoogeographic models appear significantly well calibrated

376 (Table 1). The zoogeographic model shows a better goodness of fit, higher discrimination and

377 greater classification power than the environmental model. However, both provide significant

378 complementary information about the virus distribution. Thus, our environmental model

379   suggests that waterlogged areas (swamps and swamp forests) could limit the presence of

380   Ebola virus (Fig 2). Vegetation/cropland mosaics away from the rain forest block appeared

381   unfavourable in the zoogeographic model, the only exception being a small area of the

382   northern Central African Republic savannas; this area corresponds with the discovery of

383   Ebola virus genetic sequences in rodents in gallery forests around Bohou River, 490 km

384   North of Bangui (Morvan et al. 1999).

385

386   Zoogeography was the most extended limiting factor for the presence of Ebola virus,

387   covering 75.6% of the African continent; instead, environment was the limiting factor in the

388   remaining 34.4%. These percentages turned into 55.2% and 44.8%, respectively, if

389   completely unfavourable areas for Ebola virus (favourability < 0.05) were excluded; and they

390   became 47.7% and 52.3% within the intermediate or high favourability areas (favourability >

391   0.2) (see Appendix S4 in Supporting Information). The combined model inherited, from the

392   zoogeographic model, the eastward decline of favourability values (Fig 2), where

393   zoogeographic favourability was lower than environmental favourability (Appendix S4).

394   Instead, the combined model excluded swamp forests as Ebola virus favourable areas as a

395   legacy of the environmental model.

396

397   Due to the sensitivity of the combined model, more than 92% of the 1º x 1º squares with

398   records of the Ebola virus ($n = 40$) were classified correctly in areas with favourability values

399   higher than 0.5. The combined model explained 94% of the *Zaire ebolavirus* presences ($n =$

400   33), 100% of the *Budibugyo ebolavirus* presences ($n = 2$), 75% of the *Sudan ebolavirus*

401   presences ($n = 4$), and 100% of the *Taï Forest ebolavirus* presences ($n = 1$). Up to 62.5% of

402   Ebola virus presences appeared within the highly favourable areas (i.e. value > 0.8) (Fig 4a),

403   largely coinciding with non-flooded rain forests. The 0.8-favourability threshold significantly

17

404 separated virus occurrences in humans from presences in other mammals (Fig 4a); the highly

405 favourable region included a significantly higher proportion of presences in non-human

406 mammals ($\chi^2_1 = 6.22$, $P < 0.05$), as well as in both humans and non-human mammals ($\chi^2_1 =$

407 8.00, $P < 0.01$). In contrast, presences recorded only in humans were significantly located

408 within the intermediate favourability areas (i.e. value between 0.2 and 0.8) ($\chi^2_1 = 19.16$, P <

409 0.001). This regionalization reveals an emergent property that arises from the combination of

410 the environmental and zoogeographic models. Residual analyses (to explore the differences

411 between predicted and observed probabilities of presence) indicate that there is a significant

412 positive correlation between Ebola virus points recorded outside the highly favourable region

413 and both rural population density ($r = 0.589$, $P = 0.001$) and distance to roads ($r = -0.466$, $P =$

414 0.014). A downscaled geographical representation of the combined model, at a 0.1° x 0.1°

415 spatial resolution, is shown in Fig 4b.

416

417   A total of 17 countries contained high favourability areas (> 0.8) for Ebola virus (Fig

418 4b). In decreasing order, according to the extent of highly favourable areas in each, these

419 were: Democratic Republic of the Congo, Gabon, Cameroon, Republic of Congo, Côte

420 d'Ivoire, Liberia, Ghana, Central African Republic, Nigeria, Equatorial Guinea, Sierra Leone,

421 Uganda, Benin, Angola, Tanzania, Guinea and Togo. Another four countries comprised areas

422 of intermediate favourability (0.2 – 0.8): Rwanda, Burundi, South Sudan and Kenya.

423

DISCUSSION

425   Spatial modelling, applied to pathogen systems, can generate statistically robust

426 predictions of the geographic distributions of the organism causing a disease, and its

427 maintenance host(s) (Peterson 2006, Purse & Golding 2015). However, species distribution

428 modelling applied to EVD transmission in Africa has so far underestimated the contribution

429 hosts and vulnerable species may make in sustaining the virus presence (Peterson et al. 2004,

430 Pigott et al. 2014) (see above). Our models are the first to analyse the contribution of

431 mammalian distribution patterns to the biogeography of Ebola virus in Africa. In our study,

432 we considered the four African Ebola virus species together because of the scarcity of

433 information available on the presence of species other than *Zaire ebolavirus.* The latter is the

434 most lethal (Kuhn et al. 2011), was responsible for the EVD outbreaks in central Africa

435 (Bausch & Schward 2014), and has also been associated with the current epidemic in West

436 Africa (Dudas & Rambaut 2014). Our models have very high sensitivity (0.94) and

437 specificity (0.90) to the presence of *Zaire ebolavirus*, but high favourability areas also

438 adequately predicted the presence of *Budibugyo ebolavirus*, and of *Taï Forest ebolavirus*

439 (sensitivity = 1). However, our models poorly predicted *Sudan ebolavirus*. With better

440 knowledge of the distribution of all Ebola virus species more robust models could have been

441 developed. In particular, models focused on the different Ebola species should be developed

442 promptly, since some mammalian chorotypes that were not entered in the model could better

443 explain the presence of these Ebola virus species e.g. chorotype NECF in the case of

444 *Budibugyo ebolavirus* (compare Figs. 1 and 3).

