Supplementary Materials for:

"The animal-origin of major human infectious diseases: What can past epidemics teach us about preventing the next pandemic?"

Guha Dharmarajan, Ruiyun Li, Emmanuel Chanda, Katharine R. Dean, Rodolfo Dirzo, Kjetill S. Jakobsen, Imroze Khan, Herwig Leirs, Zheng-Li Shi, Nathan D. Wolfe, Ruifu Yang and Nils Chr. Stenseth

A timeline of emergence of diseases of animal origin

To produce a timeline for the emergence of important animal-origin diseases, we first identified diseases of known zoonotic origin from the list of diseases considered to be a threat to global health [1] or which require urgent research [2] as identified by the World Health Organization. The final list included 15 diseases (see below). The diseases included those that are: (1) Strictly zoonotic and maintained in the human population only through transmission from a vertebrate animal host (e.g., Rift Valley fever and Hendra virus disease); (2) Diseases that are primarily maintained by zoonotic spillover but which can also be transmitted directly between humans (e.g., Ebola/Marburg virus diseases and MERS); (3) Diseases of animal origin which show very efficient human-to-human transmission (e.g., HIV infection, H5N1/H1N1 influenza). We also, included several diseases are suspected to be of zoonotic origin but for which the vertebrate animal reservoir remains unconfirmed (e.g., Ebola virus disease, SARS and COVID-19). For each disease we identified: (1) Year of initial identification; (2) The location from where

the pathogen was first reported. In some cases, the location of initial report is not necessarily in the areas currently affected by the disease (e.g., Marburg virus disease); (3) The countries where local transmission has been reported (i.e., we excluded countries associated only with travel-related cases, if this information was known). The diseases included in the figure and source references are detailed below: (a) Crimean-Congo hemorrhagic fever [3-6]; (b) Dengue fever [3, 7, 8]; (c) Marburg virus disease [3, 9, 10]; (d) Lassa fever [3, 11, 12]; (e) Rift Valley fever [13-15]; (f) Hendra virus disease [3, 16, 17]; (g) Highly Pathogenic Asian Avian Influenza A subtype H5N1 [3, 18, 19]; (h) Nipah virus disease [3, 16, 20]; (i) HIV/AIDS [3, 21, 22]; (j) Zika fever [23-25]; (k) Ebola virus disease [3, 26, 27]; (l) Sudden Acute Respiratory Syndrome (SARS) [3, 28, 29]; (m) Influenza A virus subtype H1N1 [30-32]; (n) Middle East Respiratory Syndrome (MERS) [33-35]; (o) Coronavirus Disease 2019 (COVID-19) [36-39].

A framework to prioritize species and geographical areas for zoonotic disease surveillance

To help prioritize zoonotic disease surveillance there is a need to identify species of specific concern that are understudied and/or eco-geographical regions where disease emergence risk is high. Here we focus only on mammals, as disease emergence risk from mammalian species is high [40]. We first tested the relative importance of three variables that have been posited to affect the likelihood of a species harboring zoonotic pathogens: (a) phylogenetic proximity; (b) spatial overlap with humans; (c) pathogen richness. To carry out these analyses we first downloaded data relating to pathogen diversity and the

zoonotic status (i.e., the identification of pathogens that can infect humans) status of mammals from a comprehensive database reported previously [40], and available at https://figshare.com/articles/dataset/Zoonotic_hosts_and_land_use_change_PREDICTS_code_and_data/7624289. Our analyses included a total of 505 species reported by Gibb et al., for which we were able to obtain phylogenetic and distribution data (see below).

To calculate phylogenetic proximity of each species to humans we obtained the mammalian phylogeny from the PHYLACINE database reported previously [41], and available at https://megapast2future.github.io/PHYLACINE_1.2/. We then generated a majority-rule consensus tree using the R package APE, and computed the consensus edges using the R package PHYTOOLS. Using the final ultrametric tree we calculated the cophenetic distances between all species and humans using APE.

