ELECTRONIC SUPPLEMENTARY MATERIAL: ROLE OF INTER-RELATED POPULATION-LEVEL HOST TRAITS IN DETERMINING PATHOGEN RICHNESS AND ZOONOTIC RISK

TIM C.D. LUCAS[1*†], HILDE M. WILKINSON-HERBOTS[2] AND KATE E. JONES[3*]

[1]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, United Kingdom. 0000-0003-4694-8107
[2]Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, United Kingdom.
[3]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, United Kingdom and Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, United Kingdom. 0000-0001-5231-3293
[†]Current address: Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, OX3 7FZ, United Kingdom.
[*]Corresponding authors: Tim C.D. Lucas: timcdlucas@gmail.com, Kate E. Jones: kate.e.jones@ucl.ac.uk

TABLE S1. A summary of all symbols used along with their units and default values. The justifications for parameter values are given in the main manuscript.

Symbol	Explanation	Units	Value
ρ	Number of pathogens		2
x, y	Group (colony) index		
p	Pathogen index i.e. $p \in \{1, 2\}$ for pathogens 1 and 2		
q	Disease class i.e. $q \in \{1, 2, 12\}$		
S_x	Number of susceptible individuals in group x		
I_{qx}	Number of individuals infected with disease(s) $q \in \{1, 2, 12\}$ in group x		
R_x	Number of individuals in group x in the recovered with immunity class		
N	Total Population size		$2{,}000 - 32{,}000$
m	Number of groups		$5 - 80$
n	Group size		$100 - 1600$
a	Area	km^2	$2{,}500 - 40{,}000$
d	Density	individuals.km^{-2}	$0.2 - 3.2$
β	Transmission rate		$0.1 - 0.3$
α	Coinfection adjustment factor		0.1
γ	Recovery rate	year^{-1}.individual^{-1}	1
ξ	Dispersal rate	year^{-1}.individual^{-1}	0.01
Λ	Birth rate	year^{-1}.individual^{-1}	0.05
μ	Death rate	year^{-1}.individual^{-1}	0.05
g	Population growth rate		1
k_y	Degree of node y (number of groups that individuals from group y can disperse to)		
δ	Waiting time until next event	years	
e_i	The rate at which event i occurs	year^{-1}	
R_0	The (single subpopulation) basic reproduction number	infections	

S1. Supplementary materials and methods: Stochastic simulations

We examined the model using stochastic, continuous-time simulations implemented in R [1]. The implementation is available as an R package on GitHub [2]. The model can be written as a continuous-time Markov chain. The Markov chain contains the random variables $((S_x)_{x=1...m}, (I_{x,q})_{x=1...m,\, q\in\{1,2,12\}}, (R_x)_{x=1...m})$. Here, $(S_x)_{x=1...m}$ is a length m vector of the number of susceptibles in each group $(I_{x,q})_{x=1...m,q\in\{1,2,12\}}$ is a length $m \times 3$ vector describing the number of individuals of each disease class ($q \in \{1, 2, 12\}$) in each group. Finally, $(R_x)_{x=1...m}$ is a length m vector of the number of individuals in the recovered class. The model is a Markov chain where extinction of either pathogen species and extinction of the host population are absorbing states. The expected time for the host population to go extinct is much larger than the duration of the simulations.

At any time, suppose the system is in state $((s_x), (i_{x,q}), (r_x))$. At each step in the simulation we calculate the rate at which each possible event might occur. One event is then randomly chosen, weighted by its rate

$$p(\text{event } i) = \frac{e_i}{\sum_j e_j}, \tag{1}$$

where e_i is the rate at which event i occurs and $\sum_j e_j$ is the sum of the rates of all possible events. Finally, the length of the time step, δ, is drawn from an exponential distribution

$$\delta \sim \text{Exp}\left(\sum_j e_j\right). \tag{2}$$

We can now write down the rates of all events. Assuming asexual reproduction, that all classes reproduce at the same rate and that individuals are born into the susceptible class we get

