PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Johnson PTJ, Wilber MQ. 2017 Biological and statistical processes jointly drive population aggregation: using host–parasite interactions to understand Taylor's power law. *Proc. R. Soc. B* **284**: 20171388. http://dx.doi.org/10.1098/rspb.2017.1388

Received: 21 June 2017
Accepted: 10 August 2017

Subject Category:
Ecology

Subject Areas:
ecology, health and disease and epidemiology

Keywords:
macroecology, disease ecology, superspreaders, community ecology, population regulation, feasible sets

Author for correspondence:
Pieter T. J. Johnson
e-mail: pieter.johnson@colorado.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3861841.

THE ROYAL SOCIETY PUBLISHING

Biological and statistical processes jointly drive population aggregation: using host–parasite interactions to understand Taylor's power law

Pieter T. J. Johnson[1] and Mark Q. Wilber[2]

[1]Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA
[2]Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA, 93106, USA

PTJJ, 0000-0002-7997-5390

The macroecological pattern known as Taylor's power law (TPL) represents the pervasive tendency of the variance in population density to increase as a power function of the mean. Despite empirical illustrations in systems ranging from viruses to vertebrates, the biological significance of this relationship continues to be debated. Here we combined collection of a unique dataset involving 11 987 amphibian hosts and 332 684 trematode parasites with experimental measurements of core epidemiological outcomes to explicitly test the contributions of hypothesized biological processes in driving aggregation. After using feasible set theory to account for mechanisms acting indirectly on aggregation and statistical constraints inherent to the data, we detected strongly consistent influences of host and parasite species identity over 7 years of sampling. Incorporation of field-based measurements of host body size, its variance and spatial heterogeneity in host density accounted for host identity effects, while experimental quantification of infection competence (and especially virulence from the 20 most common host–parasite combinations) revealed the role of species-by-environment interactions. By uniting constraint-based theory, controlled experiments and community-based field surveys, we illustrate the joint influences of biological and statistical processes on parasite aggregation and emphasize their importance for understanding population regulation and ecological stability across a range of systems, both infectious and free-living.

1. Introduction

For over five decades scientists have debated the driving forces underlying the macroecological scaling relationship known as Taylor's power law (TPL), which is often used to study the degree of aggregation within or among populations [1–7]. Originally outlined in a short paper to show that variance in population density increases as a power function of its mean [1], TPL has become one of the most widely verified scaling 'rules' in ecology, with empirical support from over 500 species of viruses, protists, arthropods, plants, birds, mammals and even humans [8–12]. Part of the appeal of TPL is its simplicity: plotted in log-log space ($\log s^2 = \log a + b \log m$), the variance ($s^2$) in density is a strongly linear function of the mean (m), for which the slope (b) measures the degree of aggregation [13]. Thus, higher values of b indicate that the variance in density (i.e. aggregation) is increasing more steeply with mean density relative to low values. Unlike other aggregation measures that change with the mean (e.g. the negative binomial parameter k and the variance-to-mean ratio), b captures variation across widely varying mean values and is thus especially suited for comparisons over time and space. While early debates focused on whether such patterns emerged from species' intrinsic behaviours or demographic stochasticity [2,3,10], subsequent extensions of TPL to a wide range of non-biological systems, including stock market fluctuations, precipitation patterns,

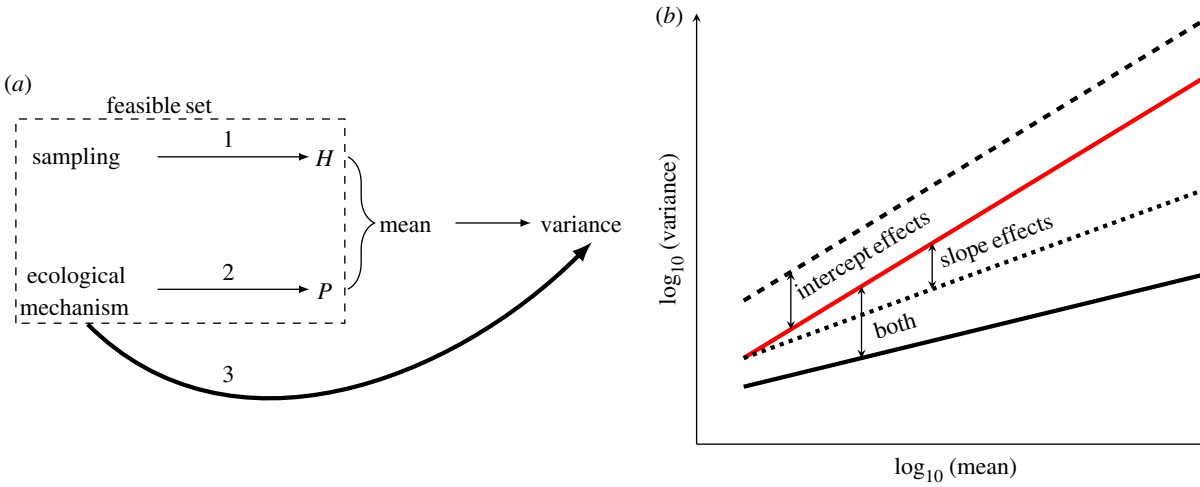

Figure 1. Conceptual diagrams displaying the two steps used in our analysis of Taylor's power law, which focuses on the scaling relationship between the sample mean and its variance. (*a*) Ecological mechanisms can affect the variance of a distribution either indirectly through the mean (arrows 1 and 2; P = total number of individuals in a sample, e.g. parasites; H = total number of units sampled, e.g. hosts, or directly; arrow 3). Feasible set theory removes these indirect effects to focus on mechanisms that directly affect variance. (*b*) Taylor's power law is often represented as the positive, linear relationship between log-transformed values of the mean and the variance. Ecological mechanisms can change the intercept, the slope or both parameters in this relationship. (Online version in colour.)

and tornado outbreaks [12,14], have illustrated that no single explanation drives this relationship. The influence of aggregation on system stability, whether in ecological food webs or human social structures, continues to motivate research on potential applications of TPL ranging from sampling agricultural pests to protecting endangered species [15,16].

One of the most recurrent questions surrounding TPL involves the degree to which variation in the slope (*b*) reflects underlying biological processes or is instead the product of statistical or sampling artefacts [17–19]. Variation in sample size, sampling method, spatial and temporal scale, and the range of population densities considered all have the potential to influence the log-mean–log-variance relationship [2,9,20,21]. Thus, rather than comparing observations to a static null value (e.g. the slope of unity from a Poisson distribution), recent evidence highlights the importance of approaches that incorporate constraints imposed by the specific dataset [5,22]. For example, Xiao *et al*. [22] used 'feasible set' theory to show that the general form and slope of TPL were an emergent function of the number of individuals and groups sampled (figure 1*a*), indicating that the classical use of a Poisson null distribution is probably inadequate. Similarly, we showed in a previous study that feasible set theory could also predict the general form of aggregated host–parasite distributions [23]. Nonetheless, these previous studies show that constraint-based approaches on their own cannot typically predict either specific slope values within empirical data or capture all of the variability in an observed distribution [23–25]. The open challenge therefore becomes identifying biological mechanisms that influence aggregation (i.e. the slope of TPL) after accounting for statistical constraints imposed by the data structure.

