

# Linking parasite populations in hosts to parasite populations in space through Taylor's law and the negative binomial distribution

Joel E. Cohen<sup>a,b,1</sup>, Robert Poulin<sup>c</sup>, and Clément Lagrue<sup>c</sup>

<sup>a</sup>Laboratory of Populations, The Rockefeller University, New York, NY 10065; <sup>b</sup>Laboratory of Populations, Earth Institute of Columbia University, New York, NY 10027; and <sup>c</sup>Department of Zoology, University of Otago, Dunedin 9054, New Zealand

Contributed by Joel E. Cohen, November 16, 2016 (sent for review July 25, 2016; reviewed by Kevin Lafferty and Ross McVinish)

**The spatial distribution of individuals of any species is a basic concern of ecology. The spatial distribution of parasites matters to control and conservation of parasites that affect human and nonhuman populations. This paper develops a quantitative theory to predict the spatial distribution of parasites based on the distribution of parasites in hosts and the spatial distribution of hosts. Four models are tested against observations of metazoan hosts and their parasites in littoral zones of four lakes in Otago, New Zealand. These models differ in two dichotomous assumptions, constituting a 2 × 2 theoretical design. One assumption specifies whether the variance function of the number of parasites per host individual is described by Taylor's law (TL) or the negative binomial distribution (NBD). The other assumption specifies whether the numbers of parasite individuals within each host in a square meter of habitat are independent or perfectly correlated among host individuals. We find empirically that the variance–mean relationship of the numbers of parasites per square meter is very well described by TL but is not well described by NBD. Two models that posit perfect correlation of the parasite loads of hosts in a square meter of habitat approximate observations much better than two models that posit independence of parasite loads of hosts in a square meter, regardless of whether the variance–mean relationship of parasites per host individual obeys TL or NBD. We infer that high local interhost correlations in parasite load strongly influence the spatial distribution of parasites. Local hotspots could influence control and conservation of parasites.**

variance function | parasite load | spatial distribution | synchrony | freshwater communities

The spatial distribution of individuals of any species is a basic concern of ecology. The spatial distribution of parasites matters to control and conservation of parasites that affect human and nonhuman populations. Despite the basic scientific and practical significance of the spatial distribution of parasites, investigations of parasite populations are often founded on their distributions and dynamic processes within and among hosts. A scientific justification for this approach is that the number of parasite individuals per host individual is likely to affect the parasite's impact on the host in theory (1) and empirically (2–5). A practical motivation for this approach is that a field investigator can collect hosts and study the parasite populations in them without the need to describe in detail the spatial distribution of the hosts or their abundance.

Why is the variation of parasite population density from 1 m<sup>2</sup> of space to another important? Kuris et al. (6) suggested that, because parasites contribute substantial biomass and productivity to estuaries, parasite ecology should be fully integrated into the general body of ecological theory. The spatial ecology of free-living species has long been a central topic in empirical and theoretical ecology but has not been fully explored for parasites. Moreover, parasites' spatial variation is likely to influence the conservation and control of parasites, especially those that affect human health, wildlife, and game. On basic scientific and practical grounds, the spatial ecology of parasites deserves fuller development.

To investigate how parasites are distributed in space, this paper develops a theoretical framework and four models that link the distribution of parasites in hosts, the distribution of hosts in space, and the distribution of parasites in space. The four models are tested against observations of the metazoan hosts in the littoral zone of four lakes in Otago, New Zealand.

Prior empirical studies of parasite populations have commonly estimated the number of parasite individuals per host individual from a sample of host individuals. For example, in a study (figure 5 in ref. 2) of macroparasites in wild vertebrate hosts and a study (figure 7A in ref. 7, p. 569) of parasitic nematodes in terrestrial mammalian hosts, the sample variance of the number of parasites per individual host was well described ( $r = 0.98$ ) by the equation  $\log(\text{sample variance}) \approx \log(a) + b \times \log(\text{sample mean})$ ,  $a > 0$ , which is a log–log form of Taylor's law (TL) (8):

$$\text{sample variance} \approx a \times (\text{sample mean})^b, \quad a > 0. \quad [1]$$

The symbol  $\approx$  signifies “approximately equals” in recognition that both sides of the equations of TL are subject to sampling variation and that the error term is left unstated. Any rescaling of population density by a positive constant (for example, multiplying population density by 10<sup>-4</sup> to convert individuals-hectare<sup>-1</sup> to individuals-meter<sup>-2</sup>) has no effect on the exponent  $b$  in 1 but changes  $a$ .

The fact that the sample variance of the number of individuals could be approximated as a power function of the sample mean of multiple sets of observations was proposed long before Taylor (8) and illustrated with entomological examples (9), a pest plant (10), and the cabbage aphid (11). Without reference to these discoveries,

## Significance

**The spatial distribution of individuals of any species is a basic concern of ecology. The spatial distribution of parasites matters to control and conservation of parasites that affect human and nonhuman populations. This paper develops a quantitative theory to predict the spatial distribution of parasites based on the distribution of parasites in hosts and the spatial distribution of hosts. The theory is tested using observations of metazoan hosts and parasites in the littoral zone of four lakes in Otago, New Zealand. We infer that the spatial distribution of parasites depends crucially on high local correlations of hosts' parasite loads. If so, local hotspots of correlated parasite loads should be considered in parasite control and conservation.**

Author contributions: J.E.C. designed research; J.E.C., R.P., and C.L. performed research; J.E.C. contributed new reagents/analytic tools; R.P. supervised data collection; C.L. collected data; J.E.C. analyzed data; and J.E.C. wrote the paper.

Reviewers: K.L., University of California, Santa Barbara; and R.M., University of Queensland.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. Email: cohen@rockefeller.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1618803114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1618803114/-DCSupplemental).

Taylor (8) brought the approximate power law relationship **1** to general attention as a widespread empirical pattern. This pattern has since become known among many ecologists as TL.

Hechinger et al. (12) investigated populations of parasites and co-occurring free-living species. They measured population density for both parasitic and free-living species in number of individuals-hectare<sup>-1</sup>. This measure of population density enabled them to compare allometric power laws in parasitic and free-living species. Measuring population density by individuals-meter<sup>-2</sup>, Lagrue et al. (13) showed that the variance and mean of population density were well approximated by **1** but that the parameters *a* and *b* differed among three so-called “lifestyles”: parasites, free-living species that were hosts of parasites (henceforth called “hosts” here), and unparasitized free-living species.

Here, we address theoretical and empirical questions about the relation of populations of parasites in hosts to populations of parasites in physical space. What are the relations between two measures of parasite population density: (i) parasite individuals per host and (ii) parasite individuals per square meter? Can mathematically transparent, empirically testable models describe accurately the variance–mean relationships of hosts per square meter, parasites per host individual, and parasites per square meter? The answer to the last question is not obvious, because any two of these distributions constrain the third.

## Theoretical Methods and Results

**General Notation and Definitions.** We investigate four models that share a common framework and three assumptions. In all of these models, we specify a single parasite species and a single host species. When a parasite species infects multiple species of hosts or when a host species is infected by multiple species of parasites, we organized the data to consider all possible pairs consisting of a single parasite species and a single host species. In our models, we analyze theoretically the set of all such single-species parasite–host pairs. In the following empirical analyses, we analyze the single-species parasite–host pairs statistically.

Assumption *i*: the number of parasites (of a selected single species) infecting a host (of a selected single species) in 1 m<sup>2</sup> of habitat is the sum of the numbers of parasites in all host individuals (of that species) in that square meter of habitat. We specify two variant forms of this assumption: one for models 1 and 2 and another for models 3 and 4. To spell out these details of assumption *i*, we now define additional variables and notation.

Let *H* be a random variable with nonnegative integer values {0, 1, 2, ...}. *H* represents the number of individuals of a particular host species per square meter of habitat. A generalist parasite may infect more than one species of host. Here, *H* refers to counts of only one selected host species. *H* is not a fixed number, but a random variable that may differ from 1 m<sup>2</sup> to another and differ over time within the same square meter. We assume that the mean and the variance of the probability distribution of *H* are positive and finite.

Let *P* be another random variable with nonnegative integer values {0, 1, 2, ...}. *P* represents the number of parasites (of a selected species) in one host individual. We assume that the mean and the variance of the probability distribution of *P* are positive and finite. The quantity *P* is sometimes called the “parasite load” (ref. 3, p. 606). If the host individual is parasitized by more than one species, we count here the individuals of only one selected parasite species. Assume *H* and *P* are independent in any square meter of habitat. Different square meter of habitat may have different distributions of *H* and *P*. Let *P<sub>i</sub>* for *i* = 1, 2, ..., *H* be random variables that represent the number of parasites in the *i*th individual host *i* = 1, 2, ..., *H* in a square meter of habitat. We assume that *P<sub>i</sub>*, *i* = 1, 2, ..., *H*, all have the distribution of *P* and are independent of *H* and independent of one another.