To calculate the degree of spatial overlap with humans we used six steps: (1) We obtained the extent of occupancy (EOO) spatial polygons for each species from the International Union for Conservation of Nature (IUCN) Red List (https://www.iucnredlist.org/resources/spatial-data-download). While using the IUCN Red List to model species distributions is fraught with difficulties (e.g., subjectivity and uniformity of coverage), such expert-lists provide the most comprehensive documentation of species distributions at global scales, and thus remain an invaluable resource [42]; (2) Previous research has shown that using the EOO directly leads to overestimation of species occupancy as the EOO could include unsuitable habitats [42]. Thus, for each species also downloaded species-specific habitat and elevation range data

from the IUCN Red List (http://apiv3.iucnredlist.org/api/v3/docs) using the R package RREDLIST; (3) We then downloaded the spatially explicit characterization of IUCN's habitat classification scheme [43], available in raster file format at https://zenodo.org/record/4058819#.X8xH19hKiUk. For each species we then calculated the proportion of usable habitat in each raster cell that fell within the EOO polygon; (4) For species with known elevation ranges, we also restricted the Area of Habitat to suitable elevations based on elevation data obtained from the USGS (https://topotools.cr.usgs.gov/gmted_viewer/gmted2010_global_grids.php); (5) We then obtained global human population data from the Socioeconomic Data and Applications Center (https://sedac.ciesin.columbia.edu/data/set/popdynamics-1-km-downscaled-pop-base-year-projection-ssp-2000-2100-rev01/data-download), and converted the population estimates to density by dividing by raster cell area; (6) Finally, for each species we calculated an index of spatial overlap with humans by summing the human density across all raster cells within the EOO weighted by the proportion of suitable habitat within each raster cell (taking into consideration habitat characteristics and elevation; see above).

We obtained data on non-human-shared pathogen richness for each mammal species from [40]. Because observed pathogen richness is expected to be correlated with research effort, Gibb et. al. controlled for effort by modelling the effect of publication effort on pathogen richness and calculating the residuals (scaled to mean 0, s.d. 1) from observed pathogen richness value of each species [40]. We used the residual richness of non-human-shared pathogens as our index of pathogen richness. This measure will tend to

84 underestimate the pathogen richness of zoonotic hosts as it does not take into
85 consideration zoonotic pathogen richness. However, we felt the measure was appropriate
86 for our analyses as our main focus was to develop a framework to identify potential
87 zoonotic hosts. Since, we expected that pathogen richness would be positively associated
88 with zoonotic host status, our approach is expected to be conservative.

89 To test if the above three variables were associated with the status of a mammalian
90 species as a zoonotic host (i.e., a species harboring zoonotic pathogens) we used the same
91 regression approach used by Gibb et al. [40], with some minor modifications. Briefly, we
92 carried out a Bayesian mixed-effects regression with a binomial error distribution using
93 Integrated Nested Laplace Approximation (INLA) as implemented in the R package R-
94 INLA. We used the host status as a Boolean dependent variable: non-hosts (species that
95 did not harbor human pathogens) and hosts (species harbored human pathogens)) being
96 classified as “0” and “1”, respectively. We used three independent variables: (1)
97 phylogenetic proximity to humans (inverse of the phylogenetic distance calculated as
98 described earlier and scaled to mean 0, s.d. 1); (2) spatial overlap with humans (using the
99 spatial overlap index calculated as described above, and scaled to mean 0, s.d. 1); (3)
100 Pathogen richness (using the residual pathogen richness as described earlier). All
101 analyses included the family and zoogeographic region as random effects We did not
102 include the order because the phylogenetic distance to humans primarily reflected order
103 level differences among the taxa. We used every unique combination Data on
104 zoogeographic regions was obtained from https://macroecology.ku.dk/resources/wallace

105 [44]. Because some species occupied multiple zoogeographic regions we used each
106 unique combination of regions as a unique random effect. Finally, the observed
107 classification of species as hosts or non-hosts of human pathogens was expected to be
108 associated with research effort [40]. We thus controlled for differences in effort using the
109 bootstrap approach described by Gibb et al. [40]. Briefly, we carried out 100 bootstrap
110 iterations, and for each iteration we refit the INLA model using the data in which non-
111 host species was randomly transitioned to host status as a Bernoulli trial with success
112 probability p equal to estimated false classification probability [for details on the
113 bootstrap analyses and calculation of the false classification probability see Gibb et al.
114 [40]]. For each fitted model we then drew 1,000 samples from the approximated joint
115 posterior distribution, and we the median and quantile ranges (67% and 95%) across all
116 samples from the bootstrap ensemble [40].