In disease ecology, aggregation of parasites within hosts is among the most foundational and universally observed patterns [11,26–30], making infection an especially appropriate process in which to investigate biological applications of TPL [31,32]. For instance, TPL has been used effectively to characterize heterogeneity in outbreak size, superspreader distributions and the competence of reservoir hosts [30,33–35]. Building from epidemiological theory, aggregation in infection emerges from factors that cause heterogeneity in host exposure or susceptibility [29,36], including variation in host traits (e.g. immunity, sex, behaviour, genetics and age/size), spatial or temporal variation in exposure, parasite-induced mortality, and interspecific competition [28,37]. Such effects can be intensity-independent, altering the intercept of the log-mean–log-variance relationship, or intensity-dependent, altering the slope, which means that the influence varies with parasite abundance (figure 1*b*). While many studies have demonstrated the utility of TPL for characterizing parasite distributions or evaluating the influence of selected factors on aggregation [11,32,38], efforts to systematically test hypothesized drivers of such variation are often limited by data inconsistencies: meta-analytic compilations of information across divergent systems, geographical areas, sampling methods and time scales frequently introduce discrepancies. Moreover, key mechanistic processes—such as host susceptibility and parasite-induced mortality—are difficult to infer from field data alone [23,39], emphasizing the necessity of complementary experiments [11,28].

Here we integrate quantitative theory, large-scale field surveys and controlled experiments to explicitly test the roles of biological processes in driving variation in parasite aggregation. Using interactions between amphibian hosts and trematode parasites as a tractable multi-host, multi-parasite system, we measured mean infection and its variance among 11 987 individual hosts and 971 populations. After showing that feasible set theory captured both the central tendency of, and nonlinear deviations from, host–parasite TPLs, we evaluated the slope and intercept effects of empirically quantified covariates related to variation among populations, species and their environments [16,20] (figure 1). We coupled field surveys with controlled infection experiments for 20 of the most common host–parasite interactions to estimate essential epidemiological outcomes, such as host competence and parasite virulence. By comparing observed patterns in aggregation to the null model provided through feasible set theory, we show that (i) aggregation varied sharply yet consistently among host and parasite species, (ii) factors that enhanced heterogeneity within host populations (e.g. variation in body size and spatial distribution in host density) increased aggregation in an intensity-independent manner by elevating the intercept of TPL, and (iii) increases in

infection competence (and especially virulence) reduced the slope of TPL, consistent with theoretical predictions but largely untested empirically. Overall, our results reveal that the shape of TPL is the joint product of both predictable, system-independent statistical properties and measurable, system-specific biological mechanisms.

2. Material and methods

(a) Field sampling: community-based profiles of parasite aggregation

Between 2009 and 2015, we collected data on amphibian hosts, snail hosts and their trematode parasite infections from 173 ponds in the East Bay region of California [40,41]. We visited each wetland twice to determine community composition, the larval density of each amphibian host species, the density of snail intermediate hosts and the corresponding information on their trematode parasites (see electronic supplementary material). For each site, we measured body size (snout–vent length) and quantified infections among 10–15 late-stage larvae or recently metamorphosed individuals from all non-endangered amphibian species (*Pseudacris regilla*, *Anaxyrus boreas*, *Lithobates catesbeianus*, *Taricha torosa* and *T. granulosa*). We focused on infection by larval trematodes (see methods in [40]), for which four species comprise 99% of observed infections: *Ribeiroia ondatrae*, *Echinostoma* spp., *Alaria marcianae* and *Cephalogoniumus americanus*. Although each parasite uses freshwater snails as first intermediate hosts and amphibians as second intermediate hosts, they vary widely in average infection abundance, spatial occurrence among ponds, and degree of host specificity [41,42].

(b) Experimental infections: quantifying host competence and parasite virulence

Over 4 years, we collected egg masses of each amphibian species from field sites, hatched them in the laboratory, and allowed larvae to develop to Gosner [43] stage 28 for anurans and Wong & Liversage [44] stage 2T for newts. For rough-skinned newts (*T. granulosa*), which lay solitary eggs that are difficult to find in natural habitats, we collected adults and allowed them to breed under simulated field conditions. Individual larvae were maintained in 1.5 l containers and exposed to a specific dosage of one of four trematode species: *Ribeiroia ondatrae*, *Echinostoma trivolvis*, *Alaria marcianae* and *Cephalogonimus americanus* (electronic supplementary material). The target dosages were 0 (control), 20, 40, 100, 200 and 500 cercariae per host with an average of 12 replicates per treatment, for a total of 1248 experimental hosts. At 20 days post-exposure, we recorded the fraction of surviving hosts in each treatment and measured the number of trematode meta- or mesocercariae per host. We measured competence as the percentage of administered parasites persisting after 20 days and verified that this measure was independent of dose (see also [42]). To calculate virulence, we used firth-adjusted logistic regression to estimate the effect of parasite dose on host survival over 20 days. The specific measure of virulence was the slope of the logistic regression, indicating the increase in the log-odds of mortality in response to a unit increase in initial parasite dose, when evaluated at the mean dose. Firth-adjusted logistic regression corrects for biased parameter estimates when sample size is moderate to small and the events which are being modelled are relatively rare [45].

(c) Analysis of aggregation

In a previous study, we demonstrated the utility of applying feasible set theory to amphibian hosts and their trematode parasites, showing that the total number of parasites P and the total number of hosts H strongly constrained the shape of a host–parasite distribution [23]. Building from those results, here we used a feasible set approach to account for inherent constraints imposed by the total number of parasites and hosts sampled on the shape of TPL (figure 1*a*; see electronic supplementary material for a full description). To generate a candidate configuration in a feasible set, P parasites are randomly partitioned into exactly H hosts. One random configuration corresponds to a potential distribution of P parasites among H hosts. This random partitioning is repeated many times, and the sampled configurations are an estimate of the feasible set given P and H [22,46]. The feasible set approach makes no mechanistic assumptions but predicts the most likely aggregation level given data constraints (see electronic supplementary material), effectively recognizing that many combinations of mechanisms can generate similar levels of aggregation [25,47]. For every observed host–parasite population in the field sampling described above, we extracted the total number of parasites P and total number of hosts H and randomly sampled 1000 configurations from the feasible set. We computed the log variance of the number of parasites per host in each configuration and took the median log variance as the feasible set-predicted log variance (see electronic supplementary material for justification). We then took the difference between the observed log-variance in infection and the predicted log-variance for all observed host–parasite populations [22] (electronic supplementary material). This residual log-variance was the response variable used for all of the analyses described below. For all analyses, log refers to $\log_{10}$.

To investigate how host, parasite and host-by-parasite interactions affected patterns of parasite aggregation, we used linear mixed effects models with an information-theoretic approach. Residual log-variance in infection (see above) was modelled as a normally distributed response and we included log-mean infection as a fixed effect with sampling year and pond identity as random intercept terms (see electronic supplementary material for justification). For a given covariate, we distinguished between two types of potential effects on the log-mean residual log-variance relationship: intercept effects and slope effects (figure 1*b*). All continuous predictor variables were scaled to have a mean of zero and a standard deviation of one. To remove the 'forbidden zone' [48], we only included populations that had $\geq$3 hosts and $\geq$4 parasites, which tended to eliminate host-by-parasite combinations with very low competence (260 populations removed). After further removing any populations with missing values for candidate predictor variables, our dataset consisted of 971 host-by-parasite distributions, 11 987 hosts, and 332 684 parasites. Here, each of these 971 points represents a unique combination of host species, parasite species, wetland identity and sample year.