445

446   Peterson et al. (2007) identified 55 groups of mammalian taxa of interest as potential

447 reservoirs, based on a detailed inspection of distributional overlap patterns with known

448 filovirus disease outbreaks. Our list of 68 species biogeographically-linked to Ebola virus

449 differs from the list of likely reservoir species presented by Peterson et al. (2007). These

450 authors generated an inventory of candidate reservoir species, assuming that the reservoir

451 supports persistent, largely asymptomatic Ebola virus infections, which *de facto* eliminates

452 most primates and ungulates. Our list, unlike Peterson's, is 33% primates (families

453 Hominidae and Cercopithecidae) and 20% ungulates (families Suidae and Bovidae).

454   Members of the fruit bat family (Pteropodidae), Molossid bats, viverrid carnivores, as well as

455   the rodent *Praomys jacksoni* are included in both lists. In our list of species, however, we also

456   included other fruit bats (*Casinycteris argynnis*, *Scotonycteris ophiodon* and *S. zenkeri*) as

457   members of the chorotypes significantly linked to the Ebola virus distribution. Of the two

458   rodent genera *Cricetomys* and *Thryonomys* included in our analyses given their prominence

459   in the bushmeat trade, *C. emini* emerged as worth investigating. Thus, unlike Peterson's, our

460   list is not a proposal of potential reservoir species; instead, we present a list of mammals

461   biogeographically associated with Ebola virus, and thus with the geographic potential of

462   being involved in the virus cycle as reservoirs, hosts or susceptible species. This is not a

463   closed list, even though we have restricted the analysis to a limited number of species. This is

464   because any species (i.e. also those ones not considered in the analysis) are members in all

465   chorotypes to a certain degree (Olivero et al. 2011). Thus, it is possible to evaluate whether

466   unassessed taxa can be considered part of the zoogeographic factor as defined in the Ebola

467   virus model.

468

469   Of the 28 mammal species with published record of probable contact with Ebola

470   virus, only 8 were not included in our 64-species list: *Rousettus aegyptiacus*, *Eidolon helvum*,

471   *Epomophorus gambianus*, *Micropteropus pusillus*, *Tadarida condylura*, *Civettictis civetta*,

472   *Papio anubis* and *Mandrillus leucophaeus* (Appendix S3). What these species have in

473   common is that a large proportion of their populations are distant from where Ebola virus has

474   been detected. We do not challenge the evidenced relationship of these species with Ebola

475   virus, but note that they do not show strong overlap with it (even though there is overlap in

476   some parts of the range). It could be argued that subspecific taxa of the same species may

477   differ in their ability to serve as Ebola virus hosts. In the case of these 8 mammal species,

478   however, all subspecies currently recognized (Kingdon et al. 2013) are partially distributed

479   outside the Ebola virus range, with the only exception of *T. c. osborni*. These species might

480   be more interesting from a phylogeographic perspective, but their distributions cannot explain

481   the geographic distribution of the African Ebola viruses in wildlife; the virus might

482   potentially be hosted by individuals of these species, but factors other than the distribution of

483   these hosts could explain the observed geographic limits of the Ebola virus in Africa (e.g.

484   other hosts or environmental conditions).

485

486   We show, as Pigott et al. (2014) clearly indicated, that the high relative contribution

487   of vegetation in the model might underscore clear limits to the presence of Ebola virus. As in

488   the Pigott et al. (2014) and Murray et al. (2015) maps (and less patently in Peterson et al.'s

489   2004), we confirm West Africa as a highly favourable region for Ebola virus. Our map also

490   suggests that the virus distribution may be even more widespread than previously suspected,

491   given that additional favourable areas for the virus are found throughout the coastal areas of

492   West and Central Africa, stretching from Cameroon to Guinea, and extending further East

493   into the East African Lakes region (i.e. Uganda). Of the 17 countries with highly favourable

494   areas (> 0.8) for Ebola virus occurrence in wildlife in our model, 16 of these are among the

495   top 17 considered as high-risk from Ebola by both Pigott et al. (2014) or Murray et al.

496   (2016)—10 of these common to both sources. The Democratic Republic of the Congo,

497   Gabon, Republic of Congo, Côte d'Ivoire, Uganda and Guinea have a record of EVD index

498   cases in humans, whereas Cameroon, Liberia, Central African Republic, Nigeria, Angola,

499   Ghana, Sierra Leone, Benin, Tanzania and Togo have not. Identifying 'at risk' countries that

500   are currently Ebola virus-free, according to various different evidence types, could be

501   fundamental in achieving adequate levels of preparedness. For example, countries such as

502   Equatorial Guinea, with 94% of its territory highly favourable for Ebola virus in our model,

503   though ranked number 20 by Pigott et al. (2014), should not escape attention. Similarly,

504   Burundi, Rwanda and Kenya, containing (like South Sudan) areas of intermediate

505   favourability —covering all of Burundi and Rwanda— but so far free from EVD cases,

506   should be targeted for further investigation.

507

508   In our map, the favourability for the disease in areas south of the Congo River is

509   remarkably lower than in other rainforest areas; a pattern not apparent in the Pigott map. A

510   possible explanation may be a lack of adequate zoogeographic conditions for the virus to

511   flourish linked to a lower diversity of potential host species in this area. There is evidence

512   that the richness and population structure of mammals found in the central part of the Congo

513   basin is considerably lower than in areas along the western, northern and eastern side of the

514   Congo River (Fa et al. 2014). The limiting role of the zoogeographic factor south to the

515   Congo River (Appendix S4) supports this possibility, but we cannot discard the chance of

516   observation bias in a lowly accessible area as the central Congo.

