

# Fundamental Factors Determining the Nature of Parasite Aggregation in Hosts

Sébastien Gourbière, Serge Morand, David Waxman

► **To cite this version:**

Sébastien Gourbière, Serge Morand, David Waxman. Fundamental Factors Determining the Nature of Parasite Aggregation in Hosts. PLoS ONE, Public Library of Science, 2015, <10.1371/journal.pone.0116893>. <hal-01214507>

**HAL Id: hal-01214507**

**<https://hal-univ-perp.archives-ouvertes.fr/hal-01214507>**

Submitted on 13 Oct 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



RESEARCH ARTICLE

# Fundamental Factors Determining the Nature of Parasite Aggregation in Hosts

Sébastien Gourbière<sup>1,2</sup>, Serge Morand<sup>3</sup>, David Waxman<sup>4\*</sup>

**1** Université de Perpignan Via Domitia, EA 4218, 'Institut de Modélisation et d'Analyse en Géo-Environnements et Santé' (IMAGES), Perpignan 66100, France, **2** University of Sussex, School of Life Sciences, Centre for the Study of Evolution, Brighton BN1 9QG, UK, **3** Institut des Sciences de l'Evolution—CNRS, Département Génétique Environnement, CC065, Université Montpellier 2, 34095, Montpellier cedex 05, France, **4** Centre for Computational Systems Biology, Fudan University, Shanghai 200433, PRC

\* [davidwaxman@fudan.edu.cn](mailto:davidwaxman@fudan.edu.cn)



## Abstract

The distribution of parasites in hosts is typically aggregated: a few hosts harbour many parasites, while the remainder of hosts are virtually parasite free. The origin of this almost universal pattern is central to our understanding of host-parasite interactions; it affects many facets of their ecology and evolution. Despite this, the standard statistical framework used to characterize parasite aggregation does not describe the processes generating such a pattern. In this work, we have developed a mathematical framework for the distribution of parasites in hosts, starting from a simple statistical description in terms of two fundamental processes: the exposure of hosts to parasites and the infection success of parasites. This description allows the level of aggregation of parasites in hosts to be related to the random variation in these two processes and to true host heterogeneity. We show that random variation can generate an aggregated distribution and that the common view, that encounters and success are two equivalent filters, applies to the average parasite burden under neutral assumptions but it does not apply to the variance of the parasite burden, and it is not true when heterogeneity between hosts is incorporated in the model. We find that aggregation decreases linearly with the number of encounters, but it depends non-linearly on parasite success. We also find additional terms in the variance of the parasite burden which contribute to the actual level of aggregation in specific biological systems. We have derived the formal expressions of these contributions, and these provide new opportunities to analyse empirical data and tackle the complexity of the origin of aggregation in various host-parasite associations.

## OPEN ACCESS

**Citation:** Gourbière S, Morand S, Waxman D (2015) Fundamental Factors Determining the Nature of Parasite Aggregation in Hosts. PLoS ONE 10(2): e0116893. doi:10.1371/journal.pone.0116893

**Academic Editor:** Carolina Barillas-Mury, NIAID, UNITED STATES

**Received:** April 1, 2014

**Accepted:** December 16, 2014

**Published:** February 17, 2015

**Copyright:** © 2015 Gourbière et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are contained within the paper.

**Funding:** This investigation received financial support from: the European Union EU-FP7-PEOPLE-Intra-European Fellowship for career development (IEF) grant no. 253483; a 'Bonus Quality Research' grant of the Université de Perpignan Via Domitia; an "Investissements d'Avenir" grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01). This work was performed within the framework of the LABEX ECOFECT (ANR-11-LABX-0048) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency

## Introduction

A fundamental aspect of the relationship between parasites and hosts is contained in the distribution of parasites amongst hosts. This distribution has repeatedly been shown to be clustered or aggregated (see [1–3] for reviews) in the sense that typically, a few hosts harbour many

(ANR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

parasites, while the remainder of the hosts are virtually parasite free. Exceptions to this pattern are so rare that aggregation is considered a fundamental aspect of the definition of parasitism [4] or has been described as the ‘First Law of Parasitism’ [5].

Aggregation has very significant implications for both hosts and parasites, since it affects their genetics and evolution, and has been recognised to have many consequences for public health and livestock management. Aggregation has been shown to affect parasite ecology by stabilizing host-parasite population dynamics [6, 7] and facilitating interspecific co-infection as a result of increased host susceptibility [8]. Aggregation also influences parasite evolution by, e.g., increasing the level of intra-specific competitive interaction and the rate of within-host adaptive parasite diversification [9]. Accordingly, aggregation of parasites amongst hosts affects the transmission of infectious human diseases [10]. Risk heterogeneity typically leads to an increase in  $R_0$  of vector-borne diseases by a factor of 2–4, and has even larger impact on the transmission of sexually-transmitted diseases, which implies that the proportion of the population that must be protected for elimination via untargeted control program is usually much higher than expected [11]. A quantitative understanding of the mechanisms which lead to the observed levels of aggregation is thus essential for our knowledge of parasite ecology and evolution. One of the most fundamental issues in the field is to what extent differences in parasite loads reflect differences in the exposure of hosts to the infective stages of a parasite, or differences in the success of a parasite in infecting its hosts (see [12], p198).

The distribution of individual parasites amongst hosts is commonly described by a negative binomial distribution since this distribution is a flexible discrete distribution with two parameters [13] whose form naturally accommodates aggregation. However, the negative binomial distribution only provides a phenomenological description of aggregation. In particular, its parameters are not explicitly linked to any representation of the processes underlying exposure of hosts to parasites and the success of the parasites in infecting the hosts. In the present work we introduce a framework, with mechanistic underpinnings, that allows a formal link between the pattern of parasites aggregation amongst hosts, and the processes that lead to this feature.

We proceed by classifying the various processes that are potentially involved in producing the distribution of parasites amongst hosts into two types: (i) the number of encounters between a host and a parasite or a source of parasites, and (ii) the number of parasites that are ultimately carried by a host, which result from a single encounter with a parasite or a source of parasites. These two processes, henceforth referred to as ‘encounters’ and ‘success’ according to Combes conceptual framework (see [12], p. 199), are described by two distributions which combine to produce the overall distribution of the number of parasites in hosts. Because this framework is established with no assumption about the overall distribution of the number of parasites amongst hosts, it gives the ability to test and explore, in a highly flexible way, the relative importance of encounters and success in generating the parasite distribution.

A common assumption in the field is that heterogeneity in hosts generates an aggregated distribution of parasites amongst hosts. Such heterogeneity can be induced by different social or sexual host behaviours that modulate the rate of encounters between susceptible and infected hosts/vectors, or that alter parasite success of infection because of, e.g., allo-grooming (see [14] and [15] for reviews). Variation in physiological status, such as variation in host condition due to nutritional stress, can lead to variations in infections [16, 17]. Finally, habitat difference can generate different risk of contact with the parasite. Typically, inhabitants leaving in the outer part of villages can be more exposed to zoonoses that are transmitted by vectors dispersing from sylvatic areas [18–20]. Thus, without denying the existence of processes that generate heterogeneity in encounters and success, we first question the need for these implicit assumptions and ask: can an aggregated parasite distribution arise without any intrinsic heterogeneity, and simply follow from random variation associated with encounters and success?

Given this question we can go further and ask: does randomness in encounters have the same consequences for the parasite distribution as randomness in success? We shall begin the analysis under the very simple assumptions that all hosts are equivalent, and that all parasites are equivalent. This constitutes a neutral model [21] where the parasite distribution in hosts results from random variation in both encounters and success. We then go a step further, and introduce intrinsic differences between hosts, to work out the implications of actual host heterogeneity, which then combines with random variation in encounters and success.

## Results

### Modelling of the parasite distribution

#### Neutral model

The key quantity in this problem is the number of parasites in (or associated with) a host, which we denote by  $N$ . Since the number of parasites varies from host to host,  $N$  is a random variable and our model describes the statistics of  $N$ . Assuming there is no vertical transmission of parasites, the essence of the model is that a host is born parasite free and then has a random number of encounters with a parasite/source of parasites. An encounter is characterised by its success, as measured by the number of parasites that are *ultimately* carried by the host, due to the encounter.

