On a mechanistic process of macroparasite aggregation

Jomar F. Rabajante[1,*], Elizabeth L. Anzia[1] & Chaitanya S. Gokhale[2]

[1]Institute of Mathematical Sciences and Physics,
University of the Philippines Los Baños, 4031 Laguna, Philippines

[2]Research Group for Theoretical Models of Eco-evolutionary Dynamics,
Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology,
August Thienemann Str. 2, 24306, Plön, Germany

Running Title: A mechanistic process of macroparasite aggregation

Corresponding author: Jomar F. Rabajante, Institute of Mathematical Sciences and Physics, University of the Philippines Los Baños, Philippines, jfrabajanate@up.edu.ph

Abstract

Parasite aggregation, a recurring pattern in macroparasite infections, is considered one of the "laws" in parasite ecology. Few hosts have a large number of parasites while most hosts have a low number of parasites. This pattern has been widely studied using phenomenological models, by using the negative binomial distribution. However, to infer the mechanisms of aggregation, a mechanistic model is essential. Here we formulate such a mechanistic model of parasite aggregation in hosts without initially assuming a negative binomial distribution. Our results show that a simple model of parasite accumulation still results in an aggregated pattern, as shown by the derived mean and variance of the parasite distribution. By incorporating the derived mean and variance to the host-parasite interaction, we can predict how aggregation affects the population dynamics of the hosts and parasites through time. Thus, our results can directly be applied to observed data as well as can be utilised in designing statistical sampling procedures. Overall, we have shown how a plausible mechanistic process can result in the often observed phenomenon of parasite aggregation occurring in numerous ecological scenarios.

Keywords: macroparasite, aggregation, negative binomial, mechanistic model, host-parasite interaction, accumulation

Key Findings

- Parasite aggregation is considered one of the “laws” in parasite ecology – few hosts harbouring a large number of parasites.
- While examples abound, there is lack of mechanistic models available to explain the phenomenon
- Taking a bottom up approach we construct a simple model of host-parasite population dynamics which naturally results in parasite aggregation – negative binomial distribution of parasites in the host population
- While providing a plausible mechanism our model can be readily deployed in field work when designing sampling methodology or analysis of available data.

1 Introduction

Parasites are ubiquitous (Zimmer, 2001). Yet, for parasitologists, sampling can be a hard problem (Cundil and Alexander, 2015; Hollingsworth et al., 2015). This is due to how the parasites are distributed among hosts (Rozsa et al., 2000). While most hosts harbour very few parasites, only a few individuals are hosts to a large number of parasites (Crofton, 1971). Normal distribution is not appropriate to represent such parasite distribution in a host community. This phenomenon, termed as parasite aggregation, is perhaps one of the few "laws" in biology because it is a recurring pattern in nature and finding exceptions to this pattern is rare (Gourbiére et al., 2015; Poulin, 2007; Shaw and Dobson, 1995). The pattern is specifically observed in macroparasite infections, which include diseases brought about by helminths and arthropods (Poulin, 2007; Shaw and Dobson, 1995). The distribution pattern of macroparasites in host populations is an important factor in addressing challenges in investigating disease transmission, such as in the case of human onchocerciasis and schistosomiasis (Basanez and Boussinesq, 1999; Churcher et al., 2005; Guilhem et al., 2012). Aggregation can also affect co-infection by various parasites, parasite-driven evolutionary pressures, and stability of host-parasite communities (Morrill and Forbes, 2012, 2016; Morrill et al., 2017; Wilson et al., 2001). Thus, the study of parasite aggregation addresses a fundamental issue in ecology.

The distribution of hosts with different numbers of parasites can be well captured by the negative binomial (Crofton, 1971; Fisher, 1941; Wilson et al., 1996). Given this fact, a number of theoretical models about host-parasite dynamics implicitly assume the negative binomial distribution, mainly based on a phenomenological (statistical) modeling framework (Adler and Kretzschmar, 1992; Anderson, 1978; Gourbiére et al., 2015; Shaw and Dobson, 1995). The phenomenological principle is based on the observation that - (i) a distribution is Poisson if showing random behavior with equal mean and variance, (ii) binomial if underdispersed with greater mean than variance, or (iii) negative binomial if overdisperesed with greater variance than the mean (Anderson and Gordon, 1982; Wilson et al., 2001). Overdispersion is observed in host-parasite interaction such as between cattle (*Bos taurus*) and warble fly (*Hypoderma bovis*), and between Nile tilapia (*Oreochromis niloticus*) and copepod *Ergasilus philippinensis* (Lopez, 2001; Wilson et al., 2001). The negative binomial distribution is hypothesized to arise due to the heterogeneity in the characteristics of the hosts and parasites, and variation in the environment. Heterogeneous exposures, infection rates and susceptibility of host individuals are observed to produce aggregated distributions

of parasites (Galvani, 2003; Wilber et al., 2017; Wilson et al., 2001). However, is heterogeneity in the characteristics of the hosts and parasites a necessary condition for aggregation? Using a mechanistic model, we show that a regular but non-extreme parasite accumulation can lead to aggregation. This proves that even in regular systems (i.e., with homogeneous proportion of parasite accumulation), aggregation is widely possible.

Traditionally, in arriving at the desired model for host-macroparasite interaction, phenomenological modelling is assumed (Anderson, 1978; Anderson and May, 1978; Crofton, 1971; Rosà and Pugliese, 2002; Yakob et al., 2014). Phenomenological models are based on data gathered, but the mechanisms underlying the phenomenon are generally hidden. There is a need for descriptive and predictive models that assume the underlying mechanistic processes of parasite dynamics, which can be used in formulating disease control programs (Hollingsworth et al., 2015). The goal of this study is to mechanistically illustrate the interaction between hosts and macroparasites without assuming a negative binomial distribution. Our approach considers parasite accumulation without direct reproduction in host individuals, which is a consequence of the complex life history of macroparasites (Auld and Tinsley, 2015; McCallum et al., 2017; Viney and Cable, 2011). For example, the Nile tilapia fish are infected by the acanthocephalans parasites via the ingestion of infected zooplankton. The parasites grow, mate and lay eggs inside the fish, and then eggs are expelled in the lake through faecal excretions. The parasites in the excretions attach to the zooplanktons, and the cycle of infection through foraging continues (de la Cruz et al., 2013; Paller et al., 2016).

Recently, various attempts have been made to model the accumulation and aggregation of parasites mechanistically. The stratified worm burden, which is based on a chain of infected compartments as in Susceptible-Infected (S-I) modeling framework, was used to model schistosomiasis infections (Gurarie et al., 2010, 2015, 2016). This model can be well simulated numerically for specific cases, but possibly difficult to implement analytic studies to find general mathematical conclusions (Gurarie et al., 2010). Moreover, the Poisson-Gamma Mixture Model has been proposed to model aggregation based on parasite accumulation (Calabrese et al., 2011). The negative binomial distribution arises naturally from a Poisson-Gamma process; however, is it possible to derive a mechanistic model of aggregation without initially assuming a distribution related to the negative binomial? In another study, a mechanistic model has been proposed based on the random variation in the exposure of hosts to parasites and the infection success of parasites (Gourbiére et al., 2015). This model anchors its

derivation and conclusion on the common assumption that heterogeneity in individual hosts leads to parasite aggregation. Moreover, in most models of macroparasite infections, the rate of disease acquisition by susceptible individuals (known as the "force of infection" β) is assumed as a parameter. However, identifying the value of β can be difficult or infeasible even in the availability of data (McCallum et al., 2017). In our proposed model, this challenge can be addressed since our assumed infection parameter, denoted by P, can be directly estimated from parasite count data gathered from host samples in empirical studies. Our model is simple and tractable, and can be incorporated as part of classical modeling frameworks (e.g., host-parasite interaction with logistic growth) (Edelstein-Keshet, 2005).

2 Model

The core concept of our model is captured in Fig. 1. The total host population has a density of X. With probability P_0, the hosts are parasite-free and the density of parasite-free hosts is then $X_0 = P_0X$. In a similar fashion, the density of hosts which have at least one parasite is $X_1 = P_1X$. The total host population is therefore equivalent to $X = X_0 + X_1 = P_0X + P_1X$. Moreoever, the class of individuals X_i which have at least $i > 1$ number of parasites is assumed as a subset of X_1 (i.e., $X_{n+1} \subseteq X_n \subseteq \ldots \subseteq X_2 \subseteq X_1$), and can be modeled as,

$$\begin{aligned} X_0 &= P_0X \\ X_1 &= P_1X \\ X_2 &= P_2X_1 = P_2P_1X \\ &\vdots \\ X_n &= P_nX_{n-1} = P_nP_{n-1}\cdots P_2P_1X \\ X_{n+1} &= P_{n+1}X_n = P_{n+1}P_nP_{n-1}\cdots P_2P_1X. \end{aligned} \tag{1}$$

From the above set of equations, we can derive the distribution of living hosts (all except the $n+1$ compartment) with different parasite load with respect to the total host population. Fig. 1 represents the classification of the states of the hosts depending on their parasite load. The compartment associated with X_i, $i \geq 1$ contains the hosts infected by at least i number of parasites. The parameter P_{i+1} is the probability that a host in X_i acquires an additional parasite. Consequently, the difference $X_i - X_{i+1}, i \geq 1$ is the population density of hosts with *exactly* i number of parasites. The population density of living hosts (with maximum tolerable parasite load n) is then $X_n - X_{n+1}$.