$$s_x \to s_x + 1 \quad \text{at a rate of } \Lambda\left(s_x + \sum_q i_{qx} + r_x\right) \tag{3}$$

where $s_x \to s_x + 1$ is the event that the number of susceptibles in group x will increase by 1 (a single birth) and $\sum_q i_{qx}$ is the sum of all infection classes $q \in \{1, 2, 12\}$. The rates of death, given a death rate μ, and no increased mortality due to infection, are given by

$$s_x \to s_x - 1 \quad \text{at a rate of } \mu s_x, \tag{4}$$

$$i_{qx} \to i_{qx} - 1 \quad \text{at a rate of } \mu i_{qx}, \tag{5}$$

$$r_x \to r_x - 1 \quad \text{at a rate of } \mu r_x. \tag{6}$$

We modelled transmission as being density-dependent. This assumption was more suitable than frequency-dependent transmission as we were modelling a disease transmitted by saliva or urine in highly dense populations confined to caves, buildings or potentially a small number of tree roosts. We were notably not modelling a sexually transmitted disease (STD) as spillover of STDs from bats to humans is likely to be rare. Infection of a susceptible with either Pathogen 1 or 2 is therefore given by

$$i_{1x} \to i_{1x} + 1, \quad s_x \to s_x - 1 \quad \text{at a rate of } \beta s_x (i_{1x} + i_{12x}), \tag{7}$$

$$i_{2x} \to i_{2x} + 1, \quad s_x \to s_x - 1 \quad \text{at a rate of } \beta s_x (i_{2x} + i_{12x}), \tag{8}$$

while coinfection, given the coinfection adjustment factor α, is given by

$$i_{12,x} \to i_{12,x} + 1, \quad i_{1x} \to i_{1x} - 1 \quad \text{at a rate of } \alpha\beta i_{1x} (i_{2x} + i_{12x}), \tag{9}$$

$$i_{12,x} \to i_{12,x} + 1, \quad i_{2x} \to i_{2x} - 1 \quad \text{at a rate of } \alpha\beta i_{2x} (i_{1x} + i_{12x}). \tag{10}$$

Note that lower values of α give lower rates of coinfection as in Castillo-Chavez et al. [3].

The rate of migration from group y (with degree k_y) to group x, given a dispersal rate ξ is given by

$$s_x \to s_x + 1, \quad s_y \to s_y - 1 \quad \text{at a rate of } \frac{\xi s_y}{k_y}, \tag{11}$$

$$i_{qx} \to i_{qx} + 1, \quad i_{qy} \to i_{qy} - 1 \quad \text{at a rate of } \frac{\xi i_{qy}}{k_y}, \tag{12}$$

$$r_x \to r_x + 1, \quad r_y \to r_y - 1 \quad \text{at a rate of } \frac{\xi r_y}{k_y}. \tag{13}$$

Note that the dispersal rate does not change with infection. As above, this is due to the low virulence of bat viruses. Finally, recovery from any infectious class occurs at a rate γ

$$i_{qx} \to i_{qx} - 1, \ \ r_x \to r_x + 1 \ \text{ at a rate of } \ \gamma i_{qx}. \tag{14}$$

S2. SUPPLEMENTARY MATERIALS AND METHODS: DETERMINISTIC MODEL

We can study the model analytically if we restrict the population to a single subpopulation. We first study the endemic pathogen. We then have a typical SIR model with vital dynamics [4].