517

518   Serological surveys among human populations support a rainforest origin of the potential

519   reservoir of Ebola virus (Gonzalez et al. 1989, 2000). Living or spending significant time

520   deep in the forest is positively correlated with seropositivity in Pygmies and in other human

521   groups (Gonzalez et al. 2000, Becquart et al. 2010, Schoepp et al. 2014). This is consistent

522   with our finding of a highest favourability for Ebola virus in the rainforest, and with the

523   recorded concurrence of cases of spillover from animals to humans there even though this

524   information was not used to build the model. However, epidemic outbreaks also originated

525   outside of the rain forest block. This is the case for the first known outbreaks in humans,

526   recorded in Zaire and Sudan in 1976 (WHO 1976, 1978a), of the last and most important

527   episode in West Africa (Bausch & Schward 2014), and of some other cases in Uganda

528   (Albariño et al. 2013), Congo (Pourrut et al. 2005), the Democratic Republic of the Congo

529   (CDC 2009) and Sudan (Onyango et al. 2007). All of them have in common that the human

530   patient zero were not documented to be related to cases of EVD in animals, and that they

531   occurred in areas of intermediate favourability for Ebola virus according to our model. Thus,

532   a main contribution of our model is the finding of a geographical context for exploring which

533   mechanisms differentiate Ebola virus transmission to humans in highly favourable areas

534   compared to areas of intermediate favourability for Ebola virus is of high epidemiological

535   relevance. In the latter areas, gallery forests may play a significant role in harbouring and

536   spreading the Ebola virus.

537

538   EVD epidemics seem to be related more with human behaviour, which increases the risk

539   of contact with the reservoir, than with the emergence of a highly pathogenic viral strain

540   (Gonzalez et al. 2000). The significant relation between human presence (i.e. population

541   density and roads) and EVD cases in areas only moderately favourable to Ebola virus

542   suggests that EVD transmission to humans in these areas could be influenced by

543   anthropogenic factors. A more frequent contact between humans and forest fauna would

544   amplify the chances of Ebola virus transmission to humans, even where environmental or

545   zoogeographic conditions are not the most favourable for the virus. About half of the EVD

546   index cases reported in the region of intermediate favourability happened in the limits of

547   highly favourable areas (Fig 4), which could have facilitated contacts. Even so, some index

548   cases have been located away from the forest domain. One possible explanation for the

549   arrival of *Zaire Ebolavirus* in West Africa, during the 2013 Guinean outbreak, far from its

550   usual haunts in central Africa, is that the virus was introduced by traveller bats (Bausch &

551   Schward 2014). Sáez et al. (2015) also suggested that the index case in Guinea may have

552   been infected by a colony of insectivorous free-tailed bats (*Mops condylurus*). Do

553   Chiropterans have a more significant role in zoonotic spillover in the intermediate-

554   favourability areas, compared with the deep rainforest haunts of Ebola virus? In the first

555   documented case of EVD in Sudan in 1976, the index case was located (by the World Health

556   Organization) in a cotton factory far from the forest block, where the only wild significantly

557   abundant species was an insectivorous bat species (WHO 1978b). Additional analysis and

558   study of the human-wildlife interface are still required to delineate areas in which human

559   populations are at-risk of zoonotic transmission of Ebola virus. In this task, our model could

560   be used to delineate the geographic context of the analysis. As risk factors for zoonotic

561   transmission of Ebola virus are better understood, it will be possible to incorporate this

562   information into future risk mapping assessments and to develop mitigation or prevention

563   measures.

564

565

566   **ACKNOWLEDGEMENTS**

570

571   **REFERENCES**

572   Acevedo P, Real R (2012) Favourability: Concept, distinctive characteristics and potential

573   usefulness. *Naturwissenschaften* 99: 515-522.

574   Albariño CG, Shoemaker T, Khristova ML, Wamala JF, Muyembe JJ, Balinandi S,

575   Tumusiime A, Campbell S, Cannon D, Gibbons A, Bergeron E, Bird B, Dodd K,

576   Spiropoulou C, Erickson BR, Guerrero L, Knust B, Nichol ST, Rollin PE, Ströher U

577   (2013) Genomic analysis of filoviruses associated with four viral hemorrhagic fever

578   outbreaks in Uganda and the Democratic Republic of the Congo in 2012. *Virology* 442:

579   97-100.

580 Alexander JS, McNamara J, Rowcliffe JM, Oppong J, Milner-Gulland EJ (2015) The role of

581   bushmeat in a West African agricultural landscape. *Oryx* 49: 643-651.

582 Barbosa AM, Real R, Muñoz AR, Brown JA (2013) New measures for assessing model

583   equilibrium and prediction mismatch in species distribution models. *Diversity and*

584   *Distributions* 29: 1333-1338.

585 Barbosa AM, Real R (2012) Applying fuzzy logic to comparative distribution modelling: a

586   case study with two sympatric amphibians. *The Scientific World Journal* 2012; ID

587   428206. doi: 10.1100/2012/428206.

588 Baroni-Urbani C (1980) A statistical table for the degree of coexistence between two species.

589   *Oecologia* 44: 287-289.

590 Baroni-Urbani C, Buser MW (1976) Similarity of binary data. *Systematic Zoology* 25: 251-

591   259.

592 Baroni-Urbani C, Rufo S, Vigna-Taglianti A (1978) Materiali per una biogeografia italiana

593   fondata su alcuni generi di Coleotteri, Cicindelidi, Carabidi e Crisomelidi. *Estratto della*

594   *Memorie della Societa Entomologica Italiana* 56: 35‑92.

595 Bausch DG, Schward L (2014) Outbreak of Ebola Virus Disease in Guinea: Where ecology

596   meets economy. *PLOS Neglected Tropical Diseases* 2014; 8:e3056. doi:

597   10.1371/journal.pntd.0003056.

598 Becquart P, Wauquier N, Mahlakõiv T, Nkoghe D, Padilla C, Souris M, Ollomo B, Gonzalez

599   J-P, De Lamballerie X, Kazanji M, Leroy EM (2010) High prevalence of both humoral

600   and cellular immunity to Zaire ebolavirus among rural populations in Gabon. *PLOS ONE*

601   2010; 5: e9126. doi: 10.1371/journal.pone.0009126.

602   Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and

603   powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*

604   57: 289-300.