We denote the number of encounters of a host with a parasite/source of parasites by  $\mathcal{E}$ , and the success of the  $j$ 'th encounter by  $S_j$  where  $j = 1, 2, \dots, \mathcal{E}$ . The number of parasites in a host,  $N$ , is then given by a sum of successes of the different encounters of the host with a parasite/source of parasites:

$$N = \sum_{j=1}^{\mathcal{E}} S_j. \quad (1)$$

We take the  $S_j$  and  $\mathcal{E}$  to be independent random variables whose possible values are 0, 1, 2, ... and the value of the sum is taken to zero if  $\mathcal{E}$  equals zero. The  $S_j$  are all taken to have the same probability distribution, and hence have identical expected values and identical variances, which we write as  $E[S]$  and  $\text{Var}(S)$ , respectively.

Equation (1) constitutes a fairly general viewpoint for the infection of hosts by a single parasite species. While Eq. (1) looks simple, this appearance is deceptive;  $N$  is a quantity which is composed of a random number ( $\mathcal{E}$ ) of random variables (the  $S_j$ ) and hence constitutes a *compound random variable* [22], which requires two probability distributions to characterise it. A number of different biological scenarios may be considered for Eq. (1). The variable,  $\mathcal{E}$ , can represent the number of independent encounters of a given host with individual parasites. Alternatively,  $\mathcal{E}$  can represent the number of encounters of a host with sources of parasites such as: infected sites (for air, soil or water-borne diseases), infected vectors (for vector-borne diseases) or infected hosts for directly transmitted parasites. Similarly, the variable  $S_j$  could represent the viability of a single parasite, and hence would only take two values, namely 0 or 1, corresponding to death or survival of the parasite. However, the variable  $S_j$  could also account for multiple infections/contacts and/or within host parasite viability and reproduction, in which case  $S_j$  is capable of taking values larger than 1, and the discrete distribution it follows would reflect this feature.

It is worth mentioning that the causes of the random variability in success can be associated with the parasite or with the host. Such variability can be due to a finite random sampling of parasite diversity in each host, just as it can reflect random variability in host resistance or condition. But, whatever the definition of  $\mathcal{E}$  and  $S_j$  and their causes, the key assumption of this

neutral model is that the levels of random variability in encounters and success are the same for all hosts. In other words, encounters and success are realizations of random variables whose distribution is common to all individuals, and hosts are thus ‘statistically equivalent’. Accordingly, there is no intrinsic difference between hosts and these variations can be seen as a form of demographic stochasticity, whose effects on parasite distribution can be evaluated from the neutral model.

### Heterogeneous model

The above model can now be expanded to incorporate host heterogeneity, by allowing different types of hosts, labelled  $t = 1, 2 \dots h$ . Consider a randomly captured host of type  $t$ . The number of parasites in the host is then related to the following: (i) the number of encounters, specific to a host of type  $t$ , that the host has with a parasite/source of parasites, and which we write as  $\mathcal{E}_t$ , and (ii) the success that is associated with each encounter of a host of type  $t$ , which we write as  $S_j^{(t)}$ . The number of parasites in a randomly captured host can be written as

$$N = \sum_{t=1}^h H_t \sum_{j=1}^{\mathcal{E}_t} S_j^{(t)}. \tag{2}$$

In this equation the  $H_t$  indicate the type of host captured. Only one of the  $H_t$ 's takes the value of unity, while the remainder are zero, and the probability with which  $H_t = 1$  is  $p_t$ . Thus explicitly,  $(H_1, H_2, \dots, H_h)$  constitutes a multinomial random variable with parameters 1 and  $(p_1, p_2, \dots, p_h)$ .

As in the neutral model, there are various alternative biological scenarios that can lead to heterogeneity in the rate at which different hosts encounter parasites. The rate of encounters can vary with the level of host foraging activities [23, 24], the characteristics of the host habitat [25–27], the spatial and temporal co-distributions of hosts and vectors [28–31]. Similarly, the rate of parasite success in the host can depend on the individual level of immunity [32–34] or physiological status [35–38].

### Results for the specific models

We shall establish general results for some summary statistics of the distribution of parasites in hosts, that emerges from the neutral and heterogeneous models introduced above.

#### Neutral model

All results given below, for the neutral model, are established in [Methods A](#)

We begin with [Eq. \(1\)](#), which relates the number of parasites of an individual host,  $N$ , to the number of encounters that they have with a parasite/source of parasites,  $\mathcal{E}$ , and the ultimate success of these parasites on different encounters, i.e., the  $S_j$ . A direct consequence of [Eq. \(1\)](#) combined with the assumptions made above is that the expected value of the number of parasites of a host,  $E[N]$ , is related to the expected number of encounters,  $E[\mathcal{E}]$ , and the expected level of success,  $E[S]$ , as

$$E[N] = E[\mathcal{E}]E[S]. \tag{3}$$

Additionally, the relationship between the variance of  $N$  and the various statistics of  $\mathcal{E}$  and  $S_j$  can be shown to be

$$\text{Var}(N) = \text{Var}(\mathcal{E})(E[S])^2 + \text{Var}(S)E[\mathcal{E}]. \tag{4}$$

Equations (3) and (4) are standard results for compound random variables (see e.g., [22]), and can be found in Methods A, as part of the analysis.

It is important to point out that Eqs. (3) and (4) apply for any distributions of  $\mathcal{E}$  and  $S_j$ , and hence for any distribution of  $N$  that results from Eq. (1). In particular, Eqs. (3) and (4) apply whether or not  $N$  has a negative binomial distribution. This is important since, using our framework, it can be shown that conditions need to be satisfied, by the distributions of encounters and success, in order to obtain a negative binomial distribution (see Methods A). These conditions appear to be rather specific for the distribution of successes, as we show in the second example in Methods A.

As expected, the mean number of contacts and successes combine in a symmetric way to produce the mean number of parasites in a host [12]. However, the asymmetric way that the summary statistics of encounters and success enter the result for the variance in the number of parasites in a host, Eq. (4), means that these two processes cannot be regarded as having equivalent effects on the variance.

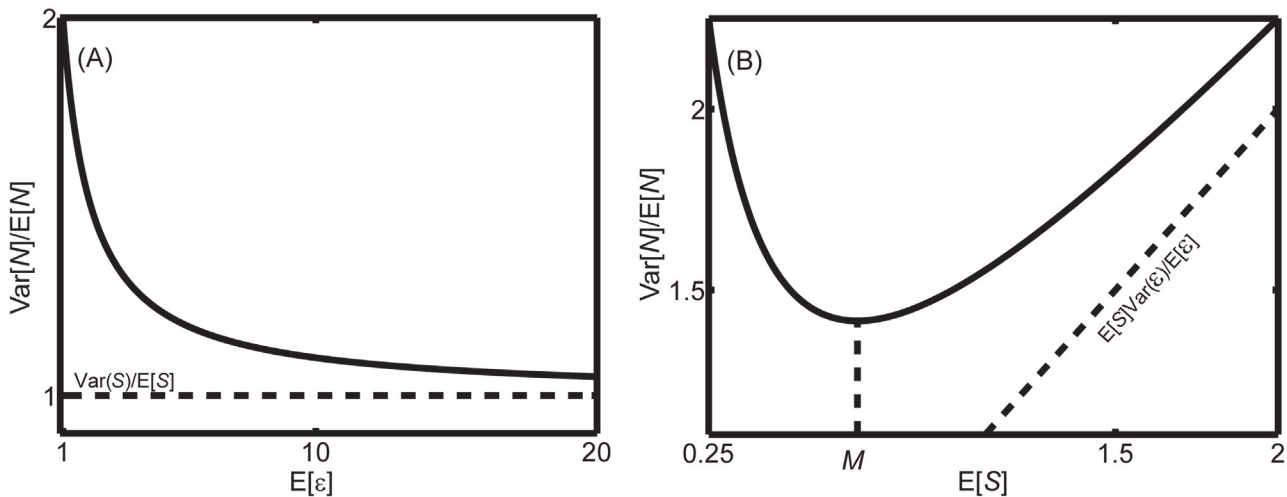
We note that the variance of  $N$  depends on the variance of the number of contacts and the variance of success, however, these variances are weighted differently in Eq. (4), namely by the squared mean of success and the mean of contact, respectively. Such an asymmetry of the weightings allows the distribution of parasites in hosts to take a variety of forms. We thus used a measure of aggregation that is applicable to a general distribution, namely the variance-to-mean-ratio of the number of parasites in hosts. Values of the variance-to-mean-ratio that are greater than 1 are associated with aggregation [39]. Using Eqs. (3) and (4) this ratio can be expressed as the sum of two terms, involving the variance-to-mean-ratio in encounters and success:

$$\frac{\text{Var}(N)}{E[N]} = \frac{\text{Var}(\mathcal{E})}{E[\mathcal{E}]} E[S] + \frac{\text{Var}(S)}{E[S]}. \tag{5}$$

The two contributions in this equation, of the variance-to-mean-ratios of encounters and success, are weighted differently; the average success multiplies the contribution of the variance-to-mean-ratio in encounters. The simplicity of this quantitative outcome of our neutral model leads to two general insights into the emergence of aggregation from random variation associated with encounters and success. First, aggregation decreases with the mean number of contacts,  $E[\mathcal{E}]$ , until the level of aggregation reaches an asymptotic level of  $\text{Var}(S)/E[S]$ , and only reflects randomness in success (Fig. 1A). Second, the mean level of success,  $E[S]$ , can have a more complex effect on aggregation than the mean number of contacts because of its presence in two places in Eq. (5) (Fig. 1B).