To determine whether host and parasite species had consistent effects on the shape of TPL, we tested whether including intercept and slope fixed effects for host or parasite identity improved model fit relative to log-mean infection alone using stepwise model selection and delta Akaike's information criterion (AIC) [49]. Because this approach does not provide mechanistic insights into why particular host or parasite species affect aggregation, we subsequently substituted host and parasite identity terms for measurements of specific attributes related to host life history, characteristics of the host population, and epidemiological properties of the host-parasite interaction. At the host population level, we included log-mean host body size, heterogeneity in host body size (i.e. the residual variance in host body size after accounting for mean host body size), log-host density, and spatial heterogeneity in host density within the pond (i.e. the residual variance in amphibians captured per dipnet sweep after accounting for mean density; see [50] and electronic supplementary material for calculations). At the parasite level, we included parasite body mass,

Table 1. The five best-fit models used to examine the effect of host traits, parasite traits and host–parasite interactions on the shape of Taylor's power law. For all models, the response variable was the residual log-variance in infection after controlling for the feasible set. All models included random effects of study site and sample year. For continuous predictors, we report the estimated model coefficient and its 95% confidence interval based on a parametric bootstrap ($n = 500$). For factor predictors, we report the test statistic and p-value from a parametric bootstrapped likelihood ratio test comparing the model in a given row with all of the listed variables included to a reduced model without the factor of interest ($n = 500$). Interaction terms are indicated using a colon between variables and italicized text (with an asterisk) indicates that either the bootstrapped p-value <0.05 for factor variables or the bootstrapped 95% CI does not overlap zero for continuous variables. All p-values and confidence intervals shown are adjusted for multiple comparisons within a given model using a false discovery rate and false coverage-statement rate of $q = 0.05$ [53,54], where the number of comparisons is the number of variables shown for a given model. A lower value of AIC (ΔAIC) indicates a better model.

model description (number)	model fixed effects		AIC	ΔAIC
best model with host and parasite identity (1)	logmean	−0.071, (−0.175, 0.02)	623.78	4
	*host**	$20.594_{df=4}$, $p = 0.003$		
	*parasite**	$25.004_{df=3}$, $p = 0.003$		
	*logmean:parasite**	$24.621_{df=3}$, $p = 0.003$		
best model with quantitative host predictors and parasite identity (2)	logmean	−0.06, (−0.165, 0.044)	619.78	0
	*body size variation**	0.029, (0.004, 0.052)		
	spatial heterogeneity	0.022, (−0.001, 0.047)		
	*mean body size**	0.036, (0.015, 0.06)		
	*parasite**	$21.685_{df=3}$, $p = 0.004$		
	*logmean:parasite**	$23.519_{df=3}$, $p = 0.004$		
best model with host identity and quantitative parasite predictors (3)	*logmean**	−0.163, (−0.20, −0.129)	623.38	3.60
	*host**	$20.697_{df=4}$, $p = 0.005$		
	parasite mass	−0.025, (−0.068, 0.021)		
	*cercariae shed per snail**	0.060, (0.030, 0.088)		
	inf. snail density	−0.025, (−0.061, 0.011)		
	*logmean:inf. snail density**	−0.058, (−0.099, −0.021)		
	logmean:parasite mass	−0.038, (−0.086, 0.002)		
best model with quantitative host and parasite predictors, virulence (4)	*logmean**	−0.167, (−0.202, −0.134)	630.99	11.21
	*body size variation**	0.028, (0.007, 0.053)		
	*spatial heterogeneity**	0.023, (0.0005, 0.045)		
	*cercariae shed per snail**	0.052, (0.018, 0.081)		
	*parasite mass**	−0.035, (−0.068, −0.006)		
	*inf. snail density**	−0.042, (−0.08, −0.004)		
	virulence	0.018, (−0.036, 0.068)		
	*logmean:virulence**	−0.067, (−0.105, −0.03)		
best model with quantitative host and parasite predictors, competence (5)	*logmean**	−0.157, (−0.192, −0.122)	631.57	11.79
	*body size variation**	0.028, (0.005, 0.05)		
	spatial heterogeneity	0.022, (−0.002, 0.048)		
	*cercariae shed per snail**	0.058, (0.029, 0.088)		
	parasite mass	−0.026, (−0.067, 0.016)		
	*inf. snail density**	−0.049, (−0.086, −0.013)		
	competence	0.013, (−0.031, 0.053)		
	*logmean:competence**	−0.065, (−0.096, −0.024)		

average number of cercariae released per snail, and average density of snails infected with a given parasite across sites that supported the infection (see electronic supplementary material, table S1 and [40,51,52]). We subsequently replaced host and parasite predictors with the log-transformed experimental estimates of host competence and parasite virulence for each host-by-parasite combination. We expected increasing competence or virulence to decrease the slope of TPL (figure 1) [36]. We did not include both variables simultaneously because they were positively correlated ($r = 0.59$, $p < 0.001$) and no host–parasite combination had low competence and high virulence. In the final step, we used forward model selection to re-incorporate any influential host or parasite variables and identify the best overall models (table 1; see electronic supplementary material, tables S2–S5 for full model selection). All mixed effects model analysis was performed in R (v. 3.2.4) using the 'lme4' package [55], and feasible set