605   Bierkens MFP, Finke PA, de Willigen P (2000) *Upscaling and downscaling methods for*

606   *environmental research*. Kluwer Academic Publishers, Dordrecht.

607   Bombi P, d'Amen M (2012) Scaling down distribution maps from atlas data: a test of

608   different approaches with virtual species. *Journal of Biogeography* 39: 640-651.

609   Caillud D, Levréro F, Cristescu R, Gatti S, Dewas M, Douadl M, Gutier-Hion A, Raymond M,

610   Ménard N (2006) *Gorilla* susceptibility to Ebola virus: The cost of sociality. *Current Biology* 16:

611   489-491.

612   Center for Disease Control and Prevention (CDC) (2009) *Ebola Hemorrhagic Fever,*

613   *Information Packet.* Special Pathogens Branch Division of High-Consequence Pathogens

614   and Pathology. National Center for Emerging Zoonotic Infectious Diseases, Centers for

615   Disease Control and Prevention, U.S. Department of Health and Human Services,

616   Atlanta, USA.

617   Cliff AD, Haggett P (1995) The epidemiological significance of islands. *Health & Place* 1:

618   199-209.

619   Dubois D, Prade H (1980) *Fuzzy sets and systems: theory and applications*. Academic Press,

620   New York.

621   Dudas G, Rambaut A (2014) Phylogenetic analysis of Guinea 2014 EBOV Ebolavirus

622   outbreak. *PLOS Currents Outbreaks* 2014. doi:

623   10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

624   Estrada A, Real R, Vargas JM (2008) Using crisp and fuzzy modelling to identify

625   favourability hotspots useful to perform gap analysis. *Biodiversity and Conservation* 17:

626   857‑871.

627   Fa JE, Olivero J, Farfán MA, Márquez AL, Vargas JM, Real R, Nasi R (2014) Integrating

628   sustainable hunting in biodiversity protection in Central Africa: Hot spots, weak Spots,

629   and strong spots. *PLOS ONE* 2014; 9: e112367. doi: 10.1371/journal.pone.0112367.

630   Fa JE, Olivero J, Real R, Farfán MA, Márquez AL, Vargas JM, Ziegler S, Wegmann M,

631   Brown D, Margetts B, Nasi R (2015) Disentangling the relative effects of bushmeat

632   availability on human nutrition in central Africa. *Scientific Reports* 5: 8168. doi:

633   10.1038/srep08168.

634   Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in

635   conservation presence-absence models. *Environmental Conservation* 24: 38-49.

636   Formenty P, Boesch C, Wyers M, Steiner C, Donati F, Dind F, Walker F, Le Guenno B (1999) Ebola

637   virus outbreak among wild chimpanzees living in a rain forest of Côte d'Ivoire. *Journal of*

638   *Infecious Diseases* 179: 120-126.

639   Gonzalez J-P, Josse R, Johson ED, Merlin M, Georges AJ, Abandja J, Danyod M, Delaporte

640   E, Dupont A, Ghogomu A, Kouka-Bemba D, Madelon MC, Sima A, Meunier DMY

641   (1989) Antibody prevalence against haemorrhagic fever viruses in randomized

642   representative Central African populations. *Research Virology*140: 319-331.

643   Gonzalez J-P, Nakoune E, Slenczka W, Vidal P, Morvan JM (2000) Ebola and Marburg virus

644   antibody prevalence in selected populations of the Central African Republic. *Microbes*

645   *and Infection* 2: 39-44.

646   Groseth A, Feldmann H, Strong JE (2007) The ecology of Ebola virus. *Trends in*

647   *Microbiology* 15: 408-416.

648   Guernier V, Hochberg ME, Guégan J-F (2004) Ecology drives the worldwide distribution of

649   human diseases. *PLoS Biology* 2: e141.

650   Hayman DTS, Emmerich P, Yu M, Wang L-F, Suu-Ire R, Fooks AR, Cunningham AA,

651   Wood JLN (2010) Long-term survival of an urban fruit bat seropositive for Ebola and

652   Lagos bat viruses. *PLoS ONE* 5: e11978.

653   Hayman DTS, Yu M, Crameri G, Wang L-F, Suu-Ire R, Wood JLN, Cunningham AA (2012)

654   Ebola virus antibodies in fruit bats, Ghana, West Africa. *Emerging Infectious Diseases*

655   18: 1207-1209.

656   Hosmer DW, Lemeshow S. (2000) *Applied Logistic Regression*. Wiley-Interscience, New

657   York, USA.

658   Hurlbert AH, Jetz W (2007) Species richness, hotspots, and the scale dependence of range

659   maps in ecology and conservation. *PNAS* 104: 13384–13389.

660   International Union for Conservation of Nature (IUCN). *IUCN Red List of Threatened*

661   *Species*. *Version 2012.1*. http://www.iucnredlist.org.

662   Kingdon J, Happold D, Butynski T, Hoffman M, Happold M, Kalina J (2013) *Mammals of*

663   *Africa: 6 Vols.* Bloomsbury Publishing, London, UK.

664   Ksiazek TG, Rollin PE, Jahrling PB, Johnson E, Dalgard DW, Peters CJ (1992) Enzyme

665   immunosorbent assay for Ebola virus antigens in tissues of infected primates. *Journal of*

666   *Clinical Microbiology* 30: 947-950.

667   Ksiazek TG, West CP, Rollin PE, Jahrling PB, Peters CJ (1999) ELISA for the detection of

668   antibodies to Ebola viruses. *The Journal of Infectious Diseases* 179: S192-S198.

669   Kuhn JH, Becker S, Ebihara H, Geisbert TW, Johnson KM, Kawaoka Y, Lipkin WI, Negredo

670   AI, Netesov SV, Nichol ST, Palacios G, Peters CJ, Tenorio A, Volchkov VE, Jahrling PB

671   (2010) Proposal for a revised taxonomy of the family Filoviridae: classification, names of

672   taxa and viruses, and virus abbreviations. *Archives of Virology* 155: 2083-2103.

