On the mechanistic roots of an ecological law: parasite aggregation

Jomar F. Rabajante^{1,2,*}, Elizabeth L. Anzia¹ & Chaitanya S. Gokhale³

¹Institute of Mathematical Sciences and Physics,

University of the Philippines Los Baños, 4031 Laguna, Philippines

²Faculty of Education,

University of the Philippines Open University, 4031 Laguna, Philippines

³Research 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

*jfrabajante@up.edu.ph

Abstract

Parasite aggregation, a recurring pattern in macroparasite infections, is considered one of the “laws” of parasite ecology. Few hosts have a large number of parasites while most hosts have a low number of parasites. Phenomenological models of host-parasite systems thus use the negative-binomial distribution. However, to infer the mechanisms of aggregation, a mechanistic model that does not make any a priori assumptions is essential. Here we formulate a mechanistic model of parasite aggregation in hosts without 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 statistics in host-parasite interactions, 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 inform the designing of 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, thus providing a basis for a “law” of ecology.

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

Introduction

Parasites are ubiquitous [1]. Nevertheless, for parasitologists, sampling can be a hard problem [2, 3]. This is due to the distribution of parasites among hosts [4]. While most hosts harbour very few parasites, only few individuals are hosts to a large number of parasites [5]. 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 [6, 7, 8]. The pattern is specifically observed in macroparasite infections, which include diseases brought about by helminths and arthropods [7, 8]. Knowing the macroparasite distribution in host populations is essential for addressing challenges in investigating disease transmission. Examples are human onchocerciasis and schistosomiasis [9, 10, 11]. Aggregation can also affect co-infection by various parasites, parasite-driven evolutionary pressures, and stability of host-parasite communities [12, 13, 14, 15]. 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 [5, 16, 17]. Given this fact, several theoretical models about host-parasite dynamics implicitly assume the negative-binomial distribution, mainly based on a phenomenological (statistical) modelling framework [18, 6, 8, 19]. The phenomenological principle is based on the observation that - (i) a distribution is Poisson if showing random behaviour with equal mean and variance, (ii) binomial if underdispersed with higher mean than variance, or (iii) negative-binomial if overdispersed with higher variance than the mean [12, 20]. Overdispersion is observed in host-parasite interaction, such as between cattle (*Bos taurus*) and warble fly (*Hypoderma*

47 *bovis*), and between Nile tilapia (*Oreochromis niloticus*) and copepod *Ergasilus philip-*
 48 *pinensis* [12, 21]. The negative-binomial distribution is hypothesized to arise due to
 49 the heterogeneity in the characteristics of the host-parasite interaction, and environ-
 50 mental variation. Heterogeneous exposures, infection rates and susceptibility of host
 51 individuals can produce aggregated distributions of parasites [12, 22, 23, 24]. How-
 52 ever, is heterogeneity in the characteristics of the hosts and parasites a sufficient
 53 condition for aggregation [25]? Using a mechanistic model, we show that a regular
 54 but non-extreme parasite accumulation can lead to aggregation, and neither to Uni-
 55 form nor Poisson distribution. We show that even in regular stochastic host-parasite
 56 systems (i.e., with homogeneous conditional probabilities of macroparasite accumu-
 57 lation), aggregation is widely possible.

58 Traditionally, in arriving at the desired model for host-macroparasite interaction,
 59 phenomenological modelling is assumed [18, 26, 5, 27, 28]. Phenomenological mod-
 60 els are based on data gathered, but typically, the mechanisms underlying the phe-
 61 nomenon are hidden. There is a need for descriptive and predictive models that
 62 assume the underlying mechanistic processes of parasite dynamics, useful in formu-
 63 lating disease control programs [2]. The goal of this study is to mechanistically illus-
 64 trate the interaction between hosts and macroparasites without assuming a negative-
 65 binomial distribution. Our approach considers parasite accumulation without direct
 66 reproduction in host individuals, which is a consequence of the complex life history
 67 of macroparasites [29, 30, 31]. For example, the Nile tilapia fish are infected by the
 68 acanthocephalans parasites via the ingestion of infected zooplankton. The parasites
 69 grow, mate and lay eggs inside the fish, and then eggs are expelled in the lake through
 70 faecal excretions. The parasites in the excretions are passively consumed by the in-
 71 termediate hosts (zooplanktons), and the cycle of infection through foraging continues
 72 [32, 33].

73 1.1 The gap in existing models

74 Recently, various attempts have been made to model the accumulation and aggre-
 75 gation of parasites mechanistically. The stratified worm burden, which is based on
 76 a chain of infected compartments as in Susceptible-Infected (S-I) modeling frame-
 77 work, was used to model schistosomiasis infections [34, 35, 36]. This model can
 78 be well simulated numerically for specific cases, but possibly difficult to implement
 79 analytic studies to find general mathematical conclusions [34]. Although there are
 80 several studies that have considered addressing this issue (e.g., simplifying infinite
 81 dimensional differential equations) [37], our model proposes an alternative and sim-
 82 ple approach.

