Scaling body size fluctuations

Andrea Giometto a,b, Florian Altermatt a, Francesco Carrara a, Amos Maritan a, and Andrea Rinaldo a,d,1

*Laboratory of Ecolhydrology, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; aDepartment of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, CH-8600 Dübendorf, Switzerland; bDepartment of Physic ed Astronomia, Università di Padova, I-35131 Padua, Italy; and cDipartimento di Ingegneria Civile, Edile ed Ambientale, Università di Padova, I-35131 Padova, Italy

Contributed by Andrea Rinaldo, January 29, 2013 (sent for review November 27, 2012)

The size of an organism matters for its metabolic, growth, mortality, and other vital rates. Scale-free community size spectra (i.e., size distributions regardless of species) are routinely observed in natural ecosystems and are the product of intra- and interspecies regulation of the relative abundance of organisms of different sizes. Intra- and interspecies distributions of body sizes are thus major determinants of ecosystems’ structure and function. We show experimentally that single-species mass distributions of unicellular eukaryotes covering different phyla exhibit both characteristic sizes and universal features over more than four orders of magnitude in mass. Remarkably, we find that the mean size of a species is sufficient to characterize its size distribution fully and that the latter has a universal form across all species. We show that an analytical physiological model accounts for the observed universality, which can be synthesized in a log-normal form for the intraspecies size distributions. We also propose how ecological and physiological processes should interact to produce scale-invariant community size spectra and discuss the implications of our results on allometric scaling laws involving body mass.

Why should a continuous, gap-free spectrum of organismic sizes emerge from the ecological and evolutionary processes that shape their ecosystems? The origins and the implications of the absence of preferential body sizes, which is routinely observed across a variety of ecosystems regardless of broad differences in climatic and environmental conditions (1–5), have been attracting much interest from field and theoretical ecologists (6–13). Scale invariance, epitomized by power-law probability distributions (9, 12, 14–19), requires regularities of the component parts (the species’ size distributions) making up the whole [the community size spectra (i.e., the probability distributions of size regardless of species)]. In particular, a necessary condition for scaling community size spectra is the lack of peaks that pinpoint frequent occurrences, and therefore excess abundance (and vice versa) within any given range of sizes. Such features are particularly interesting if robust to environmental fluctuations because their dynamic origin could lie in the self-organization of complex adaptive systems (6, 9, 15).

Body size distributions in natural ecosystems are strongly related to the life history of the organisms and to the dynamics of their living communities (18). Thereby, they modulate structure and function of the ecosystem at any scale. Size spectra, which display the relative abundance of organisms of different sizes within or across species, convey a synoptic and possibly taxon-independent image of ecological communities (1, 2, 20, 21). As such, they have long been attracting much interest in ecology because they hold important predictive power (e.g., fish stock projections from planktonic size spectra) (2, 5). Because examples and counterexamples of scaling spectra abound (2, 3, 5, 7, 21–24), it is an unsettled issue as to whether scaling size spectra represent some central tendency of statistically stationary states of natural ecosystems. For instance, the operational computation of mean phytoplankton size was shown to depend on the sample size typically (7) and scaling relationships were documented for interspecific plant biomass (20, 25, 26), whereas some terrestrial ecosystems exhibit ubiquitous gaps in size and uneven relative abundances of organisms (1, 21).

Single species inhabiting communities, however, do exhibit a species-specific mean and variance of their sizes, as even common sense suggests. Thus, there naturally exists the mean size of a particular species, as usually implied by most, if not all, biological scaling laws (10, 27–31), wherein one typical mass subsumes a whole distribution of sizes. One thus wonders how evolutionary and ecological processes interact to modulate species’ abundances, the range of sizes proper to each functional group, and the number of species existing within a given niche or range of sizes to concoct regular, taxon-independent, continuous size spectra. Moreover, one expects that the existence of a range of possible sizes for a species (and how such a range varies for different mean sizes) has to be taken into account when addressing scaling laws in biology (e.g., allometric scaling laws) (10, 11, 27–29, 32–39).