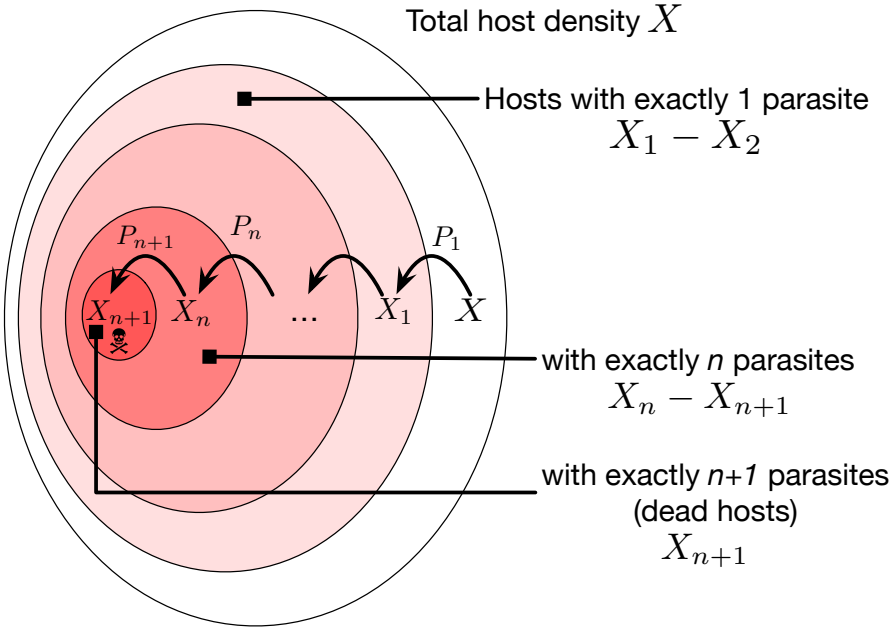

Figure 1: Hierarchical compartment model of macroparasite accumulation in hosts. The total host density is X. The density of parasite-free hosts is denoted by X_0. The population density of hosts with *at least* i number of parasites is X_i, $i \geq 1$ ($X_{i+1} \subseteq X_i$). When a host in X_i acquires a parasite, we are assured that the host is also a host with at least $i+1$ parasites. A living host can harbour a maximum of n parasites. The parameter $P_i, i \geq 1$ can be interpreted as the net probability of parasite acquisition by a host in state X_{i-1}. Refer to Table SI.1 in SI.1.1 for the description of the state variables and parameters.

Thus the death of hosts due to the parasite is captured by the transition into the X_{n+1} state given by $X_{n+1} = P_{n+1}X_n = \prod_{i=1}^{n+1} P_i X$, the number of dead hosts (last compartment in Eqs. (1)).

Our model, described as such, does not follow the classical input-output modeling framework (e.g., stratified worm burden model which is an extension of the standard S-I epidemiological models). Such a classical modeling framework could be intractable when modeling aggregation (Gurarie et al., 2010). The number of variables and equations in the stratified worm burden model can increase as the number of states (compartments) increases. Here, we model the states in the compartment diagram using proportions so we can easily analyze the distribution of parasites using probabilities. The dynamics of parasite infection can then be summarized using the properties of the derived distributions, such as the mean and variance. Also, in input-output models, the transfer from one state (e.g., susceptible) to another (e.g., state with 3 parasites) needs to pass through intermediate diseased states (e.g., states with 1 and 2 parasites, respectively). In our model, a host can acquire more than 1 parasite, and this can be modeled by adjusting the parameter P_i. This is consistent with the experimental approaches and comparable to the data gathered. For example, the proportion P_i is directly computable from parasite load data from sampled hosts as compared to computing the force of infection β in S-I models (Gandon and Day, 2009).

The quantities relating to the host and parasite proportions, X_i, P_i and n can be dynamic (e.g., may change over time). Here, we assume that X_i changes following a logistic growth with parasitism, and P_i is a function of parasite population Y (hence, also a function of time). This allows us to connect our model to experimental data. The parameter values (e.g., P_i) can be determined from the samples gathered at a certain time instant, and the pattern of the temporal evolution of the parameter values can be inferred from time series data.

Depending on the exact transmission mode of the parasite, P_i can have different functional forms. A basic assumption would be that P_i is a function of parasite encounter and transmission rates. For example, we could have $P_i = \frac{Yp}{nK+c}$ for all $i \neq 0$. Here, the total ecological carrying capacity for the parasite population is assumed to be $nK + c$, where K is the host carrying capacity and n is the maximum number of parasites that a living host can harbour without dying. The parameter c is the quantitative representation of the environment where the parasites can survive outside the hosts. Thus we can interpret $\frac{Y}{nK+c}$ as the parasite encounter probability with p being the parasite transmission probability. We focus on cases where parasites highly de-

pend on the hosts to survive, and without loss of generality, we assume a small value $c = 1$. Assuming homogeneity in parasite transmission, p is kept constant. Moreover, the host basal growth rate is set to a constant r_H, assuming that macroparasites do not affect reproduction of hosts.

For the parasite population, the total parasite density in host population is $Y = \sum_{i=1}^{n} Y_i$ where $Y_i = i(X_i - X_{i-1}), i \geq 1$. The distribution of parasites is according to the following:

$$\begin{aligned} Y_1 &= (P_1 - P_2P_1)\, X \\ Y_2 &= 2\left(\prod_{i=1}^{2} P_i - \prod_{i=1}^{3} P_i\right) X \\ Y_3 &= 3\left(\prod_{i=1}^{3} P_i - \prod_{i=1}^{4} P_i\right) X \\ &\vdots \\ Y_n &= n\left(\prod_{i=1}^{n} P_i - \prod_{i=1}^{n+1} P_i\right) X. \end{aligned} \tag{2}$$

Based on the set of Eqs. (2), the parasite population density in host population can be written as $Y = \sum_{i=1}^{n} Y_i = \widetilde{E[N]}X$. Here, N is the random variable representing the parasite load in a living host, and $\widetilde{E[N]}$ is the approximate mean of N.

The carrying capacity of the host population is assumed to be equal to K. The death rate of the hosts is assumed to be $\left(\frac{Yp}{nK+1}\right)^{n+1}$, which is derived from the equation representing X_{n+1}. Moreover, the per capita parasite reproduction rate is assumed to be r_p. Together with the carrying capacity for the parasites $(nK+1)$ the population dynamics between the hosts and parasites can be modeled assuming logistic growth as follows (refer to Table SI.1 for the description of the state variables and parameters):

$$\frac{dX}{dt} = \overbrace{r_H X\left(1 - \frac{X}{K}\right)}^{\text{host logistic growth}} - \overbrace{\left(\frac{Yp}{nK+1}\right)^{n+1} X}^{\text{host death due to parasites}} \tag{3}$$

$$\frac{dY}{dt} = \overbrace{r_P \widetilde{E[N]} X\left(1 - \frac{Y}{nX+1}\right)}^{\text{parasite logistic growth}}. \tag{4}$$

3 Results

The distribution of the parasites in the living hosts is represented by $\widetilde{E[N]}$. The general expression for this distributions as discussed above is given by,

$$\begin{aligned}\widetilde{E[N]} &= \sum_{i=1}^{n} iP_i^i\,(1-P_i) \\ &= \frac{Yp}{nK+1}\left(1-\frac{Yp}{nK+1}\right)A \end{aligned} \tag{5}$$

where A is derived from the derivative of a geometric series (see SI.1.2). If the parasite transmission $P_i = \frac{Yp}{nK+1} = 1$ (for $i \neq 0$), then the distribution of the parasites in the living hosts has $\widetilde{E[N]} = 0$ since all hosts that are harbouring at least $n+1$ number of parasites are dead. Now, supposing, $P = P_i = \frac{Yp}{nK+1} < 1$, we have,

$$\widetilde{E[N]} = \frac{nP^{n+2} - (n+1)P^{n+1} + P}{1-P}. \tag{6}$$

as the mean, and the variance, $\widetilde{Var[N]}$, derived in the SI.1.3.