673 Kuhn JH, Dodd LE, Wahl-Jensen V, Radoshitzky SR, Bavari S, Jahrling PB (2011)

674 Evaluation of perceived threat differences posed by filovirus variants. *Biosecurity and*

675 *Bioterrorism* 9: 361-371.

676 Lahm SA, Kombila M, Swanepoel R, Barnes RFW (2007) Morbidity and mortality of wild

677 animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994-2003.

678 *Transactions of the Royal Society of Tropical Medicine and Hygiene* 101: 64-78.

679 Legendre P, Legendre L (1998) *Numerical Ecology*. Second English edition. Elsevier

680 Science, Amsterdam, Netherlands.

681 Leroy EM, Rouquet P, Formenty P, Souquière S, Kilbourne A, Froment J-M, Bermejo M,

682 Smit S, Karesh W, Swanepoel R, Zaki SR, Rollin PE (2004a) Multiple Ebola virus

683 transmission events and rapid decline of Central African wildlife. *Science* 303: 387-390.

684 Leroy EM, Telfer P, Kumulungui B, Yaba P, Rouquet P, Roques P, Gonzalez J-P, Jsiazek

685 TG, Rollin PE, Nerrienet E (2004b) A serological survey of Ebola virus infection in

686 Central African nonhuman primates. *Journal of Infectious Diseases* 90: 1895-1899.

687 Leroy EM, Epelboin A, Mondonge V, Pourrut X, Gonzalez J-P, Muyembe-Tamfum J-J,

688 Formenty P (2009) Human Ebola outbreak resulting from direct exposure to fruit bats in

689 Luebo, Democratic Republic of Congo, 2007. *Vector-Borne and Zoonotic Diseases* 9:

690 723-728.

691 Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the

692 performance of predictive distribution models. *Global Ecology and Biogeography* 17:

693 145-151.

694 Morvan JM, Deubel V, Gounon P, Nakouné E, Barrière P, Murri S, Perpète O, Selekon B,

695 Coudrier D, Gautier-Hion A, Colyn M, Volehkov V (1999) Identification of Ebola virus

696 sequences present as RNA or DNA in organs of terrestrial small mammals of the Central

697 African Republic. *Microbes and Infection* 1: 1193-1201.

698   Muñoz A-R, Real R (2006) Assessing the potential range expansion of the exotic monk

699   parakeet in Spain. *Diversity and Distributions*12: 656-665.

700   Murray KA, Preston N, Allen T, Zambrana-Torrelio C, Hosseini PR, Daszak P (2015) Global

701   biogeography of human infectious diseases. *PNAS* 112: 12746-12751.

702   Olival KJ, Hayman DT (2014) Filoviruses in bats: Current knowledge and future directions.

703   *Viruses* 6: 1759-1788.

704   Olivero J, Real R, Márquez AL (2011) Fuzzy chorotypes as a conceptual tool to improve

705   insight into biogeographic patterns. *Systematic Biology* 60: 645-660.

706   Olivero J, Márquez AL, Real R (2011) Integrating fuzzy logic and statistics to improve the

707   reliable delimitation of biogeographic regions and transition zones. *Systematic Biology*

708   62: 1-21.

709   Onyango CO, Opoka ML, Ksiazek TG, Formenty P, Ahmed A, Tukei PM, Sand RC, Ofula

710   VO, Konongoi SL, Coldren RL, Grein T, Legros D, Bell M, De Cock KM, Bellini WJ,

711   Towner JS, Nichol ST, Rollin PE (2007) Laboratory diagnosis of Ebola hemorrhagic

712   fever during an outbreak in Yambio, Sudan, 2004. *Journal of Infectious Diseases* 196:

713   193-198.

714   Peterson AT, Bauer JT, Mills JN (2004) Ecologic and geographic distribution of Filovirus

715   disease. *Emerging Infectious Diseases*10: 40-47.

716   Peterson AT, Papeş M, Carroll DS, Leirs H, Johnson KM (2007) Mammal taxa constituting

717   potential coevolved reservoirs of filoviruses. *Journal of Mammalogy* 88: 1544-1554.

718   Peterson AT (2006) Ecological niche modeling and spatial patterns of disease transmission.

719   *Emerging Infectious Diseases* 12: 1822–1826.

720   Pigott DM, Golding N, Mylne A, Huang Z, Henry AJ, Weiss DJ, Brady OJ, Kraemer MUG,

721   Smith DL, Moyes CL, Bhatt S, Gething PW, Horby PW, Bogoch II, Brownstein JS,

722   Mekaru ST, Tatem AJ, Khan K, Hay SI (2014) Mapping the zoonotic niche of Ebola

723   virus disease in Africa. eLife 2014; 10.7554/eLife.04395. doi:

724   http://dx.doi.org/10.7554/eLife.04395.

725   Plyusnin A, Morzunov S (2001) Virus evolution and genetic diversity of hantaviruses and

726   their rodent hosts. *Current Topics in Microbiology and Immunology* 256: 47-75.

727   Pourrut X, Kumulungui B, Wittmann T, Moussavou G, Délicat A, Yaba P, Nkoghe D,

728   Gonzalez J-P, Leroy EM (2005) The natural history of Ebola virus in Africa. *Microbes*

729   *and Infection* 7: 1005-1014.

730   Purse BV, Golding N (2015) Tracking the distribution and impacts of diseases with

731   biological records and distribution modelling. *Biological Journal of the Linnean Society*

732   115: 664-677.

733   Real R, Barbosa AM, Vargas JM (2006) Obtaining environmental favourability functions

734   from logistic regression. *Environmental Ecological Statistics* 13: 237-245.