83 The Poisson-Gamma Mixture Model has been proposed to model aggregation
 84 based on parasite accumulation [38]. The negative-binomial distribution arises natu-
 85 rally from a Poisson-Gamma process; however, is it possible to derive a mechanistic
 86 model of aggregation without initially assuming a distribution related to the negative-
 87 binomial? The answer to this question is “yes”, and there are existing studies that
 88 consider bottom-up mechanistic models [39, 37, 40]; however, these models are very
 89 few compared to the phenomenological approaches available. Our model is pro-
 90 posed as an alternative approach, but with unique features, that can be added to the
 91 list of available models useful for studying diverse types of macroparasites. One such
 92 feature of our model is that we can predict over a wide-range of parameter space
 93 when underdispersed and random parasite distributions (not just the overdispersed
 94 distribution) will emerge. Our model also predicts the emergence of the Geometric
 95 distribution, confirming previous predictions [41], which can arise even in the absence
 96 of heterogeneity.

97 In another study, a mechanistic model has been proposed based on the random
 98 variation in the exposure of hosts to parasites and in the infection success of parasites
 99 [6]. This model anchors its derivation and conclusion on the common assumption that
 100 heterogeneity in individual hosts and/or parasites leads to parasite aggregation. In our
 101 proposed model, we can have further insights on the resulting distribution of parasites

102 given homogeneous or heterogeneous biological, ecological and environmental fac-
103 tors affecting the host-parasite interaction.

104 Moreover, in most models of macroparasite infections, the rate of disease acqui-
105 sition by susceptible individuals (known as the “force of infection” β) is assumed as
106 a parameter. However, identifying the value of β can be difficult or infeasible even in
107 the availability of data [29]. In our proposed model, this challenge can be addressed
108 since our assumed infection parameter, denoted by P , can be directly estimated from
109 parasite count data gathered from host samples in empirical studies.

110 Our model provides an alternative framework which is simple and tractable. Be-
111 sides being useful in estimating required sampling size in the field, our work can be
112 incorporated as part of classical modeling frameworks (e.g., host-parasite interaction
113 with logistic growth) [42].

114 Proposed model

115 The core concept of our model is captured in Fig. 1. The total host population has
116 a density of X . With probability P_0 , the hosts are parasite-free, and the density of
117 parasite-free hosts is then $X_0 = P_0X$. Similarly, the density of hosts which have at
118 least one parasite is $X_1 = P_1X$. The total host population is therefore equivalent to
119 $X = X_0 + X_1 = P_0X + P_1X$. Moreover, the class of individuals X_i which have at least
120 $i > 1$ number of parasites is assumed as a subset of X_1 (i.e., $X_{n+1} \subseteq X_n \subseteq \dots \subseteq$
121 $X_2 \subseteq X_1$), and can be modeled as,

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

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 , $i \geq 1$ ($X_{i+1} \subseteq X_i$). 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 the Table in SI for the description of the state variables and parameters.

122 From the above set of equations, we can derive the distribution of living hosts (all
123 except the $n + 1$ compartment) with different parasite load concerning the total host
124 population. Fig. 1 represents the classification of the states of the hosts depending
125 on their parasite load. The compartment associated with X_i , $i \geq 1$ contains the hosts
126 infected by at least i number of parasites. The parameter P_{i+1} characterizes the
127 net probability that a host in X_i acquires an additional parasite (that is, probability
128 of catching a parasite less the probability of removing a parasite from the host, e.g.,
129 through treatment or parasite death). Consequently, the difference $X_i - X_{i+1}$, $i \geq 1$
130 is the population density of hosts with *exactly* i number of parasites. The population
131 density of living hosts (with maximum tolerable parasite load n) is then $X_n - X_{n+1}$.

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 modelling framework (e.g., stratified worm burden model, which is an extension of the standard S-I epidemiological models). Such a classical modelling framework could be intractable when modelling aggregation [34]. 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 efficiently 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., the state with three 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 one parasite, captured by adjusting the parameter P_i . This approach is consistent with the experimental approaches and comparable to the data gathered. For example, the proportion $P_i = \frac{X_{n+1}}{X_n}, n > 1$ is directly computable from parasite load data from sampled hosts as compared to computing the force of infection β in S-I models [43].

The quantities relating to the host and parasite proportions, X_i , P_i and n can be dynamic (e.g., may change over time). For example, 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 temporal connection allows us to relate our model to experimental data. The parameter values (e.g., P_i) can be determined from the samples gathered at a specific 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. An underlying 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 depend on the hosts to survive, e.g., we assume a small value of $c = 1$. Homogeneity in parasite transmission is implemented by keeping p constant.