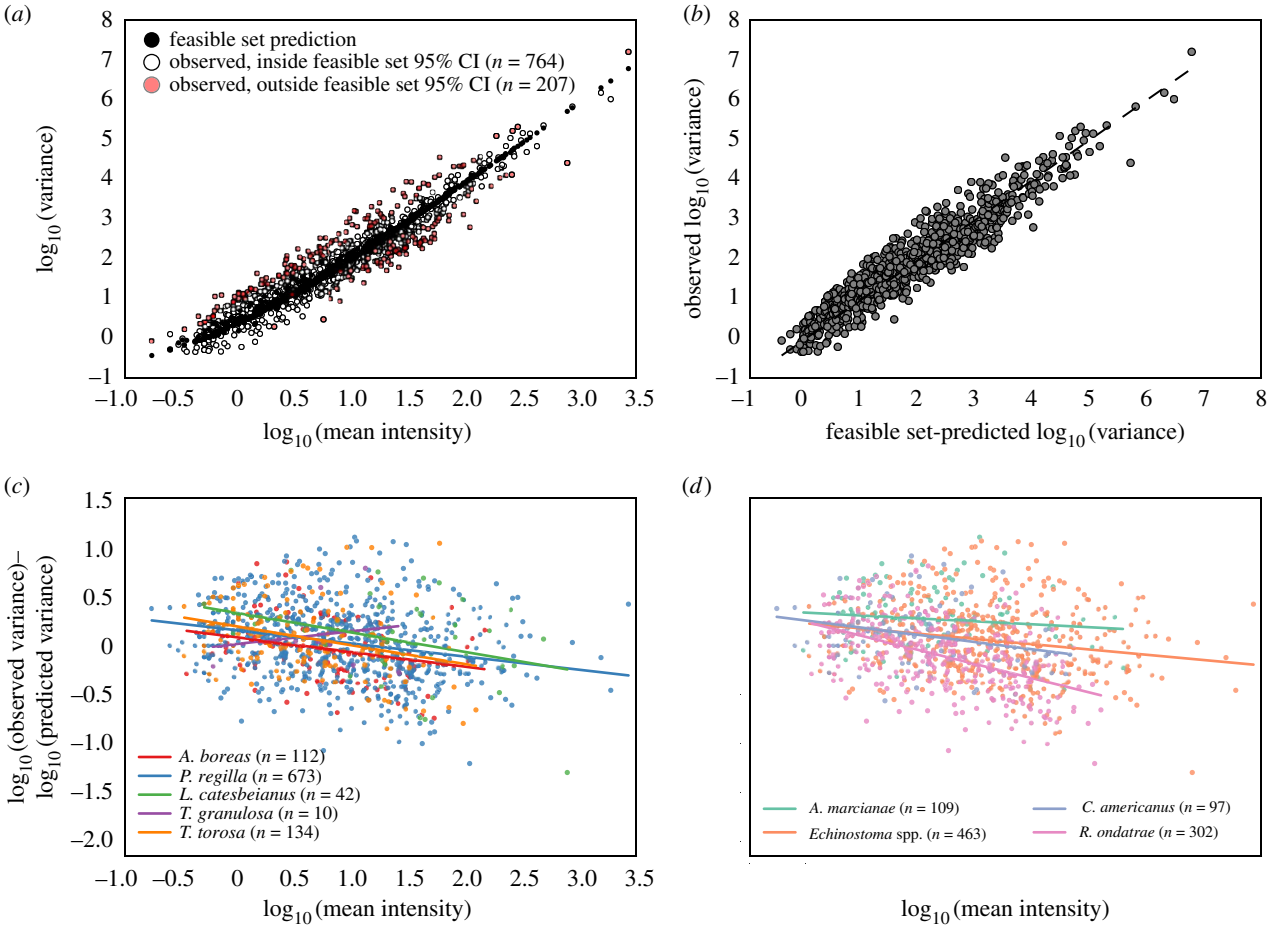

Figure 2. Using feasible set theory to account for constraints in empirical patterns of parasite aggregation within amphibian hosts. (*a*) The observed relationship between the log-mean and log-variance of infection for 971 host–parasite distributions (white and red points), for which each point corresponds to a unique combination of host species, parasite species, wetland identity and sample year. Black points represent the predicted log-variance values from feasible set theory. White points fall within the 95% confidence interval (CI) of the feasible set-predicted log variance and red points fall outside this interval (95% CIs are based on 1000 draws of feasible set-predicted log variance given *P* and *H*). Note that the quadratic shape is a prediction from feasible set theory. (*b*) The feasible set-predicted log-variance values plotted against the observed log-variance for all 971 populations. The black dashed line is the one-to-one line, along which all points would fall if feasible set theory perfectly predicted the observed log variance. (*c*) The residual variation (observed–predicted) in the data after accounting for the feasible set; each point is coloured according to amphibian host species identity. The lines give the best-fit regressions. While host species have different intercepts, the slopes are not significantly different (intercept $\chi^2_{df=4} = 20.594$, $p < 0.001$; slope $\chi^2_{df=4} = 8.513$, $p = 0.072$). (*d*) Same as (*c*) but points are coloured according to parasite species. Parasite species identity affects both the intercept and slope of the residual variation (intercept $\chi^2_{df=3} = 25.004$, $p < 0.01$; slope $\chi^2_{df=3} = 24.621$, $p < 0.01$).

analysis was performed in Python (v. 2.7.12) using the 'pypartitions' package [46]. Scripts are available in the electronic supplementary material.

3. Results

By using a feasible set approach to account for constraints imposed by the number of sampled hosts and parasites, we were able to describe 90% of the variation in the log-mean–log-variance relationship among 332 684 parasites from 11 987 hosts (figure 2*a*,*b*). However, while the feasible set captured the central tendency of the host–parasite TPL, *P* and *H* alone were not sufficient to describe the observed patterns of parasite aggregation (21% of the observed data points fell outside of the respective 95% CIs for the feasible set-predicted log variance; figure 2*a*). Importantly, we detected strong and consistent effects of host and parasite species identity on empirical patterns of aggregation even after controlling for feasible set predictions. The TPL intercept—but not its slope—varied among amphibian species (likelihood ratio test [LRT] intercept: $\chi^2_{df=4} = 20.594$, $p < 0.001$; LRT slope: $\chi^2_{df=4} = 8.513$, $p = 0.082$; figures 2*c* and 3*a*). For instance, bullfrogs (*L. catesbeianus*) and newts (*Taricha granulosa* and *T. torosa*) had, on average, higher variance for a given infection intensity relative to other species (figure 3*a*). The effects of parasite identity, in contrast, manifested on both the slope and the intercept (figures 2*d* and 3*b*; LRT slope: $\chi^2_{df=3} = 24.621$, $p < 0.01$; LRT intercept: $\chi^2_{df=3} = 25.004$, $p < 0.01$), and were more important than host effects (ΔAIC without host: 12.6; ΔAIC without parasite: 139.3; compared with model 1 in table 1). Infection by the highly virulent trematode *R. ondatrae* was associated with the greatest reduction in slope (figure 3*b*). Including both host and parasite identity collectively accounted for 29.5% of the residual variance in aggregation not described by the feasible set and reduced AIC by 136.25 units compared with a model with only the log-mean of infection. These patterns were broadly consistent across sampling years (figure 3).

Replacing host and parasite species identity with biological traits offered insights into the potential mechanisms underlying observed patterns. Host populations with a larger