A negative binomial distribution describing the probability distribution of the number of successes before the m-th failure, where ρ is the probability of success can be written as $NB(m,\rho)$. The mean and variance of $NB(m,\rho)$ are $\frac{m\rho}{1-\rho}$ and $\frac{m\rho}{(1-\rho)^2}$, respectively. From Eq. (6), $\widetilde{E[N]}$ is equivalent to the mean of a negative binomial distribution with $m = nP^{n+1} - (n+1)P^n + 1$ and $\rho = P$. For the non-truncated negative binomial distribution, we consider $n \to \infty$ (Crofton, 1971; Shonkwiler, 2016). In the next section (3.1), we discuss that as $n \to \infty$, $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ respectively converge to the mean and variance of a geometric distribution ($NB(1,\rho)$).

3.1 Constant host-parasite encounter probability

Suppose the parasite encounter probability $\left(\frac{Y}{nK+1}\right)$ is fixed. This implies that whatever the values of Y and $nK+1$, $P = \frac{Yp}{nK+1}$ is always constant. We can also interpret P as the constant geometric mean of the parasite acquisition probabilities P_i, $i \geq 1$.

For a large n and $P < 1$, $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ approximate the mean and variance of $N \in \{0,1,2,...,n\}$, respectively. $E[N]$ can be interpreted as the average number of parasites in a host with variance $Var[N]$. From Eq. (6) as $n \to \infty$, $\widetilde{E[N]} = E[N]$ is equivalent to the mean of a negative binomial distribution with $m = 1$ and $\rho = P$, which characterizes the mean of a geometric distribution. Let us denote this mean

and variance as

$$E[\infty]_P = \frac{P}{1-P} \text{ and } Var[\infty]_P = \frac{P}{(1-P)^2}. \quad (7)$$

The variance-to-mean ratio is $\frac{1}{1-P}$, which increases as P increases (Table SI.2).

The implication of this result is that even if the host can sustain a large number of parasites ($n \to \infty$) for $P < 1$, we have finite mean and variance at the population level. Also, the variance-to-mean ratio is greater than 1 characterizing an over-dispersed distribution of parasite load in the host population. For example, $E[\infty]_{0.5}$ implies that the average number of parasites in a host is 1 with variance $Var[\infty]_{0.5} = 2$. $E[\infty]_{0.9}$ implies that the average number of parasites in a host is 9 with variance $Var[\infty]_{0.9} = 90$.

As stated earlier, $E[\infty]_P$ is clearly an approximation. The error due to the approximation can be estimated and is shown in Fig. 2A. For a wide range of parameter values of the maximum tolerable parasite load of the host (n) and of the probability of parasite acquisition by a host (P) we see that the mean of the geometric distribution is a good estimate for $\widetilde{E[N]}$. Fig. 2A further illustrates that as n approaches infinity, the error becomes smaller. While, a small value for n and a higher P produce higher error, this could also be an improbable scenario in nature. The condition where the probability of acquiring more parasites is high and the number of tolerable parasites is low, would result in higher infection rates and potentially lead to the extinction of the hosts. This is illustrated in Fig. 3 where high parasite-driven host death drives the host population extinct $X_1^* = 0$.

The same is true with the approximation of $\widetilde{Var[N]}$ by $Var[\infty]_P$. Fig 2B shows that as $n \to \infty$, the error becomes smaller. This illustrates that the variance of the geometric distribution is a good estimate for $\widetilde{Var[N]}$ (Fig. 3).

In the supplementary information Figs. SI.1 and SI.2, we present examples of parasite load distribution in host population with differing values of P. An intermediate value of P results in aggregated distribution. However, the distribution becomes more negatively skewed as P increases which shows high host mortality due to harbouring high parasite load. This is the reason why as P increases, the errors in approximating $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ using the geometric distribution also increase.

3.1.1 Population dynamics

We now analyze in detail the population dynamics between the hosts and parasites (Eqs. (3) and (4)). Since the probability of parasite acquisition by a host P is constant,

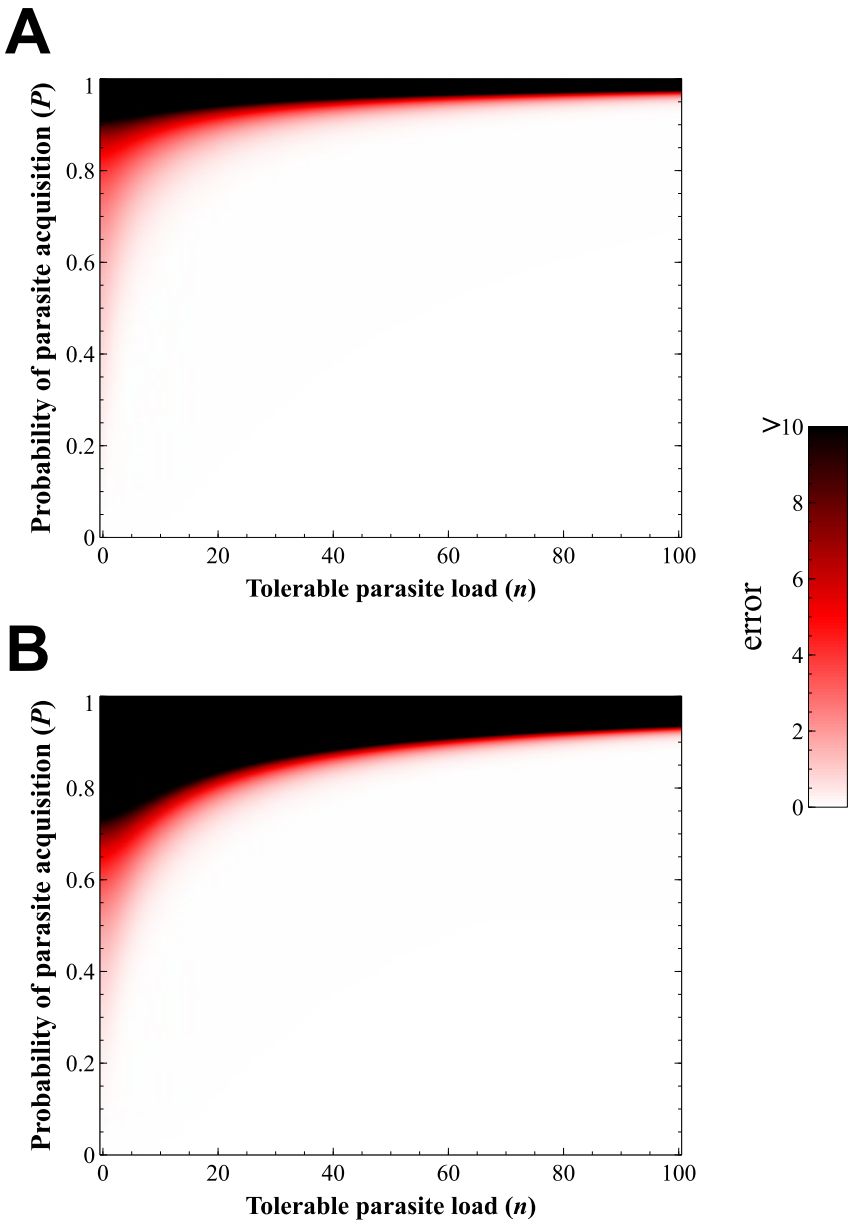

Figure 2: Approximation of $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ using the mean and variance of the geometric distribution, respectively. This shows how well the geometric distribution represents the parasite load distribution in hosts, especially when n is high and P is relatively low. (A) Absolute difference between $\widetilde{E[N]}$ and $E[\infty]_P$, $error = \left|\widetilde{E[N]} - E[\infty]_P\right|$. (B) Absolute difference between $\widetilde{Var[N]}$ and $Var[\infty]_P$, $error = \left|\widetilde{Var[N]} - Var[\infty]_P\right|$.

we can decouple the host-parasite dynamics. Analyzing first the host population (Eq. (3)), we have a logistic growth function with death or harvesting term ($D = P^{n+1}X$) (Clark, 2010). In Fig. 3, the blue curve represents the logistic growth function G. The red line represents the death function D. The intersection of the two curve is an equilibrium.

There are two equilibrium points if $r_H > P^{n+1}$, where one is unstable ($X_1^* = 0$) and the other is stable $\left(X_2^* = \frac{(r_H - P^{n+1})K}{r_H}\right)$ (Fig. 3A). If $r_H \leq P^{n+1}$, the zero equilibrium point is stable and the only steady state of the dynamics (Fig. 3B), implying an eventual extinction of the host population. For the parasites to exploit the maximum growth rate of the hosts ($r_H K/4$) (Fig. 3), the parasite-driven host death rate should be $P^{n+1} = r_H/2$.