117 To develop a framework for spatial prioritization of zoonotic host surveillance efforts we
118 developed a macroecological model (MEM) of zoonotic host diversity [45, 46]. Briefly,
119 we first estimated the underlying diversity of zoonotic host species by summing all
120 species-specific AOH rasters (see details above). We then fitted a MEM to these data
121 using a random forests algorithm (as implemented in the R package RANGER). The
122 independent variables used included those associated with: (1) terrain (elevation, slope,
123 aspect and roughness) calculated from elevation data (see above) using the R package
124 RASTER; (2) the bioclimatic variables calculated using the R packages DISMO and
125 ENVIREM using climate data (downloaded from http://chelsa-climate.org/); (3) landcover

(downloaded from https://luh.umd.edu/data.shtml) which included four major habitat types, namely forest (primary and secondary forested land), non-forest (primary and secondary non-forested land), agriculture (all crops, pasture and rangeland) and urban. We also included two variables associated with secondary habitats, namely the mean age of secondary habitat (years) and mean secondary biomass density (kg C/m^2); (4) Human population density (see above). Prior to analyses we reduced the number of variables by only retaining one of any pair of highly correlated variables ($R > 0.70$). We controlled for spatially inequality in research effort using the bootstrap approach described above by fitting 100 independent random forest models, with each bootstrap iteration randomly transitioning non-host to host status based on the false classification probability (see above). We report the average value of all 100 random forest model predictions.

While zoonotic host diversity is an important variable affecting zoonotic disease risk [47, 48], human density also plays an important role. Thus, areas with high zoonotic host diversity could have low risk of disease emergence in humans if human density (and thus encounter risk) is low. To prioritize areas based on both zoonotic host diversity and human density we generated a composite raster consisting of 16 risk categories based on the pairwise combination of the quantiles of the zoonotic host diversity and human population density rasters. The composite raster was plotted on an additive (cyan-magenta-yellow) color scale to visually emphasize differences in the two axes considered (Fig. 3).

Supplementary references

1. WHO. Ten threats to global health in 2019 Geneva, Switzerland: World Health Organization; 2020 [July 1, 2020]. Available from: https://www.who.int/news-room/feature-stories/ten-threats-to-global-health-in-2019

2. WHO. Prioritizing diseases for research and development in emergency contexts Geneva, Switzerland: World Health Organization; 2020 [July 1, 2020]. Available from: https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts.

3. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990-3. doi: 10.1038/nature06536. PubMed PMID: 18288193; PubMed Central PMCID: PMCPMC5960580.

4. Spengler JR, Estrada-Pena A, Garrison AR, Schmaljohn C, Spiropoulou CF, Bergeron E, et al. A chronological review of experimental infection studies of the role of wild animals and livestock in the maintenance and transmission of Crimean-Congo hemorrhagic fever virus. Antiviral Res. 2016;135:31-47. Epub 2016/10/28. doi: 10.1016/j.antiviral.2016.09.013. PubMed PMID: 27713073; PubMed Central PMCID: PMCPMC5102700.

5. Gargili A, Estrada-Pena A, Spengler JR, Lukashev A, Nuttall PA, Bente DA. The role of ticks in the maintenance and transmission of Crimean-Congo hemorrhagic fever virus: A review of published field and laboratory studies. Antiviral Res. 2017;144:93-119. Epub 2017/06/06. doi: 10.1016/j.antiviral.2017.05.010. PubMed PMID: 28579441; PubMed Central PMCID: PMCPMC6047067.

6. CDC. Crimean-Congo Hemorrhagic Fever (CCHF): Outbreak Distribution Map Atlanta, GA: Centers for Disease Control and Prevention; 2014 [July 22, 2020]. Available from: https://www.cdc.gov/vhf/crimean-congo/outbreaks/distribution-map.html.

7. Holmes E, Twiddy S. The origin, emergence and evolutionary genetics of dengue virus. Infection, Genetics and Evolution. 2003;3(1):19-28. doi: 10.1016/s1567-1348(03)00004-2.

8. CDC. Dengue around the world Atlanta, GA: Centers for Disease Control and Prevention; 2020 [July 20, 2020]. Available from: https://www.cdc.gov/dengue/areaswithrisk/around-the-world.html.

9. Swanepoel R, Smit SB, Rollin PE, Formenty P, Leman PA, Kemp A, et al. Studies of reservoir hosts for Marburg virus. Emerg Infect Dis. 2007;13(12):1847-51. Epub 2008/02/09. doi: 10.3201/eid1312.071115. PubMed PMID: 18258034; PubMed Central PMCID: PMCPMC2876776.

10. CDC. Chronology of Marburg Hemorrhagic Fever Outbreaks Atlanta, GA: Centers for Disease Control and Prevention; 2014 [July 22, 2020]. Available from: https://www.cdc.gov/vhf/marburg/resources/outbreak-table.html.

11. Monath TP. Lassa fever: review of epidemiology and epizootiology. Bull World Health Organ. 1975;52(4-6):577-92. Epub 1975/01/01. PubMed PMID: 782738; PubMed Central PMCID: PMCPMC2366662.