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

$$\begin{aligned} Y_1 &= (P_1 - P_2 P_1) X \\ Y_2 &= 2 \left(\prod_{i=1}^2 P_i - \prod_{i=1}^3 P_i \right) X \\ Y_3 &= 3 \left(\prod_{i=1}^3 P_i - \prod_{i=1}^4 P_i \right) X \\ &\vdots \\ Y_n &= n \left(\prod_{i=1}^n P_i - \prod_{i=1}^{n+1} P_i \right) X. \end{aligned} \quad (2)$$

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 .

2.1 Application to host-parasite population dynamics

The host basal growth rate is set to a constant r_H , assuming that macroparasites do not affect reproduction of hosts. 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 modelled assuming logistic growth as follows (refer to Table SI.1 in SI for the description of the state variables and parameters):

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

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

Results

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

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

where A is derived from the derivative of a geometric series (see SI). 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,

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

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

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 $n \rightarrow \infty$ [5, 44]. In the next section (3.1), we discuss that as $n \rightarrow \infty$, $\widetilde{E}[N]$ 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 $N \in \{0, 1, 2, \dots, n\}$, respectively. $E[N]$ can be interpreted as the average number of parasites in a host with variance $Var[N]$. From Eq. (6) as $n \rightarrow \infty$, $\widetilde{E}[N] = E[N]$ 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 in SI).

This result implies that even if the host can sustain a large number of parasites ($n \rightarrow \infty$) 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 overdispersed 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 an approximation. The error due to the approximation can be estimated, 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 reasonable 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 a

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 $E[\infty]_P$, $error = |\widetilde{E[N]} - E[\infty]_P|$. (B) Absolute difference between $\widetilde{Var[N]}$ and $Var[\infty]_P$, $error = |\widetilde{Var[N]} - Var[\infty]_P|$.

higher error, this could also be an improbable scenario in nature. The condition where there is a high probability of acquiring more parasites but the host can only tolerate low parasite burden would result in morbid infection rates and potentially lead to the mortality and eventual extinction of the hosts. This possibility 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. 2 B shows that as $n \rightarrow \infty$, the error becomes smaller. This result 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 the host population with differing values of P . If the value of P is intermediate, it 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. Therefore, 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, 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 ($D = P^{n+1}X$) [45]. 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 curves 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 ($X_2^* = \frac{(r_H - P^{n+1})K}{r_H}$) (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 $\widetilde{E}[N] = 0$;

Figure 3: Illustration of host population dynamics. The population dynamics of hosts is 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 density of host increases. If $G < D$ (blue curve is below the red line) then the population density 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 .

250 and a stable coexistence of hosts and parasites at $Y_2^* = nX^* + 1$ where X^* is the
 251 host equilibrium. The equilibrium state Y_1^* is unstable, and Y_2^* is stable. The condition
 252 $\widetilde{E[N]} = 0$ for Y_1^* only happens if $P = 0$ (proof in SI).

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

$$E[N] + \sigma\sqrt{\text{Var}[N]} < n \quad (8)$$

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

264 **3.2 Variable host-parasite encounter probability:** 265 **Population dynamics**

266 Until now, we assumed that the parasite-encounter probability P as a constant. How-
 267 ever, $P = \frac{Y}{nK+1}$ (Eqs. (3) and (4)) is a function of the dynamic variable Y , the
 268 parasite density. Given the dynamics of the parasite, we can have three possible
 269 equilibrium points. Two of the equilibrium points are trivial, one where no host and
 270 parasite exist ($X^* = 0, Y^* = Y_0$), and another at the carrying capacity of the hosts
 271 ($X^* = K, Y^* = Y_0 = 0$) 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

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

Figure 4: Illustration of Eq. (10), the intersection of the curves formed by the left- (blue line) and right-hand (red curve) sides of Eq. (10). Left-hand side is based on $G = r_H \left(1 - \frac{X}{K}\right)$, the right-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 $D = \left(\frac{(E[N] + \sigma\sqrt{Var[N]})X+1}{nK+1}p\right)^{n+1}$.

This leads to

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

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 $Y^* = (E[N] + \sigma\sqrt{Var[N]})X^* + 1$.

278 Since $E[N] + \sigma\sqrt{Var[N]} < n$, the new equilibrium state, if it exists, is $X^* = \gamma > \alpha$,
 279 where α is the equilibrium state if $Y^* = nX^* + 1$ (Fig. 4). This relation means that
 280 aggregation limits the parasitism-driven stress affecting the host population. Sample
 281 numerical simulations of host-parasite population dynamics are shown in Fig. 5 with
 282 varying values of σ . The value of $E[N] + \sigma\sqrt{Var[N]}$ converges to n if the value of
 283 σ is increased. In the example, notice that an aggregated parasite distribution with
 284 $\sigma = 2$ increases the parasite population rapidly but without causing too much harm to
 285 the host population (Fig. 5).