Now, analyzing the parasite population (Eq. (4)), there are two possible equilibrium states: the parasite can go extinct, $Y_1^* = 0$ which happens if $X = 0$ or $\widetilde{E[N]} = 0$, and a stable coexistence of hosts and parasites at $Y_2^* = nX^* + 1$ where X^* is the host equilibrium. The equilibrium state Y_1^* is unstable, and Y_2^* is stable. The condition $\widetilde{E[N]} = 0$ for Y_1^* only happens if $P = 0$ (proof in SI.1.4).

Over time the total parasite population density in living host population converges to $Y_2^* = nX^* + 1$. The expected number of parasites in the hosts then tends to n (parasites in the environment represented by c are not included since they are outside the living hosts). One might think that this limiting case may not be the situation if parasite distribution in living hosts is aggregated. If aggregation affects the carrying capacity of the parasite population, the parameter n in the denominator $nX + 1$ in Eq. 4 can be replaced by

$$E[N] + \sigma\sqrt{Var[N]} < n \tag{8}$$

where σ represents the contribution of the variance to the average number of parasites in a host. Hence, if the distribution associated with parasite aggregation is considered, $\widetilde{E[N]} \rightarrow E[N] + \sigma\sqrt{Var[N]}$ as time $t \rightarrow \infty$. In the next section, we investigate the case when P is not constant through time.

3.2 Variable host-parasite encounter probability

Until now we assumed that the parasite encounter probability P as a constant. However, $P = \frac{Y}{nK+1}$ (Eqs. (3) and (4)) is a function of the dynamic variable Y, the parasite density. Given the dynamics of the parasite, we can have three possible equilibrium points. Two of the equilibrium points are trivial, one where no host and par-

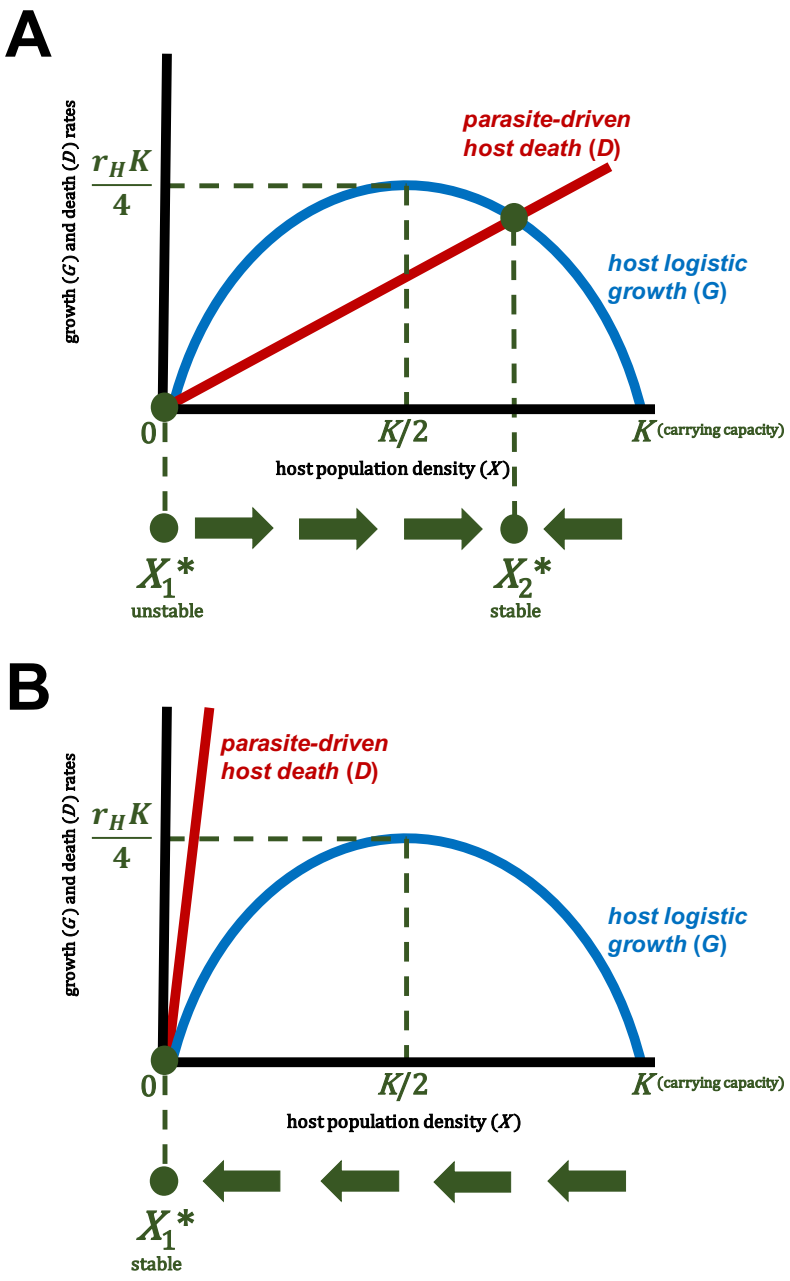

Figure 3: Illustration of the population dynamics of hosts based on Eq. (3). Suppose the host logistic growth function is $G = r_H X\left(1 - \frac{X}{K}\right)$ (blue curve) and parasite-driven death rate function is $D = P^{n+1}X$ (red line). The possible maximum growth rate of the hosts is $r_H K/4$ where r_h is the host basal growth (reproduction) rate and K is the carrying capacity of the host population X. If $G > D$ (blue curve is above the red line) then the population of host increases. If $G < D$ (blue curve is below the red line) then the population of host decreases. An intersection of the growth curve (blue) and the death rate line (red) is an equilibrium point. (A) There are two equilibrium points: the unstable X_1^* and the stable X_2^*. (B) There is one equilibrium point: the stable X_1^* that represents parasitism-driven extinction of hosts. The red line with steep slope which possibly leads to the extinction of hosts and parasites can arise due to low value of n and/or large value of P.

asite exist $(X^* = 0, Y^* = Y_0)$, and another at the carrying capacity of the hosts $(X^* = K, Y^* = Y_0 = 0)$ with $\widetilde{E[N]} = 0$ representing the disease-free state.

The third equilibrium point posits a coexistence of hosts and parasites and can be derived from

$$\begin{cases} r_H\left(1 - \frac{X^*}{K}\right) = \left(\frac{Y^* p}{nK+1}\right)^{n+1} \\ Y^* = nX^* + 1. \end{cases} . \tag{9}$$

This leads to

$$r_H\left(1 - \frac{X^*}{K}\right) = \left(\frac{nX^* + 1}{nK + 1} p\right)^{n+1} , \tag{10}$$

which we can analyze by investigating the intersection of the curves formed by the left and right hand sides of this equation. Suppose this intersection is $X^* = \alpha > 0$ (Fig. 4). The equilibrium point is then $(X^* = \alpha, Y^* = n\alpha + 1)$ where α satisfies Eq. (10). This equilibrium point is stable (Fig. 4).

If parasite aggregation affects the carrying capacity of the parasite population, we can replace Y^* in Eq. 9 by an implicit equation $Y^* = \left(E[N] + \sigma\sqrt{Var[N]}\right) X^* + 1$. Since $E[N] + \sigma\sqrt{Var[N]} < n$, the new equilibrium state, if it exists, is $X^* = \gamma > \alpha$, where α is the equilibrium state if $Y^* = nX^* + 1$ (Fig. 4). This means that aggregation limits the parasitism-driven stress affecting the host population. Sample numerical simulations of host-parasite population dynamics are shown in Fig. 5 with varying values of σ. The value of $E[N] + \sigma\sqrt{Var[N]}$ converges to n if the value of σ is increased. In the example, notice that an aggregated parasite distribution with $\sigma = 2$ increases the parasite population rapidly but without causing too much harm to the host population (Fig. 5).