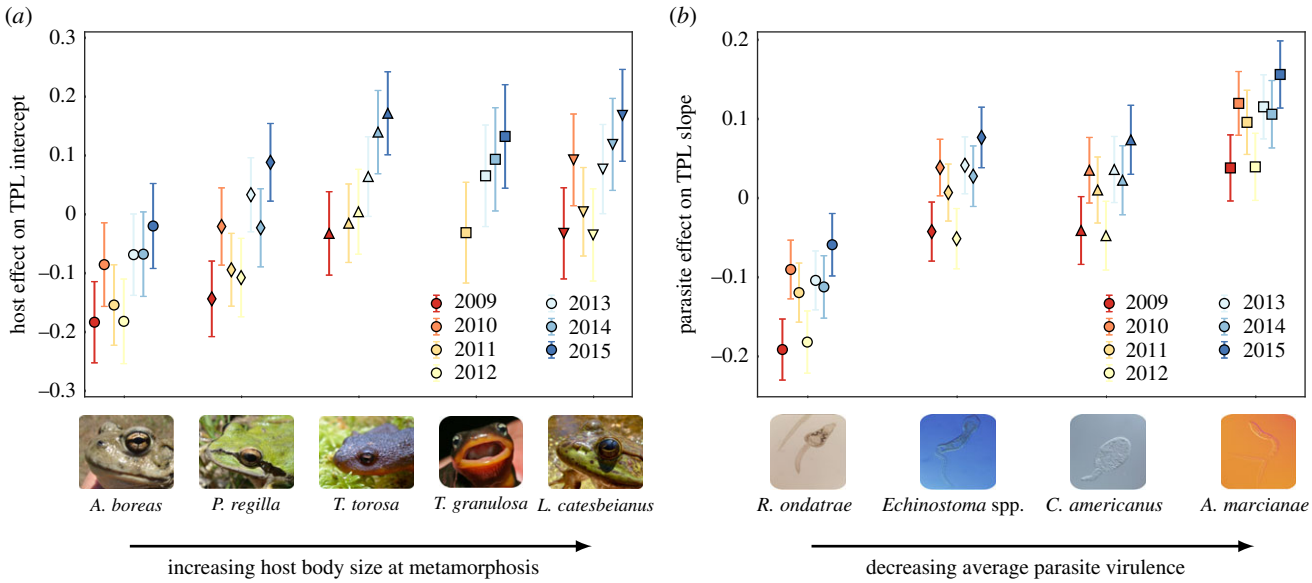

Figure 3. The effects of host and parasite species identity on Taylor's power law across the 7-year sampling period. While (*a*) parasite species and (*b*) host species had differing effects on the estimated slope or intercept of the log-mean–log-variance relationship, these effects were broadly consistent over time. Symbol shapes correspond to the identity of parasite or host species and colour indicates the sample year. For a given host or parasite, points are staggered with respect to the *x*-axis for visual clarity. Effect sizes and standard errors were extracted from a linear mixed-effects model.

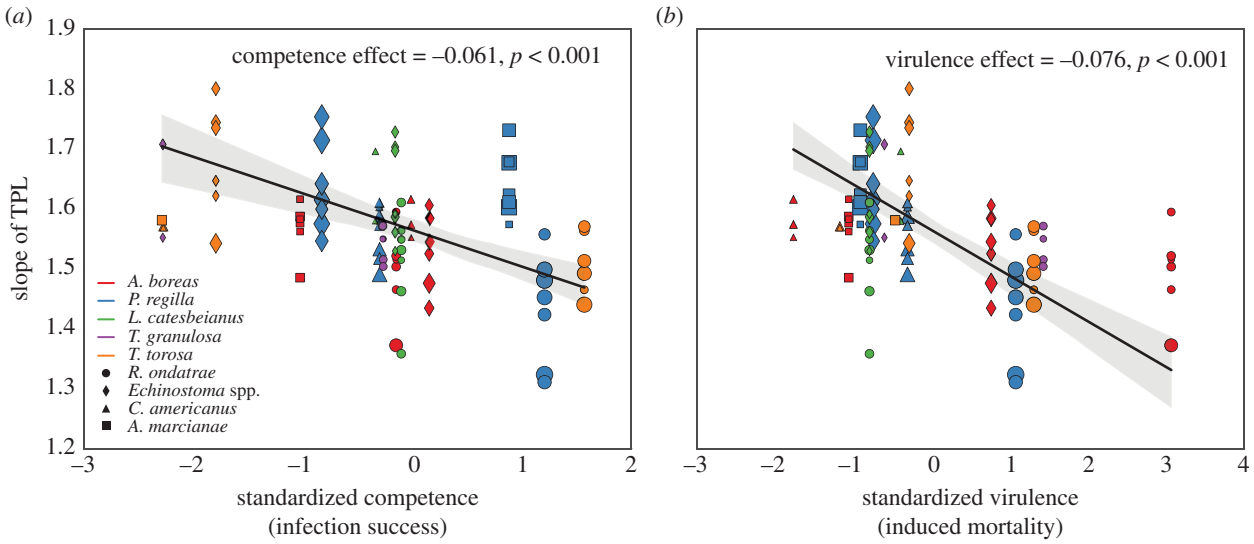

Figure 4. Effects of infection (*a*) competence and (*b*) virulence on the slope of Taylor's power law (TPL). Competence and virulence for each host–parasite combination were measured experimentally using controlled exposure dosages. Laboratory-derived estimates of (standardized) virulence and competence for each host-parasite combination are plotted against the empirically observed slope of the log-mean–log-variance relationship from sampled field sites. Each point represents the estimated TPL slope based on a linear mixed effect model with year, host and parasite species. The colour and shapes of each point indicates the host and parasite species, respectively. Multiple points of the same shape and colour correspond to multiple years of observation. The size of each point reflects the number of data points for that combination of host, parasite and year. The black line represents the best-fit line from a weighted least-squares regression with virulence or competence as the predictor variable with the 95% confidence interval indicated by the grey region. On the upper right of each plot is the estimated coefficient from this regression and its corresponding *p*-value. (Online version in colour.)

average body size, more among-host variation in body size and a higher degree of spatial heterogeneity exhibited greater parasite aggregation (table 1, model 2). These were intercept effects and collectively reduced model AIC by 4 units relative to host species identity. In contrast, parasite traits had both intercept and slope effects, consistent with the effects of parasite species identity. Parasites that, on average, released more cercariae per infected snail showed an intensity-independent increase in aggregation among amphibian hosts (table 1, model 3). Moreover, increases in parasite body mass or the average density of infected snails were associated with slope-dependent decreases in aggregation (table 1, model 3).

Including traits related to the host-by-parasite interaction accounted for additional variation in parasite aggregation. Based on experimental exposures, infection competence and virulence both varied widely among the 20 tested host-parasite combinations (figure 4). *Ribeiroia ondatrae* was the most virulent parasite, yet its effects ranged from a 52% increase in *per capita* mortality risk in *A. boreas* to a 0.4% increase for *L. catesbeianus*. Incorporating these estimates into the models revealed that both terms reduced the slope of TPL (table 1, models 4–5; figure 4*a*,*b*). Host–parasite pairs with high competence (figure 4*a*) or associated with higher host mortality (figure 4*b*) both exhibited substantially

lower slopes of TPL than those with low competence and low virulence. Based on the competing-model approach, the combination of host-population characteristics (body size heterogeneity and spatial heterogeneity), parasite traits (cercariae released per infected snail, parasite body mass, and infected snail density) and attributes of the host-by-parasite interaction (competence or virulence) accounted for 28% of the residual variance in parasite aggregation.

4. Discussion

Despite the long history and diverse applications of Taylor's power law, the identity and importance of its contributing processes continue to be debated [3,4,10,22]. In part this stems from the concurrent challenges of accounting for statistical constraints of the data [22] while also obtaining sufficiently comparable information to test alternative mechanisms simultaneously. Many studies assume that slope values greater than one—which would be the expected value according to a Poisson distribution where the mean equals the variance—are indicative of aggregation [11,34,56,57], yet the static slope provided by a Poisson is an inadequate null for determining whether this level of aggregation is 'unusual' given the observed data [23]. The current combination of multi-species field surveys with experimental quantification of common host–parasite interactions revealed that patterns of parasite aggregation among 971 populations were an emergent property of the host species, the parasite species, and the host-by-parasite interactions. The use of feasible set theory, which predicts the most likely aggregation patterns given data constraints, indicated that host–parasite TPLs are tightly constrained by how many hosts and parasites comprise each population [22], effectively allowing us to describe 90% of the variation in TPL shape. Unlike conventional log-mean–log-variance approaches, feasible set theory also accounts for expected non-linearities in empirical TPLs [20,22] and obviates the need for including sample size as an additional predictor [38] (figure 1*a*).