Let *S* be the number of individuals of the selected parasite species in all individuals of the selected host species per square

meter of habitat. The symbol *S* is a mnemonic for sum of parasites in 1 m<sup>2</sup> of habitat space; *S* recalls “Sum over Space.” When *H* = 0, define *S* = 0. When *H* > 0, models 1 and 2 assume that the total number of parasites in 1 m<sup>2</sup> of habitat is a sum of a random number *H* of independent random variables *P<sub>i</sub>*, *i* = 1, 2, ..., which are each independent of *H*. In brief,  $S = P_1 + \dots + P_H$ .

At the opposite extreme, models 3 and 4 assume that the numbers of parasites per host individual in a square meter of habitat are perfectly correlated among all host individuals, although they are independent of the number of hosts *H*. Then, the total number of parasites in 1 m<sup>2</sup> of habitat is a product  $Z = H \times P$  of independent random variables. If hosts are absent (i.e., *H* = 0), then parasites are necessarily absent (i.e., *Z* = 0) in parallel with models 1 and 2.

These two equations,  $S = P_1 + \dots + P_H$  and  $Z = H \times P$ , are variant forms of assumption *i*: the number of parasites in a square meter is the sum of the parasites in all host individuals in that square meter (to repeat: for a specified parasite species and a specified host species). Because *H* and *P* have finite, positive means and variances, so do *S* and *Z*.

We now prepare assumption *ii*. For a random variable *X* that has a finite positive population mean and finite positive population variance, let the population mean be  $\mu_X = E(X)$  and population variance be  $\nu_X = Var(X) = E[(X - E(X))^2]$ . Thus,  $\mu_H$ ,  $\nu_H$ ,  $\mu_P$ ,  $\nu_P$ ,  $\mu_S$ , and  $\nu_S$  denote the population means and population variances of *H*, *P*, and *S*, respectively. Because these moments are all finite and positive, their logarithms exist and are not equal to  $\pm\infty$ .

**Variance functions.** A variance function is a standard statistical concept (14). Suppose that the probability distributions of *H* and *P* depend on a parameter  $\theta$ , such as temperature, nutrient concentration, light availability, or other factors that vary from 1 m<sup>2</sup> to another. Then, the moments  $\mu_H$ ,  $\nu_H$ ,  $\mu_P$ ,  $\nu_P$ ,  $\mu_S$ , and  $\nu_S$  also depend on the parameter  $\theta$ . The equation  $\nu_H(\theta) = f(\mu_H(\theta))$  means that the variance of *H* given the parameter  $\theta$  equals a function *f* of the mean of *H* given the parameter  $\theta$ . When this equation holds, then *f*(.) is called the variance function (or sometimes, the variance–mean function) of the family of random variables  $\{H(\theta)\}_\theta$  parameterized by  $\theta$ . More succinctly, *f*(.) is called the variance function of *H*, because it maps the mean to the variance. For example, if *H*( $\theta$ ) satisfies TL (**1**), then the variance function of  $\{H(\theta)\}_\theta$  is a power function  $f(x) = ax^b$ . In the usual statement of TL, as in **1** above,  $\theta$  is left unstated. For simplicity, our notation omits explicit specification that the population means and population variances of *H*, *P*, *S*, and *Z* depend on a parameter  $\theta$ .

In empirical tests of TL, the sample mean and sample variance differ from one block of observations to another, and  $\theta$  could be interpreted as a label of each block. The interpretations of  $\theta$  may be illustrated by published examples. In one test of TL (figure 5 in ref. 2, p. S118), each value of  $\theta$  specified 1 of 263 pairs of (mean abundance per host, variance of abundance per host) of macro-parasites in wildlife host populations. In another test of TL (figure 7A in ref. 7, p. 569), each value of  $\theta$  specified one pair of (mean abundance per host, variance of abundance per host) of adult nematode worms recorded from individual guts of 66 terrestrial mammalian species (*n* = 104 values of  $\theta$  corresponding to 104 reported samples). In a third test of TL (figure 1 in ref. 15, p. 543), each value of  $\theta$  specified one pair of (mean abundance per host, variance of abundance per host) of helminth parasites of fish from 410 samples (with 180 parasitic helminth species and 68 fish host species from 62 different published papers). In our earlier test of TL (13), each value of  $\theta$  specified one pair of (mean population density per square meter, variance of population density per square meter) from a specified lake sampled in a specified season counting a specified species of parasite (253 mean–variance pairs) or host (151 mean–variance pairs).

Assumption *ii*: *H*, the number of host individuals-meter<sup>-2</sup> (of a specified species), satisfies TL. Here,  $\theta$  may be interpreted as a label associated with each square meter or the collection of square

meter in each of a set of samples. Explicitly, for some constants  $a > 0$ ,  $b$  ( $b$  is not necessarily positive),

$$\nu_H = a\mu_H^b \tag{2}$$

(TL variance function for host individuals per square meter).

Assumption *iii*: the mean number of parasites per host is a power function of the mean host density. Explicitly, for some constants  $f > 0$ ,  $g$  ( $g$  is not necessarily positive),

$$\mu_P = f\mu_H^g \tag{3}$$

(host–parasite density scaling).

If  $g = 0$  in Eq. 3, then  $\mu_P = f$  (i.e., mean parasite density per host  $\mu_P$  is constant regardless of  $\mu_H$ ). Depending on whether  $g > 0$  or  $g < 0$ , the mean density of parasites per host increases or decreases with increasing mean host density per square meter. If the value of  $g$  depends on quadrat size, then the model results may depend on the spatial scale of observation.

This assumption is a flexible quantitative formulation of the possibility that there may be no relation ( $g = 0$ ) between the mean number of parasites per host and the mean host density per square meter; that there may be a negative relation ( $g < 0$ : greater host abundance per square meter is associated with a reduced mean parasite burden per host) as an effect of herd immunity, dilution, or body size (bigger hosts are rarer and can accommodate more parasites); or that there may be a positive relation ( $g > 0$ : greater host abundance is associated with an increased parasite burden per host) as an effect of contagion or reduced host resistance from crowding. We introduce this assumption in the models and ask the data to reveal the relationship, while leaving the mechanism of the relationship (if  $g \neq 0$ ) for future research.

**Four alternative models.** Four models differ in each of two assumptions, each of which has two alternatives. Thus, four models may be summarized by a  $2 \times 2$  table (Table 1). The first assumption specifies whether the variance function of parasites per host  $P$  comes from TL or the negative binomial distribution (NBD). The NBD has traditionally been widely confirmed (3) and assumed for the abundance of metazoan parasites in individual hosts since the work in ref. 16. The second assumption specifies whether  $P_i$ , the number of parasites in the  $i$ th host,  $i = 1, 2, \dots, H$ , is independent (implying zero correlation) among host individuals in a square meter or identical for all hosts in this square meter (implying correlation one). These two alternatives correspond to the complete absence of synchrony and perfect synchrony, respectively, of the parasite loads of hosts in  $1 \text{ m}^2$ . (In models of statistical physics, the analogous difference is called “annealed” vs. “quenched.”) For brevity, we do not analyze here the obvious possibility of correlations among  $P_i$  that are intermediate between zero and one.

**Models 1 and 2: Independence of Parasites per Host.** Let  $P_i$ ,  $i = 1, 2, \dots$  be independently and identically distributed (iid) random variables with the distribution of  $P$ , also independent of  $H$ . It is well known (equation 7.2 in ref. 17, p. 119; equation 3 in ref. 18, p. 122; and ref. 19, p. 110 gives a detailed elementary derivation)

that, if  $S = P_1 + \dots + P_H$ , where all  $P_i$  are iid with the distribution of  $P$ , then

$$\mu_S = \mu_H \mu_P \tag{4}$$

(product rule for means) and

$$\nu_S = \mu_H \nu_P + \nu_H \mu_P^2 \tag{5}$$

(variance under independence).

We now use Eq. 5 to express the variance of  $S$  as a function of the mean of  $S$  under two alternative assumptions about the variance function of  $P$ . Model 1 assumes that the parasites per host satisfy TL. Model 2 assumes that the parasites per host satisfy the NBD.