286 Discussion

287 Negative-binomial distribution is commonly used in modeling macroparasite infections
 288 [46, 47]. Here, we propose a mechanistic model of parasite aggregation without ini-
 289 tially assuming a statistical distribution. Our results show that the emergent values of
 290 the mean and variance of the macroparasite distribution indeed denote a negative-
 291 binomial. We have shown that accumulation of macroparasites (e.g., through forag-
 292 ing) is sufficient for aggregation to arise under a wide range of conditions (Fig. 2).
 293 The complex life cycle of parasites, a relatively large maximum tolerable parasite load
 294 n , and relatively moderate parasite acquisition probability P_i are essential factors in
 295 parasite aggregation. Thus the parasites have the opportunity to reproduce through
 296 infecting hosts but without killing many hosts. It would be rare in nature to find hosts
 297 with low tolerance n with parasites having high P_i resulting in non-aggregated parasite
 298 distribution since these hosts (as well as the parasites, if specific) would be extinct.

299 Our model is useful in predicting the conditions resulting in overdispersed (aggre-
 300 gated), underdispersed, and random parasite distributions by investigating the values
 301 of the parasite acquisition parameter P_i . Overdispersion is observed when the vari-
 302 ance of the parasite load in hosts is higher than the mean, while underdispersion has
 303 a mean higher than its variance. Random patterns arise if mean equals the variance.
 304 This prediction is valuable in parasitology when performing empirical studies, espe-
 305 cially when designing statistical sampling procedures. If the parasites are aggregated

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 $p = 0.8$.

306 in the host population, then a large number of samples is expected to be needed to
307 select those hosts with high parasite load at the tail of the distribution [48].

308 The main advantage of our model compared to the classical input-output mod-
309 elling framework is its simplicity without losing important biological details. Therefore,
310 we can model host-parasite interactions using minimal models, but still, consider-
311 ing aggregation. The designed model framework has few variables since the effect
312 of parasite distribution can be summarised using its moments (mean and variance).
313 The framework supports traditional population dynamic models, such as the logistic
314 host-parasite interaction model, which are amenable to numerical and analytic math-
315 ematical investigations. Remarkably, our assumed parameters, especially P_i , can be
316 directly calculated from available empirical data. Moreover, our model is more gen-
317 eral than the stratified worm burden in terms of parasite acquisition [34]. In stratified
318 worm burden, a host can only acquire one parasite at a time. In our model, hosts can
319 acquire more than one parasite since compartments associated with X_i are defined
320 as hosts with “at least” (not “exactly”) i parasite load, a scenario impossible in the
321 stratified worm burden model.

322 Multiple indices are typically proposed to measure parasite aggregation. One ex-
323 ample is the negative-binomial parameter k , which is equivalent to m in $NB(m, \rho)$.
324 This is defined by the equation $Var[N] = E[N] + E^2[N]/k$. As k decreases to zero
325 (e.g., less than one but positive), the parasite distribution is said to become more
326 aggregated. This low k can also infer heterogeneity in infection factors [49]. If k ap-
327 proaches infinity, the Poisson distribution results. Increase in k indicates a movement
328 toward “randomness” [49, 50]. When $k = 1$, the parasite population follows a geo-
329 metric distribution. Based on our results, we have identified parameters that affect
330 the aggregation index k , most notably are the parameters P_i and n . This implies that
331 P_i can be used as an alternative measure of aggregation that biologists can use in
332 studying patterns of macroparasite distribution in hosts.

333 Given a set of parasite counts gathered from host samples, one can calculate the
334 estimates for P_i , $\widetilde{E[N]}$ and $\widetilde{Var[N]}$. If we want to calculate the expected value of $N \in$
335 $\{0, 1, 2, \dots, n\}$ without assuming a large n , we can simply normalize the probabilities

336 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
 337 to obtain the variance of N , $Var[N]$. The variables and parameters X_i , Y_i , P_i , $E[N]$
 338 and $Var[N]$ can be functions of time, and a time series analysis to study the temporal
 339 pattern of these variables and parameters can be implemented. If P_i 's, the host-
 340 parasite encounter probabilities, are statistically equal, we can assume a fixed P for
 341 each unit of time by taking the geometric mean of P_i ($i \geq 1$), that is, $P = (\prod_{i=1}^n P_i)^{1/n}$.