Here, we have precisely measured the intraspecies size distributions of 13 species of protists in isolation or in competition (40–42), covering a relatively broad set of field conditions (Materials and Methods). Examples of such distributions as functions of the linear size in standard environmental conditions are shown in Fig. 1. The corresponding transformed distributions as functions of volume span over four orders of magnitude and are shown in Fig. 24. Let \( p_k(m) \) denote the measured size spectrum of the \( k \)th species: Such a \( p_k(m) \) measures the relative proportion of individuals of a given species \( k \) with mass belonging to \( (m, m + dm) \), assuming a continuous distribution of sizes. We tested whether \( p_k(m) \) exhibits a finite-size scaling form (22, 30, 36, 43, 44) obtained by the product of two terms, an algebraic power of size multiplied by a suitable scaling function \( F \); that is,

\[
p_k(m) = \frac{1}{m^2} F \left( \frac{m}{\langle m \rangle} \right),
\]

where \( \langle m \rangle_k \) is the mean mass of the \( k \)th species and \( F(x) \), critically, is the same scaling function for all species (dimensional analysis and normalization conditions that \( F \) must satisfy are discussed in SI Text). Eq. 1 implies that the only species dependence of the size distribution occurs through the average mass \( \langle m \rangle_k \) of species \( k \). Note that the two exponents in Eq. 1, \( \Delta \) and \( \phi \), are not independent. This follows from imposing \( \int_\Delta \! dm \, m \, p_k(m) \propto \langle m \rangle_k \) (where \( \mathcal{R} \) is the suitable range of sizes); in fact, \( \int_\Delta \! dm \, m \, p_k(m) = \int_{\Delta \phi} \! dm \, m \, F \left( \frac{m}{\langle m \rangle} \right) \propto \langle m \rangle_k^{\Delta - \phi} \phi \) is proportional to \( \langle m \rangle_k \) only if the two exponents satisfy \( (2 - \Delta) \phi = 1 \).


The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. E-mail: andrea.rinaldo@epfl.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1301552110/-/DCSupplemental.
Environmental factors are capable of affecting the size distribution of any given species, although they do not alter its scaling nature because both the exponents and the scaling function \( F \) are unchanged.

The observed regularities are compelling, given the profound biological diversity of the species (belonging to the phyla/divisions of Ciliophora, Euglenozoa, Chlorophyta, and Cryptophyta) considered in this study (50). In addition, protists and unicellular algae are of key ecological significance. In fact, they are the basic food source of almost all aquatic food webs, and unicellular algae are responsible for almost 50% of the worldwide biomass production (51). Additionally, the observed universality of eukaryote intraspecies size distributions holds in a range of more than four orders of magnitude in mass. This suggests the existence of a simple underlying mechanism responsible for the empirical patterns observed. One possible explanation for the reported universality may be found in the physiological processes that determine the size of all unicellular species, namely, cellular growth and cell division. We found that a simple mathematical model of these processes (52–54) can justify the scaling form of unicellular eukaryotes’ size distributions without the need to specify further biological details (SI Text). We therefore suggest that the universal features of intraspecies size distributions emerge from fundamental physiological constraints.

The detailed identification of the scaling function \( F \) in Eq. 2 is interesting but inessential for our tenet because the collapse of the distributions suffices in documenting the universality sought after. However, a log-normal functional form for \( p(m) \) provides admisible and rooted in a theoretical framework for the time evolution of the distribution of body sizes in ecological time scales (Fig. 2D, Inset and SI Text). In this context, the size distribution of organisms of a given species is the stationary distribution of an Ornstein–Uhlenbeck process (55) in the variable \( x = \log(m/(m_i)) \) (56, 57). A log-normal form for \( p(m) \) can also be recovered as a particular case of the physiological model cited above; thus, the two models are not mutually exclusive (SI Text).

