**Electronic supplementary material**

**Author**: Akira Terui, Keita Ooue, Hirokazu Urabe, Futoshi Nakamura

**Title:** Parasite infection induces size-dependent host dispersal: consequences for parasite persistence

**Journal**: Proceedings of the Royal Society B: Biological Sciences

**Article DOI**: 10.1098/rspb.2017.1491

**Derivation of Bernoulli observation process** The Bernoulli distribution yields a “success” (recapture) with probability *p* and a “failure” (unrecapture) with probability 1 – *p*. In our sampling design, fish may be recaptured with probability φ*sD*. However, fish may not be recaptured for several reasons, which include the following states:

1. Alive and stayed at the 1200-m study section, but not detected [(1−φ)*sD*]

2. Alive but left behind the 1200-m study section [*s*(1−*D*)]

3. Dead (1−*s*)

Since these states are mutually exclusive, the probability of “unrecapture” is the sum of the three probabilities, which can be reduced to 1−φ*sD*. Therefore, the observation process can be described as the Bernoulli probability function *g*: *g*(*k*; φ, *s*, *D*) = (φ*sD*)*k*(1−φ*sD*)1−*k*for *k* ϵ {0, 1}.

**Estimation of capture probability** Let n*jk* denote the number of fish captured in the *k*th removal taken from the *j*th subsection. We modeled the matrix of observed fish numbers n*jk* as a multinomial distribution, n*jk*|N*j*, **π** ~ Multinomial(N*j*, **π**), conditional on the unknown number of fish N*j* in the *j*th subsection and the conditional probabilities of capture **π** associated with the three-pass removal. We let θ*j* represent the probability of capture during a single removal from the *j*th subsection. Then π*jk* = θ*j­*(1−θ*j*)*k*−1 specifies the probability of capture of fish during the *k*th removal, given that they have not been captured in earlier removals. The parameter θ*j* was assumed to vary among subsections as: logit(θ*j*) = α + ε*j*. The parameter ε*j* is the random variation among subsections and was normally distributed as ε*j* ~ Normal(0, σε). The probability of capture with the three-pass removal at subsection *j* (φ*j*) was given by φ*j* = ∑θ*j*(1−θ*j*)*k*−1 (*k* = 1–3). We assumed that the unknown number of fish N*j* follows a Poisson distribution as: N*j* ~ Poisson(λ).

 The model was fitted to the data with JAGS ver. 4.1.0 and the package ‘‘*runjags*’’ in R 3.3.1. Vague priors were assigned to the parameters: i.e., normal distributions for log(λ) and α (mean = 0, variance = 104) and truncated normal distributions for σε andσλ (mean = 0, variance = 104, range = 0–100). Three Markov chain Monte Carlo chains were run jointly with the dispersal model in the main text.

**Connection with traditional epidemiological parameters** The basic reproductive number *R0* is a key epidemiological parameter that describes the average number of secondary infections caused by a single infected host in a wholly susceptible host group [1]. In our study system, however, it was impossible to define the statistic because of the lack of direct transmission among host individuals.

However, the *R\** statistic, a group-level equivalent of the basic reproductive number *R0* [2], can be expressed as a function of parameters presented in our simulation model: *R\** = (*s*ω*KC*)/*E,* where ω = 1 − *dx*. The parameter ω is the average emigration rate, and thus, ω*K* represents the expected number of infected hosts emigrating from a parasite-occupied patch per time step. This measure quantifies the average number of habitat patches (or host groups) “infected” by a single parasite-occupied patch in a wholly susceptible system. Since the parameters *s*, *C*, *E* were fixed in our model, higher values of γ and *K* correspond to higher values of *R\** (i.e., increase along the gradient of *x*-axis in each panel of Fig. 4).

**References**

1. Poulin R. 2007 *Evolutionary ecology of parasites*. Oxfordshire, UK, Princeton University Press.

2. Cross P.C., Lloyd-Smith J.O., Johnson P.L.F., Getz W.M. 2005 Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecology Letters* **8**, 587-595.

**Table S1** Results of the Bayesian model that explains individual-level variation in dispersal kernel (using a data set excluding two “super dispersers” with ≥ 180 m dispersal distance; see Fig. S1 for raw data). Posterior probability represents the proportion of parameter estimates (MCMC samples) assigned to be either negative or positive. Parameters with a posterior probability of > 0.95 are shown in bold. SE: standard error.

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | Estimate | SE | Posterior probability |
| Negative | Positive |
| Intercept (μ) | 2.402 | 0.095 | - | - |
| Infection (β1) | 0.186 | 0.144 | 0.11 | 0.89 |
| Body size (β2) | 0.143 | 0.096 | 0.07 | 0.93 |
| Infection · Size (β3) | **0.613** | 0.200 | 0.00 | **1.00** |



**Fig. S1** Relationships between body size (fork length) and distance moved for uninfected (a) and infected fish host (b). Dots represent fish individuals.