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

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

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

349 Here, we have determined the resulting distribution of parasites in the host popula-
 350 tion using homogeneous conditional probabilities of parasite acquisition. If the values
 351 of the acquisition probabilities become heterogeneous, such as if $P_1 < P_2 < \dots < P_n$,
 352 $P_1 > P_2 > \dots > P_n$ or P_i 's are completely arbitrary, then our derived formulas to es-
 353 timate $E[N]$ and $Var[N]$ are not applicable (e.g., Eq. (6)). However, our qualitative
 354 analysis to investigate the dynamics of Eqs. (3) and (4) could still hold, and a numeri-
 355 cal study would then be the feasible way forward. To find appropriate formulae (if data
 356 are not available) for $\widetilde{E[N]}$ and $\widetilde{Var[N]}$ would then be a challenge for the future.

357 Various future studies can stem from our model design. One can include the exis-
 358 tence of alternative, or intermediate hosts and vectors, especially that neglected tropi-
 359 cal diseases are common due to vector-borne macroparasite infections [2]. Extending
 360 our model is possible, to include how spatial aspects and different treatment strate-
 361 gies affect the population dynamics of the hosts and parasites, and to infer the rea-
 362 sons why the value of k is dynamic as observed in empirical studies [51, 52, 46, 53].
 363 Inclusion of multiple parasites [54, 55], different models of host-parasite interactions

364 in food webs [56, 57], or the explicit inclusion of population dynamics together with
365 host-parasite co-evolution [58, 59, 60, 61] are other possible directions. Testing if
366 such complexities retain the observed phenomenon of parasite aggregation would
367 then be a genuine test of the “law”.

368 **Acknowledgements**

369 JFR appreciates the time spent at the Max Planck Institute for Evolutionary Biology
370 developing the project. JFR and ELA are supported by the Institute of Mathematical
371 Sciences and Physics, University of the Philippines Los Baños. JFR is supported
372 by The Abdus Salam International Centre for Theoretical Physics Associate Scheme.
373 CSG is supported by the Max Planck Society.

374 **Declarations**

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

378 **Author Contributions**

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

382 **Data Availability**

383 The relevant codes for generating the figures are available at
384 <https://github.com/tecoevo/parasiteaggregation>

References

- [1] Zimmer, C. *Parasite Rex: Inside the Bizarre World of Nature's Most Dangerous Creatures* (Atria Books, 2001).
- [2] Hollingsworth, T. D. *et al.* Seven challenges for modelling indirect transmission: Vector-borne diseases, macroparasites and neglected tropical diseases. *Epidemics* **10**, 16–20 (2015).
- [3] Cundil, B. & Alexander, N. D. E. Sample size calculations for skewed distributions. *BMC Medical Research Methodology* **15** (2015).
- [4] Rozsa, L., Reiczigel, J. & Majoros, G. Quantifying parasites in samples of hosts. *The Journal of parasitology* **86**, 228–232 (2000).
- [5] Crofton, H. D. A quantitative approach to parasitism. *Parasitology* **62**, 179–193 (1971).
- [6] Gourbière, S., Morand, S. & Waxman, D. Fundamental factors determining the nature of parasite aggregation in hosts. *PLoS ONE* **10** (2015).
- [7] Poulin, R. Are there general laws in parasite ecology? *Parasitology* **134**, 763–776 (2007).
- [8] Shaw, D. J. & Dobson, A. P. Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S133 (1995).
- [9] Guilhem, R., Šimková, A., Morand, S. & Gourbière, S. Within-host competition and diversification of macro-parasites. *Journal of The Royal Society Interface* **9**, 2936–2946 (2012).
- [10] Basanez, M. & Boussinesq, M. Population biology of human onchocerciasis. *Philosophical Transactions of the Royal Society B* **354**, 809–826 (1999).

- 409 [11] Churcher, T. S., Ferguson, N. M. & nez, M.-G. B. Density dependence and
410 overdispersion in the transmission of helminth parasites. *Parasitology* **131**, 121–
411 132 (2005).
- 412 [12] Wilson, K. *et al.* Heterogeneities in macroparasite infections: Patterns and pro-
413 cesses. In Hudson, P., Rizzoli, A., Grenfell, B., Heesterbeek, H. & Dobson, A.
414 (eds.) *The Ecology of Wildlife Diseases*, chap. 2, 6–44 (Oxford University Press,
415 Oxford, 2001).
- 416 [13] Morrill, A. & Forbes, M. R. Random parasite encounters coupled with condition-
417 linked immunity of hosts generate parasite aggregation. *International Journal for*
418 *Parasitology* **42**, 701–706 (2012).
- 419 [14] Morrill, A. & Forbes, M. R. Aggregation of infective stages of parasites as an
420 adaptation and its implications for the study of parasite-host interactions. *The*
421 *American Naturalist* **187**, 000–000 (2016).
- 422 [15] Morrill, A., Forbes, M. R. & Dargent, F. Explaining parasite aggregation: more
423 than one parasite species at a time. *International Journal for Parasitology* **47**,
424 185–188 (2017).
- 425 [16] Fisher, R. A. The negative binomial distribution. *Annals of Eugenics* **11**, 182–187
426 (1941).
- 427 [17] Wilson, K., Grenfell, B. T. & Shaw, D. J. Analysis of aggregated parasite distribu-
428 tions: A comparison of methods. *Functional Ecology* **10**, 592–601 (1996).
- 429 [18] Anderson, R. M. The regulation of host population growth by parasitic species.
430 *Parasitology* **76**, 119–157 (1978).
- 431 [19] Adler, F. R. & Kretzschmar, M. Aggregation and stability in parasite-host models.
432 *Parasitology* **104**, 199–205 (1992).
- 433 [20] Anderson, R. M. & Gordon, D. M. Processes influencing the distribution of para-
434 site numbers within host populations with special emphasis on parasite-induced
435 host mortalities. *Parasitology* **85**, 373–398 (1982).