$$\frac{dS}{dt} = \Lambda n - \mu S - \beta S I_1, \tag{15}$$

$$\frac{dI_1}{dt} = \beta S I_1 - (\gamma + \mu)I_1, \tag{16}$$

$$\frac{dR}{dt} = \gamma I_1 - \mu R. \tag{17}$$

S2.0.1. *Equilibrium values of single pathogen model.* To find the equilibrium number of susceptibles, S^*, we set $\frac{dI_1}{dt} = 0$. We then have the trvial. steady state of $S^* = n$ and $I_1^* = 0$. To find the endemic steady state ($I_1^* \neq 0$) we write

$$\beta S^* I_1 = (\gamma + \mu)I_1, \tag{18}$$

$$S^* = \frac{\gamma + \mu}{\beta}. \tag{19}$$

To find I^* we substitute Equation 19 into Equation 15 and set $\frac{dI_1}{dt} = 0$ giving

$$0 = \Lambda n - \mu S^* - \beta S^* I_1^*, \tag{20}$$

$$0 = \Lambda n - \frac{\mu(\gamma + \mu)}{\beta} - (\gamma + \mu)I_1^*, \tag{21}$$

$$I_1^* = \frac{\Lambda n}{\gamma + \mu} - \frac{\mu}{\beta}. \tag{22}$$

S2.0.2. *Second pathogen.* We assume that Pathogen 2 is introduced when the endemic pathogen is at equilibrium. The introduction of Pathogen 2 will cause the system to shift to a new equilibrium. However, we continue to use S^* and I_1^* to mean the values above. Initially, the dynamics of Pathogen 2 are governed by

$$\frac{dI_2}{dt} = \beta S^* I_2 + \alpha \beta I_1^* I_2 - (\gamma + \mu)I_2 \tag{23}$$

where $\beta S^* I_2$ is the number of infections per unit time, $\alpha \beta I_1^* I_2$ is the number of coinfections per unit time and $(\gamma + \mu)I_2$ is the number of deaths and recoveries per unit time. I_2 is therefore the number of individuals infected with Pathogen 2 regardless of whether they are infected with Pathogen 1. In this sense it is the combined size of classes I_2 and I_{12} as used in the stochastic model (Section S1). For simplicity here we simply label this group I_2. Substituting in Equations 19 and 22, and noting that $R_0 = \frac{\beta n}{\gamma + \mu}$ we get

$$\frac{dI_2}{dt} = (\gamma + \mu)I_2 + \alpha \beta \left(\frac{\Lambda n}{\gamma + \mu} - \frac{\mu}{\beta} \right) I_2 - (\gamma + \mu)I_2, \tag{24}$$

$$\frac{dI_2}{dt} = \alpha \left(\frac{\beta \Lambda n}{\gamma + \mu} - \mu \right) I_2, \tag{25}$$

$$\frac{dI_2}{dt} = \alpha (\Lambda R_0 - \mu) I_2. \tag{26}$$

This is greater than zero when

$$\alpha (\Lambda R_0 - \mu) I_2 > 0 \tag{27}$$

As I_2 is positive, this is greater than zero when $\Lambda R_0 - \mu > 0$ and $\alpha > 0$. As $\Lambda = \mu$ due to the assumption of stable population size, $\Lambda R_0 - \mu > 0$ is true as long as $R_0 > 1$. R_0 is greater than one (as we have assumed that Pathogen 1 is endemic), Therefore, $\frac{dI_2}{dt}$ is greater than zero provided α is greater than zero.

TABLE S2. Raw data for geographic range size simulations. The population parameters are shown along with the number of invasions and the number of simulations. β is the transmission rate, n is group size, m is the number of groups and N is the total population size.

β	n	m	Area (×1000 km²)	N (×1000)	Density (km⁻²)	Invasions	Both Extinct
0.1	400	20	2.5	8	3.2	2	0
0.1	400	20	5.0	8	1.6	3	0
0.1	400	20	10.0	8	0.8	2	0
0.1	400	20	20.0	8	0.4	3	0
0.1	400	20	40.0	8	0.2	2	0
0.2	400	20	2.5	8	3.2	3	0
0.2	400	20	5.0	8	1.6	3	0
0.2	400	20	10.0	8	0.8	1	0
0.2	400	20	20.0	8	0.4	4	0
0.2	400	20	40.0	8	0.2	1	0
0.3	400	20	2.5	8	3.2	3	0
0.3	400	20	5.0	8	1.6	3	0
0.3	400	20	10.0	8	0.8	3	0
0.3	400	20	20.0	8	0.4	5	0
0.3	400	20	40.0	8	0.2	9	0

TABLE S3. Raw data for group size simulations. The population parameters are shown along with the number of invasions and the number of simulations. β is the transmission rate, n is group size, m is the number of groups and N is the total population size.