**Model 1: Independent number P of parasites per host and power law variance function (TL) of P.** Assume  $P$  obeys TL [i.e., there exist constants  $c > 0$ ,  $d$  ( $d$  is not necessarily positive),  $f > 0$ ,  $g$  ( $g$  is not necessarily positive)], such that

$$\nu_P = c\mu_P^d \tag{6}$$

(TL for parasites per host individual).

Then, we prove in *SI Appendix* that the exact predicted variance of parasites per square meter is

$$\nu_S = cf \frac{d-1}{g+1} \mu_S^{\frac{1+dg}{g+1}} + af \frac{2-b}{g+1} \mu_S^{\frac{b+2g}{g+1}} \tag{7}$$

(model 1 exact variance function of parasites per square meter).

The variance  $\nu_S$  of the number  $S$  of parasites per square meter is a sum of two power functions of the mean of  $S$ . The exponents of  $\mu_S$  in Eq. 7 depend exclusively on the exponents  $b$ ,  $d$ , and  $g$  and do not at all depend on the coefficients  $a$ ,  $c$ , and  $f$ .

As  $\mu_S \rightarrow \infty$ ,  $S$  obeys TL with an exponent equal to the larger of the two exponents of  $\mu_S$  in Eq. 7. At the other extreme, as  $\mu_S \rightarrow 0$ ,  $S$  obeys TL with an exponent equal to the smaller of the two exponents of  $\mu_S$ . The exponent of  $\mu_S$  in the first term is  $(1 + dg)/(g + 1)$ , and the exponent of  $\mu_S$  in the second term is  $(b + 2g)/(g + 1)$ . Because the denominators are the same, which exponent of  $\mu_S$  is bigger depends only on the numerators. Other examples of TL in hosts and parasites lead us to expect the exponent  $b$  of TL Eq. 2 for hosts-meter<sup>-2</sup> to exceed 1 and the exponent  $d$  of TL Eq. 6 for parasite individuals per host individual to be between 1 and 2. If these two expectations hold true, then  $1 + dg < b + 2g$ ; therefore, the second term on the right side of Eq. 7 would be expected to dominate as  $\mu_S \rightarrow \infty$ , and the first term would be expected to dominate asymptotically as  $\mu_S \rightarrow 0$ .

**Properties of the NBD.** To specify the NBD (ref. 20, p. 306), let  $\rho > 0$  be a positive real number. When  $\rho$  is not an integer, the NBD is sometimes called the Pólya distribution. Let  $p$  be a positive probability,  $0 < p \leq 1$ , and let  $q = 1 - p$ ,  $0 \leq q < 1$ . A random variable  $X$  taking only the nonnegative integer values  $0, 1, 2, \dots$  has the NBD if and only if

$$\Pr[X = x | \rho, p] \equiv f(x | \rho, p) = \binom{x + \rho - 1}{\rho - 1} q^\rho p^x, \quad x = 0, 1, 2, \dots$$

The mean and variance of  $X$  are

$$E(X) = \frac{\rho q}{p}, \quad \text{Var}(X) = \frac{\rho q}{p^2}.$$

Because  $p \leq 1$ ,  $E(X) \leq \text{Var}(X)$ . If  $p < 1$ , then  $E(X) < \text{Var}(X)$ . A distribution is said to be overdispersed if  $E(X) < \text{Var}(X)$ . The NBD is overdispersed if  $p < 1$ . In the Poisson distribution, by contrast, the mean equals the variance.

**Table 1. Assumptions that differentiate models 1–4**

Correlation of parasites $P_i$ among hosts in $1 \text{ m}^2$	Variance function of the number $P_i$ of parasites per host individual	
	TL (power function)	Negative binomial (quadratic function)
All $P_i$ uncorrelated	Model 1	Model 2
All $P_i$ perfectly correlated	Model 3	Model 4



A family of NBDs is a collection of NBDs in which one or both of its parameters  $\rho, p$  vary. The variance function of the NBD depends on which parameter is assumed to vary. If  $\theta = \rho$  varies and  $p$  is constant, then the variance is proportional to the mean  $Var(X(\rho)) = p^{-1}E(X(\rho))$ . This equation is a special case of TL 1 with coefficient  $a = p^{-1}$  and exponent  $b = 1$ . If  $\rho$  is constant and  $\theta = p$  varies, then the variance is a quadratic function of the mean with no constant term:

$$Var(X(p)) = E(X(p)) + \rho^{-1}[E(X(p))]^2 = E(X(p))[1 + \rho^{-1}E(X(p))]. \quad [8]$$

This example shows the importance of being explicit about the meaning of  $\theta$  when more than one parameter may vary.

We now show that, when  $\rho$  is constant and  $\theta = p$  varies, a family of NBDs is not consistent with TL, except asymptotically in the extremes of large  $E(X)$  and small  $E(X)$ . As is standard, we use the notation  $x \ll y$  to mean that  $x$  is much smaller than  $y$  or that  $y$  is much larger than  $x$ . When Eq. 8 holds and  $0 < \rho \ll E(X)$ , then  $1 \ll E(X)/\rho$ ; therefore,  $E(X) \ll \rho^{-1}[E(X)]^2$ , and the second term on the right side of Eq. 8 is much larger than the first term. Asymptotically for large  $E(X)$ , TL 1 holds with exponent  $b = 2$ .

When Eq. 8 holds and  $\rho \gg E(X) > 0$ , then  $1 \gg E(X)/\rho$ ; therefore,  $E(X) \gg [E(X)]^2/\rho$ , and the first term on the right side of Eq. 8 is much larger than the second term. Asymptotically for small  $E(X)$ , TL 1 holds with exponent  $b = 1$ . Over the whole range of  $E(X)$  from very small to very large, obviously TL cannot hold with constant  $b$ .

In Eq. 8,  $\rho^{-1}$  is the same as  $k$  in equation 1 in ref. 8, p. 732. Taylor (8) remarks: “Unfortunately  $k$  is not always independent of [the sample mean]  $m$ ” [i.e., the model of constant  $k$  (or  $\rho$ ) and changing  $p$  in a family of NBDs does not hold empirically in general].

For a family of NBDs with constant  $\rho$  and varying  $p$  (equation 6 in ref. 21, p. 162),

$$\rho^{-1} = \frac{Var(X) - E(X)}{[E(X)]^2}. \quad [9]$$

This relationship is not compatible with TL 1. In a family of NBDs with constant  $\rho > 0$  and varying  $p$ ,  $0 < p \leq 1$ ,  $\log Var(X)$  is a strictly convex function of  $\log E(X)$ . We prove this statement in *SI Appendix*, using a result of ref. 31.

The self-contradictory assumption that  $\rho$  could be constant while TL holds appeared in at least two papers by Taylor and coworkers (ref. 22 and equation 2 in ref. 23, p. 1056) and has been repeated by others (equation 2 in ref. 24, p. 306). We prove in *SI Appendix* that using the right side of *SI Appendix*, Eq. S17 as “the estimate of  $1/k$  [= our  $\rho$ ] assuming  $k$  to be constant over all population densities” as in ref. 23, p. 1056, paragraph ii, b is self-contradictory (24). The number of parasites per host may be described exactly by a family of NBDs with constant  $\rho$  and varying  $p$ , or it may be described exactly by TL, but it cannot be described exactly by both over all population densities.

**Model 2: Independent number P of parasites per host and NBD variance function of P.** Model 2 assumes that  $P$  obeys the variance function Eq. 8 of a family of NBDs with constant  $\rho$  and varying  $p$ . Using the symbols in Eq. 6, Eq. 8 is equivalent to

$$\nu_P = \mu_P + \rho^{-1}\mu_P^2 \quad [10]$$

(negative binomial variance function with constant  $\rho$  and varying  $p$ ).

Then, instead of Eq. 7, the variance of  $S$  is a sum of a linear term plus two power functions:

$$\nu_S = \mu_S + \rho^{-1}f^{\frac{1}{1+g}}\mu_S^{\frac{1+2g}{1+g}} + af^{\frac{2-b}{1+g}}\mu_S^{\frac{b+2g}{1+g}} \quad [11]$$

(model 2 exact variance function of parasites per square meter).

The exponents of all three terms are independent of  $\rho$ .