12. CDC. Lassa Fever: Outbreak Distribution Map Atlanta, GA: Centers for Disease Control and Prevention; 2014 [July 22, 2020]. Available from: https://www.cdc.gov/vhf/lassa/outbreaks/index.html.

13. Daubney R, Hudson JR, Garnham PC. Enzootic hepatitis or rift valley fever. An undescribed virus disease of sheep cattle and man from east africa. The Journal of Pathology and Bacteriology. 1931;34(4):545-79. doi: 10.1002/path.1700340418.

14. Linthicum KJ, Britch SC, Anyamba A. Rift Valley Fever: An Emerging Mosquito-Borne Disease. Annu Rev Entomol. 2016;61:395-415. Epub 2016/03/18. doi: 10.1146/annurev-ento-010715-023819. PubMed PMID: 26982443.

15. CDC. RVF Distribution Map Atlanta, GA: Centers for Disease Control and Prevention; 2020 [July 22, 2020]. Available from: https://www.cdc.gov/vhf/rvf/outbreaks/distribution-map.html.

16. Eaton BT, Broder CC, Middleton D, Wang LF. Hendra and Nipah viruses: different and dangerous. Nat Rev Microbiol. 2006;4(1):23-35. Epub 2005/12/17. doi: 10.1038/nrmicro1323. PubMed PMID: 16357858; PubMed Central PMCID: PMCPMC7097447.

17. WHO. Hendra virus infection Geneva, Switzerland: World Health Organization; 2020 [July 22, 2020]. Available from: https://www.who.int/health-topics/hendra-virus-disease#tab=tab_1.

18. Van Kerkhove MD, Mumford E, Mounts AW, Bresee J, Ly S, Bridges CB, et al. Highly pathogenic avian influenza (H5N1): pathways of exposure at the animal-human interface, a systematic review. PLoS One. 2011;6(1):e14582. Epub 2011/02/02. doi: 10.1371/journal.pone.0014582. PubMed PMID: 21283678; PubMed Central PMCID: PMCPMC3025925.

19. WHO. Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2020 Geneva, Switzerland: World Health Organization; 2020 [July 22, 2020]. Available from: https://www.who.int/influenza/human_animal_interface/2020_10_07_tableH5N1.pdf?ua=1.

20. WHO. Nipah virus Geneva, Switzerland: World Health Organization; 2018 [July 22, 2020]. Available from: https://www.who.int/news-room/fact-sheets/detail/nipah-virus.

21. Holmes EC. On the origin and evolution of the human immunodeficiency virus (HIV). Biol Rev Camb Philos Soc. 2001;76(2):239-54. Epub 2001/06/09. doi: 10.1017/s1464793101005668. PubMed PMID: 11396848.

22. IHME. Global Health Data Exchange Seattle, WA: Institute for Health Metrics and Evaluation; 2020 [July 22, 2020]. Available from: http://ghdx.healthdata.org/gbd-results-tool.

23. Simpson DIH. Zika virus infection in man. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1964;58(4):335-8. doi: 10.1016/0035-9203(64)90200-7.

24. Kindhauser MK, Allen T, Frank V, Santhana RS, Dye C. Zika: the origin and spread of a mosquito-borne virus. Bull World Health Organ. 2016;94(9):675-86C. Epub 2016/10/07. doi: 10.2471/BLT.16.171082. PubMed PMID: 27708473; PubMed Central PMCID: PMCPMC5034643.

25. WHO. Countries and territories with current or previous Zika virus transmission Geneva, Switzerland: World Health Organization; 2019 [July 22, 2020]. Available from: https://www.who.int/emergencies/diseases/zika/countries-with-zika-and-vectors-table.pdf.

26. Muyembe-Tamfum JJ, Mulangu S, Masumu J, Kayembe JM, Kemp A, Paweska JT. Ebola virus outbreaks in Africa: past and present. Onderstepoort J Vet Res. 2012;79(2):451. Epub 2013/01/19. doi: 10.4102/ojvr.v79i2.451. PubMed PMID: 23327370.

27. CDC. 2014-2016 Ebola Outbreak in West Africa Atlanta, GA: Centers for Disease Control and Prevention; 2019 [July 22, 2020]. Available from: https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html.

28. Wang LF, Shi Z, Zhang S, Field H, Daszak P, Eaton BT. Review of bats and SARS. Emerg Infect Dis. 2006;12(12):1834-40. Epub 2007/03/01. doi: 10.3201/eid1212.060401. PubMed PMID: 17326933; PubMed Central PMCID: PMCPMC3291347.

29. WHO. Cumulative Number of Reported Probable Cases of SARS Geneva, Switzerland: World Health Organization; 2003 [July 22, 2020]. Available from: https://www.who.int/csr/sars/country/2003_07_11/en/.