After controlling for the feasible set predictions, our results identified strong, joint effects of host and parasite species identity on infection aggregation, which collectively accounted for approximately 30% of the residual variation. These effects were consistent across a large spatial extent and over 7 years of data collection, suggesting that aggregation patterns were a function of both the species under consideration as well as its specific environment (in this case the host; see [34]). While parasite identity affected the slope and intercept of the log-mean–log-variance relationship, host species only affected the intercept (see also [10]), emphasizing the importance of testing the biological processes responsible for varying aggregation strength across different environments (i.e. host-parasite combinations). Although many studies have focused on variables that affect either the intercept or more often the slope of TPL, consideration of both intercept and slope effects on TPL provides a more conceptually integrated framework. For instance, analyses that constrain the intercept to zero (a mean infection of 1), compare log-variance estimates at a fixed value of infection or extract residuals from the log-mean–log-variance regression all preclude joint assessments of slope and intercept changes in TPL [38,57,58]. While such approaches may be necessary if the intercept is sensitive to variation in sampling method and scale, particularly when compiling data among many sources, the systematic and consistent sampling of the current study afforded an opportunity to extend the range of inference.

By linking empirical censuses of parasite aggregation with field-based measurements of population-, species- and environment-level traits, we show that characteristics of the host population—including variance in body size and spatial heterogeneity—increased parasite aggregation in an intensity-independent manner. Hosts populations with larger average body sizes exhibited greater aggregation, consistent with both the longer duration over which parasites were likely acquired and hypothesized life-history trade-offs between developmental time (or body size) and investments in immune defences [42,59–62]. Characteristics of the parasite, such as cercariae size, number of parasites released per snail host and the density of infected snails per wetland, also affected parasite aggregation. These results stemmed from the tendency of more virulent parasites to be larger in size but fewer in number [63]. Thus, low-virulence parasites, such as *Alaria marcianae*, positively affected infection heterogeneity in amphibian hosts both because its small cercariae were unlikely to induce host mortality (which reduces aggregation) and because of the inherently greater variance associated with the large number of infectious stages released per snail (figure 2*d*).

Perhaps more importantly, the current findings highlighted the influence of species interactions in determining infection heterogeneity among populations [58,64]. Previous studies have debated whether aggregation generally and the slope of TPL specifically is an intrinsic property of the species being studied or instead the product of variation in habitat, sample size and sampling methods [3,8–10]. Our results advance this debate by illustrating that, while species identity matters, aggregation is more accurately captured by variables reflective of the species-by-environment interaction. In the case of parasites, this entailed the joint roles of host and parasite species identity and the epidemiological outcomes of this interaction—namely, infection competence and virulence. Both variables had slope effects on aggregation, albeit for different reasons. Higher virulence and parasite-induced mortality likely decreased aggregation by truncating variance around mean parasite abundance, especially at higher infection loads [36,47]. These effects associated most strongly with infection by *R. ondatrae*, which both in the current study and past research induced mortality in a strongly dose-dependent manner [65]. For competence, host–parasite combinations with low competence were associated with larger TPL slopes relative to those with high competence (figure 4*a*). This variation is probably the product of differences in the infection process: generating a high average infection load when competence is low requires a substantial increase in parasite exposure, which will incur its own increase in variability either spatially among hosts or temporally among exposure events. Interestingly, parasite identity remained a more influential covariate than either competence or virulence (table 1), which could stem from biological processes not captured by these epidemiological predictors, such as evolutionary history, or the potential for a species-factor to better allow for nonlinear effects than a continuous variable [66].

The aggregation of parasites within hosts is considered an 'ecological law' in disease ecology, such that the slope of the log-mean–log-variance relationship for parasites per host often exceeds unity [26,29,30]. For instance, Shaw & Dobson [11] reported a slope of 1.55 across 263 macroparasite–host

populations (s.e. = 0.037, 95% CI = [1.48, 1.62]), similar to the slope of 1.59 in the present study of 971 populations (s.e. = 0.017, 95% CI = [1.56, 1.63]) and to the mean value of approximately 1.52 reported by Xiao *et al*. [22] for free-living species (s.e. and 95% CI not available). However, efforts to identify the drivers of such variation have yielded surprisingly few generalities. While meta-analytic analyses have identified correlates associated with aggregation, including parasite specificity, coinfection, host life history, parasite developmental stage, and transmission mode [11,28,56,62,67], testing the individual and combined roles of such factors is often hindered by data availability. In a synthetic comparison of published data on aggregation by 180 parasite taxa, Poulin [38] concluded that there may be 'little point to seeking universal causes for the remaining variation' in parasite aggregation beyond that explained by mean infection load. As reflected by our own analyses, some of this outcome is probably methodological: the large influence of statistical processes on the perceived pattern of aggregation indicates that adequate accounting of numerical constraints via methods such as feasible set theory is often necessary before testing for biological drivers. In addition, challenges associated with a meta-analytic approach may restrict opportunities to broadly test the influence of a wide range of consistently measured mechanistic variables. For instance, in his comparative study of parasite aggregation in fish hosts, Poulin [38] was only able to include variation in host body size, parasite taxon and developmental stage as predictors; moreover, by focusing on the residuals of the log-mean–log-variance relationship without including covariate interactions with log-mean, that study was only able to test for effects on the intercept of Taylor's power law (not concurrent effects on the slope). The dataset assembled in this study is unique in the number of directly and consistently measured covariates at the population, species and species-interaction levels, with replication across multiple host and parasite species over a 7-year period, which helps explain the detection of effects not observed in previous research. Nonetheless, an important frontier will be to test these variables in other, comparably well-studied multi-host, multi-parasite systems to assess their generality and further incorporate variables such as host immunity, behavioural variation and phylogenetic history.

The ubiquity of Taylor's power law across biological and non-biological systems has led to the question of whether the shape of the log-mean–log-variance relationship contains a signature of system-specific processes beyond statistical properties of the dataset. Our analyses show that when this pattern is put in the context of the constraints imposed by the data, we can identify consistent effects of putative mechanisms on the shape of host–parasite TPLs, including both intensity-dependent and intensity-independent processes. The high predictability of parasite aggregation from constraint-based theory and the associated effects of biological mechanisms on the slope and intercept of this relationship have important implications for predicting parasite aggregation and its effects on population regulation, ecological stability and disease management [15,68,69]. While we used interactions between hosts and macroparasites as a useful and clearly delineated model system, both the results and the overall approach employed here are broadly applicable to aggregation in other biological and non-biological systems.

Ethics. The described data were collected with the approval of the University of Colorado's Institutional Animal Care and Use Committee (protocols 1002.02 1302.01) and in accordance with sampling protocols approved by the California Department of Fish and Wildlife (SC-3683 and SC-10560), the US Fish and Wildlife Service (TE-181714), Santa Clara County Parks, East Bay Regional Parks District, East Bay Municipal Utility District, California State Parks and other local landowners.

Data accessibility. The data and metadata associated with this article are available and have been uploaded as part of the supplementary material. They are also available on Dryad [70].

Authors' contributions. P.T.J.J. conceived of the initial idea for the study, coordinated field and experimental studies, and organized the data; M.Q.W. and P.T.J.J. performed data analyses; and both authors contributed to the writing and editing of the manuscript.

Competing interests. We have no competing interests.

Funding. This work was supported by grants from the National Science Foundation (DEB-0841758, DEB-1149308 and DGE-1144085), the National Institutes of Health (R01GM109499), and the David and Lucile Packard Foundation.