β	n	m	Area (×1000 km²)	N (×1000)	Density (km⁻²)	Invasions	Both Extinct
0.1	100	20	2.5	2	0.8	4	12
0.1	200	20	5.0	4	0.8	5	0
0.1	400	20	10.0	8	0.8	2	0
0.1	800	20	20.0	16	0.8	0	0
0.1	1600	20	40.0	32	0.8	55	0
0.2	100	20	2.5	2	0.8	3	8
0.2	200	20	5.0	4	0.8	6	0
0.2	400	20	10.0	8	0.8	0	0
0.2	800	20	20.0	16	0.8	39	0
0.2	1600	20	40.0	32	0.8	95	0
0.3	100	20	2.5	2	0.8	1	9
0.3	200	20	5.0	4	0.8	4	0
0.3	400	20	10.0	8	0.8	7	0
0.3	800	20	20.0	16	0.8	67	0
0.3	1600	20	40.0	32	0.8	100	0

TABLE S4. Raw data for number of groups simulations. The population parameters are shown along with the number of invasions and the number of simulations. β is the transmission rate, n is group size, m is the number of groups and N is the total population size.

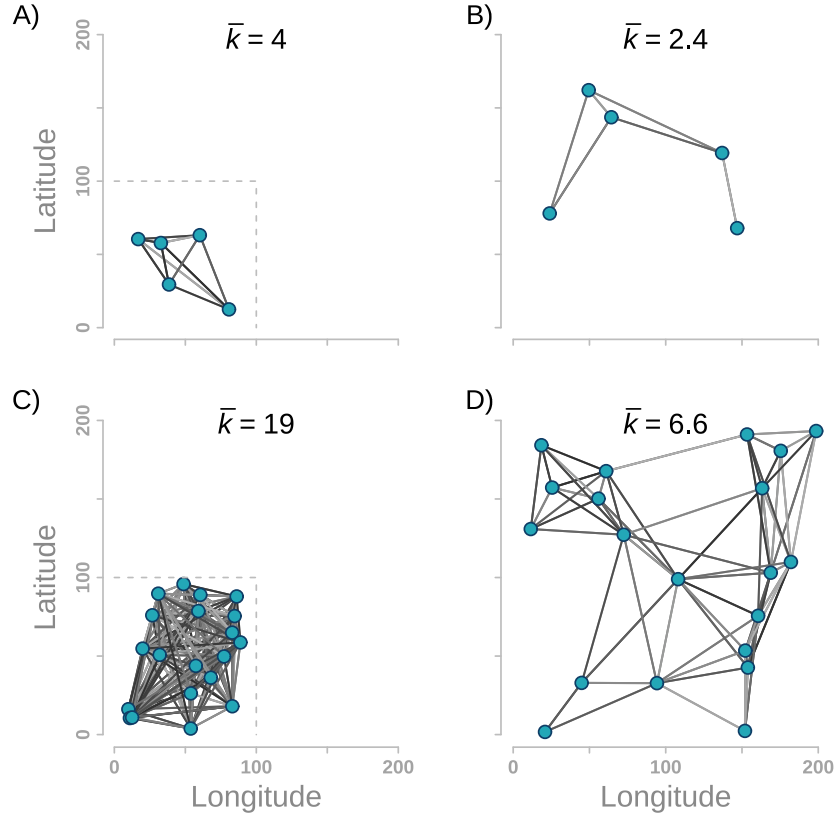| β	n	m	Area $(\times 1000 \text{ km}^2)$	N $(\times 1000)$	Density (km^{-2})	Invasions	Both Extinct
0.1	400	5	2.5	2	0.8	0	3
0.1	400	10	5.0	4	0.8	0	0
0.1	400	20	10.0	8	0.8	2	0
0.1	400	40	20.0	16	0.8	2	0
0.1	400	80	40.0	32	0.8	7	0
0.2	400	5	2.5	2	0.8	2	1
0.2	400	10	5.0	4	0.8	1	0
0.2	400	20	10.0	8	0.8	0	0
0.2	400	40	20.0	16	0.8	3	0
0.2	400	80	40.0	32	0.8	11	0
0.3	400	5	2.5	2	0.8	1	4
0.3	400	10	5.0	4	0.8	2	0
0.3	400	20	10.0	8	0.8	7	0
0.3	400	40	20.0	16	0.8	15	0
0.3	400	80	40.0	32	0.8	17	0