735   Real R, Olivero J, Vargas JM (2008) Using chorotypes to deconstruct biogeographical and

736   biodiversity patterns: the case of breeding waterbirds in Europe. *Global Ecology and*

737   *Biogeography* 17: 735-746.

738   Romero D, Olivero J, Brito JC, Real R (in press) Comparison of approaches to combine

739   species distribution models based on different sets of predictors. *Ecography* In press; doi:

740   10.1111/ecog.01477.

741   Rouquet P, Froment J-M, Bermejo M, Kilbourn A, Karesh W, Reed P, Kumulungui B, Yaba P,

742   Délicat A, Rollin PE, Leroy EM (2005) Wild animal mortality monitoring and human Ebola

743   outbreaks, Gabon and Republic of Congo, 2001-2003. *Emerging Infectious Diseases* 11: 283-290.

744   Sáez AM, Weiss S, Nowak K, Lapeyre V, Zimmermann F, Düx A, Kühl H, Kaba M, Regnaut

745   S, Merkel K, Sachse A, Thiesen U, Villányi L, Boesch C, Dabrowski PW, Radonić A,

746   Nitsche A, Leendertz SAJ, Petterson S, Becker S, Krähling B, Couacy-Hymann E,

747   Akoua-Koffi C, Weber N, Schaade L, Fahr J, Borchert M, Gogarten JF, Calvignac-

748   Spencer S, Leendertz FH (2015) Investigating the zoonotic origin of the West African

749   Ebola epidemic. EMBO *Molecular Medicine* 7: 17-23.

750   Schoepp RJ, Rossi CA, Khan SH, Goba A, Fair JN (2014) Undiagnosed acute viral febrile

751   illnesses, Sierra Leone. *Emerging Infectious Diseases* 20: 1176-1182.

752   Smith KF, Guégan J-F (2010) Changing geographic distributions of human pathogens.

753   *Annual Review of Ecology, Evolution and Systematics* 41: 231-250.

754   Sneath PHA, Sokal RR (1973) *Numerical Taxonomy. The Principles and Practices of*

755   *Numerical Classification*. Freeman, San Francisco, USA.

756   Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. 1. Methodology. *Biological*

757   *Journal of the Linnean Society* 10: 199-228.

758   World Health Organization (WHO) (1978a) Ebola haemorrhagic fever in Sudan, 1976.

759   *Bulletin of the World Health Organization* 56: 247-270.

760   World Health Organization (WHO) (1978b) Ebola haemorrhagic fever in Zaire, 1976.

761   *Bulletin of the World Health Organization* 56: 271-293.

762   World Health Organization (WHO) (2002) Suspected acute haemorrhagic fever syndrome, Gabon.

763   *Weekly Epidemiological Record* 77: 213.

764   World Health Organization (WHO) (2003a) Outbreak(s) of Ebola haemorrhagic fever, Congo and

765   Gabon, October 2001 - July 2002. *Weekly Epidemiological Record* 26: 223-228.

766   World Health Organization (WHO) (2003b) Outbreak(s) of Ebola haemorrhagic fever in the Republic

767   of the Congo, January-April 2003. *Weekly Epidemiological Record* 33: 285-296.

768

769   **Fig 1. Animal species for which there is a record of naturally occurring Ebola virus**

770   **infection either from serological or PCR positivity, or from increased mortality**

771   **attributed to EVD.** Top map: localities with a record of Ebola virus presence in wildlife.

772   Bottom map: 1º x 1º squares with a record of Ebola virus presence (outlined in white); spatial

773   model of Ebola virus (spatial favourability increases from white to black). Maps were

774   generated using ArcGIS.

775

776   **Fig 2. Modelling of the environmental/zoogeographic favourability for the presence of**

777   **Ebola virus in wildlife.** The environmental model is based on *Terra-Firme* Rain Forests

778   (TFRF), Natural Vegetation/Cropland Mosaics (NVCM) and Annual Temperature Range

779   (ATR, with increasing values from yellow to red). The zoogeographic model is composed by

780   four types of mammalian distributions or chorotypes (see Fig 3). The two models are

781   combined according to fuzzy-logic, requiring both environmentally and zoogeographically

782   favourable conditions. Maps were generated using ArcGIS.

783

784   **Fig 3. Classification of 96 mammal species according to distributional similarities (S =**

785   **Baroni-Urbani & Buser similarity index).** *: Branches defining significant chorotypes (P <

786   0.05). For every chorotype, accumulated favourability values (increasing from white to black)

787   of all species are mapped. Acronyms name chorotypes. **NWCF**: North-Western Congolian

788   Forest; **SECF**: South-Eastern Congolian Forest; **NCF**: Northern Congolian Forest; **SCF**:

789   Southern Congolian Forest; **CSBF**: Cross-Sanaga-Bioko Forest; **RF**: Rain Forest; **WAF**:

790   West-African Forest; **SS**: Sub-Saharan; **RA**: *Rousettus aegyptiacus*; **TB**: *Tadarida*

791   *bemmeleni*; **MO**: *Mus oubanguii*; **AS**: *Allochrocebus solatus*; **CD**: *Cercopithecus dryas*; **MG**:

792   *Mus goundae*; **PM**: *Praomys mutoni*; **SK**: *Sylvisorex konganensis*. Solid circles indicate the

793   64 mammals that are significant members (degree > 0.5) in chorotypes explaining the

794   zoogeographic model of Ebola virus (RF, WAF, NWCF and MG). Red circles indicate

795   published records of probable contact with the Ebola virus. Maps were generated using

796   ArcGIS.

797

798   **Fig 4. Combination of the environmental and the zoogeographic models of Ebola virus**

799   **at the original 1º × 1º resolution (A) and downscaled to a 0.1º × 0.1º resolution.** Red tones

800   distinguish regions with high (value > 0.8), intermediate (0.2-0.8), and low favourability (<