Any increase in  $E[S]$  (at a fixed value of  $\text{Var}(S)$ ) decreases the variance-to-mean-ratio of success, and thus lowers the level of aggregation. However, increasing  $E[S]$  simultaneously increases the contribution of randomness in encounters by magnifying the differences between hosts that had low and high levels of contacts with parasites. The balance between these two antagonistic effects leads to an intermediate level of average success of  $E[S] = \sqrt{\text{Var}(S)E[\mathcal{E}]/\text{Var}(\mathcal{E})}$  that corresponds to a lowest level of aggregation of  $2\sqrt{\text{Var}(S)\text{Var}(\mathcal{E})/E[\mathcal{E}]}$ , while at larger  $E[S]$  the level of aggregation asymptotically changes as  $E[S]\text{Var}(\mathcal{E})/E[\mathcal{E}]$ .

Although we aim at providing a mechanistic alternative to the negative binomial distribution to describe parasite distributions, it is worth mentioning that the standard measure of aggregation, the parameter  $k$  of the negative binomial distribution, can be expressed in terms of the summary statistics of  $\mathcal{E}$  and  $S_j$  (see Methods A). This could be used when the specific conditions for the distribution to be a negative binomial distribution are fulfilled (see Methods A).



**Figure 1.** This figure illustrates how aggregation varies with either host-parasite encounters or parasite success (in infecting hosts). In Panel A we have adopted the values  $\text{Var}(S)/E[S] = 1$  and  $\text{Var}(\mathcal{E})E[S] = 1$ . We show how parasite aggregation varies with the mean number of encounters. The level of aggregation decreases with the number of encounters, and asymptotically approaches a value that depends only on the variance-to-mean ratio of parasite success, i.e.,  $\text{Var}(S)/E[S]$ . In Panel B we have adopted the values  $\text{Var}(\mathcal{E})/E[\mathcal{E}] = 1$  and  $\text{Var}(S) = 0.5$ . We show how parasite aggregation varies with the mean parasite success in infecting hosts. The level of aggregation initially decreases with the average success of parasites in infecting their hosts until a minimum is reached at a value of  $E[S]$  of  $M = \sqrt{\text{Var}(S)E[\mathcal{E}]/\text{Var}(\mathcal{E})}$ , as indicated on the abscissa. The level of aggregation then starts to increase, with an asymptotically achieved slope that is directly proportional to the variance-to-mean-ratio of encounters, i.e.,  $\text{Var}(\mathcal{E})/E[\mathcal{E}]$ .

doi:10.1371/journal.pone.0116893.g001

### Heterogeneous model

All results given below, for the heterogeneous model, are established in [Methods B](#).

The above results can be expanded to multiple types of hosts by using [Eq. \(2\)](#) that relates the number of parasites of an individual host,  $N$ , to random variables associated with encounters and success, for different host types. The mean and variance of the number of encounters are then conditional on the host type, and we write these as  $E[\mathcal{E}|t]$  and  $\text{Var}(\mathcal{E}|t)$ , respectively. Similarly, we denote the mean and variance of the success of a parasite to infect hosts of type  $t$  by  $E[S|t]$  and  $\text{Var}(S|t)$ , respectively. We find that the expected number of parasites in a host is a weighted average over contributions from different possible host types:

$$E[N] = \sum_{t=1}^h p_t E[\mathcal{E}|t] E[S|t] \tag{6}$$

(see [Methods B](#) for details) where  $p_t$  denotes the frequency of host type  $t$ . Thus, the mean number of parasites in a host depends on the mean number of encounters and success in each host type and on the way these are correlated across host types. Intuitively, we expect the total number of parasites to be larger when the hosts with the highest rate of encounters with parasites are the more susceptible ones. This appears more obviously in an alternative expression for  $E[N]$  that is completely equivalent to [Eq. \(6\)](#), namely

$$E[N] = \mu m + \text{Cov}(\mathcal{E}, S). \tag{7}$$

Here  $\mu = \sum_{t=1}^h p_t E[\mathcal{E}|t]$  and  $m = \sum_{t=1}^h p_t E[S|t]$  represent the means of encounters and successes, across different host types, while  $\text{Cov}(\mathcal{E}, S) = \sum_{t=1}^h p_t (E[\mathcal{E}|t] - \mu)(E[S|t] - m)$  measures how encounter and success are correlated with each other across different host types. [Equation \(7\)](#) clearly shows that the mean number of parasites is linearly related to the covariance between encounters and success.

A negative correlation between encounters and infection success can emerge as a result of processes related to host group size. The rate of encounter with directly transmitted parasites is positively correlated with the size of the group in many host species [40]. But at the same time, infection success of ectoparasites may decrease with the number of individuals through enhanced allo-grooming behaviour [31, 41]. Alternatively, a positive correlation can result from variation associated with home range. A small home range size in mammals is usually associated with increased host densities that favours parasite encounters and simultaneously affects host condition, which has a positive effect on infection success [16].

The variance of the distribution of parasites in hosts,  $\text{Var}(N)$ , that incorporates random individual variation and actual heterogeneity in contacts and success can also be derived (see [Methods B](#)) and is given by

$$\begin{aligned} \text{Var}(N) = & \sum_{t=1}^h p_t \left[ \text{Var}(\mathcal{E}|t)(E[S|t])^2 + E[\mathcal{E}|t]\text{Var}(S|t) \right] \\ & + \sum_{t=1}^h p_t \left( E[\mathcal{E}|t]E[S|t] - \sum_{u=1}^h p_u E[\mathcal{E}|u]E[S|u] \right)^2. \end{aligned} \tag{8}$$

This result indicates that there are contributions to  $\text{Var}(N)$  from individual variation that is within host types (first sum), and also from effects of heterogeneity between host types (second sum).

The expression in [Eq. \(8\)](#) can be approximated, to provide a more explicit relationship between the variance in parasite numbers in hosts and various correlations involving encounters and success. We achieve this by determining the leading corrections to the results of the homogeneous model (see [Methods B](#) for details). This leads to an equation for the variance in the number of parasites in a host of the form

$$\text{Var}(N) \approx \sigma^2 m^2 + \mu v^2 + \text{quadratic correlation terms} \tag{9}$$

where  $\sigma^2 = \sum_{t=1}^h p_t \text{Var}(\mathcal{E}|t)$  and  $v^2 = \sum_{t=1}^h p_t \text{Var}(S|t)$  stand for means, across all host types, of the variances of encounter and success. The leading two terms in [Eq. \(9\)](#) are equivalent to the terms in the homogeneous model, [Eq. \(4\)](#), so that conclusions on the asymmetric contributions of the levels of contact and success remain, although they now apply to averages across host types. Obviously, the quadratic correlation terms can potentially make the whole expression of [Eq. \(9\)](#) substantially more complicated, as one should expect from the level of complexity incorporated into the heterogeneous model. However, these additional terms can be fully identified from [Eqs. \(B3\) and \(B9\) of Methods B](#), and they are related to the variances and covariances of the basic summary statistics (mean and variance) describing the distribution of encounters and success conditional on host types (see [Eq. \(B10\) in Methods B](#)).

The variance-to-mean-ratio of the number of parasites amongst hosts can, using [Eqs. \(7\) and \(9\)](#), be approximated as

$$\frac{\text{Var}(N)}{E[N]} = \frac{\sigma^2 m^2 + \mu v^2}{\mu m} + \text{quadratic correlation terms} \tag{10}$$

where, as above, the quadratic correlation terms take into account leading effects of between-type correlations. Thus again, the leading terms in [Eq. \(10\)](#) on the right hand side are equivalent to the results of the homogeneous model ([Eq. \(5\)](#)), although they now apply to means across host types. Accordingly, the highest levels of aggregation are expected when strong levels of random individual variation in encounters ( $\sigma^2$ ) are associated with high average rates of



success ( $m$ ) as the latter magnifies the former. Similarly, aggregation increases when there is large random individual variation in success ( $v^2$ ) are associated with high average rates of encounters ( $\mu$ ). It is not obvious how to identify host-parasite systems that could serve as examples of such co-variations since random individual variation in encounters and success are typically not measured in the field. However, there are a few published experimental results that suggest that rodents exposed to water-borne, air-borne or even soil-transmitted parasites could be good candidates for measurement of the level of such demographic stochasticity (see [discussion](#)). As expected, parasites will also be strongly aggregated when hosts with a high average rate of encounters are those on which parasites are successful, since this would lead to high mean parasite load and a lower variance to the mean ratio.