- 436 [21] Lopez, N. C. Parasitic crustaceans in fishes from some philippine lakes. In San-
437 tiago, C. B., Cuvin-Aralar, M. L. & Basiao, Z. U. (eds.) *Conservation and Eco-
438 logical Management of Philippine Lakes in Relation to Fisheries and Aquacul-
439 ture*, 75–79 (Southeast Asian Fisheries Development Center, Aquaculture De-
440 partment, Iloilo, Philippines; Philippine Council for Aquatic and Marine Research
441 and Development, Los Banos, Laguna, Philippines; and Bureau of Fisheries and
442 Aquatic Resources, Quezon City, Philippines, Philippines, 2001).
- 443 [22] Galvani, A. P. Immunity, antigenic heterogeneity, and aggregation of helminth
444 parasites. *Journal of Parasitology* **89**, 232–241 (2003).
- 445 [23] Wilber, M. Q., Johnson, P. T. J. & Briggs, C. J. When can we infer mechanism
446 from parasite aggregation? a constraint-based approach to disease ecology.
447 *Ecology* **98**, 688–702 (2017).
- 448 [24] Janovy, J. J. & Kutish, W. A model of encounters between host and parasite
449 populations. *Journal of Theoretical Biology* **134**, 391–401 (1988).
- 450 [25] Leung, B. Aggregated parasite distributions on hosts in a homogeneous environ-
451 ment: examining the poisson null model. *International Journal for Parasitology*
452 **28**, 1709–1712 (1998).
- 453 [26] Anderson, R. M. & May, R. M. Regulation and stability of host-parasite population
454 interactions: I. regulatory processes. *The Journal of Animal Ecology* **47**, 219–
455 247 (1978).
- 456 [27] Yakob, L. *et al.* Modelling parasite aggregation: disentangling statistical and eco-
457 logical approaches. *International Journal of Parasitology* **44**, 339–342 (2014).
- 458 [28] Rosà, R. & Pugliese, A. Aggregation, stability, and oscillations in different models
459 for host-macroparasite interactions. *Theoretical Population Biology* **61**, 319–334
460 (2002).
- 461 [29] McCallum, H. *et al.* Breaking beta: deconstructing the parasite transmission
462 function. *Philosophical Transactions of the Royal Society B* **372** (2017).

- 463 [30] Auld, S. K. J. R. & Tinsley, M. C. The evolutionary ecology of complex lifecycle
464 parasites: linking phenomena with mechanisms. *Heredity* **114**, 125–132 (2015).
- 465 [31] Viney, M. & Cable, J. Microparasite life histories. *Current Biology* **21**, R767–
466 R774 (2011).
- 467 [32] de la Cruz, C. P. P., Paller, V. G. V., Jr., M. Z. B. & Avila, A. R. B. Distribution
468 pattern of *Acanthogyrus* sp. (acanthocephala: Quadrigyridae) in nile tilapia (*Ore-
469 ochromis niloticus* L.) from sampaloc lake, philippines. *Journal of Nature Studies*
470 **12**, 1–10 (2013).
- 471 [33] Paller, V. G. V., Resurreccion, D. J. B., de la Cruz, C. P. P. & Bandal, M. Z. Acan-
472 thocephalan parasites (*Acanthogyrus* sp.) of nile tilapia (*Oreochromis niloticus*)
473 as biosink of lead (pb) contamination in a philippine freshwater lake. *Bulletin of*
474 *Environmental Contamination and Toxicology* **96**, 810–815 (2016).
- 475 [34] Gurarie, D., King, C. H. & Wang, X. A new approach to modelling schistosomiasis
476 transmission based on stratified worm burden. *Parasitology* **137**, 1951–1965
477 (2010).
- 478 [35] Gurarie, D., King, C. H., Yoon, N. & Li, E. Refined stratified-worm-burden models
479 that incorporate specific biological features of human and snail hosts provide
480 better estimates of schistosoma diagnosis, transmission, and control. *Parasites*
481 *& Vectors* **9** (2016).
- 482 [36] Gurarie, D. *et al.* Modelling control of schistosoma haematobium infection: pre-
483 dictions of the long-term impact of mass drug administration in africa. *Parasites*
484 *& Vectors* **8** (2015).
- 485 [37] Kretzschmar, M. Comparison of an infinite dimensional model for parasitic dis-
486 eases with a related 2-dimensional system. *Journal of Mathematical Analysis*
487 *and Applications* **176**, 235–260 (1993).