Acknowledgements. Our sincerest thanks to the many individuals who helped collect these data over the years, including especially D. Calhoun and T. McDevitt-Galles. For access to properties and logistical support during field sampling, we thank East Bay Regional Parks District, East Bay Municipal Utility District, Santa Clara County Parks, Blue Oak Ranch Reserve, California State Parks, The Nature Conservancy, Open Space Authority, Mid-peninsula Open Space and many private landowners. For discussions and feedback helpful in shaping the manuscript, we thank C. Briggs, J. Cohen, members of the Macroecology of Infectious Disease Research Coordination Network (funded by NSF DEB 131223) and three anonymous reviewers.

References

1. Taylor LR. 1961 Aggregation, variance and the mean. *Nature* **169**, 732–735. (doi:10.1038/189732a0)
2. Anderson RM, Gordon DM, Crawley MJ, Hassell MP. 1982 Variability in the abundance of animal and plant species. *Nature* **296**, 245–248. (doi:10.1038/296245a0)
3. Downing JA. 1986 Spatial heterogeneity: evolved behaviour or mathematical artefact? *Nature* **323**, 255–257. (doi:10.1038/323255a0)
4. Kilpatrick AM, Ives AR. 2003 Species interactions can explain Taylor's power law for ecological time series. *Nature* **422**, 65–68. (doi:10.1038/nature01471)
5. Cohen JE, Xu M. 2015 Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. *Proc. Natl Acad. Sci. USA* **112**, 7749–7754. (doi:10.1073/pnas.1503824112)
6. 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**, 1791–1796. (doi:10.1073/pnas.1422475112)
7. Cohen JE, Poulin R, Lagrue C. 2017 Linking parasite populations in hosts to parasite populations in space through Taylor's law and the negative binomial distribution. *Proc. Natl Acad. Sci. USA* **114**, E47–E56. (doi:10.1073/pnas.1618803114)
8. Taylor LR, Woiwod IP, Perry JN. 1978 The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* **47**, 383–406. (doi:10.2307/3790)
9. Taylor LR, Woiwod IP, Perry JN. 1980 Variance and the large scale spatial stability of aphids, moths and birds. *J. Anim. Ecol.* **49**, 831–854. (doi:10.2307/4230)
10. Taylor LR, Perry JN, Woiwod IP, Taylor RAJ. 1988 Specificity of the spatial power-law exponent in ecology and agriculture. *Nature* **332**, 721–722. (doi:10.1038/332721a0)

11. Shaw DJ, Dobson AP. 1995 Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S133. (doi:10.1017/S0031182000075855)
12. Eisler Z, Bartos I, Kertész J. 2008 Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142. (doi:10.1080/00018730801893043)
13. Xu M, Schuster WS, Cohen JE. 2015 Robustness of Taylor's law under spatial hierarchical groupings of forest tree samples. *Popul. Ecol.* **57**, 93–103. (doi:10.1007/s10144-014-0463-0)
14. Tippett MK, Cohen JE. 2016 Tornado outbreak variability follows Taylor's power law of fluctuation scaling and increases dramatically with severity. *Nat. Commun.* **7**, 10668. (doi:10.1038/ncomms10668)
15. Cohen JE. 2014 Taylor's law and abrupt biotic change in a smoothly changing environment. *Theor. Ecol.* **7**, 77–86. (doi:10.1007/s12080-013-0199-z)
16. Xu M, Kolding J, Cohen JE. 2016 Taylor's power law and fixed-precision sampling: application to abundance of fish sampled by gillnets in an African lake. *Can. J. Fish. Aquat. Sci.* **7**, 1–14.
17. Giometto A, Formentin M, Rinaldo A, Cohen JE, Maritan A. 2015 Sample and population exponents of generalized Taylor's law. *Proc. Natl Acad. Sci. USA* **112**, 7755–7760. (doi:10.1073/pnas.1505882112)
18. Tokeshi M. 1995 On the mathematical basis of the variance–mean power relationship. *Res. Popul. Ecol.* **37**, 43–484. (doi:10.1007/BF02515760)
19. Soberón J, Loevinsohn M. 1987 Patterns of variations in the numbers of animal populations and the biological foundations of Taylor's law of the mean. *Oikos* **48**, 249–252. (doi:10.2307/3565509)
20. 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. (doi:10.1111/oik.02622)
21. Cohen MJ *et al.* 2016 Do geographically isolated wetlands influence landscape functions? *Proc. Natl Acad. Sci. USA* **113**, 1978–1986. (doi:10.1073/pnas.1512650113)
22. Xiao X, Locey KJ, White EP, Kerkhoff AEAJ, Day ET. 2015 A process-independent explanation for the general form of Taylor's law. *Am. Nat.* **186**, E51–E60. (doi:10.1086/682050)
23. Wilber MQ, Johnson PT, Briggs CJ. 2016 When can we infer mechanism from parasite aggregation? A constraint-based approach to disease ecology. *Ecology* **98**, 688–702. (doi:10.1002/ecy.1675)
24. Xu M. 2015 Taylor's power law: before and after 50 years of scientific scrutiny. *ArXiv Prepr.* (https://arxiv.org/abs/1505.02033)
25. Locey KJ, White EP. 2013 How species richness and total abundance constrain the distribution of abundance. *Ecol. Lett.* **16**, 1177–1185. (doi:10.1111/ele.12154)
26. Crofton HD. 1971 A model of host–parasite relationships. *Parasitology* **63**, 343–364. (doi:10.1017/S0031182000079890)
27. Croll NA, Anderson RM, Gyorkos TW, Ghadirian E. 1982 The population biology and control of *Ascaris lumbricoides* in a rural community in Iran. *Trans. R. Soc. Trop. Med. Hyg.* **76**, 187–197. (doi:10.1016/0035-9203(82)90272-3)
28. Wilson K, Bjørnstad ON, Dobson AP, Merler S, Poglayen G, Randolph SE, Read AF, Skorping A. 2002 Heterogeneities in macroparasite infections: patterns and processes. In *The ecology of wildlife diseases* (eds PJ Hudson, A Rizzoli, BT Grenfell, JAP Heesterbeek, AP Dobson), pp. 6—44. Oxford, UK: Oxford University Press.
29. Poulin R. 2007 Are there general laws in parasite ecology? *Parasitology* **134**, 763–776. (doi:10.1017/S0031182006002150)
30. Stephens PR *et al.* 2016 The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distributions and impacts. *Ecol. Lett.* **19**, 1159–1171. (doi:10.1111/ele.12644)
31. Keeling M, Grenfell B. 1999 Stochastic dynamics and a power law for measles variability. *Phil. Trans. R. Soc. Lond. B* **354**, 769–776. (doi:10.1098/rstb.1999.0429)
32. Morand S, Guégan J-F. 2000 Distribution and abundance of parasite nematodes: ecological specialisation, phylogenetic constraint or simply epidemiology? *Oikos* **88**, 563–573. (doi:10.1034/j.1600-0706.2000.880313.x)
33. Woolhouse MEJ, Taylor LH, Haydon DT. 2001 Population biology of multihost pathogens. *Science* **292**, 1109–1112. (doi:10.1126/science.1059026)
34. Morand S, Krasnov B. 2008 Why apply ecological laws to epidemiology? *Trends Parasitol.* **24**, 304–309. (doi:10.1016/j.pt.2008.04.003)
35. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
36. 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)
37. Galvani AP. 2003 Immunity, antigenic heterogeneity, and aggregation of helminth parasites. *J. Parasitol.* **89**, 232–241. (doi:10.1645/0022-3395(2003)089[0232:IAHAA0]2.0.CO;2)
38. Poulin R. 2013 Explaining variability in parasite aggregation levels among host samples. *Parasitology* **140**, 541–546. (doi:10.1017/S0031182012002053)
39. Martin LB, Burgan SC, Adelman JS, Gervasi SS. 2016 Host competence: an organismal trait to integrate immunology and epidemiology. *Integr. Comp. Biol.* **56**, 1225–1237. (doi:10.1093/icb/icw064)
40. Johnson PTJ, Preston DL, Hoverman JT, Richgels KLD. 2013 Biodiversity decreases disease through predictable changes in host community competence. *Nature* **494**, 230–233. (doi:10.1038/nature11883)
41. Johnson PTJ, Wood CL, Joseph MB, Preston DL, Haas SE, Springer YP. 2016 Habitat heterogeneity drives the host-diversity-begets-parasite-diversity relationship: evidence from experimental and field studies. *Ecol. Lett.* **19**, 752–761. (doi:10.1111/ele.12609)
42. Johnson PTJ, Hoverman JT. 2012 Parasite diversity and coinfection determine pathogen infection success and host fitness. *Proc. Natl Acad. Sci. USA* **109**, 9006–9011. (doi:10.1073/pnas.1201790109)
43. Gosner KL. 1960 A simplified table for staging anuran embryos and larvae with notes on identification. *Herpetologica* **16**, 183–190.
44. Wong CJ, Liversage RA. 2004 Limb developmental stages of the newt *Notophthalmus viridescens*. *Int. J. Dev. Biol.* **49**, 375–389. (doi:10.1387/ijdb.041910cw)
45. Firth D. 1993 Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38. (doi:10.1093/biomet/80.1.27)
46. Locey KJ, McGlinn DJ. 2013 Efficient algorithms for sampling feasible sets of macroecological patterns. *PeerJ Prepr.* **1**, e78v1.
47. Wilber MQ, Weinstein SB, Briggs CJ. 2016 Detecting and quantifying parasite-induced host mortality from intensity data: method comparisons and limitations. *Int. J. Parasitol.* **46**, 59–66. (doi:10.1016/j.ijpara.2015.08.009)
48. Taylor LR, Woiwod IP. 1982 Comparative synoptic dynamics. I. Relationships between inter- and intra-specific spatial and temporal variance/mean population parameters. *J. Anim. Ecol.* **51**, 879–906. (doi:10.2307/4012)
49. Burnham KP, Anderson DR. 2004 Multimodel inference understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304. (doi:10.1177/0049124104268644)
50. Koprivnikar J, Riepe TB, Calhoun DM, Johnson PTJ. In press. Whether larval amphibians school does not affect the parasite aggregation rule: testing the effects of host spatial heterogeneity in field and experimental studies. *Oikos*. (doi:10.1111/oik.04249)
51. Lambden J, Johnson PTJ. 2013 Quantifying the biomass of parasites to understand their role in aquatic communities. *Ecol. Evol.* **3**, 2310–2321. (doi:10.1002/ece3.635)
52. Hannon ER, Calhoun DM, Chadalawada S, Johnson PTJ. In review. Circadian rhythms of trematode parasites: applying mixed model to test underlying patterns. *Parasitology*.
53. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Methodol.* **57**, 289–300.
54. Benjamini Y, Yekutieli D. 2005 False discovery rate–adjusted multiple confidence intervals for selected parameters. *J. Am. Stat. Assoc.* **100**, 71–81. (doi:10.1198/016214504000001907)
55. Bates D, Maechler M, Bolker B, Walker S. 2013 lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5.
56. Boag B, Lello J, Fenton A, Tompkins DM, Hudson PJ. 2001 Patterns of parasite aggregation in the wild European rabbit (*Oryctolagus cuniculus*). *Int. J. Parasitol.* **31**, 1421–1428. (doi:10.1016/S0020-7519(01)00270-3)