**Models 3 and 4: Identical Numbers of Parasites per Host.** The next two models assume that, when  $H > 0$ , every host individual in  $1 \text{ m}^2$  of habitat has an identical number  $P$  of parasites. This number of parasites will differ from  $1 \text{ m}^2$  to another, but the same number  $P$  of parasites resides in every host in a square meter. Then the number of parasites per square meter of habitat is  $P_1 + \dots + P_H = H \times P$ , where (as previously assumed) the number  $H$  of hosts per square meter is independent of the number  $P$  of parasites per host (conditional on the parameter  $\theta$ ). To avoid confusion with the previous two models, we write

$$Z = H \times P$$

for the number of parasites per square meter in models 3 and 4. At one extreme, models 1 and 2 assume independence (conditional on  $\theta$ ) of parasite loads in different hosts in  $1 \text{ m}^2$  (therefore, correlation 0). At an opposite extreme, models 3 and 4 assume perfect identity of parasite loads in different host individuals in  $1 \text{ m}^2$  (therefore, correlation 1).

$S$  and  $Z$  have the same mean (25),  $\mu_S = \mu_Z = \mu_H \mu_P$ , given by Eq. 4. However,  $S$  and  $Z$  have different variances. Instead of Eq. 5 for models 1 and 2, models 3 and 4 have (equation 2 in ref. 25, p. 709)

$$\nu_Z = \mu_H^2 \nu_P + \nu_H \mu_P^2 + \nu_H \nu_P. \quad [12]$$

If  $\mu_H \geq 1$ , then  $\nu_Z > \nu_S$ . Ross McVinish pointed out that  $\nu_Z \geq \nu_S$  always; moreover,  $\nu_Z = \nu_S$  if and only if there is zero probability that  $H \geq 2$  (that is, if and only if the number of individuals of the selected host species per square meter is always 0 or 1). (These statements are proved in *SI Appendix*.) Intuitively, unless hosts are rare (with zero or one host individuals per square meter, in which case correlations in the number of parasites among different host individuals living in the same square meter are impossible), the perfect correlation of the hosts' parasite loads strictly increases the variance of the number of parasites per square meter of habitat.

**Model 3: Identical numbers P of parasites per host and power law variance function (TL) of P.** Model 3 assumes that  $P$  obeys TL. In Eq. 12, we replace  $\nu_P$  using TL for parasites (Eq. 6), we replace  $\nu_H$  using TL for hosts (Eq. 2), and we replace  $\mu_P$  using host-parasite density scaling (Eq. 3). Eventually, we get

$$\nu_Z = cf^d \mu_H^{2+dg} + af^2 \mu_H^{b+2g} + acf^d \mu_H^{b+dg}.$$

To replace  $\mu_H$  in each of three terms on the right, we use *SI Appendix*, Eq. S16 three times with the three values of  $K = 2 + dg$ ,  $b + 2g$ ,  $b + dg$ . After simplification, the result is a sum of three power functions:

$$\nu_Z = cf^{\frac{d-2}{g+1}} \mu_S^{\frac{2+dg}{g+1}} + af^{\frac{2-b}{g+1}} \mu_S^{\frac{b+2g}{g+1}} + acf^{\frac{d-b}{g+1}} \mu_S^{\frac{b+dg}{g+1}} \quad [13]$$

(model 3 exact variance function of parasites per square meter).

**Model 4: Identical numbers P of parasites per host and NBD variance function of P.** Model 4 assumes that  $P$  obeys the variance function Eq. 10 of a family of NBDs with constant  $\rho$  and varying  $p$  and that, when  $H > 0$ , every host individual (of the selected species) in  $1 \text{ m}^2$  has an identical number  $P$  of parasites (of the selected species). Calculations similar to those above (making repeated use of *SI*

Appendix, Eq. S16 to go from the first line to the second line) lead to

$$\begin{aligned} \nu_Z &= f\mu_H^{2+g} + \frac{f^2}{\rho}\mu_H^{2+2g} + af\mu_H^{b+g} + (1+\rho^{-1})af^2\mu_H^{b+2g} \\ &= f^{\frac{-1}{1+g}}\mu_S^{\frac{2+g}{1+g}} + \rho^{-1}\mu_S^2 + (1+\rho^{-1})af^{\frac{2-b}{1+g}}\mu_S^{\frac{b+2g}{1+g}} + af^{\frac{1-b}{1+g}}\mu_S^{\frac{b+g}{1+g}} \quad [14] \end{aligned}$$

(model 4 exact variance function of parasites per square meter).

## Empirical and Statistical Methods

**Empirical Methods.** Lagrue et al. (13) described the field sites and the methods of collecting the data. In brief, all metazoan species in the littoral zones of four lakes in Otago, New Zealand were collected and classified as parasitic, free-living with parasites (here hosts), and free-living without parasites. Each lake was sampled multiple times (depending on the sampling method) in each of three field seasons at multiple locations in each lake. Because of the mobility of fish and the impossibility of counting entire fish populations of large areas, estimates of population densities of fish species as individuals-meter<sup>-2</sup> may be subject to larger errors than those of, for example, sessile invertebrates. We measured the population density of each parasite species as individuals-meter<sup>-2</sup> separately for each distinct combination of host species and parasite species. The data structured in this way have not been analyzed previously. We illustrate this method by an example.

In Lake Hayes in September, two species of host insects, *Oecetis* sp. and *Triplectides* sp., were infected with metacercariae of the parasite *Microphalloidea* sp. In 199 samples of *Microphalloidea* sp., 91 were found in *Oecetis* sp., and 108 were found in *Triplectides* sp. To estimate the density per square meter of the parasite *Microphalloidea* sp., we distinguished combinations of *Microphalloidea* sp. with different host species and found 24.9 individuals-meter<sup>-2</sup> *Microphalloidea* sp. in 91 samples in *Oecetis* sp. and 45 individuals-meter<sup>-2</sup> *Microphalloidea* sp. in 108 samples in *Triplectides* sp. [Another method would have been to pool all 199 samples of *Microphalloidea* sp. This method would have yielded 69.9 individuals-meter<sup>-2</sup> *Microphalloidea* sp., regardless of host. We rejected this method in preliminary analyses, because the resulting values of  $\log(\mu_H) + \log(\mu_R)$  did not approximate as well the values of  $\log(\mu_S)$  predicted from Eq. 4.] Population densities of host species were measured as individuals-meter<sup>-2</sup>, regardless of parasites.

Here, we do not use the data on free-living species without parasites. As noted above, unparasitized free-living species had different body size distributions and taxonomic distributions from both parasites and hosts (13). Based on large sample sizes and careful searches for parasites within free-living species, we think that it is unlikely that our distinction between hosts and unparasitized free-living species is artifactual. The data reported in this paper are in Dataset S1.

**Statistical Methods.** We obtained 209 measurements of seven variables for different combinations of host species and parasite species: the mean and the variance of host individuals-meter<sup>-2</sup>, the mean and the variance of parasite individuals per host individual, the mean and the variance of parasite individuals-meter<sup>-2</sup>, and the minimum number of host individuals captured in a sample. This minimum ranged from 0, when a host did not occur in a sample at a particular locality, lake, and season, to 58 hosts. This minimum sample size influenced one of the relationships analyzed below.

Following ref. 13 and many others, we tested power law relationship  $y = ax^b$  and estimated the parameters by the conventional method of taking the common logarithm (to base 10, not the natural logarithm to base  $e$ ) of  $x$  and  $y$  variables and doing least-squares regression. Our justification for using ordinary linear regression was that the variance of the sample mean is much smaller than the variance of the sample variance (ref. 26, p. EV-4, where

the pros and cons of this widespread procedure are discussed). We also tested the adequacy of each linear relationship  $\log y = \log a + b \log x$  by fitting a quadratic generalization (from ref. 27)  $\log y = \log a + b \log x + c(\log x)^2$  and checking whether the quadratic coefficient  $c$  differed significantly from zero. All significance tests used  $\alpha = 0.01$ . If the 99% confidence interval (99% CI) of  $c$  included zero, we inferred that the linear model was acceptable, because there was not compelling evidence in favor of the quadratic alternative. This procedure was conservative in the sense that, had we used a correction for multiple comparisons, the CI for  $c$  would have been wider, and it would have been harder to detect deviations from linearity. Where we accepted the linear model, we would have accepted it after making a correction for multiple comparisons. However, where we rejected the null hypothesis  $c = 0$ , the corrected significance level would have been larger than 0.01.

In all figures, data are solid dots, and theoretical curves are lines (solid, dash-dotted, dashed, or dotted). Computations used Matlab R2015a (28) running under Microsoft Windows 7.