30. CDC. Outbreak of swine-origin influenza A (H1N1) virus infection-Mexico, March-April 2009. Morbidity and mortality Weekly Report. 2009;58(17):467.

31. Novel Swine-Origin Influenza AVIT, Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med. 2009;360(25):2605-15. Epub 2009/05/09. doi: 10.1056/NEJMoa0903810. PubMed PMID: 19423869.

32. WHO. Situation updates - Pandemic (H1N1) 2009 Geneva, Switzerland: World Health Organization; 2010 [July 22, 2020]. Available from: https://www.who.int/csr/disease/swineflu/updates/en/.

33. Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. The Lancet. 2015;386(9997):995-1007. doi: 10.1016/s0140-6736(15)60454-8.

34. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367(19):1814-20. Epub 2012/10/19. doi: 10.1056/NEJMoa1211721. PubMed PMID: 23075143.

35. WHO. WHO MERS Global Summary and Assessment of Risk Geneva, Switzerland: World Health Organization; 2019 [July 22, 2020]. Available from: https://apps.who.int/iris/bitstream/handle/10665/326126/WHO-MERS-RA-19.1-eng.pdf?ua=1.

36. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33. Epub 2020/01/25. doi: 10.1056/NEJMoa2001017. PubMed PMID: 31978945; PubMed Central PMCID: PMCPMC7092803.

37. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26(4):450-2. Epub 2020/04/15. doi: 10.1038/s41591-020-0820-9. PubMed PMID: 32284615; PubMed Central PMCID: PMCPMC7095063.

38. CSSE. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Available at: https://github.com/CSSEGISandData/COVID-19 [accessed on July 10, 2020]. 2020.

39. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. 2020;20(5):533-4. doi: 10.1016/s1473-3099(20)30120-1.

40. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, et al. Zoonotic host diversity increases in human-dominated ecosystems. Nature. 2020. Epub 2020/08/08. doi: 10.1038/s41586-020-2562-8. PubMed PMID: 32759999.

41. Faurby S, Davis M, Pedersen RO, Schowanek SD, Antonelli A, Svenning JC. PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. Ecology. 2018;99(11):2626. Epub 2018/07/11. doi: 10.1002/ecy.2443. PubMed PMID: 29989146.

42. Rondinini C, Di Marco M, Chiozza F, Santulli G, Baisero D, Visconti P, et al. Global habitat suitability models of terrestrial mammals. Philos Trans R Soc Lond B Biol Sci. 2011;366(1578):2633-41. Epub 2011/08/17. doi: 10.1098/rstb.2011.0113. PubMed PMID: 21844042; PubMed Central PMCID: PMCPMC3140734.

43. Jung M, Dahal PR, Butchart SHM, Donald PF, De Lamo X, Lesiv M, et al. A global map of terrestrial habitat types. Sci Data. 2020;7(1):256. Epub 2020/08/08. doi: 10.1038/s41597-020-00599-8. PubMed PMID: 32759943; PubMed Central PMCID: PMCPMC7406504.

44. Holt BG, Lessard JP, Borregaard MK, Fritz SA, Araujo MB, Dimitrov D, et al. An update of Wallace's zoogeographic regions of the world. Science. 2013;339(6115):74-8. Epub 2012/12/22. doi: 10.1126/science.1228282. PubMed PMID: 23258408.

45. Biber MF, Voskamp A, Niamir A, Hickler T, Hof C. A comparison of macroecological and stacked species distribution models to predict future global terrestrial vertebrate richness. Journal of Biogeography. 2019;47(1):114-29. doi: 10.1111/jbi.13696.

46. Calabrese JM, Certain G, Kraan C, Dormann CF. Stacking species distribution models and adjusting bias by linking them to macroecological models. Global Ecology and Biogeography. 2014;23(1):99-112. doi: 10.1111/geb.12102. PubMed PMID: WOS:000327608500009.

47. Han BA, Kramer AM, Drake JM. Global patterns of zoonotic disease in mammals. Trends Parasitol. 2016;32(7):565-77. Epub 2016/06/19. doi: 10.1016/j.pt.2016.04.007. PubMed PMID: 27316904; PubMed Central PMCID: PMCPMC4921293.

48. Han BA, Schmidt JP, Bowden SE, Drake JM. Rodent reservoirs of future zoonotic diseases. Proc Natl Acad Sci U S A. 2015;112(22):7039-44. Epub 2015/06/04. doi: 10.1073/pnas.1501598112. PubMed PMID: 26038558; PubMed Central PMCID: PMCPMC4460448.