A yet unproven but reasonable ansatz would posit that this behavior might apply to multicellular or arbitrarily complex organisms as well, resulting in even broader validity. An indication supporting this statement is the experimental size distribution of a multicellular organism that we measured with the same (details concerning constraints on the exponents are provided in SI Text). To verify the hypothesis, we plot \( m^2 p_i(m) \) vs. \( m/(m_i)^2 \) for all 13 protist species (Fig. 2B) and vary \( \Delta \) and \( \phi \) until a satisfactory data collapse (45) is observed. The best collapse is found for \( \Delta = 1.0 \) (and therefore \( \phi = 1.0 \); Fig. 2B). A quantitative method (46) to produce the best collapse yields \( \Delta = 1.01 \pm 0.05 \) (Fig. 2B, Inset and SI Text). A relevant consequence of Eq. 1, where \( \Delta = \phi = 1 \), is that the jth moment \( (m_j) \) is proportional to \( (m_i) \) \( (\text{where } j = 1, 2, 3, \ldots) \). In particular, the variance of the species’ sizes does increase with the mean size. The proportionality of successive moments ratios to \( (m_i) \) provides an independent test that further corroborates the validity of Eq. 1 (Fig. 2C and SI Text). We have thus found that a single parameter, the average mass of a species, is sufficient to characterize its size distribution fully. This is far from trivial because, in general, a probability distribution is determined by all its moments (47).

Environmental factors are capable of affecting the size distribution of any given species (48, 49). To test the effects of environmental conditions on Eq. 1, we question whether the measured size distributions might still be described by the universal functional form:

\[
p(m) = \frac{1}{f} F \left( \frac{m}{m_i} \right),
\]

where \( \langle m \rangle \) is the mean mass, critically determined (i) by the species or (ii) by phenotypic plasticity due to environmental factors. To that end, we have investigated a set of manipulated ecological time scales. Although obviously far from exhausting field-like scenarios, a sizeable plasticity was observed (Fig. 3A and SI Text). Crucially, once rescaled by the actual mean body size of the sample, whether constrained by temperature or by competition, all distributions collapse again and the scaling exponent estimate proves to be unaffected (\( \Delta = 1.01 \pm 0.10 \); Fig. 3B, Inset). Significantly, therefore, environmental factors are capable of affecting the size distribution of any given species, although they do not alter its scaling nature because both the exponents and the scaling function \( F \) are unchanged.

The observed regularities are compelling, given the profound biological diversity of the species (belonging to the phyla/divisions of Ciliophora, Euglenozoa, Chlorophyta, and Cryptophyta) considered in this study (50). In addition, protists and unicellular algae are of key ecological significance. In fact, they are the basic food source of almost all aquatic food webs, and unicellular algae are responsible for almost 50% of the worldwide biomass production (51). Additionally, the observed universality of eukaryote intraspecies size distributions holds in a range of more than four orders of magnitude in mass. This suggests the existence of a simple underlying mechanism responsible for the empirical patterns observed. One possible explanation for the reported universality may be found in the physiological processes that determine the size of all unicellular species, namely, cellular growth and cell division. We found that a simple mathematical model of these processes (52–54) can justify the scaling form of unicellular eukaryotes’ size distributions without the need to specify further biological details (SI Text). We therefore suggest that the universal features of intraspecies size distributions emerge from fundamental physiological constraints.

The detailed identification of the scaling function \( F \) in Eq. 2 is interesting but inessential for our tenet because the collapse of the distributions suffices in documenting the universality sought after. However, a log-normal functional form for \( p(m) \) provides admisible and rooted in a theoretical framework for the time evolution of the distribution of body sizes in ecological time scales (Fig. 2D, Inset and SI Text). In this context, the size distribution of organisms of a given species is the stationary distribution of an Ornstein–Uhlenbeck process (55) in the variable \( x = \log(m/(m_i)) \) (56, 57). A log-normal form for \( p(m) \) can also be recovered as a particular case of the physiological model cited above; thus, the two models are not mutually exclusive (SI Text).

A yet unproven but reasonable ansatz would posit that this behavior might apply to multicellular or arbitrarily complex organisms as well, resulting in even broader validity. An indication supporting this statement is the experimental size distribution of a multicellular organism that we measured with the same
methods in our laboratory conditions. In fact, we found that the size distribution of a multicellular species (Cephalodella sp.) showed a very good collapse with the protist size distributions once rescaled according to Eq. 2 (Fig. 4).