## Empirical Results

**Descriptive Summary of Relationships in Data.** The mean and the variance of host abundances  $H$  per square meter were distinctly bimodal (SI Appendix, Fig. S1, two upper left diagonal histograms). The less abundant mode corresponded to the larger and less abundant fishes, whereas the more abundant mode corresponded to the smaller and more abundant invertebrate hosts. The tightest relationships among six main variables (excluding minimum sample size) were those between the mean and the corresponding variance of each of three measures of abundance:  $H$  (host individuals per square meter),  $P$  (parasites per host individual), and  $S$  (parasites per square meter) (SI Appendix, Fig. S1, off-diagonal scatterplots). In addition, there were clear positive associations between the mean hosts per square meter and mean parasites per square meter and between the variance of hosts per square meter and variance of parasites per square meter.

**Empirical Tests of the Framework Assumptions.** The variance and mean of the number  $H$  of hosts per square meter are described well by TL (Fig. 1A) ( $R^2 = 0.9876$ ). This finding is qualitatively consistent with the finding of table 2 in ref. 13 that TL described well ( $R^2 = 0.9810$ ) what they called “free-living parasitized” species but differs slightly in parameter estimates. Whereas ref. 13 estimated slope = 2.0193 with 95% CI = 1.9739, 2.0646 and intercept = 0.2903, we estimated slope = 2.0856 with 99% CI = 2.0434, 2.1278 and intercept = -0.014318. The discrepancy is because of a different way of organizing the data as described in *Empirical and Statistical Methods*.

On average, the larger the mean number of hosts per square meter, the smaller the mean number of parasites per host (Fig. 1B). On log-log coordinates, the slope -0.24575 of a linear approximation to this relationship is not statistically distinguishable from -1/4, which is a scaling exponent that plays a major role in the metabolic theory of ecology (ref. 29, p. 1775), and a quadratic approximation is not a significant improvement over a linear relationship. The scatter around a linear relationship is the largest among the relationships examined here ( $R^2 = 0.1849$ ), and the error variance 1.229 on the log<sub>10</sub> scale is more than an order of magnitude.

This negative relationship between the mean number of hosts per square meter and the mean number of parasites per host is qualitatively consistent with a finding (figure 3 in ref. 30) in which host density was calculated by pooling individuals of all host species used by a parasite species. Our analysis involves one host species and one parasite species.

The mean number of parasites per square meter is very close to the product of the mean number of hosts per square meter times the number of parasites per host (Fig. 1C) as predicted by Eq. 4.

The variance and mean of  $S$ , the number of parasites per square meter, are described well ( $R^2 = 0.9838$ ) by the empirical TL for parasites per square meter (Fig. 1D):

$$v_S = h\mu_S^k \quad [15]$$

(TL for the number of parasites per square meter).

This finding is qualitatively consistent with the finding of table 2 in ref. 13 that TL described parasitic species well ( $R^2 = 0.9708$ ) but differs slightly in parameter estimates. Whereas ref. 13 estimated slope = 2.1020 with 95% CI = 2.0568, 2.1473 and intercept = 0.4333, we estimated slope = 2.1166 with 99% CI = 2.0675, 2.1657 and intercept = 0.26315. The discrepancy is because of a different way of organizing the data as described in *Empirical and Statistical Methods*.

A consequence of the good agreement with TL with values of the intercept not far from 1 is that the point (0, 0) in Fig. 1D roughly separates mean values of  $S$  greater than 1 (on the right) from mean values of  $S$  less than 1 (on the left) at the same time that it separates variances of  $S$  greater than 1 (above) from variances of  $S$  less than 1 (below).

The parameter estimates of all linear and some quadratic relationships in the text, some of their 99% CIs, and measures of goodness of fit ( $R^2$ , error variance) are in Table 2. The parameters of the quadratic polynomials are reported only when an  $F$  test

showed (by  $p_0 < 0.01$ ) that the quadratic model was a statistically significant improvement over the linear model.

**Empirical Results for Model 1.** TL approximated ( $R^2 = 0.9142$ ) the variance of the number  $P$  of parasites per host individual as a function of the mean number of parasites per host individual (Fig. 2A), but on log-log scales, a convex (curved upward) quadratic relationship was significantly better than TL.

For model 1, using the empirical estimates of the parameters from Table 2 gives the exponent of  $\mu_S$  in the first term on the right side of Eq. 7 as  $(1 + dg)/(g + 1) = 0.8116$  and the exponent of  $\mu_S$  in the second term as  $(b + 2g)/(g + 1) = 2.1135$ . These estimates confirm the expectations presented above under theoretical results. Model 1 does not predict for  $S$ , the number of parasites per square meter, the observed TL, a power law with a single exponent, over the full range of  $\mu_S$  from very small to very large.

Another way to arrive at the same conclusion is to observe that the exponent of  $\mu_S$  in the first term of Eq. 7 equals the exponent of  $\mu_S$  in the second term if and only if  $1 + dg = b + 2g$  or  $b - 1 = (d - 2)g$ . Substituting the empirical estimates of these parameters from Table 2 evaluates the left side as  $b - 1 = 1.0856$  and the right side as  $(d - 2)g = 0.1036$ , different by an order of magnitude.

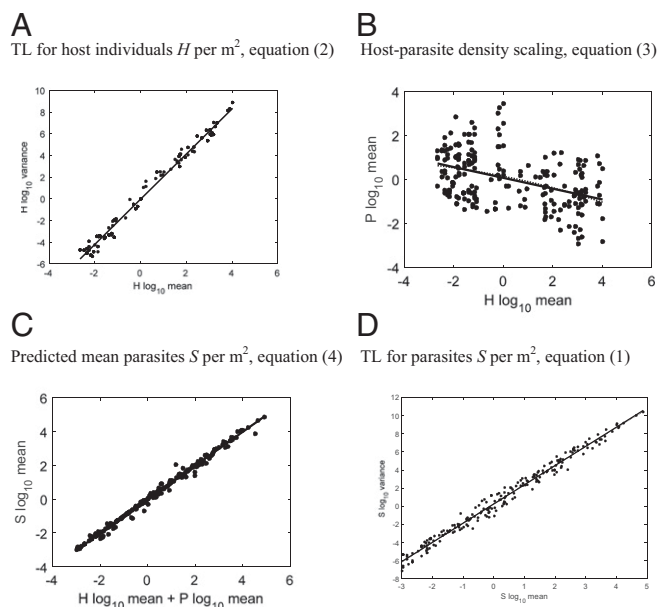
For mean densities or variances of the parasites per square meter greater than one ( $S \log_{10} \text{ mean} > 0$ ), the variance of the number of parasites per square meter is well approximated by a sum of two power functions of the mean number of parasites per square meter (Fig. 3A) as predicted by Eq. 7. Both the predicted slope 2.1135 and the level of the predicted variance  $v_S$  are close to the fitted slope  $k = 2.1166$  and level of the observed variance. However, for lower mean densities and variances, the predicted variance is higher than the empirical variance. On log-log scales, the sum of two power functions of the mean number of parasites per square meter is convex.

**Empirical Results for Model 2.** The empirical number of parasites per host individual seems statistically to have a strictly convex variance function on log-log scales (Fig. 2). The quadratic variance function Eq. 10 of the NBD is closer to the empirical variance function (Fig. 2B and C) than the straight line predicted on log-log scales from TL (Fig. 2A).

For each combination of host species and parasite species (or life stage), the number of hosts used to estimate the mean and variance of the number of parasites per host varied among four lakes and three seasons sampled. When all 209 combinations of host species and parasite species (or life stage) are plotted (Fig. 2B), regardless of the number of hosts sampled, the deviations from the quadratic variance function Eq. 10 of the NBD are greater than when only the 23 host-parasite pairs that had a minimum of 15 hosts sampled in every lake and season are included (Fig. 2C). This comparison suggests that small sample sizes may be at least partly responsible for the deviations from the mean-variance relation of the NBD. This inference does not exclude the possibility that other factors, such as location or season, may be correlated with sample size and may partially explain the deviations from the quadratic variance function Eq. 10 of the NBD.

**Testing the Predicted Variance of the Number of Parasites per Square Meter.**

To test whether the empirical variance of the number of parasites per square meter is well described by the variance predicted by Eq. 11 requires estimates of the parameters on the right side of Eq. 11. Table 2 gives (after rounding to four decimal places)  $a = 0.9676$ ,  $b = 2.0856$ ,  $f = 1.2175$ ,  $g = -0.2458$ , and  $\rho^{-1} = 2$ . The predictions are calculated using the unrounded parameter estimates. Then the exponents of  $\mu_S$  in the three terms on the right side of Eq. 11 are 1,  $(1 + 2g)/(1 + g) = 0.6742$ , and  $(b + 2g)/(1 + g) = 2.1135$ .