801   0.2) for the presence of Ebola virus. White areas are completely unfavourable for Ebola virus

802   (< 0.05). Circles represent squares with cases of serological or PCR positivity of Ebola virus

803   infection in animals, animal mortality attributed to EVD, and zoonotic transmission to

804   humans. G: Guinea; SL: Sierra Leone; L: Liberia; CI: Cote d'Ivoire; GH: Ghana; T: Togo;

805   Be: Benin; N: Nigeria; CA: Cameroon; CAR: Central African Republic; S: South Sudan; U:

806   Uganda; R: Rwanda; B: Burundi; DRC: Democratic Republic of Congo; A: Angola; CO:

807   Congo; GA: Gabon; EG: Equatorial Guinea. Maps were generated using ArcGIS.

808

809 **Table 1. Model assessment and comparison based on calibration, goodness of fit, and**

810 **discrimination and classification capacities.** In a calibrated model, H-L should be non-

811 significant; for higher goodness of fit, -2logL should be lower; for better discrimination, AUC

812 should be higher; for better classification, sensitivity, specificity, CCR and Kappa should be

813 higher whereas under-prediction and over-prediction should be lower.

814

	Model	
Indices	Environmental	Zoogeographic
H-L*	χ^2_7=13.02, P>0.05	χ^2_7=7.03, P>0.05
-2logL**	226.26	225.48
AUC***	0.943	0.965
Sensitivity	0.925	0.975
Specificity	0.885	0.889
Kappa	0.180	0.196
CCR****	0.886	0.890
Under-prediction	0.00135	0.00045
Over-prediction	0.886	0.877

815 * Hosmer-Lemeshow calibration index

816 ** -2 x ln(Likelihood)

817 ***Area Under the receiver-operating-characteristic (ROC) Curve

818 **** Correct Classification Rate

819

Supporting Information - online

820

821 **Appendix S1.** Referenced locations of Ebola virus presence in wildlife.

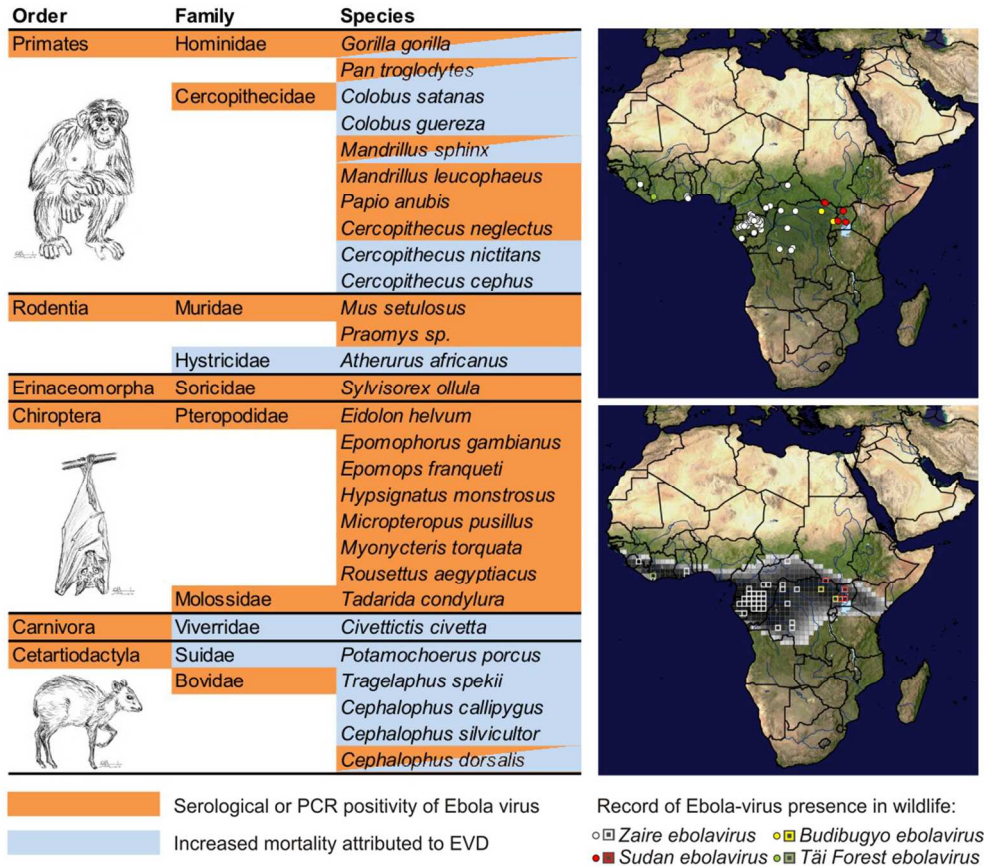822 **Appendix S2.** Predictor-variable description and sources.

823 **Appendix S3.** List and features of the 216 mammal species considered in this study.

824 **Appendix S4.** Contribution of environment and zoogeography to the Ebola virus distribution

825 model based on combination of explanatory factors.

826

827

Order	Family	Species
Primates	Hominidae	*Gorilla gorilla*
		Pan troglodytes
	Cercopithecidae	*Colobus satanas*
		Colobus guereza
		Mandrillus sphinx
		Mandrillus leucophaeus
		Papio anubis
		Cercopithecus neglectus
		Cercopithecus nictitans
		Cercopithecus cephus
Rodentia	Muridae	*Mus setulosus*
		Praomys sp.
	Hystricidae	*Atherurus africanus*
Erinaceomorpha	Soricidae	*Sylvisorex ollula*
Chiroptera	Pteropodidae	*Eidolon helvum*
		Epomophorus gambianus
		Epomops franqueti
		Hypsignatus monstrosus
		Micropteropus pusillus
		Myonycteris torquata
		Rousettus aegyptiacus
	Molossidae	*Tadarida condylura*
Carnivora	Viverridae	*Civettictis civetta*
Cetartiodactyla	Suidae	*Potamochoerus porcus*
	Bovidae	*Tragelaphus spekii*
		Cephalophus callipygus
		Cephalophus silvicultor
		Cephalophus dorsalis