As explained above, such a positive correlation could be associated with a small host home range, while a negative correlation is expected in host groups of large size.

The quadratic correlation terms of [Eq. \(10\)](#) are directly related to those appearing in [Eq. \(B12\)](#) of [Methods B](#), and thus also correspond to variances and covariances of basic summary statistics of the distributions of encounters and success. These expressions make up an explicit link between the patterns of aggregation and basic statistics on the level of host random individual variation and actual heterogeneity in the processes of encounters and success. Although no general results can be derived from these quadratic correlations because of their remaining complexity, [Eq. \(9\)](#) and [\(10\)](#) provide the theoretical background necessary to gain a better knowledge of the parasite distribution, if combined with a good empirical knowledge of the undoubtedly system-specific correlations between encounters and success.

## Discussion

Aggregation of parasites amongst hosts is one of the rare phenomena in biology that has been described as a 'law' [[1](#), [5](#)]. The negative binomial distribution has been of central importance to establish evidence for aggregation in data [[5](#), [42](#)], as well as to provide theoretical predictions on ecological and evolutionary consequences of parasite distributions among hosts [[6](#), [9](#), [43](#)]. However, a good fit to a negative binomial distribution does not provide any information about the causes of aggregation since a number of different biological phenomena have been proposed to generate this flexible distribution [[44](#)], as well as different combinations of statistical laws [[13](#)]. A good fit cannot then be interpreted as support of a unique hypothesis about any underlying mechanisms. In the present work, we have introduced and developed simple mechanistic models of parasite aggregation, which are taken to originate from distributions corresponding to the two main types of factors thought to generate aggregation: (i) the encounters between host and parasites, and (ii) the success of parasites once in contact with the host [[12](#)]. This provides a biologically intuitive mathematical framework to quantitatively investigate the importance of random individual variation and actual heterogeneities in encounters and success on parasite aggregation. Here we discuss the main outcomes of the model and the way they could be tested empirically.

A first outcome of the present work is that random individual variation can potentially produce an aggregated distribution of parasites amongst hosts. Such prediction is consistent with previous empirical and theoretical finding, where causes other than intrinsic heterogeneity amongst hosts generate aggregated distributions of parasites. The rate of mice infection by blacklegged ticks depends substantially on 'bad luck', i.e. inhabiting a home range with high vector density [[28](#)]. Aggregated distributions have also been shown to emerge when parasites are homogeneously distributed in the environment, provided that the probability of infection is related to the distance between the host and the source of parasites [[45](#)]. Additionally, clumped infections, i.e. infections of several larvae at the same time, have a strong impact on

the level of aggregation [46]. We finally note that similar aggregated distributions of parasites [47] and parasitoids [48] have been generated using compound distributions representing other more specific demographic processes. One cannot thus a priori rule out the null hypothesis that random variation is partly responsible for the observed levels of aggregation (see [1–3] for reviews), and the causal relationships between sources of true heterogeneity amongst hosts and the observed distribution of parasites should thus be better quantified by taken into account random individual variation in encounters and success. An obvious need is to evaluate the level of ‘neutral aggregation’ that can be explained in absence of actual host heterogeneity. Our neutral model allows for such evaluation provided that it can be parameterized from independent experimental assessments of the levels of random individual variation in encounters and success. Dose-infectivity curves derived from artificial infection experiments could allow the measurement of variability in infection success in the same way as they have been used to assess the distribution of host susceptibility [49–51]. Some of these authors [51] then used a modelling framework that assumed a constant number of encounters with parasites, and a flexible statistical distribution to describe host variation in the rate of parasite acquisition. Fitting this relationship to dose-infectivity profile they obtained the maximum likelihood estimates of the parameters of the host susceptibility distribution. Setting up similar experiments with all host individuals originating from a single type, e.g. isogenic mice, the outcome would provide a distribution that gives a measure of random individual variation in parasite success. To estimate random individual variation in encounters may be more challenging, but could be investigated whenever hosts can be kept in parasite free environments and the rate of exposure controlled experimentally. For instance, individually marked rainbow trout were introduced at regular time intervals into cages so that the rate of exposure to trematode parasites, under natural conditions, could be controlled [52]. Similar experimental designs are conceivable for rodents that can be breed and exposed at controlled rate to water-borne (such as the fluke *Schistosoma mansoni*, [53]), to food borne (such as the acanthocephalan *Moniliformis moniliformis* through the consumption of infected prey, [54]) or even soil-transmitted parasites (such as *Nippostrongylus brasiliensis*, [55]). Interestingly, our results also suggest that to assess individual variation in one of these two components (encounter or success), the experimental design should be set up with a high average level of the other component. We note that when the average number of encounters is large, the level of parasite aggregation converges toward the level of random variation in success ( $\text{Var}(N)/E[N] \sim \text{Var}(S)/E[S]$ ). Similarly, when the level of parasite success is large, the level of parasite aggregation is linearly related to the level of random variation in contacts ( $\text{Var}(N)/E[N] \sim \text{Var}(\mathcal{E})E[S]/E[\mathcal{E}]$ ). Under these conditions, the desired quantities could thus be more easily assessed.

A second significant outcome of our model is in providing simple predictions in the relative effect of randomness in encounters and success on the parasite distribution. Although the mean values of encounter and success combine multiplicatively to give the mean number of parasites in hosts, as expected when we take the view that the ‘filter’ of encounter and the ‘filter’ of compatibility are equivalent ‘apertures’ that limit the acquisition of parasites by hosts [12], the latter representation is misleading when trying to understand the variability of infection and thus aggregation. As a consequence of the natural sequential order in which encounter and success contribute to determine the parasite load of a host, the two ‘filters’ are not equivalent in their effects on the variance-to-mean-ratio in the number of parasites in hosts. While aggregation decreases with the average number of host-parasite encounters and converges towards  $\text{Var}(S)/E[S]$ , it varies in a non-linear way with the average parasite success in hosts. This results in an intermediate level of success, equal to  $\sqrt{E[\mathcal{E}]\text{Var}(S)/\text{Var}(\mathcal{E})}$ , leading to the lowest level of aggregation. Such a prediction could be tested in the above experimental settings by controlling the level of average

success in hosts through the use of various strains of parasites [56], manipulation of the physiological status of the host [57] or change in its level of immunity [58].

Importantly, the asymmetry between the effect of encounters and success is also apparent in the contributions of actual host heterogeneity. Although some general insights can be drawn from our results, effects are then more complex and hard to anticipate without a specific system at hand. Investigation of those effects should be started on very simple systems with a known (and simple) source of heterogeneity, like resistant, sensitive or even tolerant host genotypes [59]. In such context, and in an experimental set up similar to those discussed above, one would be able to measure the host type specific average rates of contact and success as well as the level of random variation in success and contact for each type. This could provide estimates of terms appearing in Eqs. (9) and (10) would give the opportunity to handle standard sensitivity or elasticity analysis to clarify the contributions of random individual variation and actual differences between hosts in generating aggregation in specific systems.

To conclude, the mechanistic framework that we have developed in this paper formally links the pattern of parasite aggregation to random variation and heterogeneities in the processes of host-parasite encounters and the success of infection. This simple but flexible framework should improve our ability to gain a better understanding of the origins and implications of aggregation in parasite ecology, evolution and the control of infectious diseases.

## Methods

### A. Derivation of results for the neutral model

In this subsection we give a derivation of results for the neutral model. We first determine general results (Eqs. (3) and (4) of the main text) before showing particular results related to the use of the negative binomial distribution. We note that there may sometimes need to be constraints on the distributions of  $\mathcal{E}$  and  $S_j$ .

#### General results

We begin with a discrete random variable  $N$ , which takes the values 0, 1, 2, .... The *probability generating function* for  $N$  is defined by  $G_N(\lambda) = E[\lambda^N] = \sum_{n=0}^{\infty} \text{Prob}(N = n)\lambda^n$  where  $E[\dots]$  denotes an expected (or average value) and  $\lambda$  is a real variable. All information about the distribution of  $N$  is contained in  $G_N(\lambda)$  [22]. Henceforth we shall refer to probability generating functions simply as *generating functions*.