4 Discussion

Negative binomial distribution is commonly used in modeling macroparasite infections (Pennycuick, 1971; Wegner et al., 2008). Here, we propose a mechanistic model of parasite aggregation without initially assuming a statistical distribution. Our results show that the emergent values of the mean and variance of the macroparasite distribution indeed denote a negative binomial. We have shown that accumulation of macroparasites (e.g., through foraging) is sufficient for aggregation to arise under a wide range of conditions (Fig. 2). The complex life cycle of parasites, a relatively large

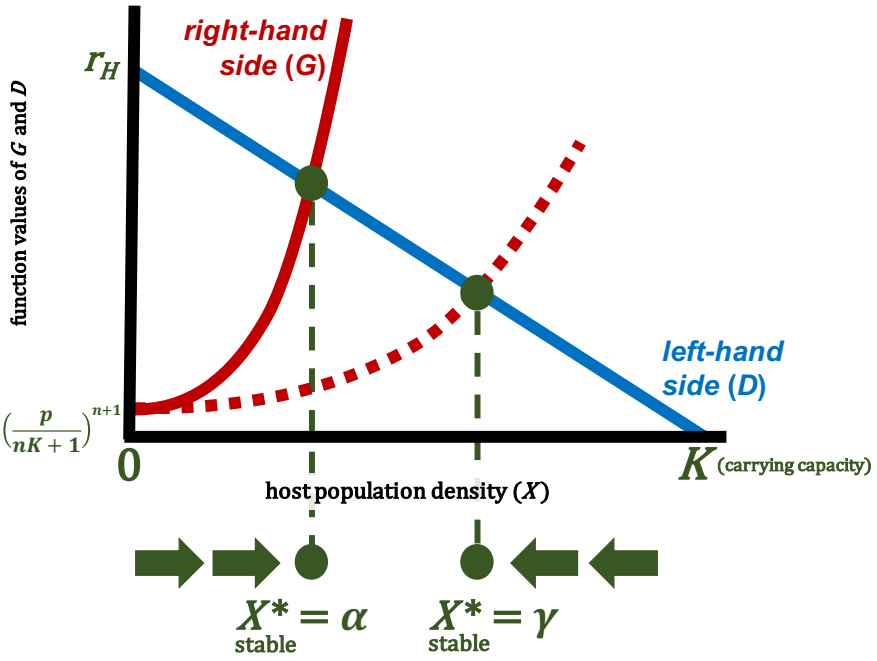

Figure 4: Illustration of the intersection of the curves formed by the left- (blue line) and right-hand (red curve) sides of Eq. (10). Right-hand side is based on $G = r_H\left(1 - \frac{X}{K}\right)$, the left-hand side is based on $D = \left(\frac{nX+1}{nK+1}p\right)^{n+1}$. If $G > D$ (blue line is above the red curve) then the population of host increases. If $G < D$ (blue line is below the red curve) then the population of host decreases. The intersection of the two curves is a non-zero stable equilibrium point (X^*). The broken red curve represents the function $D = \left(\frac{\left(E[N]+\sigma\sqrt{Var[N]}\right)X+1}{nK+1}p\right)^{n+1}$.

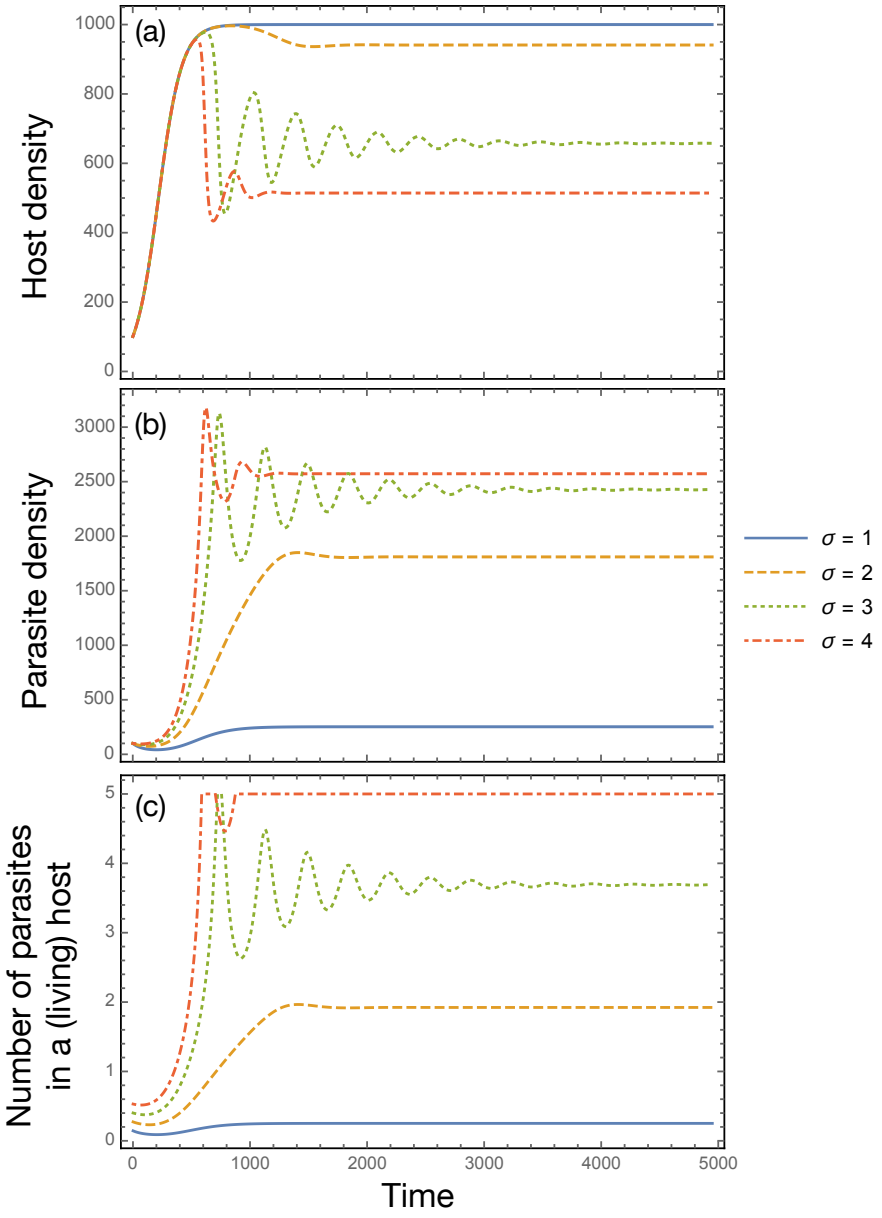

Figure 5: **Host-parasite dynamics for different values of σ, the intensity of how much the variance affects the average number of parasites per host.** The value of σ affects (A) host density and (B) parasite density. (C) The average number of parasites in a (living) host ($E[N]+\sigma\sqrt{Var[N]}$) converges to the maximum possible (here $n = 5$) as the variance becomes more of a contributing factor (increase in σ). For low contribution of the variance (low σ), the hosts are typically parasite free, however increasing σ strengthens the host-parasite interactions leading to the typical oscillations (or an internal fixed point where both population coexist). Parameter values used in the numerics: $K = 1000$, $r_H = 0.01$, $r_P = 0.1$, and $p = 0.8$.

maximum tolerable parasite load n, and relatively moderate parasite acquisition probability P_i are important factors in parasite aggregation. This means that parasites have the opportunity to reproduce through infecting hosts but without killing many hosts. It would be rare in nature to find hosts having low n with parasites having high P_i resulting in non-aggregated parasite distribution since these hosts (as well as the parasites, if specific) are expected to be extinct.

Our model design can be used to predict what conditions result in over-dispersed (aggregated), under-dispersed, and random parasite distributions by investigating the values of the parasite acquisition parameter P_i. Over-dispersion is observed when the variance of the parasite load in hosts is higher than the mean, while under-dispersion has a mean higher than its variance. Random pattern arises if mean equals the variance. This prediction is valuable in parasitology when performing empirical studies, especially when designing statistical sampling procedures. If the parasites are aggregated in the host population, then a large number of samples is expected to be needed to select those hosts with high parasite load at the tail of the distribution (Shvydka et al., 2018).

The main advantage of our model compared to the classical input-output modelling framework is its simplicity without losing important biological details, such as we can model host-parasite interaction using minimal models but still aggregation is considered. The designed model framework has few variables since the effect of parasite distribution can be summarised using its moments (mean and variance). The framework supports traditional population dynamic models, such as the logistic host-parasite interaction model, which are amenable to numerical and analytic mathematical investigations. Remarkably, our assumed parameters, especially P_i, can be directly calculated from available empirical data. Moreover, our model is more general than the stratified worm burden in terms of parasite acquisition (Gurarie et al., 2010). In stratified worm burden, a host can only acquire one parasite at a time. In our model, hosts can acquire more than one parasite since compartments associated with X_i are defined as hosts with “at least” (not “exactly”) i parasite load, a scenario impossible in the stratified worm burden model.

There are multiple indices that are proposed to measure parasite aggregation. One example is the negative binomial parameter k, which is equivalent to m in $NB(m, \rho)$. This is defined by the equation $Var[N] = E[N] + E^2[N]/k$. As k decreases to zero (e.g, less than 1 but positive), the parasite distribution is said to become more aggregated. This low k can also infer heterogeneity in infection factors (Bolker, 2008). If k approaches infinity, the Poisson distribution results. Increases in k are used to in-

dicate a movement toward "randomness" (Bolker, 2008; Young and Young, 1990). In the case of parasite population following a geometric distribution, $k = 1$. Based on our results, we have identified parameters that affect the aggregation index k, most notably are the parameters P_i and n. This implies that P_i can be used as an alternative measure of aggregation that biologists can use in studying patterns of macroparasite distribution in hosts.