**Fig. 1.** Tests of assumptions of the models. (A) Test of TL for hosts  $H$  per square meter. The solid line is the log-log form of TL (Eq. 2) for host individuals per square meter. It is superposed on the dotted line of the quadratic generalization of TL. (B) Test of host-parasite density scaling. The mean number of parasites per host  $P$  is a decreasing power law function of the mean number of hosts per square meter  $H$ . The solid line is the log-log form of the power law (Eq. 3) model of host-parasite density scaling. The dotted line is a quadratic generalization, which is not significantly better. (C) Test of the product rule. The mean number of parasites per square meter is closely approximated by the product of the mean number of hosts per square meter times the mean number of parasites per host individual as predicted by the product rule (Eq. 4). The intercept does not differ significantly from zero, and the slope does not differ significantly from one. (D) Test of TL for  $S$ , the number of parasites per square meter. The power law relationship, linear on log-log coordinates (solid black line), approximates well the relation between the sample mean and sample variance of  $S$  (solid black dots). Table 2 gives parameter estimates of all linear and some quadratic relationships in the text, some of their 99% CIs, and measures of goodness of fit ( $R^2$ , error variance). All panels have 209 data points.



**Table 2. Parameter estimates and associated statistics**

Subject, equation, and parameter	Estimate	Lower CI	Upper CI	$R^2$	$p$	Error variance	$p_0$
Hosts per 1 m <sup>2</sup>							
2							
Slope $b$	2.0856	2.0434	2.1278	0.9876	2.42E-199	0.2520	0.8716
Intercept	-0.0143	-0.1053	0.0767				
$a$	0.9676	0.7847	1.1931				
Parasites per host							
6							
Slope $d$	1.5784	1.4910	1.6657	0.9142	2.30E-112	0.3524	1.91E-06
Intercept	0.8899	0.7832	0.9967				
$c$	7.7610	6.0695	9.9238				
Quadratic	0.0961	0.0451	0.1470				
Slope	1.5508	1.4666	1.6349				
Intercept	0.7470	0.6204	0.8735				
Host-parasite scaling							
3							
Slope $g$	-0.2458	-0.3390	-0.1525	0.1849	8.14E-11	1.2290	0.4985
Intercept	0.0855	-0.1155	0.2864				
$f$	1.2175	0.7665	1.9338				
Product rule for means							
4 log-log							
Slope	1.0006	0.9857	1.0155	0.9933	1.12E-226	0.0262	0.8989
Intercept	-0.0131	-0.0426	0.0163				
Parasites per 1 m <sup>2</sup>							
15							
Slope $k$	2.1166	2.0675	2.1657	0.9838	3.19E-187	0.2872	0.4065
Intercept	0.2632	0.1658	0.3605				
$h$	1.8329	1.4650	2.2933				

There were 209 observations for each regression here. The parameters  $a$ ,  $c$ ,  $f$ , and  $h$  equal 10 raised to the power of the intercept on the row immediately above.  $p$  is the probability of the observed linear relationship under the null hypothesis that the slope is zero.  $p_0$  is the probability of the observed improvement in fit of a quadratic regression over a linear regression under an  $F$  test of the null hypothesis that the quadratic coefficient is zero.

For low densities ( $\mu_S < 1$ ), the predicted variance  $\nu_S$  is too high. For larger densities, the predicted variance  $\nu_S$  is close to the observed variance (Fig. 3B).

The NBD gives a better model of the variance function of parasites per host (Fig. 2B and C) than TL (Fig. 2A). Neither TL (model 1) nor NBD (model 2) accurately approximates the empirical variance of the number of parasites per area at low mean densities and low variances. Both models successfully describe the observed variances at high densities of parasites per square meter.

**Empirical Results for Model 3.** Model 3 assumes perfect correlation of the parasite loads in different hosts in the same square meter, leading to variance function Eq. 13. In this case,  $\nu_Z$ , the predicted variance of  $Z$ , is linearly related to and exceeds only slightly the observed sample variance over the whole range of sample variance of  $S$  (Fig. 3C). Model 3 also predicts a frequency histogram of the marginal distribution of the variance that resembles the observed frequency histogram of the marginal distribution of the variance (SI Appendix, Fig. S2).

**Empirical Results for Model 4.** Model 4 assumes perfect correlation of the parasite loads in different hosts in a square meter, leading to variance function Eq. 14. In this case,  $\nu_Z$ , the predicted variance of  $Z$ , is linearly related to and exceeds only slightly the observed sample variance over the whole range of sample variance of  $S$  (Fig. 3D). Model 4 also predicts a frequency histogram of the marginal distribution of the variance that resembles the observed frequency histogram of the marginal distribution of the variance (SI Appendix, Fig. S2).

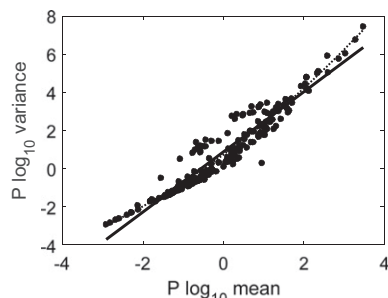
**Summary Comparisons of Four Models.** The variance functions for  $S$  or  $Z$ , the number of parasites per square meter, of all four models

have the same general mathematical form: they are a sum of powers of  $\mu_S$ , the mean number of parasites per square meter, with different exponents. They differ in the number of summands and the values of the coefficients and exponents. A wide range of exponents should produce curvature on a log-log variance plot, because the largest exponent dominates when  $\mu_S$  is large and the smallest exponent dominates when  $\mu_S$  is small. If all of the exponents in a variance function were identical, then the variance function would be a simple power law, like TL, and a log-log variance plot should be exactly linear (apart from sampling fluctuations in the sample mean and sample variance). This observation suggests that the smaller the range of exponents in the variance function of a model, the more nearly that model's variance function should resemble the empirical TL for the number of parasites per square meter.

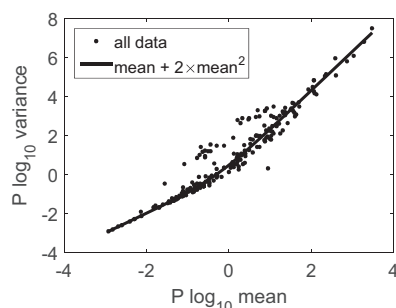
To test this suggestion, the exponents of every term in each model are assembled in Table 3. The ranges of each model's exponents (i.e., the largest exponent minus the smallest exponent) are shown below the exponents along with two measures of the lack of fit between the variances predicted by each model and the observed sample variances. The first measure is the SD of the residuals (differences) between the  $\log_{10}$  sample variance of  $S$  and the  $\log_{10}$  predicted variance of  $S$ . Here, the prediction is based on the sample mean of  $S$  associated with each sample variance, when this sample mean is inserted into the formula for the variance of  $S$  derived for each model. To convert this measure on the  $\log_{10}$  scale to the original scale on which the variance of  $S$  is measured, the last line of Table 3 shows  $10^{\text{SD}}$ .

The ranges of exponents of models 1 and 2 are roughly 10 times larger than the range of exponents of model 3 and three or four times larger than the range of exponents of model 4. As the argument above suggests that they should be, the SDs of models 1 and 2 are roughly three times the SDs of models 3 and 4 on the

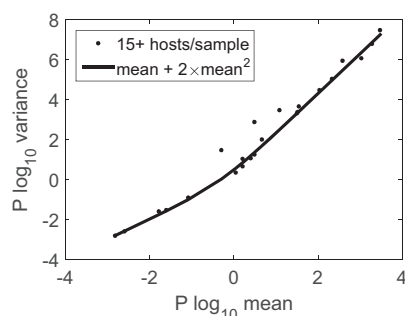
### A TL (6) for parasites $P$ per host individual



### B Negative binomial model variance function (10) for $P$ , all data



### C Negative binomial model variance function (10) for $P$ , sample size $\geq 15$



**Fig. 2.** Variance function of parasites  $P$  per host individual. Variance of the number  $P$  of parasites per host ( $\cdot$ ) as a function of the mean number of parasites per host. (A) The quadratic generalization of TL (dotted line) provides a significantly better fit than TL (Eq. 6) (solid line). All 209 data points, regardless of sample size, are included. (B) Variance function of  $P$  ( $\cdot$ ) compared with  $\text{mean} + 2 \times \text{mean}^2$  (solid black line). The NBD of the number  $P$  of parasites per host (which was posited in models 2 and 4) predicts that  $\text{variance} = \text{mean} + 2 \times \text{mean}^2$  according to Eq. 10. The value  $\rho^{-1} = 2$  was estimated by numerical experimentation. Calculations involving the NBD were carried out on the original scale of measurement (parasite individuals per host) and are plotted here on log-log scales for comparability with other figures. Data ( $\cdot$ ) are from all hosts without regard to the minimum number of hosts per estimate of the mean and variance of  $P$ . Number of mean-variance pairs (data points), 209; root-mean-squared error (RMSE) on the original scale of measurement (parasite individuals per host) =  $3.22 \times 10^5$ . (C) Minimum sample size of 15 hosts per estimate of mean and variance of  $P$ . Number of mean-variance pairs (data points), 23; RMSE on the original scale of measurement (parasite individuals per host) = 165.