We take  $N$  to represent the number of parasites in a randomly picked host, as given by Eq. (1) of the main text, namely  $N = \sum_{j=1}^{\mathcal{E}} S_j$  where  $\mathcal{E}$  is the number of exposures of a host to a parasite/source of parasites and  $S_j$  is the ultimate success of a parasite on the  $j$ 'th exposure. The value of  $N$  is taken to be 0 if  $\mathcal{E} = 0$ . We assume  $\mathcal{E}$  and the  $S_j$  are all independent random variables that can take the values 0, 1, 2, .... We take all  $S_j$  to have identical distributions, with expected value  $E[S]$  and variance  $\text{Var}(S)$ . The generating function of  $N$ , that follows from Eq. (1), is [22]

$$G_N(\lambda) = G_{\mathcal{E}}(G_S(\lambda)). \tag{A1}$$

Differentiating Eq. (A1) once or twice with respect to  $\lambda$  and then setting  $\lambda = 1$  leads to results for  $E[N]$  and  $E[N(N - 1)]$  that correspond to the following relations between the means and variances of  $N$ ,  $\mathcal{E}$  and  $S$ :

$$\begin{aligned} E[N] &= E[\mathcal{E}]E[S] \\ \text{Var}(N) &= E[\mathcal{E}]\text{Var}(S) + \text{Var}(\mathcal{E})(E[S])^2. \end{aligned} \tag{A2}$$

**Particular results on the negative binomial distribution**

To establish some particular but informative results, we take  $N$  to have a negative binomial distribution. The form of this distribution that we use follows Anderson and May [6]. It has parameters  $m$  and  $k$  and is defined by

$$\begin{aligned} \text{Prob}(N = n) &\equiv B(n; m, k) \\ &= \binom{k}{k+m}^k \frac{\Gamma(k+n)}{\Gamma(k)\Gamma(n+1)} \left(\frac{m}{k+m}\right)^n, \quad n = 0, 1, 2, \dots \end{aligned} \tag{A3}$$

where  $\Gamma(x)$  denotes Euler’s Gamma function. This distribution has a mean of  $E[N] = m$  and a variance of  $\text{Var}(N) = m^2/k + m$ . The parameter  $k$  is often used as a measure of aggregation of the distribution and it can be approximated as

$$k = \frac{(E[N])^2}{\text{Var}(N) - E[N]}. \tag{A4}$$

**Constraints on the distribution of  $\mathcal{E}$  and  $S$  that lead to a negative binomial distribution**

It is convenient to express the generating function of  $N$  in terms of the parameter

$$r = m/k \tag{A5}$$

and then find that

$$G_N(\lambda) = E[\lambda^N] = (1 + r - \lambda r)^{-k}. \tag{A6}$$

When  $N$  has the distribution of Eq. (A3), the equation of the generating function, Eq. (A1), imposes a relationship between the generating functions  $G_{\mathcal{E}}(\lambda)$  and  $G_S(\lambda)$ , namely

$$(1 + r - \lambda r)^{-k} = G_{\mathcal{E}}(G_S(\lambda)). \tag{A7}$$

Equivalently, this equation expresses a relationship between the probability distributions of  $\mathcal{E}$  and  $S_j$ . We note that if we specify either  $G_{\mathcal{E}}(\lambda)$  or  $G_S(\lambda)$ , we can exploit Eq. (A7) to determine the other generating function. This generating function can then be used to determine the corresponding probability distribution. Note however, that this procedure does not always work: it sometimes gives rise to an invalid probability distribution, because some of the ‘probabilities’ obtained are negative. We give two illustrative examples of this usage of Eq. (A7). In the first example we determine the distribution of  $\mathcal{E}$ , after assuming the distribution of  $S$ . In the second example, the distribution of  $S$  is determined after assuming the distribution of  $\mathcal{E}$ . In this case we find that a probability distribution for  $S$  follows only for restricted values of some of the parameters.

**Example 1**

If  $S$  can only take the values 0 and 1, and these occur with probabilities  $1 - \alpha$  and  $\alpha$  respectively, then  $G_S(\lambda) = 1 - \alpha + \alpha\lambda$ . It follows from Eq. (A7) that  $(1 + r - \lambda r)^{-k} = G_{\mathcal{E}}(1 - \alpha + \alpha\lambda)$ . Solving this equation for  $G_{\mathcal{E}}(\lambda)$  yields  $G_{\mathcal{E}}(\lambda) = (1 + r/\alpha - \lambda r/\alpha)^{-k}$  which, on comparison with Eq. (A6), corresponds to  $\mathcal{E}$  having a negative binomial distribution with parameters  $m/\alpha$  and  $k$ :

$$\text{Prob}(\mathcal{E} = \varepsilon) = B(\varepsilon; m/\alpha, k), \quad \varepsilon = 0, 1, 2, \dots \tag{A8}$$

This example leads to  $E[\mathcal{E}] = m/\alpha$ ,  $E[S] = \alpha$ ,  $\text{Var}(\mathcal{E}) = m/\alpha + m^2/(\alpha^2 k)$  and  $\text{Var}(S) = \alpha(1 - \alpha)$ .

**Example 2**

If  $\mathcal{E}$  has a Poisson distribution, with  $\text{Prob}(\mathcal{E} = n) = \beta^n e^{-\beta}/n!$  for  $n = 0, 1, 2, \dots$  then  $G_{\mathcal{E}}(\lambda) = \exp(\beta\lambda - \beta)$ . Using this result in Eq. (A7) yields  $(1 + r - \lambda r)^{-k} = \exp(\beta G_S(\lambda) - \beta)$ , and this can be directly solved for  $G_S(\lambda)$  with the result  $G_S(\lambda) = 1 - (k/\beta) \ln(1 + r - \lambda r)$ . Expanding  $G_S(\lambda)$  in powers of  $\lambda$  leads to

$$\text{Prob}(S = s) = \begin{cases} 1 - \frac{k}{\beta} \ln(1 + m/k), & s = 0 \\ \frac{k}{\beta} \left(\frac{m}{m+k}\right)^s \frac{1}{s}, & s = 1, 2, 3, \dots \end{cases} \tag{A9}$$

This distribution can be viewed as a mixture of two distributions: a unit probability mass located at  $s = 0$  with weight  $1 - (k/\beta) \ln(1 + m/k)$ , and a logarithmic distribution with overall weight  $(k/\beta) \ln(1 + m/k)$ . The values of the parameters  $m, k$  and  $\beta$  cannot take any values, but must be chosen so the probabilities of different values of  $S$  (in particular  $S = 0$ ) are all non-negative. For example, at fixed values of  $k$  and  $m$ , the parameter  $\beta$  must have some restrictions placed upon it; in this case we can only consider values of  $\beta$  which yield  $1 - (k/\beta) \ln(1 + m/k) \geq 0$  so that  $\text{Prob}(S = 0) \geq 0$ . Defining

$$\beta_{\min} = k \ln(1 + m/k) \tag{A10}$$

we thus require  $\beta \geq \beta_{\min}$  for the probabilities of all values of  $S$  to be non-negative.

This example leads to  $E[\mathcal{E}] = \beta, E[S] = m/\beta, \text{Var}(\mathcal{E}) = \beta$  and  $\text{Var}(S) = (m/\beta)(1 + m/k - m/\beta)$ .

**B. Derivation of results for the heterogeneous model**

In this subsection we present a derivation of results for the heterogeneous model that was introduced in this work.

In the heterogeneous model, there are  $h$  different types of hosts. In general, each host type has a different distribution of encounters with a parasite/source of parasites, and the parasite has a different distribution of success on each host type. The number of parasites,  $N$ , in a randomly picked host is taken to be

$$N = \sum_{t=1}^h H_t \sum_{k=1}^{\mathcal{E}_t} S_k^{(t)}. \tag{B1}$$

In this equation  $\mathbf{H} \stackrel{\text{def}}{=} (H_1, H_2, \dots, H_h)$  constitutes a multinomial random variable with parameters 1 and  $(p_1, p_2, \dots, p_h)$ . Each  $H_t$  can only take the values 0 and 1, and in a realisation of  $\mathbf{H}$  only one of the  $H_t$ 's equals 1 (corresponding to the particular host type obtained) while the rest are 0. The probability with which  $H_t = 1$  is  $p_t$ .