Given a set of parasite counts gathered from host samples, one can calculate the estimates for P_i, $\widetilde{E[N]}$ and $\widetilde{Var[N]}$. If we want to calculate the expected value of $N \in \{0, 1, 2, ..., n\}$ without assuming a large n, we can simply normalize the probabilities associated with N. That is, $E[N] = \frac{\widetilde{E[N]}}{1-\prod_{i=1}^{n+1} P_i}$. Similarly, we can do this normalization to obtain the variance of N, $Var[N]$. The variables and parameters X_i, Y_i, P_i, $E[N]$ and $Var[N]$ can be functions of time, and a time series analysis to study the temporal pattern of these variables and parameters can be implemented. If P_i's, the host-parasite encounter probabilities, are statistically equal, we can assume a fixed P for each unit of time by taking the geometric mean of P_i ($i \geq 1$), that is, $P = (\prod_{i=1}^{n} P_i)^{1/n}$.

The mortality rate due to parasitism in Eq. (3) can be modified to include the cases where a fraction of the hosts with $i < n$ number of parasites could also die due to infection. This rate can be formulated as bX where b is as follows:

$$b = \sum_{i=1}^{n} \omega_i \left(\prod_{j=1}^{i} P_j - \prod_{j=1}^{i+1} P_j \right) + \prod_{k=1}^{n+1} P_k. \tag{11}$$

The parameter ω_i is the fraction of $X_i - X_{i+1}$ killed by the parasites. Our conclusions from the qualitative analysis, especially given a fixed $P = P_i$ for $i \geq 1$, still remain true since we can suppose the mortality rate b as the slope of the parasite-driven host death in Fig. 3.

Here, we have shown the resulting distribution of parasites in the host population using homogeneous parasite acquisition probability. If the values of the acquisition probabilities become heterogeneous, such as if $P_1 < P_2 < \cdots < P_n$, $P_1 > P_2 > \cdots > P_n$ or P_i's are completely arbitrary, then our derived formulas to estimate $E[N]$ and $Var[N]$ are not applicable (e.g., Eq. (6)). However, our qualitative analysis to investigate the dynamics of Eqs. (3) and (4) could still hold and a numerical study would then be the feasible way forward. To find appropriate formulae (if data are not available) for $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ would then be a challenge for the future.

Various future studies can stem from our model design. One can include the existence of alternative or intermediate hosts and vectors, especially that neglected tropical diseases are commonly due to vector-borne macroparasite infections (Hollingsworth

et al., 2015). Our model can also be extended to include how spatial aspects and different treatment strategies affect the population dynamics of the hosts and parasites, and to infer the reasons why the value of k is dynamic as observed in empirical studies (Boag et al., 2001; Crompton et al., 1984; Pennycuick, 1971; Scott, 1987). Inclusion of multiple parasites (Hafer and Milinski, 2015), different models of host-parasite interactions in food webs (Flor, 1956), or the explicit inclusion of population dynamics together with host-parasite co-evolution (Gokhale et al., 2013; Rabajante et al., 2016; Song et al., 2015) are possible directions. Testing if such complexities retain the observed phenomenon of parasite aggregation would be a true test of the "law".

5 Acknowledgements

JFR appreciates the time spent at the Max Planck Institute for Evolutionary Biology developing the project.

6 Financial Support

JFR and ELA are supported by the Institute of Mathematical Sciences and Physics, University of the Philippines Los Baños. CSG is supported by the Max Planck Society.

7 Declarations

The authors declare that there are no conflicts of interest. The funding sources had no involvement in the study design, in the analysis and interpretation of results, in the writing of the manuscript and in the decision to submit the article for publication.

8 Author Contributions

JFR conceptualized the problem. JFR and CSG designed the model. JFR and ELA implemented the simulations. All authors contributed in the analysis of the model, interpretation of results, writing of the manuscript and preparation of figures.

9 Data Availability

https://github.com/tecoevo/parasiteaggregation

References

F. R. Adler and M. Kretzschmar. Aggregation and stability in parasite-host models. *Parasitology*, 104:199–205, 1992.

R. M. Anderson. The regulation of host population growth by parasitic species. *Parasitology*, 76:119–157, 1978.

R. M. Anderson and D. M. Gordon. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology*, 85:373–398, 1982.

R. M. Anderson and R. M. May. Regulation and stability of host-parasite population interactions: I. regulatory processes. *The Journal of Animal Ecology*, 47:219–247, 1978.

S. K. J. R. Auld and M. C. Tinsley. The evolutionary ecology of complex lifecycle parasites: linking phenomena with mechanisms. *Heredity*, 114:125–132, 2015.

M. Basanez and M. Boussinesq. Population biology of human onchocerciasis. *Philisophical Transactions of the Royal Society B*, 354:809–826, 1999.

B. Boag, J. Lello, A. Fenton, D.M. Tompkins, and P.J. Hudson. Patterns of parasite aggregation in the wild european rabbit (*Oryctolagus cuniculus*). *International Journal for Parasitology*, 31(13):1421 – 1428, 2001.

B. M. Bolker. Probability and stochastic distributions for ecological modeling. In *Ecological models and data in R*, pages 103–146. Princeton University Press, Princeton NJ, 2008.

J. M. Calabrese, J. L. Brunner, and R. S. Ostfeld. Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended poisson-gamma mixture model. *PLoS ONE*, 6(12), 2011.

T. S. Churcher, N. M. Ferguson, and M.-G. Basá nez. Density dependence and overdispersion in the transmission of helminth parasites. *Parasitology*, 131:121–132, 2005.

C. W. Clark. *Mathematical Bioeconomics*. John Wiley & Sons Inc., third edition, 2010.

H. D. Crofton. A quantitative approach to parasitism. *Parasitology*, 62(2):179–193, April 1971.

D. W. T. Crompton, A. E. Keymer, and S. E. Arnold. Investigating over-dispersion; *Moniliformis* (acanthocephala) and rats. *Parasitology*, 88:317–331, 1984.

B. Cundil and N. D. E. Alexander. Sample size calculations for skewed distributions. *BMC Medical Research Methodology*, 15, 2015.

C. P. P. de la Cruz, V. G. V. Paller, M. Z. Bandal Jr., and A. R. B. Avila. Distribution pattern of *Acanthogyrus* sp. (acanthocephala: Quadrigyridae) in nile tilapia (*Oreochromis niloticus* l.) from sampaloc lake, philippines. *Journal of Nature Studies*, 12(1):1–10, 2013.

L. Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, 2005.

R. A. Fisher. The negative binomial distribution. *Annals of Eugenics*, 11(1):182–187, 1941.

H. H. Flor. The complementary genetic systems in flax and flax rust. *Advances in Genetics*, 8:29–54, 1956.

A. P. Galvani. Immunity, antigenic heterogeneity, and aggregation of helminth parasites. *Journal of Parasitology*, 89(2):232–241, 2003.

S. Gandon and T. Day. Evolutionary epidemiology and the dynamics of adaptation. *Evolution*, 63(4):826–838, 2009.

C. S. Gokhale, A. Papkou, A. Traulsen, and H. Schulenburg. Lotka-Volterra dynamics kills the Red Queen: population size fluctuations and associated stochasticity dramatically change host-parasite coevolution. *BMC Evolutionary Biology*, 13:254, 2013.

S. Gourbiére, S. Morand, and D. Waxman. Fundamental factors determining the nature of parasite aggregation in hosts. *PLoS ONE*, 10(2), 2015.

R. Guilhem, A. Šimková, S. Morand, and S. Gourbière. Within-host competition and diversification of macro-parasites. *Journal of The Royal Society Interface*, 9:2936–2946, 2012.

D. Gurarie, C. H. King, and X. Wang. A new approach to modelling schistosomiasis transmission based on stratified worm burden. *Parasitology*, 137:1951–1965, 2010.

D. Gurarie, N. Yoon, E. Li, M. Ndeffo-Mbah, D. Durham, A. E. Phillips, H. O. Aurelio, J. Ferro, A. P. Galvani, and C. H. King. Modelling control of schistosoma haematobium infection: predictions of the long-term impact of mass drug administration in africa. *Parasites & Vectors*, 8, 2015.

D. Gurarie, C. H. King, N. Yoon, and E. Li. Refined stratified-worm-burden models that incorporate specific biological features of human and snail hosts provide better estimates of schistosoma diagnosis, transmission, and control. *Parasites & Vectors*, 9, 2016.

N. Hafer and M. Milinski. When parasites disagree: evidence for parasite-induced sabotage of host manipulation. *Evolution*, 69(3):611–620, 2015.