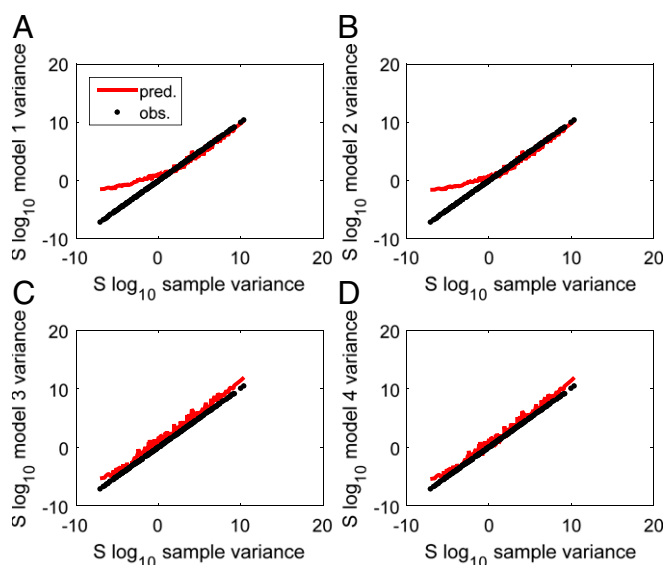
$\log_{10}$  scale, and  $10^{\text{SD}}$  is more than an order of magnitude larger for models 1 and 2 than for models 3 and 4. This qualitative difference between models 1 and 2 on the one hand and models 3 and 4 on the other hand is reflected in the systematic difference in shape between the theoretical and observed variance functions in Fig. 3. Model 3, the best fitting model, has the smallest range of exponents

and the lowest values of SD and  $10^{\text{SD}}$ , but its advantage over model 4 is small. These results suggest that the decisive difference between the more successful models 3 and 4 and the less successful models 1 and 2 is the assumption in models 3 and 4 that parasite loads of different hosts in  $1 \text{ m}^2$  are highly (here, perfectly) correlated by contrast with the assumption in models 1 and 2 that parasite loads of different hosts in  $1 \text{ m}^2$  are uncorrelated. This difference matters far more than whether the variance function of parasites per host obeys TL or NBD.

### Discussion

Motivated by a desire to embed the ecology of parasites more firmly within the framework of general ecology, we developed data and models to link the distribution of parasites in hosts with the distribution of parasites in space.

**Discussion of Empirical Results.** Empirically, we confirmed TL for hosts per square meter (Fig. 1A) and parasites per square meter (Fig. 1B) after organizing field data by pairing each host species with each parasite species as the basic unit of analysis. We also found empirically that the log variance of parasites per host was better described as a strictly convex function of the log



**Fig. 3.** Predictions of the variance of  $S$ , the number of parasites per square meter, from four models. For each value of the horizontal axis, which is the  $\log_{10}$  sample variance of  $S$  in all four panels, the black dot falling along the diagonal shows the same value on the vertical axis as a standard of perfect agreement for comparison with the predicted  $\log_{10}$  variance of  $S$  from each model shown by the red continuous curve. The predicted variance of  $S$  is computed from the observed sample mean of  $S$  according to the variance function of  $S$  derived theoretically for each model using the parameter estimates  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $f$ ,  $g$ , and  $\rho$  estimated independently from other relationships. There is no curve fitting or adjustment of parameters between the sample variance of  $S$  (black dots) and the theoretical variance of  $S$  (red curve). The red and black points would be superimposed if the model's predicted variance of  $S$  matched perfectly the sample variance of  $S$  and if there were no sampling variability in the observations. (A) Model 1 assumes that parasite numbers per host  $P$  are independent with TL variance function. (B) Model 2 assumes that parasite numbers per host  $P$  are independent with negative binomial variance function. (C) Model 3 assumes that parasite numbers  $P$  are identical in all hosts with TL variance function. (D) Model 4 assumes that parasite numbers  $P$  are identical in all hosts with negative binomial variance function. Only models 3 and 4 predict a variance of  $S$  close and linearly related to the observed sample variance of  $S$  over the whole range of the observed sample variance of  $S$ . All panels have 209 data points. obs., observed; pred., predicted.



**Table 3. Exponents of  $\mu_5$  and their numerical values, the ranges of the exponents, and summaries of the deviations between  $\log_{10}$  sample variance of  $S$  and  $\log_{10}$  predicted variance of  $S$  in models 1–4**

Exponent of $\mu_5$	Model 1	Model 2	Model 3	Model 4
$(1+d \times g)/(1+g)$	0.8116			
$(b+2 \times g)/(1+g)$	2.1135	2.1135	2.1135	2.1135
1		1		
$(1+2 \times g)/(1+g)$		0.6742		
$(2+d \times g)/(1+g)$			2.1374	
$(b+d \times g)/(1+g)$			2.2509	
$(2+g)/(1+g)$				2.3258
$(2+2 \times g)/(1+g)$				2
$(b+g)/(1+g)$				2.4393
Range of exponents (high minus low)	1.3019	1.4393	0.1374	0.4393
SD of residuals between $\log_{10}$ sample variance of $S$ and $\log_{10}$ predicted variance of $S$	1.6874	1.6313	0.5540	0.5605
$10^{\text{SD of } \log_{10} \text{ residuals}}$	48.6810	42.7898	3.5807	3.6353

mean parasites per host (Fig. 2), contrary to TL, but in accordance with the NBD. The nonlinear (log–log) relationship became clearer when small sample sizes with fewer than 15 observations were excluded (Fig. 2C). This convexity in the log variance of the number of parasite individuals per host as a function of the log mean of the number of parasite individuals per host differs from some prior findings (2, 7, 15) but is consistent with the variance function of the NBD, which has been widely confirmed (3, 16) or assumed for the distribution of parasite individuals per host individual.

We showed empirically that the product rule Eq. 4 holds. To high precision, the mean number of parasites per square meter is the product of the mean number of parasites per host times the mean number of hosts per square meter (Fig. 1C). This agreement is not a tautology or accounting identity. The empirical agreement with the product rule is compatible with the assumption of independence (conditional on the square meter or value of  $\theta$ ) between  $P$  (parasites per host) and  $H$  (hosts per square meter) or perfect correlation of  $P$  among hosts in a square meter.

Prompted by the goal of developing a theory to relate the number of parasites per host individual to the number of parasites per square meter of habitat, we posited on theoretical grounds a relationship (Eq. 3) called “host–parasite density scaling,” which was consistent with our data (Fig. 1B). This negative relationship summarizes the broad tendency of the mean parasite density per host to decline as the mean host density per square meter increases. For mathematical convenience in working with the power law of TL, we picked a power law for the mathematical form of this relationship, recognizing that the widely scattered data are compatible with other ways of expressing it. Except for a qualitatively similar finding from different analyses of the raw data (30), we are not aware that host–parasite density scaling (Eq. 3) has been previously posited theoretically or supported empirically.

The tendency of the mean parasite density per host to decline as the mean host density per square meter increases may be caused by multiple mechanisms, including herd immunity, dilution (when a constant input of infectious propagules is distributed over a larger number of potential hosts), or host body size (bigger hosts are rarer, and each individual host can accommodate more parasites). Determining which of these mechanisms or

others accounts for negative host–parasite density scaling remains a project for future research.

**Discussion of Theoretical Results.** Our four theoretical models make three assumptions. The first assumption is that the total number of parasites in  $1 \text{ m}^2$  is the sum of the numbers of parasites in all of the hosts in that  $1 \text{ m}^2$ . Models 1 and 2 assume that the number of parasites in one host individual is independent of the numbers of parasites in all other host individuals in the same  $1 \text{ m}^2$ . Models 3 and 4 assume that the numbers of parasites in every host individual are identical, although independent of the numbers of hosts in the  $1 \text{ m}^2$ . The second assumption is that the mean numbers of parasites per host are a power law function of the mean numbers of hosts per square meter. The third assumption is that the numbers of hosts per square meter obey TL. We verified the second and third assumptions empirically.