To indicate that expected values (and other statistics) are conditional on host type, we shall use notation where  $E[\mathcal{E}|t]$  and  $\text{Var}(\mathcal{E}|t)$  denote the mean and variance of the number of encounters, when conditional on the host type  $t$ .

Proceeding, we note that with  $N$  given by Eq. (B1), and  $a_1, a_2, \dots, a_h$  a set of non random numbers, we can determine the generating function of  $N$ , using the result  $E\left[\lambda^{\sum_{t=1}^h H_t a_t}\right] = \sum_{t=1}^h p_t \lambda^{a_t}$ .

This leads to

$$G_N(\lambda) = E\left[\lambda^{\sum_{t=1}^h H_t \sum_{k=1}^{\mathcal{E}_t} S_k^{(t)}}\right] = \sum_{t=1}^h p_t E\left[\lambda^{\sum_{k=1}^{\mathcal{E}_t} S_k^{(t)}}\right] = \sum_{t=1}^h p_t G_{\mathcal{E}_t}(G_{S^{(t)}}(\lambda)). \tag{B2}$$

We shall determine the mean and variance of  $N$  from this expression, and it is helpful to establish some notation that we shall use. We define

$$\begin{aligned}
 \mu &= \sum_{t=1}^h p_t E[\mathcal{E}|t] \\
 m &= \sum_{t=1}^h p_t E[S|t] \\
 \sigma^2 &= \sum_{t=1}^h p_t \text{Var}(\mathcal{E}|t) \\
 v^2 &= \sum_{t=1}^h p_t \text{Var}(S|t) \\
 C_{a,b,c,d} &= \sum_{t=1}^h p_t (E[\mathcal{E}|t] - \mu)^a (\text{Var}(\mathcal{E}|t) - \sigma^2)^b (E[S|t] - m)^c (\text{Var}(S|t) - v^2)^d
 \end{aligned} \tag{B3}$$

### Mean of $N$

We determine the mean of  $N$  by differentiating  $G_N(\lambda)$  once with respect to  $\lambda$  and then setting  $\lambda = 1$ . This yields

$$E[N] = \sum_{t=1}^h p_t E[\mathcal{E}|t] E[S|t]. \tag{B4}$$

Equation (B4) can be rewritten in the alternative form

$$E[N] = \mu m + C_{1,0,1,0} \tag{B5}$$

where  $C_{a,b,c,d}$  is defined in Eq. (B3) and explicitly,  $C_{1,0,1,0}$  is the covariance

$$C_{1,0,1,0} = \sum_{t=1}^h p_t (E[\mathcal{E}|t] - \mu)(E[S|t] - m). \tag{B6}$$

### Variance of $N$

We first determine the expected value  $E[N(N - 1)]$  by differentiating  $G_N(\lambda)$  (Eq. (B2)) twice with respect to  $\lambda$  and then setting  $\lambda = 1$ . This yields

$$E[N(N - 1)] = \sum_{t=1}^h p_t E[\mathcal{E}(\mathcal{E} - 1)|t] (E[S|t])^2 + \sum_{t=1}^h p_t E[\mathcal{E}|t] E[S(S - 1)|t]. \tag{B7}$$

From this result and Eq. (B4) we obtain

$$\begin{aligned}
 \text{Var}(N) &= \sum_{t=1}^h p_t \text{Var}(\mathcal{E}|t) (E[S|t])^2 + \sum_{t=1}^h p_t E[\mathcal{E}|t] \text{Var}(S|t) \\
 &\quad + \sum_{t=1}^h p_t \left( E[\mathcal{E}|t] E[S|t] - \sum_{i=1}^h p_i E[\mathcal{E}|i] E[S|i] \right)^2.
 \end{aligned} \tag{B8}$$

Hence having multiple types leads to an extra level of averaging (as indicated by the sums in Eq. (B8)) plus an additional positive term in the variance reflecting an aspect of variation between different host types (the final sum in Eq. (B8)).

The expression in Eq. (B8) is somewhat complicated, but we can approximate it, under the assumption that deviations *between* different host types that are higher than quadratic order can be neglected. Using the *delta technique* (see e.g., [60]), we obtain the expression

$$\begin{aligned} \text{Var}(N) \approx & \sigma^2 m^2 + \mu v^2 + (\sigma^2 + \mu^2) C_{0,0,2,0} + m^2 C_{2,0,0,0} \\ & + 2\mu m C_{1,0,1,0} + 2m C_{0,1,1,0} + C_{1,0,0,1} \end{aligned} \quad (\text{B9})$$

which contains a leading term that is derived from average over all types, and various quadratic correlation terms that arise from between type variation.

We can write the  $C_{a,b,c,d}$  that appear in Eq. (B9) in a more suggestive form, using  $T$  to denote a random variable whose value corresponds to the type of host captured. We then have

$$\begin{aligned} C_{0,0,2,0} &= \text{Var}(E[S|T]) \\ C_{2,0,0,0} &= \text{Var}(E[\mathcal{E}|T]) \\ C_{1,0,1,0} &= \text{Cov}(E[\mathcal{E}|T], E[S|T]) \\ C_{0,1,1,0} &= \text{Cov}(\text{Var}(\mathcal{E}|T), E[S|T]) \\ C_{1,0,0,1} &= \text{Cov}(E[S|T], \text{Var}(E[S|T])). \end{aligned} \quad (\text{B10})$$

Using the above forms for the  $C_{a,b,c,d}$  we can write Eqs. (B5) and (B9) as

$$E[N] = \mu m + \text{Cov}(E[\mathcal{E}|T], E[S|T]) \quad (\text{B11})$$

$$\begin{aligned} \text{Var}(N) \approx & \sigma^2 m^2 + \mu v^2 + (\sigma^2 + \mu^2) \text{Var}(E[S|T]) + m^2 \text{Var}(E[\mathcal{E}|T]) \\ & + 2\mu m \text{Cov}(E[\mathcal{E}|T], E[S|T]) + 2m \text{Cov}(\text{Var}(\mathcal{E}|T), E[S|T]) \\ & + \text{Cov}(E[S|T], \text{Var}(E[S|T])). \end{aligned} \quad (\text{B12})$$

## Acknowledgments

One of us (DW) gratefully acknowledges a Visiting Professorship from Universite de Perpignan Via Domitia.

## Author Contributions

Conceived and designed the experiments: SG DW. Wrote the paper: SG SM DW.

## References

1. Shaw DJ, Dobson AP (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* 111:S 111–S 133. doi: [10.1017/S003118200075855](https://doi.org/10.1017/S003118200075855) PMID: [8632918](https://pubmed.ncbi.nlm.nih.gov/8632918/)
2. Morand S, Krasnov BR (2008) Why apply ecological laws to epidemiology? *Trends in Parasitology* 24: 304–309. doi: [10.1016/j.pt.2008.04.003](https://doi.org/10.1016/j.pt.2008.04.003) PMID: [18514576](https://pubmed.ncbi.nlm.nih.gov/18514576/)
3. Gaba S, and Gourbière S (2008) To delay once or twice: the effect of hypobiosis and freeliving stage arrestment on the stability of host-parasite interactions. *Journal of the Royal Society Interface* 5 (25):919–28. doi: [10.1098/rsif.2007.1282](https://doi.org/10.1098/rsif.2007.1282) PMID: [18182366](https://pubmed.ncbi.nlm.nih.gov/18182366/)
4. Crofton HD (1971) A quantitative approach to parasitism. *Parasitology* 62, 179–193. doi: [10.1017/S003118200071420](https://doi.org/10.1017/S003118200071420)