T. D. Hollingsworth, J. R. C. Pulliam, S. Funk, J. E. Truscott, V. Isham, and A. L. Lloyd. Seven challenges for modelling indirect transmission: Vector-borne diseases, macroparasites and neglected tropical diseases. *Epidemics*, 10:16–20, 2015.

N. C. Lopez. Parasitic crustaceans in fishes from some philippine lakes. In C B Santiago, M L Cuvin-Aralar, and Z U Basiao, editors, *Conservation and Ecological Management of Philippine Lakes in Relation to Fisheries and Aquaculture*, pages 75–79. Southeast Asian Fisheries Development Center, Aquaculture Department, Iloilo, Philippines; Philippine Council for Aquatic and Marine Research and Development, Los Banos, Laguna, Philippines; and Bureau of Fisheries and Aquatic Resources, Quezon City, Philippines, Philippines, 2001.

H. McCallum, A. Fenton, P. J. Hudson, B. Lee, B. Levick, R. Norman, S. E. Perkins, M. Viney, A. J. Wilson, and J. Lello. Breaking beta: deconstructing the parasite transmission function. *Philosophical Transactions of the Royal Society B*, 372, 2017.

A. Morrill and M. R. Forbes. Random parasite encounters coupled with condition-linked immunity of hosts generate parasite aggregation. *International Journal for Parasitology*, 42:701–706, 2012.

A. Morrill and M. R. Forbes. Aggregation of infective stages of parasites as an adaptation and its implications for the study of parasite-host interactions. *The American Naturalist*, 187:000–000, 2016.

A. Morrill, M. R. Forbes, and F. Dargent. Explaining parasite aggregation: more than one parasite species at a time. *International Journal for Parasitology*, 47:185–188, 2017.

V. G. V. Paller, D. J. B. Resurreccion, C. P. P. de la Cruz, and M. Z. Bandal. Acanthocephalan parasites (*Acanthogyrus* sp.) of nile tilapia (*Oreochromis niloticus*) as biosink of lead (pb) contamination in a philippine freshwater lake. *Bulletin of Environmental Contamination and Toxicology*, 96(6):810–815, 2016.

L. Pennycuick. Frequency distributions of parasites in a population of three-spined sticklebacks, gasterosteus aculeatus l., with particular reference to the negative binomial distribution. *Parasitology*, 63(3):389–406, 1971.

R. Poulin. Are there general laws in parasite ecology? *Parasitology*, 134(Pt 6):763–776, 2007.

J. F. Rabajante, J. M. Tubay, H. Ito, T. Uehara, S. Kakishima, S. Morita, J. Yoshimura, and D. Ebert. Host-parasite red queen dynamics with phase-locked rare genotypes. *Science Advances*, 2(3), 2016. doi: 10.1126/sciadv.1501548.

R. Rosà and A. Pugliese. Aggregation, stability, and oscillations in different models for host-macroparasite interactions. *Theoretical Population Biology*, 61:319–334, 2002.

L. Rozsa, J. Reiczigel, and G. Majoros. Quantifying parasites in samples of hosts. *The Journal of parasitology*, 86(2):228–232, April 2000.

M. E. Scott. Temporal changes in aggregation: a laboratory study. *Parasitology*, 94 (3):583–595, 1987.

D. J. Shaw and A. P. Dobson. Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology*, 111:S111–S133, 1995.

J. S. Shonkwiler. Variance of the truncated negative binomial distribution. *Journal of Econometrics*, 195:209–210, 2016.

S. Shvydka, V. Sarabeev, V.D. Estruch, and C. Cadarso-Suarez. Optimum sample size to estimate mean parasite abundance in fish parasite surveys. *Helminthologia*, 55(1):52–59, 2018.

Y. Song, C. S. Gokhale, A. Papkou, H. Schulenburg, and A. Traulsen. Host-parasite coevolution in populations of constant and variable size. *BMC Evolutionary Biology*, 15:212, 2015.

M. Viney and J. Cable. Microparasite life histories. *Current Biology*, 21:R767–R774, 2011.

K. M. Wegner, M. Kalbe, M. Milinski, and T. B.H. Reusch. Mortality selection during the 2003 european heat wave in three-spined sticklebacks: effects of parasites and mhc genotype. *BMC Evolutionary Biology*, 8(1):124, 2008.

M. Q. Wilber, P. T. J. Johnson, and C. J. Briggs. When can we infer mechanism from parasite aggregation? a constraint-based approach to disease ecology. *Ecology*, 98(3):688–702, 2017.

K. Wilson, B. T. Grenfell, and D. J. Shaw. Analysis of aggregated parasite distributions: A comparison of methods. *Functional Ecology*, 10(5):592–601, 1996.

K. Wilson, O. N. BjÃžrnstad, A. P. Dobson, S. Merler, G. Poglayen, S. E. Randolph, A. F. Read, and A. Skorping. Heterogeneities in macroparasite infections: Patterns and processes. In P. Hudson, A. Rizzoli, B. Grenfell, H. Heesterbeek, and A. Dobson, editors, *The Ecology of Wildlife Diseases*, chapter 2, pages 6–44. Oxford University Press, Oxford, 2001.

L. Yakob, R. J. S. Magalh aes, D. J. Gray, G. Milinovich, N. Wardrop, R. Dunning, J. Barendregt, F. Bieri, G. M. Williams, and A. C.A. Clements. Modelling parasite aggregation: disentangling statistical and ecological approaches. *International Journal of Parasitology*, 44:339–342, 2014.

L. J. Young and J. H. Young. A spatial view of the negative binomial parameter k when describing insect populations. *Conference on Applied Statistics in Agriculture*, 1990.

C. Zimmer. *Parasite Rex: Inside the Bizarre World of Nature's Most Dangerous Creatures*. Atria Books, 2001.

Table SI.1: Description of the state variables and parameters. All state variables and parameters are non-negative.

Notation	Description
X	total host population density
X_0	population density of hosts without parasite
$X_i, i \geq 1$	population density of hosts with *at least* i number of parasites
$X_i - X_{i+1}, i \geq 1$	number of hosts having *exactly* i number of parasites
Y	total parasite population density
Y_i	population density of parasites associated with hosts having exactly i number of parasites
r_H	basal host population growth rate
r_P	basal parasite population growth rate
K	carrying capacity of host population
c	quantitative measure of the environment where the parasites can survive apart from the hosts
p	parasite transmission probability
n	maximum parasite load of a host without dying
P_0	proportion of X without parasite
$P_i, i \geq 1$	proportion of X_{i-1} infected by at least i number of parasites; net probability of parasite acquisition by a host
G	growth function
D	death function

SI.1 Supplementary Information

SI.1.1 List of state variable and parameters

Since we have

$$X_0 + (X_1 - X2) + (X_2 - X_3) \ldots + (X_n - X_{n+1}) + X_{n+1} = X$$
$$P_0 + (P_1 - P_2P_1) + \left(\prod_{i=1}^{2} P_i - \prod_{i=1}^{3} P_i\right) + \ldots + \left(\prod_{i=1}^{n} P_i - \prod_{i=1}^{n+1} P_i\right) + \prod_{i=1}^{n+1} P_i = 1,$$

the mean of the random variable $N+1$ is

$$E[N+1] = 0P_0 + 1(P_1 - P_2P_1) + 2\left(\prod_{i=1}^{2} P_i - \prod_{i=1}^{3} P_i\right) + \ldots + n\left(\prod_{i=1}^{n} P_i - \prod_{i=1}^{n+1} P_i\right) + w\prod_{i=1}^{n+1} P_i$$

The random variable $N+1 \in \{0, 1, 2, \ldots, n, w\}$ (where $w \geq n+1$) represents the parasite load in living and parasitism-driven dead hosts. For simplicity, we let $w = n+1$ since hosts with equal or more than $n+1$ parasites are dead due to parasitism. Let us define $\widetilde{E[N]} = E[N+1] - (n+1)\prod_{i=1}^{n+1} P_i = E[N+1] - (n+1)\left(\frac{Yp}{nK+1}\right)^{n+1}$, which is the approximate mean of N (where $N \in \{0, 1, 2, \ldots, n\}$) is as discussed in the main text.

SI.1.2 Derivation of A

We assume that the parasite acquisition probability is $P = \frac{Yp}{nK+1} < 1$. Using the geometric series, we know that,

$$1 + P + P^2 + \ldots + P^n = \frac{1 - P^{n+1}}{1 - P}.$$

Then taking the derivative of this geometric series results in

$$1 + 2P + \ldots + nP^{n-1} = \frac{nP^{n+1} - (n+1)P^n + 1}{(1-P)^2}.$$

The expression for A is:

$$\begin{aligned} A &= \sum_{i=1}^{n} (iP)^{i-1} \\ &= \frac{n(P)^{n+1} - (n+1)(P)^n + 1}{(1-P)^2}. \end{aligned} \tag{SI.1}$$

Table SI.2: Description of the notations used in representing the mean and variance of the parasite distribution.