- 488 [38] Calabrese, J. M., Brunner, J. L. & Ostfeld, R. S. Partitioning the aggregation
489 of parasites on hosts into intrinsic and extrinsic components via an extended
490 poisson-gamma mixture model. *PLoS ONE* **6** (2011).
- 491 [39] Tallis, G. M. & M.K., L. Stochastic models of populations of helminthic parasites
492 in the definitive host. i. *Mathematical Biosciences* **4**, 39–48 (1969).
- 493 [40] Isham, V. Stochastic models of host-macroparasite interaction. *The Annals of*
494 *Applied Probability* **5**, 720–740 (1995).
- 495 [41] Duerr, H. & Dietz, K. Stochastic models for aggregation processes. *Mathematical*
496 *Biosciences* **165**, 135–145 (2000).
- 497 [42] Edelstein-Keshet, L. *Mathematical Models in Biology* (Society for Industrial and
498 Applied Mathematics, 2005).
- 499 [43] Gandon, S. & Day, T. Evolutionary epidemiology and the dynamics of adaptation.
500 *Evolution* **63**, 826–838 (2009).
- 501 [44] Shonkwiler, J. S. Variance of the truncated negative binomial distribution. *Journal of Econometrics* **195**, 209–210 (2016).
- 502
503 [45] Clark, C. W. *Mathematical Bioeconomics* (John Wiley & Sons Inc., 2010), third
504 edn.
- 505 [46] Pennycuik, L. Frequency distributions of parasites in a population of three-
506 spined sticklebacks, *gasterosteus aculeatus* L., with particular reference to the
507 negative binomial distribution. *Parasitology* **63**, 389–406 (1971).
- 508 [47] Wegner, K. M., Kalbe, M., Milinski, M. & Reusch, T. B. Mortality selection during
509 the 2003 european heat wave in three-spined sticklebacks: effects of parasites
510 and mhc genotype. *BMC Evolutionary Biology* **8**, 124 (2008).
- 511 [48] Shvydka, S., Sarabeev, V., Estruch, V. & Cadarso-Suarez, C. Optimum sample
512 size to estimate mean parasite abundance in fish parasite surveys. *Helminthologia* **55**, 52–59 (2018).
513

- 514 [49] Bolker, B. M. Probability and stochastic distributions for ecological modeling. In
515 *Ecological models and data in R*, 103–146 (Princeton University Press, Prince-
516 ton NJ, 2008).
- 517 [50] Young, L. J. & Young, J. H. A spatial view of the negative binomial parame-
518 ter k when describing insect populations. *Conference on Applied Statistics in*
519 *Agriculture* (1990).
- 520 [51] Scott, M. E. Temporal changes in aggregation: a laboratory study. *Parasitology*
521 **94**, 583–595 (1987).
- 522 [52] Boag, B., Lello, J., Fenton, A., Tompkins, D. & Hudson, P. Patterns of parasite
523 aggregation in the wild european rabbit (*Oryctolagus cuniculus*). *International*
524 *Journal for Parasitology* **31**, 1421 – 1428 (2001).
- 525 [53] Crompton, D. W. T., Keymer, A. E. & Arnold, S. E. Investigating over-dispersion;
526 *Moniliformis* (acanthocephala) and rats. *Parasitology* **88**, 317–331 (1984).
- 527 [54] Hafer, N. & Milinski, M. When parasites disagree: evidence for parasite-induced
528 sabotage of host manipulation. *Evolution* **69**, 611–620 (2015).
- 529 [55] Anzia, E. L. & Rabajante, J. Antibiotic-driven escape of host in a parasite-induced
530 red queen dynamics. *Royal Society Open Science* **5**, 180693 (2018).
- 531 [56] Lester, R. J. G. & R., M. Does moving up a food chain increase aggregation in
532 parasites? *Journal of the Royal Society Interface* **13**, 20160102 (2016).
- 533 [57] Jatulan, E. O., Rabajante, J. F., Banaay, C. G. B., Fajardo, A. C. F. J. & Jose, E. C.
534 A mathematical model of intra-colony spread of american foulbrood in european
535 honeybees (*Apis mellifera* L.). *PLoS ONE* **10**, e0143805 (2015).
- 536 [58] Gokhale, C. S., Papkou, A., Traulsen, A. & Schulenburg, H. Lotka-Volterra
537 dynamics kills the Red Queen: population size fluctuations and associated
538 stochasticity dramatically change host-parasite coevolution. *BMC Evolutionary*
539 *Biology* **13**, 254 (2013).