The observation of more than 20 orders of magnitude in organismic sizes in natural ecosystems (2, 27) leads to the conclusion that there has been little long-term impediment to the development, on evolutionary time scales, of any particular size. On ecological time scales, however, a characteristic size emerges as a fundamental property of a species, determined by biological constraints and by biotic and abiotic interactions. Such a characteristic size, in turn, modulates the entire size distribution of the species.

Our results have important implications on scaling laws in ecology. They decouple the effects of biotic and abiotic interactions, which regulate abundances and can affect a species’ mean size, from the individuals’ physiology, which shapes intraspecies size distributions. Our replicated, controlled experiments corroborate and extend comparative field findings of marine microbial size spectra to broader size and taxonomic diversity (22). We speculate that such behavior can reasonably be expected to extend over broader domains. Then, theoretical linkages of diverse empirical macroecological relationships, traditionally treated as independent (30, 58), would be substantiated. In fact, because single-species size distributions would be characterized by specific mean values and variances, a precise requirement would be cast on the number of species existing at stationarity within a niche of size, and on the related abundances, to produce community size spectra that lack characteristic scales (SI Text). Ecological interactions among species would consequently adapt to produce thinning laws (i.e., to control the relative abundance of species, given their characteristic sizes) (29, 32). The fact that these thinning laws have been shown to be robust to perturbations further emphasizes their universal character (59).

Finally, a distribution of the form in Eq. 2 implies that the variance of the species’ sizes increases quadratically with the mean.

Fig. 2. Evidence for a universal single-species size distribution. (A) Volume probability distributions of 13 protist species, spanning four orders of magnitude in mass. (B) Data collapse of \( m^3 p(m) \) vs. \( m \langle m \rangle^2 \); the best collapse is observed for \( \Delta = \phi = 1.0 \). (Inset) Minimum of the functional \( E(\Delta) \) provides the best estimate for the exponent and the associated error (46) (SI Text). (C) Proportionality of successive moments of \( m \) to \( \langle m \rangle \) is an independent verification of the hypothesis in Eq. 1. (D) Fit of a Gaussian scaling function \( F \) as a function of log \( m/\langle m \rangle \) (dashed blue line) contrasting the ensemble average size distribution (red line); the orange region is the 99.7% confidence interval around the average. The scaling function yields a log-normal form for \( p_m(m) \) (SI Text).
size. Because the characteristic mass of a species is frequently adopted as the independent variable in allometric scaling laws (10, 11, 27–29, 32, 34–39), its increasing variance must have an impact on the scaling of the dependent variable, such as metabolic rates in Kleiber’s law. We thus pose the basis for a reexamination of allometric relations by considering appropriate fluctuations in both the dependent (31, 33) (metabolic rate) and independent (mass) variables.

Materials and Methods

Protist Cultures. Replicated single-species cultures of 13 different protists and unicellular algae (all called “protists”) were initialized with three species of freshwater bacteria (Serratia fonticola, Breviacciulus brevis, and Bacillus subtilis) as a food resource in a climatized room at 20 °C under constant fluorescent light 3 wk before the measurements. Previous studies (41, 42) support that the composition and size spectra of these communities are rather stable over this time period; thus, we can assume that cultures were at their carrying capacity while we performed the measurements. Single-species cultures were grown in 500-mL Schott flasks containing a nutrient medium made of sterilized local spring water and Protozoan Pellets (Carolina Biological Supply) at a density of 0.45 g/L. These are referred to as standard conditions concentration (EUG_4)]

Environmental Stress and Competition. In addition to the single-species cultures grown in standard conditions as described in the previous section (which are referred to in Fig. 3 as CHI_1, EUG_1, and EUP_1), we grew CHI, EUG, and EUP at other temperatures and nutrient conditions, or in competition with each other. In the latter case, the two competing species were always sufficiently separated in their size spectrum so that their two distributions did not overlap. We studied the following conditions with at least three replicates each:

i) Single species at 15 °C (EUG_2, EUP_2)

ii) Single species at 25 °C (CHI_2, EUG_3, and EUP_3)

iii) Competition at 20 °C: Two species compete for resources and have been initialized at half of their carrying capacity in 10-mL well plates [Chlamydomonas sp. with D. campylum (CHI_3), Chlamydomonas sp. with Colpidium sp. (CHI_4), and E. gracilis with the rotifer Cephalodella sp. (EUG_3)]

iv) E. gracilis at a low protist medium concentration [0.045 g/L (i.e., 1/10th of the standard conditions concentration [EUG_4])]