5. Poulin R (2007) Are there general laws in parasite ecology? *Parasitology* 134, 763–776. doi: [10.1017/S0031182006002150](https://doi.org/10.1017/S0031182006002150) PMID: [17234043](https://pubmed.ncbi.nlm.nih.gov/17234043/)
6. Anderson RM, May RM (1978) Regulation and stability of host–parasite population interactions. I. Regulatory process. *J. Anim. Ecol.* 47, 219–247 doi: [10.2307/3933](https://doi.org/10.2307/3933)
7. Adler FR, Kretzschmar M (1992) Aggregation and stability in parasite-host models. *Parasitology* 1992; 104: 199–205. doi: [10.1017/S0031182000061631](https://doi.org/10.1017/S0031182000061631) PMID: [1594287](https://pubmed.ncbi.nlm.nih.gov/1594287/)
8. Boag B, Lello J, Fenton A, Tompkins DM, Hudson PJ (2001) Patterns of parasite aggregation in the wild European rabbit *Oryctolagus cuniculus*. *International Journal for Parasitology* 31: 1421–1428. doi: [10.1016/S0020-7519\(01\)00270-3](https://doi.org/10.1016/S0020-7519(01)00270-3) PMID: [11595228](https://pubmed.ncbi.nlm.nih.gov/11595228/)
9. Rascalou G, Gourbière S (2014) Competition, virulence, host body mass and the diversification of macro-parasites. *Journal of the Royal Society Interface* 11: 20131108. doi: [10.1098/rsif.2013.1108](https://doi.org/10.1098/rsif.2013.1108) PMID: [24522783](https://pubmed.ncbi.nlm.nih.gov/24522783/)
10. Anderson RM, May RM (1991) *Infectious diseases of humans; dynamics and control*. Oxford, University Press.
11. Woolhouse MEJ, Dye C, Etard JF, Smith T (1997) Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc Natl Acad Sci U S A* 94: 338–342. doi: [10.1073/pnas.94.1.338](https://doi.org/10.1073/pnas.94.1.338) PMID: [8990210](https://pubmed.ncbi.nlm.nih.gov/8990210/)
12. Combes C (2005). *The art of being a parasite*. Chicago: University of Chicago Press.
13. Johnson NL, Kotz S, Kemp AW (2005) *Univariate discrete distributions*. 3rd Edition. Wiley Series in Probability and Statistics. John Wiley and Sons Ltd.
14. Hart BL (1994) Behavioral defense against parasites: Interaction with parasite invasiveness. *Parasitology* 109, S139–S151 PMID: [7854847](https://pubmed.ncbi.nlm.nih.gov/7854847/)
15. Loehle C (1995) Social barriers to pathogen transmission in wild animal populations. *Ecology* 76, 326–335. doi: [10.2307/1941192](https://doi.org/10.2307/1941192)
16. Bordes F, Morand S, Kelt DA, Van Vuren DH (2009) Home Range and Parasite Diversity in Mammals. *Am. Nat. Apr*; 173(4):467–74. doi: [10.1086/597227](https://doi.org/10.1086/597227) PMID: [19232003](https://pubmed.ncbi.nlm.nih.gov/19232003/)
17. Vale PF, Choisy M, Little TJ (2013) Host nutrition alters the variance in parasite transmission potential. *Biol Lett* 9: 20121145. doi: [10.1098/rsbl.2012.1145](https://doi.org/10.1098/rsbl.2012.1145) PMID: [23407498](https://pubmed.ncbi.nlm.nih.gov/23407498/)
18. Nouvellet P, Dumonteil E, Gourbière S (2013) The Improbable Transmission of *Trypanosoma cruzi* to Human: The Missing Link in the Dynamics and Control of Chagas Disease. *PLoS Negl Trop Dis* 7(11): e2505. doi: [10.1371/journal.pntd.0002505](https://doi.org/10.1371/journal.pntd.0002505) PMID: [24244766](https://pubmed.ncbi.nlm.nih.gov/24244766/)
19. Barbu C., Dumonteil E, and Gourbière S (2011) Evaluation of spatially targeted strategies to control non-domiciliated *Triatoma dimidiata* vector of Chagas disease. *PLoS Negl. Trop. Dis.* 5(5): e1045. doi: [10.1371/journal.pntd.0001045](https://doi.org/10.1371/journal.pntd.0001045) PMID: [21610862](https://pubmed.ncbi.nlm.nih.gov/21610862/)
20. Ramirez-Sierra MJ, Herrera-Aguilar M, Gourbière S, Dumonteil E (2010) Patterns of house infestation dynamics by non-domiciliated *Triatoma dimidiata* reveal a spatial gradient of infestation and potential insect manipulation by *Trypanosoma cruzi*. *Tropical Medicine & International Health* (15): 77–86. doi: [10.1111/j.1365-3156.2009.02422.x](https://doi.org/10.1111/j.1365-3156.2009.02422.x)
21. Gotelli NJ (2006) Null versus neutral models: what's the difference? *Ecography* 29: 5.
22. Ross SM (2010) *Introduction to Probability Models*, 10th ed. Academic Press, Amsterdam
23. Hutchings MR, Millner JM, Gordon I, Kyriazakis I, Jackson F (2002) Grazing decisions of Soay sheep (*Ovis aries*) on Ste. Kilda: a consequence of parasite distribution? *Oikos* 96:235–244. doi: [10.1034/j.1600-0706.2002.960205.x](https://doi.org/10.1034/j.1600-0706.2002.960205.x)
24. Knudsen R, Curtis MA, Kristoffersen R (2004) Aggregation of helminths: the role of feeding behavior of fish hosts. *Journal of Parasitology*: 90: 1–7. doi: [10.1645/GE-3184](https://doi.org/10.1645/GE-3184) PMID: [15040660](https://pubmed.ncbi.nlm.nih.gov/15040660/)
25. Pacheco-Tucuch FS, Ramirez-Sierra MJ, Gourbière S, and Dumonteil E (2012) Public street lights increase house infestation by *Triatoma dimidiata*, vector of Chagas disease in the Yucatan peninsula. *PLoS One* 7(4): e36207. doi: [10.1371/journal.pone.0036207](https://doi.org/10.1371/journal.pone.0036207) PMID: [22558384](https://pubmed.ncbi.nlm.nih.gov/22558384/)
26. Rascalou G, Pontier D, Menu F, Gourbière F (2012) Emergence and prevalence of human vector-borne diseases in sink vector populations. *PLoS One* 7(5): e36858 doi: [10.1371/journal.pone.0036858](https://doi.org/10.1371/journal.pone.0036858) PMID: [22629337](https://pubmed.ncbi.nlm.nih.gov/22629337/)
27. Dumonteil E, Nouvellet P, Rosecrans K, Ramirez-Sierra MJ, Gamboa-León R (2013) Eco-Bio-Social Determinants for House Infestation by Non-domiciliated *Triatoma dimidiata* in the Yucatan Peninsula, Mexico. *PLoS Negl Trop Dis* 7(9): e2466. doi: [10.1371/journal.pntd.0002466](https://doi.org/10.1371/journal.pntd.0002466) PMID: [24086790](https://pubmed.ncbi.nlm.nih.gov/24086790/)
28. Calabrese JM, Brunner JL, Ostfeld RS (2011) Partitioning the Aggregation of Parasites on Hosts into Intrinsic and Extrinsic Components via an Extended Poisson-Gamma Mixture Model. *PLoS One* 6(12): e29215 doi: [10.1371/journal.pone.0029215](https://doi.org/10.1371/journal.pone.0029215) PMID: [22216216](https://pubmed.ncbi.nlm.nih.gov/22216216/)