Notation	Description
$E[N]$ and $E[N+1]$	mean of the random variable $N \in \{0, 1, 2, ..., n\}$ (living hosts) and $N+1 \in \{0, 1, 2, ..., n+1\}$ (living and dead hosts), respectively
$Var[N]$ and $Var[N+1]$	variance of the random variable $N \in \{0, 1, 2, ..., n\}$ and $N+1 \in \{0, 1, 2, ..., n+1\}$, respectively
$\widetilde{E[N]}$	pseudomean for approximating $E[N]$; equal to $E[N+1] - (n+1)\prod_{i=1}^{n+1} P_i$ which is the relative parasite population frequency in living hosts, representing the distribution of the parasites in the living hosts
$\widetilde{Var[N]}$	pseudovariance for approximating $Var[N]$
$E[\infty]_P$	mean of the geometric-distributed parasite population with parasite acquisition probability P
$Var[\infty]_P$	variance of the geometric-distributed parasite population with parasite acquisition probability P

SI.1.3 Explicit formula for the variance

Let us define $\widetilde{Var[N]} = Var[N+1] - (n+1-E[N+1])^2 \prod_{i=1}^{n+1} P_i$ where $Var[N+1] = E[(N+1)^2] - E[N+1]^2$ is the variance of the random variable $N+1$.

With the same assumption as above for P, The expression for $E[(N+1)^2]$ is

$$\begin{aligned} E[(N+1)^2] &= \sum_{i=1}^{n} i^2 P^i (1-P) + (n+1)^2 P^{n+1} \\ &= P(1-P)B + (n+1)^2 P^{n+1}. \end{aligned}$$

The expression for B can be derived using the geometric series. We know that

$$1 + 2P + \ldots + nP^{n-1} = \frac{nP^{n+1} - (n+1)P^n + 1}{(1-P)^2}.$$

Multiplying both sides by P, we have

$$P + 2P^2 + \ldots + nP^n = \frac{nP^{n+2} - (n+1)P^{n+1} + P}{(1-P)^2}.$$

Taking the derivative of the left- and right-hand sides, we arrive at the following expression for B:

$$\begin{aligned} &1 + 4P + \ldots + n^2 P^{n-1} = B = \\ &\frac{-n^2 P^{n+2} + (2n^2 + 2n - 1)P^{n+1} - (n+1)^2 P^n + P + 1}{(1-P)^3}. \end{aligned}$$

Hence, the explicit formula for $\widetilde{Var[N]}$ is

$$\begin{aligned} &\widetilde{Var[N]} = \\ &\frac{P\left(-n^2 P^{n+2} + (2n^2 + 2n - 1)P^{n+1} - (n+1)^2 P^n + P + 1\right)}{(1-P)^2} \\ &+ (n+1)^2 P^{n+1} - \left(\widetilde{E[N]} + (n+1)P^{n+1}\right)^2 \\ &- \left(n + 1 - \widetilde{E[N]} + (n+1)P^{n+1}\right)^2 P^{n+1}. \end{aligned} \tag{SI.2}$$

SI.1.4 Proof for a claim in Section 3.1

Here is the proof that $\widetilde{E[N]} = 0$ when $Y_1^* = 0$ only happens if $P = 0$. Suppose $P < 1$:

$$\widetilde{E[N]} = \frac{P\left(nP^{n+1} - (n+1)P^n + 1\right)}{1-P}$$

implies $P = 0$ or $nP^{n+1} - (n+1)\,P^n + 1 = 0$. If $nP^{n+1} - (n+1)\,P^n + 1 = 0$ then

$$nP^n\,(1-P) = 1 - P^n.$$

Note that $nP^n = \frac{1-P^n}{1-P}$ represents the geometric series:

$$nP^n = 1 + P + P^2 + \ldots + P^{n-1}.$$

However, $P^n = \frac{P^0+P^1+\ldots+P^{n-1}}{n}$ is the arithmetic average of $\{P^{n-1}, P^{n-2}, \ldots, P^1, P^0\}$. This is a contradiction since $P^n \notin \{P^{n-1}, P^{n-2}, \ldots, P^1, P^0\}$.

Suppose $P = 1$: Note that $P = \frac{Y_1^* p}{nK+1}$ is fixed. If $P = 1$, then $\frac{Y_1^* p}{nK+1} = 1$ or $Y_1^* = \frac{nK+1}{p}$. But $Y_1^* = 0$ and $p < \infty$, a contradiction.

Hence, $\widetilde{E[N]} = 0$ only if $P = 0$.

SI.1.5 Histograms given different values of P, $n = 10$

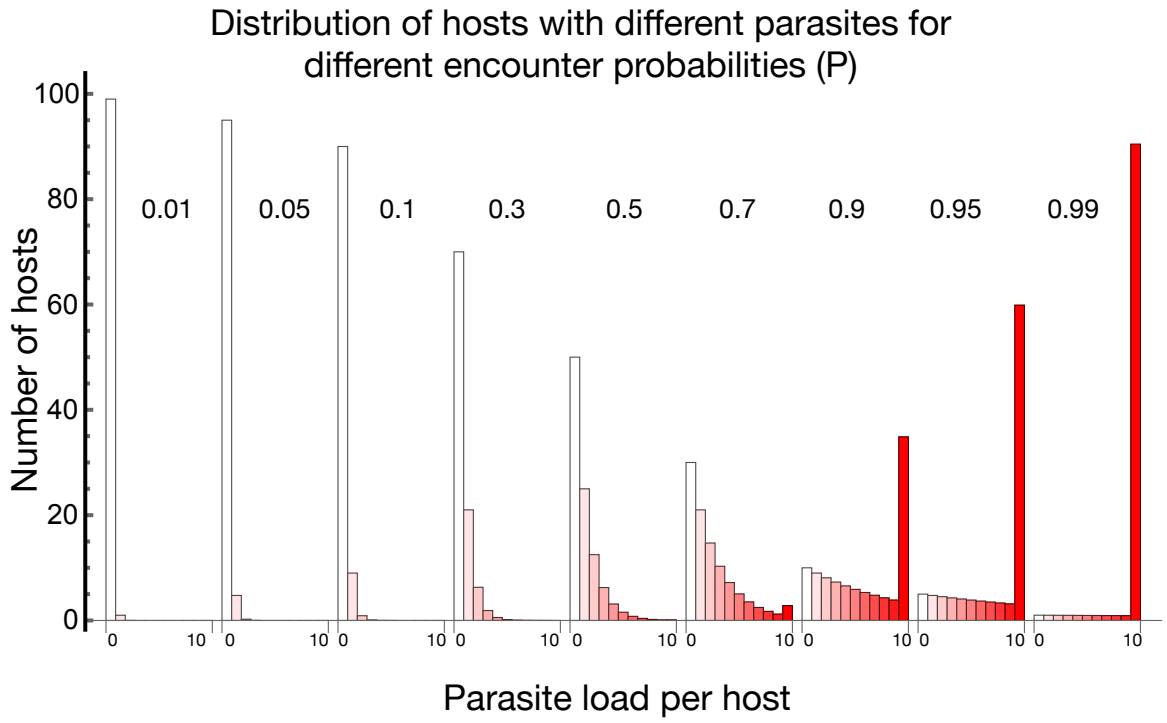

Figure SI.1: **The distributions for parasite load in hosts for different parasite acquisition probabilities P.** The distribution becomes more negatively skewed as P increases. Higher P results in high host mortality due to harbouring high parasite load. This is the reason why as P increases, the errors in approximating $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ using the geometric distribution also increase (Fig. 2).

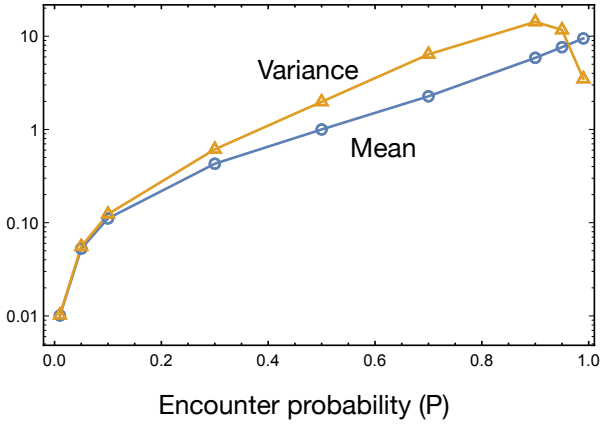

Figure SI.2: **An intermediate value of P results in parasite aggregation (variance>mean).** A value of P near 1 when $n = 10$ results in nonaggregated parasite distribution (mean>variance), which is possibly due to high mortality in hosts.