- 540 [59] Song, Y., Gokhale, C. S., Papkou, A., Schulenburg, H. & Traulsen, A. Host-
541 parasite coevolution in populations of constant and variable size. *BMC Evolu-*
542 *tionary Biology* **15**, 212 (2015).
- 543 [60] Rabajante, J. F. *et al.* Host-parasite red queen dynamics with phase-locked rare
544 genotypes. *Science Advances* **2** (2016).
- 545 [61] Cortez, M. J. V., Rabajante, J. F., Tubay, J. M. & Babierra, A. L. From epigenetic
546 landscape to phenotypic fitness landscape: Evolutionary effect of pathogens on
547 host traits. *Infection, Genetics and Evolution* **51**, 245–254 (2017).

Supplementary Material

On the mechanistic roots of an ecological law: parasite aggregation

Jomar F. Rabajante^{1,2,*}, Elizabeth L. Anzia¹ & Chaitanya S. Gokhale³

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

²Faculty of Education,
University of the Philippines Open University, 4031 Laguna, Philippines

³Research 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

*jfrabajante@up.edu.ph

SI.1 Supplementary Information

SI.1.1 List of state variable and parameters

Since we have

$$X_0 + (X_1 - X_2) + (X_2 - X_3) \dots + (X_n - X_{n+1}) + X_{n+1} = X$$

$$P_0 + (P_1 - P_2 P_1) + \left(\prod_{i=1}^2 P_i - \prod_{i=1}^3 P_i \right) + \dots + \left(\prod_{i=1}^n P_i - \prod_{i=1}^{n+1} P_i \right) + \prod_{i=1}^{n+1} P_i = 1,$$

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

$$E[N + 1] = 0P_0 + 1(P_1 - P_2 P_1) + 2 \left(\prod_{i=1}^2 P_i - \prod_{i=1}^3 P_i \right) + \dots + n \left(\prod_{i=1}^n P_i - \prod_{i=1}^{n+1} P_i \right) + w \prod_{i=1}^{n+1} P_i$$

The random variable $N + 1 \in \{0, 1, 2, \dots, n, w\}$ (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 $N \in \{0, 1, 2, \dots, n\}$) as discussed in the main text.

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 <i>at least</i> i number of parasites
$X_i - X_{i+1}, i \geq 1$	number of hosts having <i>exactly</i> 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; for simplicity, $c = 1$
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

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 $N \in \{0, 1, 2, \dots, n\}$ (living hosts) and $N + 1 \in \{0, 1, 2, \dots, n + 1\}$ (living and dead hosts), respectively
$Var[N]$ and $Var[N + 1]$	variance of the random variable $N \in \{0, 1, 2, \dots, n\}$ and $N + 1 \in \{0, 1, 2, \dots, n + 1\}$, respectively
$\widetilde{E[N]}$	pseudo-mean 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
$\widetilde{Var[N]}$	pseudo-variance 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

15 **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,

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

Then taking the derivative of this geometric series results in

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

16 The expression for A is:

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

17 **SI.1.3 Explicit formula for the variance**

18 Let us define $\widetilde{Var}[N] = Var[N+1] - (n+1 - E[N+1])^2 \prod_{i=1}^{n+1} P_i$ where $Var[N+1] =$
 19 $E[(N+1)^2] - E[N+1]^2$ 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

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

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

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

Multiplying both sides by P , we have

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

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

$$1 + 4P + \dots + n^2 P^{n-1} = B = \frac{-n^2 P^{n+2} + (2n^2 + 2n - 1) P^{n+1} - (n + 1)^2 P^n + P + 1}{(1 - P)^3}.$$

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

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

21 **Sl.1.4 Proof for a claim in Section "Constant host-parasite en-** 22 **counter probability: Population dynamics" in the main text**

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

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

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:

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

23 However, $P^n = \frac{P^0 + P^1 + \dots + P^{n-1}}{n}$ is the arithmetic average of $\{P^{n-1}, P^{n-2}, \dots, P^1, P^0\}$.

24 This is a contradiction since $P^n \notin \{P^{n-1}, P^{n-2}, \dots, P^1, P^0\}$.

25 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
26 $Y_1^* = \frac{nK+1}{p}$. But $Y_1^* = 0$ and $p < \infty$, a contradiction.

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

28 **Sl.1.5 Histograms given different values of P , $n + 1 = 10$**

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 in the main text).

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