FIGURE S1. The relationship between geographic range size and metapopulation network structure. Groups are shown by circles. Groups that are close enough for animals to disperse between (i.e. within 100 km of each other) are joined by a line. Groups are placed randomly in spaces of various sizes (grey dashed lines). A and C) the default geographic range size (10000 km^2). B and D) the largest geographic range size (40000 km^2). A and B) the smallest number of groups (five). C and D) the default number of groups (20). The mean number of connections per subpopulation, $\bar{k}$, is shown for each metapopulation.
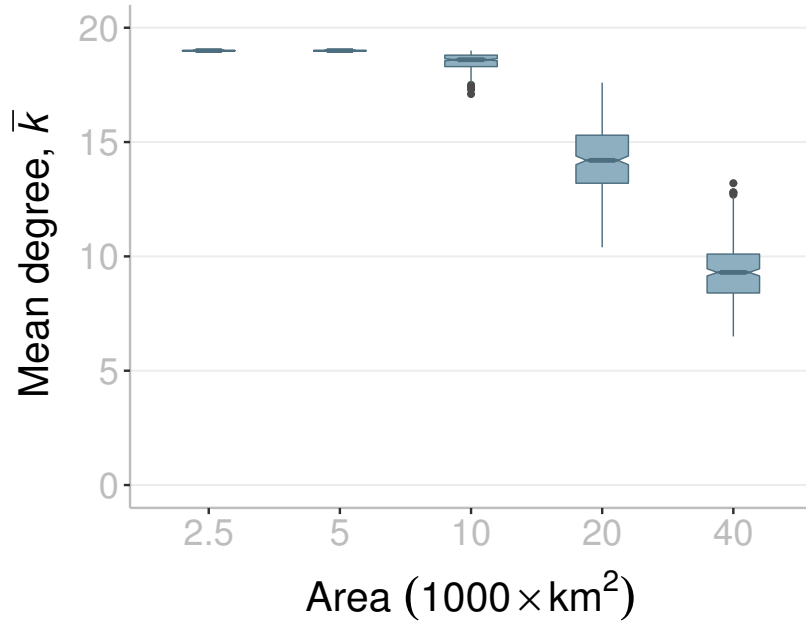
FIGURE S2. Change in average metapopulation network degree ($\bar{k}$) with increasing geographic range size. Bars show the median, boxes show the interquartile range, vertical lines show the range and grey dots indicate outlier values. Notches indicate the 95% confidence interval of the median. All simulations had 20 groups, meaning 19 is the maximum value of $\bar{k}$.
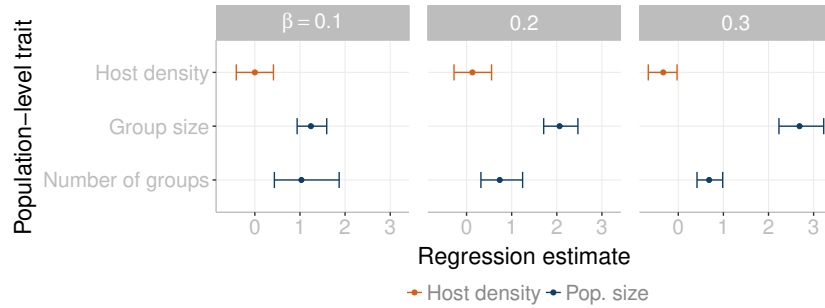