57. Lester RJG. 2012 Overdispersion in marine fish parasites. *J. Parasitol.* **98**, 718–721. (doi:10.1645/GE-3017.1)
58. Krasnov BR, Stanko M, Miklisova D, Morand S. 2006 Host specificity, parasite community size and the relation between abundance and its variance. *Evol. Ecol.* **20**, 75—91. (doi:10.1007/s10682-005-4731-5)
59. Lee KA. 2006 Linking immune defenses and life history at the levels of the individual and the species. *Integr. Comp. Biol.* **46**, 1000–1015. (doi:10.1093/icb/icl049)
60. Martin LB, Weil ZM, Nelson RJ. 2006 Refining approaches and diversifying directions in ecoimmunology. *Integr. Comp. Biol.* **46**, 1030–1039. (doi:10.1093/icb/icl039)
61. Raffel TR, Lloyd-Smith JO, Sessions SK, Hudson PJ, Rohr JR. 2011 Does the early frog catch the worm? Disentangling potential drivers of a parasite age–intensity relationship in tadpoles. *Oecologia* **165**, 1031–1042. (doi:10.1007/s00442-010-1776-0)
62. Lester RJG, McVinish R. 2016 Does moving up a food chain increase aggregation in parasites? *J. R. Soc. Interface* **13**, 20160102. (doi:10.1098/rsif.2016.0102)
63. Rohr JR, Raffel TR, Hall CA. 2010 Developmental variation in resistance and tolerance in a multi-host–parasite system. *Funct. Ecol.* **24**, 1110–1121. (doi:10.1111/j.1365-2435.2010.01709.x)
64. Ramsayer J, Fellous S, Cohen JE, Hochberg ME. 2012 Taylor's law holds in experimental bacterial populations but competition does not influence the slope. *Biol. Lett.* **8**, 316–319. (doi:10.1098/rsbl.2011.0895)
65. Johnson PTJ, Rohr JR, Hoverman JT, Kellermanns E, Bowerman J, Lunde KB. 2012 Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecol. Lett.* **15**, 235–242. (doi:10.1111/j.1461-0248.2011.01730.x)
66. Kutner M, Nachtsheim C, Neter J, Li W. 2005 *Applied linear statistical models*, 5th edn. Irwin: McGraw Hill.
67. Grear DA, Hudson P. 2011 The dynamics of macroparasite host-self-infection: a study of the patterns and processes of pinworm (Oxyuridae) aggregation. *Parasitology* **138**, 619–627. (doi:10.1017/S0031182011000096)
68. Anderson RM, May RM. 1982 Population dynamics of human helminth infections: control by chemotherapy. *Nature* **297**, 557–563. (doi:10.1038/297557a0)
69. Tompkins DM *et al.* 2002 Parasites and host population dynamics. In *The ecology of wildlife diseases* (eds PJ Hudson, A Rizzoli, BT Grenfell, JAP Heesterbeek, AP Dobson), pp. 45–62. Oxford, UK: Oxford University Press.
70. Johnson PTJ, Wilber MQ. 2017 Data from: Biological and statistical processes jointly drive population aggregation: using host–parasite interactions to understand Taylor's power law. Dryad Digital Repository. (doi:10.5061/dryad.r08t9)