Serological or PCR positivity of Ebola virus

Increased mortality attributed to EVD

Record of Ebola-virus presence in wildlife:
- *Zaire ebolavirus*   *Budibugyo ebolavirus*
- *Sudan ebolavirus*   *Täi Forest ebolavirus*

Animal species for which there is a record of naturally occurring Ebola virus infection either from serological or PCR positivity, or from increased mortality attributed to EVD.
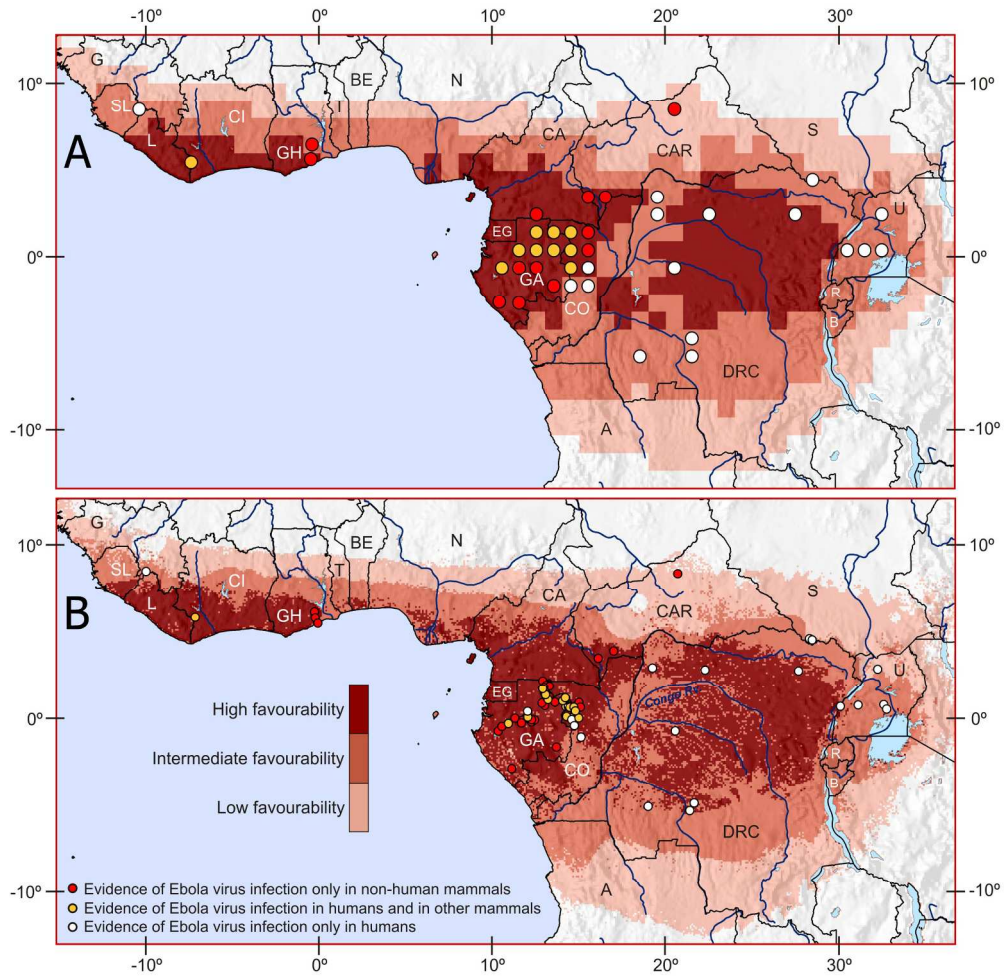201x175mm (150 x 150 DPI)
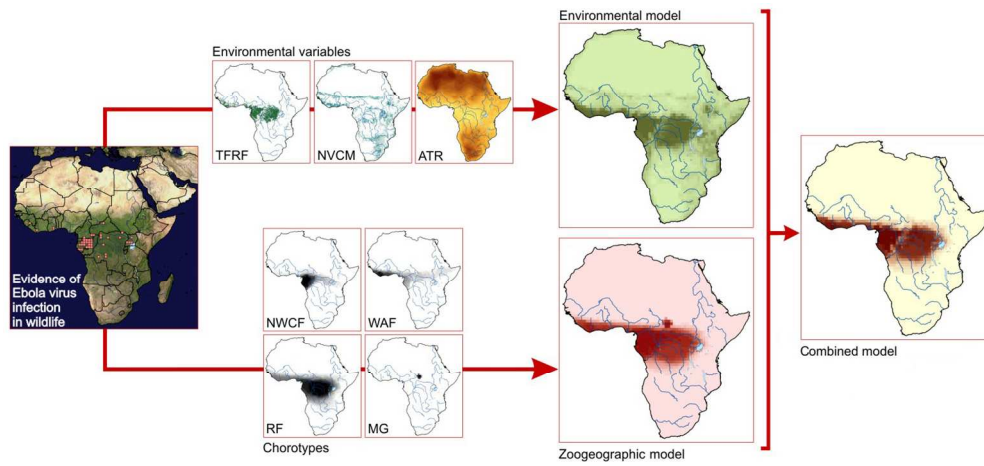
Modelling of the environmental/zoogeographic favourability for the presence of Ebola virus in wildlife.
272x125mm (150 x 150 DPI)

Modelling of the environmental/zoogeographic favourability for the presence of Ebola virus in wildlife.
157x110mm (300 x 300 DPI)

Combination of the environmental and the zoogeographic models of Ebola virus at the original 1º × 1º
resolution (A) and downscaled to a 0.1º × 0.1º resolution.
205x199mm (300 x 300 DPI)