Identifying Achilles' heel of multi-host pathogens: The concept of keystone "host" species illustrated by *Mycobacterium ulcerans* transmission

Supplementary materials

1 Model for environmental transmission and model fit

In addition of the model detailed in the main text, we have developed an alternative model where pathogen transmission is not transmitted between the different kind of hosts through a network of transmission, but only via free-living particles in the environmental aquatic reservoir (compartment B). This kind of model has been successfully applied for various systems, especially for cholera transmission (Codeço 2001; de Magny et al. 2005). We assume that contact rates between each (host) taxon and the contaminated environment are equal across the different aquatic taxa. Put in mathematical terms, that led to the following set of equations:

$$\frac{dS_i}{dt} = b_i N_i - \omega \frac{B}{B+\theta} S_i - d_i S_i \quad (1)$$

$$\frac{dI_i}{dt} = \beta \frac{B}{B+\theta} S_i - d_i I_i \quad (2)$$

$$\frac{dR_i}{dt} = \gamma I_i - d_i R_i \quad (3)$$

$$\frac{dB}{dt} = \sigma \sum_i^n I_i - \varepsilon B \quad (4)$$

where additional parameters ω, θ, σ and ε represent efficient contact rate to water, mycobacterial load needed to yields infection, bacterial production by infectious individuals into environment and mycobacterium lifespan stage into the aquatic environment, respectively.

The environmental transmission model shows also a significant correlation (figure S1), but only when taxa are considered without immunity (r=0.90, $p-value<0.05$ without immunity, and r=0.3, $p-value=0.6851$ with immunity). Moreover, this correlation coefficient is lower than in the case of the pathogen transmission through networks. Consequently, we have considered that a transmission of pathogen through ecological networks as the most likely scenario according to the field observations.

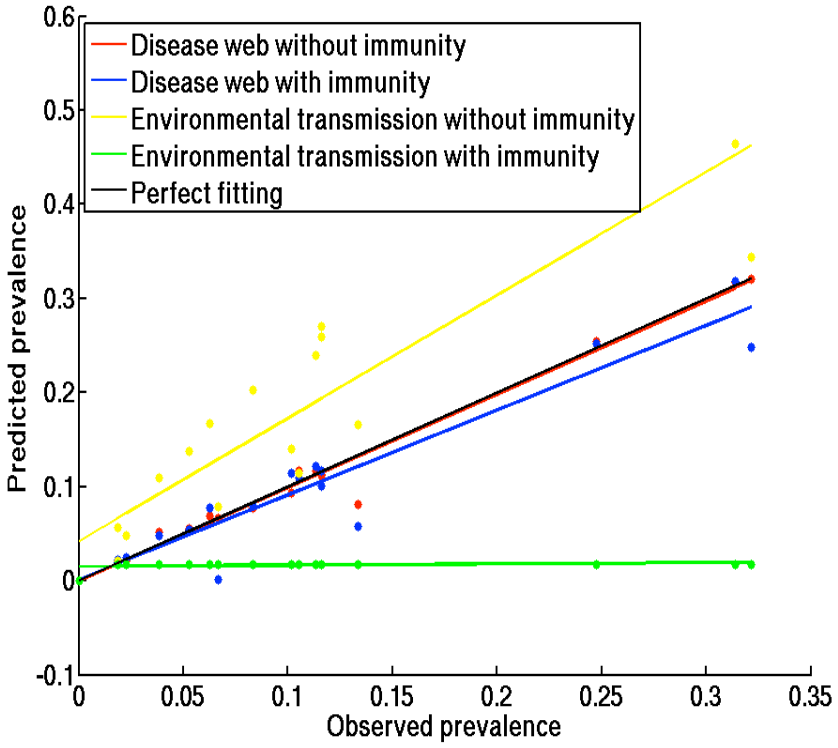

Figure 1: Fitting quality of the two models. The colored lines represent the linear regression between observed and predicted transmission prevalence means for the different hypothesis tested in the present work. The black line represents the perfect fit (*i.e.,* the linear regression without intercept and a slope coefficient equals to 1).

2 Quantifying intrinsic ecological parameters

Except taxa abundances that have been estimated from field data and parameters of networks contruction that are estimated by our mathematical model, we have to quantify the demographic parameters of each taxon, *i.e.*, birth and death rate. To keep a realistic community characterization, we used a theoretical framework that has been already assessed in other studies on aquatic ecosystems (Cohen et al. 2003) by using allometric relationships. We first quantify the body mass of each taxon by using the following relationship:

$$\log(Mi) = e - f \log(i)$$

where M_i is the body mass of each taxon rank i, and e and f two constants. Since more abundant taxa tend to live longer (West et al., 1999), we sorted each (host) taxon into minimal to maximal values according to their abundance and we assumed that each taxon represented a rank in the previous relationship (i). To quantify a taxon's growth rate, we applied the following allometric relationship (De Leo & Dobson 1996; West et al. 1999):

$$r_i = 0.6 M_i^{-0.27}$$

where r_i is the growth rate of taxa i. Since we kept population size constant ($b_i = d_i = r_i$), we have assumed throughout the manuscript that $e = 0.1$ and $f = 0.05$ (Jonnsson et al. 2005).