Under these assumptions, we found mathematically that, if the numbers of parasites per host obey TL, then the variance function of the numbers of parasites per square meter is a sum of two (model 1) or three (model 3) power functions of the mean numbers of parasites per square meter with different exponents and therefore, could not, in general, satisfy TL exactly. However, asymptotically for large mean numbers of parasites per square meter and also asymptotically for small mean numbers of parasites per square meter, the variance function approaches linearity on log–log coordinates. The slopes differ in the large and small limits when the numbers of parasites per host are uncorrelated (model 1). The slopes differ little in the large and small limits, at least for the observed parameter values, when the numbers of parasites per host are perfectly correlated (model 3).

Under the same assumptions, we showed that, if the variance function of parasites per host obeys the quadratic relationship (without constant term) of NBD, then the variance function of the numbers of parasites per square meter is a sum of three (model 2) or four (model 4) power functions of the mean numbers of parasites per square meter with different exponents and therefore, could not, in general, satisfy TL exactly. Again, however, asymptotically for large mean numbers of parasites per square meter and also asymptotically for small mean numbers of parasites per square meter, the variance function approaches linearity on log–log coordinates. The slopes differ considerably in the large and small limits when the numbers of parasites per host are uncorrelated (model 2). The slopes differ little in the large and small limits, at least for the observed parameter values, when the numbers of parasites per host are perfectly correlated (model 4).

Although TL cannot simultaneously hold exactly for  $P$ ,  $H$ , and  $S$  under our general assumptions, model 3 [which posits TL for  $P$  (parasites per host) and perfect correlation of  $P$  for different hosts in a square meter] and model 4 (which posits an NBD variance function for  $P$  and perfect correlation of  $P$  for different hosts in a square meter) reproduce reasonably well the observed TL for the distribution of parasites in space. According to model 3, TL can hold approximately for parasites per host and parasites per square meter. The widely analyzed, empirically supported NBD fits the observed variance function of parasites per host better than TL, although model 4, which incorporates the NBD variance function of parasites per host, fits the observed variance function of parasites per square meter slightly worse than model 3.

The predictions of models 3 and 4 of the variance function of the number of parasites per square meter have the right shape, unlike those of models 1 and 2, but are slightly systematically too high relative to the empirical variance of the number of parasites per square meter (Fig. 3 C and D). It is a standard fact in statistics that the variance of a sum of correlated random variables increases with the average correlation among them. Therefore the excess in the predicted variance is very likely to be caused by the assumption of perfect rather than high but imperfect average correlation of the parasite loads per host individual. A slight lowering of that assumed

level of correlation should adjust the level of the predicted variance to that observed.

## Conclusions

This analysis draws attention to the key importance of interhost correlations in parasite loads in accounting for the spatial variance of parasite population densities. Our empirical experience, not formalized here, strongly suggests that the parasite loads of different host individuals within a small area, such as 1 m<sup>2</sup>, are very likely to be more similar to each other than to the parasite loads of host individuals from a distant square meter, because there are hotspots of infection even on small spatial scales. The correlation among parasite loads of

different host individuals from the same square meter will never be 1 but will be somewhere between 0 and 1. We are unable to point to a field study that measures this correlation specifically. Future empirical research should measure directly interhost correlations in parasite loads at local (square meter) and large spatial scales.

**ACKNOWLEDGMENTS.** J.E.C. acknowledges the assistance of Priscilla K. Rogerson. R.P. and C.L. thank Anne Besson, Isa Blasco-Costa, Manna Warburton, and Kim Garrett for assistance with field collection and laboratory processing of samples. We thank the referees for constructive criticisms and Bob Lester for a useful suggestion. This work was supported by US National Science Foundation Grant DMS-1225529 (to J.E.C.). A grant from the Marsden Fund (R.P.) funded the empirical portion of this study.

1. Anderson RM, May RM (1978) Regulation and stability of host-parasite population interactions. I. Regulatory processes. *J Anim Ecol* 47:219–247.
2. Shaw DJ, Dobson AP (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: A quantitative review. *Parasitology* 111(Suppl):S111–S133.
3. Shaw DJ, Grenfell BT, Dobson AP (1998) Patterns of parasite aggregation and the negative binomial distribution. *Parasitology* 117(Pt 6):597–610.
4. Fredensborg BL, Mouritsen KN, Poulin R (2004) Intensity-dependent mortality of *Paracalliope novizealandiae* (Amphipoda: Crustacea) infected by a trematode: Experimental infections and field observations. *J Exp Mar Biol Ecol* 311:253–265.
5. Benesh DP (2011) Intensity-dependent host mortality: What can it tell us about larval growth strategies in complex life cycle helminths? *Parasitology* 138(7):913–925.
6. Kuris AM, et al. (2008) Ecosystem energetic implications of parasite and free-living biomass in three estuaries. *Nature* 454:515–518.
7. Morand S, Guégan JF (2000) Distribution and abundance of parasite nematodes: Ecological specialisation, phylogenetic constraint or simply epidemiology? *Oikos* 88(3):563–573.
8. Taylor LR (1961) Aggregation, variance and the mean. *Nature* 189(4766):732–735.
9. Bliss CI (1941) Statistical problems in estimating populations of Japanese beetle larvae. *J Econ Entomol* 34:221–232.
10. Fracker SB, Brischle HA (1944) Measuring the local distribution of Ribes. *Ecology* 25(3):283–303.
11. Hayman BI, Lowe AD (1961) The transformation of counts of the cabbage aphid (*Brevicoryne brassicae* (L.)). *N Z J Sci* 4:271–278.
12. Hechinger RF, Lafferty KD, Dobson AP, Brown JH, Kuris AM (2011) A common scaling rule for abundance, energetics, and production of parasitic and free-living species. *Science* 333(6041):445–448.
13. Lagrue C, Poulin R, Cohen JE (2015) Parasitism alters three power laws of scaling in a metazoan community: Taylor's law, density-mass allometry, and variance-mass allometry. *Proc Natl Acad Sci USA* 112(6):1791–1796.
14. Davidian M, Carroll RJ (1987) Variance function estimation. *J Am Stat Assoc* 82(400):1079–1091.
15. Poulin R (2013) Explaining variability in parasite aggregation levels among host samples. *Parasitology* 140(4):541–546.
16. Crofton HD (1971) A quantitative approach to parasitism. *Parasitology* 62:179–193.
17. Pielou EC (1977) *Mathematical Ecology* (Wiley, New York).
18. Yamamura K (1990) Sampling scale dependence of Taylor's power law. *Oikos* 59(1):121–125.
19. Ross SM (1997) *Introduction to Probability Models* (Academic, Amsterdam), 6th Ed.
20. Ghosh JK, Delampady M, Samanta T (2006) *An Introduction to Bayesian Analysis: Theory and Methods* (Springer, New York).
21. Yamamura K (2000) Colony expansion model for describing the spatial distribution of populations. *Popul Ecol* 42:161–169.
22. Taylor LR, Woiwod IP, Perry JN (1979) The negative binomial as a dynamic ecological model and the density-dependence of  $k$ . *J Anim Ecol* 48(1):289–304.
23. Perry JN, Taylor LR (1986) Stability of real interacting populations in space and time: Implications, alternatives and the negative binomial  $k_c$ . *J Anim Ecol* 55(3):1053–1068.
24. Morand S, Krasnov B (2008) Why apply ecological laws to epidemiology? *Trends Parasitol* 24(7):304–309.
25. Goodman LA (1960) On the exact variance of products. *J Am Stat Assoc* 55:708–713.
26. Cohen JE, Lai J, Coomes DA, Allen RB (2016) Taylor's law and related allometric power laws in New Zealand mountain beech forests: The roles of space, time and environment. *Oikos* 125:1342–1357.
27. Taylor LR, Woiwod IP, Perry JN (1978) The density-dependence of spatial behaviour and the rarity of randomness. *J Anim Ecol* 47(2):383–406.
28. MathWorks (2015) MATLAB R2015a (Mathworks Inc., Natick, MA).
29. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a metabolic theory of ecology. *Ecology* 85:1771–1789.
30. Lagrue C, Poulin R (2015) Bottom-up regulation of parasite population densities in freshwater ecosystems. *Oikos* 124:1639–1647.
31. Kingman JFC (1961) A convexity property of positive matrices. *J Quart Math* 12(1):283–284.