29. Nouvellet P, Dumonteil E, and Gourbière S (2011) Effects of genetic factors and infection status on wing morphology of *Triatoma dimidiata* species complex in the Yucatan peninsula, Mexico. *Infection, Genetics and Evolution* 11(6): 1243–1249 doi: [10.1016/j.meegid.2011.04.008](https://doi.org/10.1016/j.meegid.2011.04.008) PMID: [21515410](https://pubmed.ncbi.nlm.nih.gov/21515410/)
30. Gourbière S, Dorn P, Tripet F, Dumonteil E (2012) Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity* 108(3): 190–202. doi: [10.1038/hdy.2011.71](https://doi.org/10.1038/hdy.2011.71) PMID: [21897436](https://pubmed.ncbi.nlm.nih.gov/21897436/)
31. Mooring MS, Hart BL (1992) Animal grouping for protection from parasites: selfish herd and encounter-dilution effects. *Behaviour* 123:173–193. doi: [10.1163/156853992X00011](https://doi.org/10.1163/156853992X00011)
32. Galvani AP (2003) Immunity, antigenic heterogeneity, and aggregation of helminth parasites. *Journal of Parasitology* 89, 232–241. doi: [10.1645/0022-3395\(2003\)089%5B0232:IAHAAO%5D2.0.CO;2](https://doi.org/10.1645/0022-3395(2003)089%5B0232:IAHAAO%5D2.0.CO;2) PMID: [12760634](https://pubmed.ncbi.nlm.nih.gov/12760634/)
33. Goüy de Bellocq J, Ribas A, Casanova JC, Morand S (2007) Immunocompetence and helminth community of the white-toothed shrew, *Crocodyrus russula* from the Montseny Natural Park, Spain. *Eur J Wild Res* 53:315–320 doi: [10.1007/s10344-007-0094-2](https://doi.org/10.1007/s10344-007-0094-2)
34. Klein SL (2004). Hormonal and immunological mechanisms mediating sex differences in parasite infection. *Parasite Immunol* 26:247–264 doi: [10.1111/j.0141-9838.2004.00710.x](https://doi.org/10.1111/j.0141-9838.2004.00710.x) PMID: [15541029](https://pubmed.ncbi.nlm.nih.gov/15541029/)
35. Bize P, Jeanneret C, Klopfenstein A, Roulin A (2008) What makes a host profitable? Parasites balance host nutritive resources against immunity. *Am Nat* 171:107–118 doi: [10.1086/523943](https://doi.org/10.1086/523943) PMID: [18171155](https://pubmed.ncbi.nlm.nih.gov/18171155/)
36. Hoby S, Scharzenberger F, Doherr MG, Robert N, Walzer C (2006) Steroid hormone related male biased parasitism in chamois, *Rupicapra rupicapra rupicapra*. *Vet Parasitol* 138:337–348 doi: [10.1016/j.vetpar.2006.01.028](https://doi.org/10.1016/j.vetpar.2006.01.028) PMID: [16497439](https://pubmed.ncbi.nlm.nih.gov/16497439/)
37. Lafferty KD (1992) Foraging on prey that are modified by parasites. *Am. Nat.*, 140: 854–857. doi: [10.1086/285444](https://doi.org/10.1086/285444)
38. Navarro-Gonzalez N, Verheyden H, Hoste H, Cargnelutti B, Lourtet B, Merlet J, Daufresne T, Lav'n S, Hewison AJM, Morand S., Serrano E 2010. Diet quality and immunocompetence influence parasite load of roe deer in a fragmented landscape. *Eur J Wild Res* 57: 639–645 Nieberding doi: [10.1007/s10344-010-0474-x](https://doi.org/10.1007/s10344-010-0474-x)
39. Young LJ, Young JH (1998) *Statistical Ecology*. Springer, New York
40. Côté IM, Poulin R (1992) Parasitism and group size in social animals: a meta-analysis. *Behavioral Ecology*, 6: 159–165.
41. Bordes F, Blumstein DT, Morand S (2007) Rodent sociality and parasite diversity. *Biology Letters* 3: 692–694. doi: [10.1098/rsbl.2007.0393](https://doi.org/10.1098/rsbl.2007.0393) PMID: [17925270](https://pubmed.ncbi.nlm.nih.gov/17925270/)
42. Morand S, Deter J (2009) Parasitism and regulation of the host population. In *Ecology and Evolution of Parasitism*. Thomas F., Guégan J.-F., Renaud F. (Eds). Oxford University Press, Oxford.
43. Guilhem R., Šimková A., Morand S. and Gourbière S.. 2012. Within-host competition and diversification of macro-parasites. *Journal of the Royal Society Interface* 9(76):2936–2946. doi: [10.1098/rsif.2012.0358](https://doi.org/10.1098/rsif.2012.0358) PMID: [22696483](https://pubmed.ncbi.nlm.nih.gov/22696483/)
44. Poulin R (2007). *Evolutionary ecology of parasites* (2nd ed.). Princeton University Press, 342p.
45. Leung B (1998) Aggregated parasite distributions on hosts in a homogeneous environment: examining the Poisson null model. *International Journal for Parasitology* 28, 1709–1712. doi: [10.1016/S0020-7519\(98\)00128-3](https://doi.org/10.1016/S0020-7519(98)00128-3) PMID: [9846607](https://pubmed.ncbi.nlm.nih.gov/9846607/)
46. Pugliese A, Rosa R, Damaggio ML (1998) Analysis of a model for macroparasitic infection with variable aggregation and clumped infections. *Journal of Math. Biol.* 36, 419–47. doi: [10.1007/s002850050107](https://doi.org/10.1007/s002850050107) PMID: [9579031](https://pubmed.ncbi.nlm.nih.gov/9579031/)
47. Anderson RM, Gordon DM (1982) Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology* 85:373–398. doi: [10.1017/S0031182000055347](https://doi.org/10.1017/S0031182000055347) PMID: [7145478](https://pubmed.ncbi.nlm.nih.gov/7145478/)
48. May RM (1978) Host-parasitoid systems in patchy environments: A phenomenological model. *Journal of Animal Ecology* 47: 833–844. doi: [10.2307/3674](https://doi.org/10.2307/3674)
49. Regoes RR, Hottinger JW, Sygnarski L, Ebert D (2003) The infection rate of *Daphnia magna* by *Pasteuria ramose* conforms with the mass-action principle. *Epidemiology and Infection* 131(3): 957–966. doi: [10.1017/S0950268803008793](https://doi.org/10.1017/S0950268803008793) PMID: [14596538](https://pubmed.ncbi.nlm.nih.gov/14596538/)
50. Ben-Ami F, Regoes RR, Ebert D (2008) A quantitative test of the relationship between parasite dose and infection probability across different host-parasite combinations. *Proceedings of the Royal Society B* 275: 853–859. doi: [10.1098/rspb.2007.1544](https://doi.org/10.1098/rspb.2007.1544) PMID: [18198145](https://pubmed.ncbi.nlm.nih.gov/18198145/)
51. Ben-Ami F, Ebert D, Regoes RR (2010) Pathogen dose infectivity curves as a method to analyze the distribution of host susceptibility: A quantitative assessment of maternal effects after food stress and pathogen exposure. *Am. Nat.* 175: 106–115. doi: [10.1086/648672](https://doi.org/10.1086/648672) PMID: [19911987](https://pubmed.ncbi.nlm.nih.gov/19911987/)

52. Karvonen A, Hudson PJ, Seppälä O, Valtonen ET (2004) Transmission dynamics of a trematode parasite: exposure, acquired resistance and parasite aggregation. *Parasitology Research* 92:183–188. doi: [10.1007/s00436-003-1035-y](https://doi.org/10.1007/s00436-003-1035-y) PMID: [14652746](https://pubmed.ncbi.nlm.nih.gov/14652746/)
53. Khalife J, Cêtre C, Pierrot C, Capron M (2000) Mechanisms of resistance to *S. mansoni* infection: the rat model. *Parasitol Int.* 49(4):339–345. doi: [10.1016/S1383-5769\(00\)00059-3](https://doi.org/10.1016/S1383-5769(00)00059-3) PMID: [11077269](https://pubmed.ncbi.nlm.nih.gov/11077269/)
54. Richardson D.J., Brink C.D. 2011. Effectiveness of Various Anthelmintics in the Treatment of Moniliformiasis in Experimentally Infected Wistar Rats. *Vector-Borne and Zoonotic Diseases* 11 (8) 1151–1156. doi: [10.1089/vbz.2010.0151](https://doi.org/10.1089/vbz.2010.0151) PMID: [21254932](https://pubmed.ncbi.nlm.nih.gov/21254932/)
55. Yamada M, Nakazawa N, Kamata I, Arizono N (1992) Low-level infection with the nematode *Nippostrongylus brasiliensis* induces significant and sustained specific and non-specific IgE antibody responses in rats. *Immunology* 75(1): 36–40. PMID: [1537600](https://pubmed.ncbi.nlm.nih.gov/1537600/)
56. Morand S, Southgate VR, Jourdane J (2002) A model to explain the replacement of *Schistosoma intercalatum* by *Schistosoma haematobium* and the hybrid *S. intercalatum* x *S. haematobium* in areas of sympatry. *Parasitology*, 124(Pt 4), 401–408. PMID: [12003064](https://pubmed.ncbi.nlm.nih.gov/12003064/)
57. Tschirren B, Bischoff LL, Saladin V, Richner H (2007) Host condition and host immunity affect parasite fitness in a bird–ectoparasite system. *Func Ecol* 21: 372–378 doi: [10.1111/j.1365-2435.2007.01235.x](https://doi.org/10.1111/j.1365-2435.2007.01235.x)
58. Schmid-Hempel P (2009) Immune defence, parasite evasion strategies and their relevance for ‘macroscopic phenomena’ such as virulence. *Trans R Soc Lond B Biol Sci.* 364 (1513): 85–98. doi: [10.1098/rstb.2008.0157](https://doi.org/10.1098/rstb.2008.0157) PMID: [18930879](https://pubmed.ncbi.nlm.nih.gov/18930879/)
59. Råberg L, Graham AL, Read AF (2009) Decomposing health: tolerance and resistance to parasites in animals. *Philosophical Transactions of the Royal Society London B*, 364: 37–49 doi: [10.1098/rstb.2008.0184](https://doi.org/10.1098/rstb.2008.0184) PMID: [18926971](https://pubmed.ncbi.nlm.nih.gov/18926971/)
60. Bulmer MG (1979) *Principles of Statistics*. Dover, New York