FIGURE S3. GLM regression coefficients for for three sets of simulations at three transmission, β values. To examine whether host density or population size affects pathogen invasion more strongly we compare the orange and blue bars. To examine whether group size or number of groups affects pathogen invasion more strongly we compare the second and third bar in each facet.
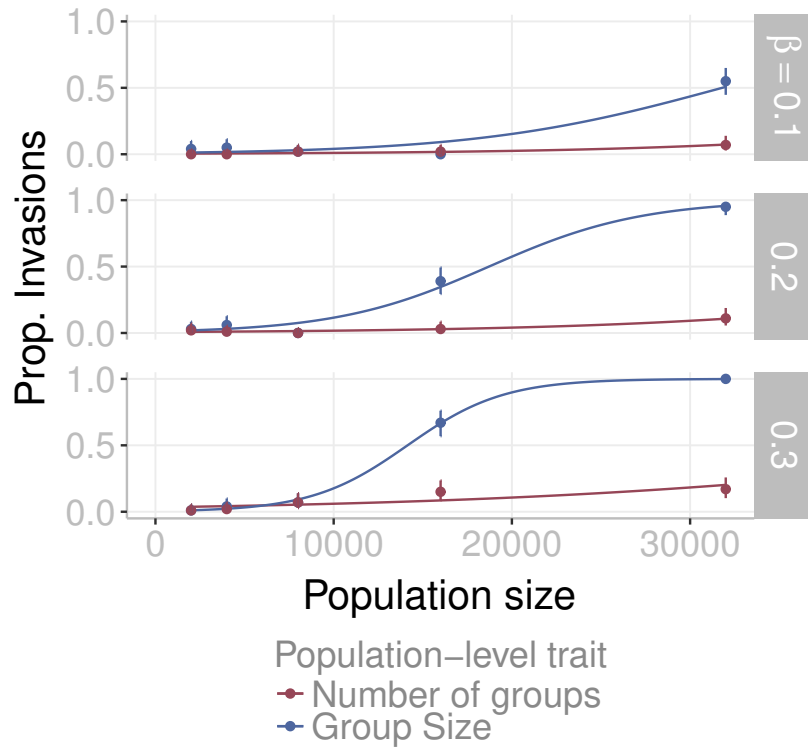
FIGURE S4. Comparison of the effect of host population size on probability of invasion when population size is altered by changing group size or number of groups. Relationships are shown separately for each transmission value, β. It can be seen that changes in group size give a much greater increase in invasion probability than changes in the number of groups. Note that this is the same data as Figure 3 but with the x-axis scaled by population size, rather than relative parameter change.

References

[1] R Development Core Team. *R: A language and environment for statistical computing.* R Foundation For Statistical Computing. Vienna, Austria, 2010. URL: http://www.R-Project.org.

[2] TCD Lucas. "MetapopEpi: Functions to run multipathogen, metapopulation epidemiological simulations" (2015). R package version 0.0.1. DOI: 10.5281/zenodo.48942. URL: https://github.com/timcdlucas/metapopepi.

[3] C Castillo-Chavez, H Hethcote, V Andreasen, S Levin, and WM Liu. "Epidemiological models with age structure, proportionate mixing, and cross-immunity". *J. Math. Biol.* 27.3 (1989), pp. 233–258. DOI: 10.1007/BF00275810.

[4] WO Kermack and AG McKendrick. "Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity". *Proc. R. Soc. A* 138.834 (1932), pp. 55–83. ISSN: 0950-1207. DOI: 10.1098/rspa.1932.0171.