3 Estimation of parameters

To estimate the parameters for network construction, we estimate the maximal likelihood by assuming that errors are normally distributed. Hence, the quantity to maximize is:

$$-LL(m(x)|s)) = log(\prod \frac{1}{\sqrt{2\pi\sigma^2}} \frac{(s(i) - m(i))^2)}{2\sigma^2}) \quad (6)$$

where s and m represent respectively collected data and simulated values, index i the different (host) taxa, x is the set of parameters and σ^2 is the variance of observed prevalence. To minimize this value, we use the classical Nelder-Menson algorithm where seeds are randomly determined 25 times and we take the best fit solution. We repeat this procedure for the 17 sites in which *M. ulcerans* has been observed.

For network model, we constraint exploration of parameters values to $1 <= \rho <= n$, $1 <= \delta <= \rho$, $10e-8 <= \beta[ii] <= 10e-1$, $10e-8 <= \beta[ij] <= 10e-1$, and this $\forall[i,j]$. For model with environmental direct aquatic transmission, we assume that $0.1 < \omega < 10$, $1 < \theta < 10e4$, $0.1 < \sigma < 10$, $0.1 < \epsilon < 10$. These values have been visually chosen to identify extreme values, from a point of view of disease transmission, to cope with the larger range of parameters' values.

4 Data

Taxon	Collected	Sampled for MU	PCR Positivity
Acarina	708	28	0.139
Ancylidae	59	4	0.000
Anura	1426	31	0.114
Atyidae	524	16	0.000
Baetidae	2863	44	0.051
Belostomatidae	401	37	0.091
Caenidae	617	28	0.083
Calopterygidae	2	1	0.000
Ceratopogonidae	261	28	0.000
Chaoboridae	36	7	0.000
Chironomidae	2793	57	0.058
Cladocera	769	13	0.000
Coenagrionidae	668	19	0.000
Conchostraca	6	1	0.000
Copepoda	1962	13	0.000
Corduliidae	167	5	0.000
Corixidae	264	15	0.000
Culicidae	1053	30	0.042
Curculionidae	11	1	0.000
Dytiscidae	358	44	0.069
Elmidae	58	5	0.500
Ephydridae	12	1	0.000
Gerridae	40	17	0.000
Gomphidae	2	1	0.000

Gyrinidae	25	2	0.000
Hebridae	10	2	0.000
Helicopsychidae	2	1	0.000
Heptageniidae	41	4	0.000
Hirudinea	89	25	0.125
Hydraenidae	432	14	0.150
Hydrometridae	8	11	0.000
Hydrophilidae	664	58	0.093
Hydropsychidae	12	2	0.000
Hygrobiidae	13	1	0.000
Lampyridae	14	5	0.000
Leptoceridae	24	6	0.000
Leptophlebiidae	15	3	0.000
Libellulidae	238	26	0.118
Lymnaeidae	12	2	0.000
Mesoveliidae	160	34	0.000
Mysidae	2	1	0.000
Naucoridae	55	18	0.143
Nematoda	17	3	0.000
Nepidae	34	21	0.139
Noteridae	334	38	0.029
Notonectidae	387	37	0.095
Oligochaeta	1264	25	0.075
Ostracoda	1081	14	0.133
Physidae	226	9	0.048
Planorbidae	525	34	0.069

Pleidae	683	24	0.000
Polycentropedidae	3	1	0.000
Polymitarcyidae	11	4	0.000
Protoneuridae	372	24	0.177
Psephenidae	1	1	0.000
Psychodidae	27	1	1.000
Pyralidae	32	3	0.667
Saldidae	3	2	0.000
Sciomycidae	3	3	0.333
Scirtidae	96	10	0.071
Sphaeriidae	8	1	1.000
Stratiomyidae	24	7	0.000
Syrphidae	4	2	0.500
Tabanidae	14	3	0.000
Thiaridae	60	8	0.000
Tipulidae	21	6	0.000
Veliidae	43	14	0.000
Viviparidae	3	3	0.000

Table 1: Data summary for *Mycobacterium ulcerans* prevalence among aquatic invertebrate taxa across 27 localities of Ghana, West Africa.

References

Codeço, C.T. (2001). Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC infectious diseases*, 1, 1.

Cohen, J.E., Jonsson, T. & Carpenter, S.R. (2003). Ecological community description using the food web, species abundance, and body size. *Proceedings of the National Academy of Sciences*, 100, 1781–1786.

Jonnsson, T., Cohen, J.E. & Carpenter, S.R. (2005). Food webs, body size and species abundance in Ecological community description. In: *Advances in ecological research* (ed. Caswell, H.). Academic Press, pp. 1–84.

De Leo, G.A. & Dobson, A.P. (1996). Allometry and simple epidemic models for microparasites. *Nature*, 379, 720–722.

de Magny, G.C., Paroissin, C., Cazelles, B., Lara, M.D., Delmas, J.-F. & Guégan, J.-F. (2005). Modeling environmental impacts of plankton reservoirs on cholera population dynamics. In: *Mathematics and applications to biology and medicine* (eds. Cancès, E. & Gerbeau, J.-F.). EDP Sciences, Paris, France.

West, G.B., Brown, J.H. & Enquist, B.J. (1999). The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science*, 284, 1677–1679.