Size Distributions. We performed size distribution measurements with a Cell Counter and Analyzer System (CASY) model TTC (Roche Applied Science). Size measurements were performed by suspending a sample taken from a protist culture in a buffer solution (CASYton), which is developed specifically to aspirate cells through a precision hole in the instrument at a constant speed (61). To perform size measurements of protists, we used capillaries with diameters of 60 μm, 150 μm, and 200 μm depending on the size of the protists under investigation. Smaller capillaries resolve better size distributions at low scale (5–20 μm), but can be blocked if larger particles pass through (it is therefore necessary to use larger capillaries to measure larger species). As a general rule and for each species, we used the smallest capillary that enabled us to separate the protist peak unequivocally from the debris in the instrument output (61) (SI Text). The size spectrum of a sample of living cells is returned by the instrument as a function of the equivalent diameter l of each cell, assuming cells to be spherical. From the definition of size distribution, p(l)dl is the fraction of individuals with an equivalent diameter in (l, l + dl) and p(m)dm is the fraction of individuals with a mass in (m, m + dm). The value of p(m) can then be calculated via the variable transformation p(l)dl = p(m)dm (m = πl^3/6). We assume a constant density equal to the density of water (7, 62), and therefore refer to volume and mass without distinction. We also assume that size distributions do not depend on time (i.e., the cultures are in a steady state characterized by size distributions of constant shape). In a typical measurement output for a culture of protists,
a peak at small sizes exist due to debris in the culture (SI Text) (61). Peaks at larger sizes are due to protists. To deconvolve the two peaks, we fit the debris peak with an exponential decay (in a region adjacent to the peak, where data lie on a straight line in a log-linear plot; SI Text) and subtracted the resulting curve from the overall spectrum. On the right side of the protist peak, we truncated the data when the measured frequency of a size channel was below 20 occurrences to separate it from the noise. Noise was uniformly distributed on all size channels with a frequency of ~10–20 counts per channel, as demonstrated by measuring pure buffer solution only. For each species, several measurements of different cultures (grown in the same conditions) were collected and summed to get an ensemble average representative of the species.

ACKNOWLEDGMENTS. We thank Regula Illi, Mathew Seymour, and Enrico Bertuzzo for support during the experiment. We gratefully acknowledge the support provided by Swiss Federal Institute of Aquatic Science and Technology (Eawag) discretionary funds, the European Research Council advanced grant program through the project River networks as ecological corridors for biodiversity, populations and waterborne disease (RINEC-237212), and Swiss National Science Foundation Projects 200021_124930/1 and 31003A_135622.
Supporting Information

Giometto et al. 10.1073/pnas.1301552110

SI Text

Meaning of Power Laws. Power laws are ubiquitous in nature. They have been shown to describe the similarity of the part and the whole of many objects in nature, ranging from coastlines, clouds, or mountain ranges (1) to natural or artificial networks like rivers (2) and up to complex social interactions (3–5). In general, a homogeneous power function [e.g., \( f(x) = Cx^\phi \) with \( C, \alpha \) constants] is intrinsically self-similar: If \( x \) is rescaled (multiplied by a constant), then \( f(x) \) is still proportional to \( x^\phi \), albeit with a different constant of proportionality. Such functions reproduce themselves on rescaling, and therefore lack natural scales, do not harbor a characteristic unit, and are said to be scale-free or true on all scales. Power-law probability distributions of size may imply an infinite mean, unless a finite range of sizes is assumed. In such a case and depending on the value of the scaling exponent, the mean, variance, and progressive moments of the distributions diverge in the infinite range. To be finite, they must depend on a finite interval of sizes sampled; thus, the mean makes no sense as a property of a population (6). This analog to the syndrome of infinite variance (1) (i.e., the progressive divergence of the variance of a self-similar or self-affine signal as the sample size is enlarged) is widely held as the typical signature of scale invariance.

Finite-Size Scaling Distributions. A general account of finite-size scaling (7) in ecology is provided by Banavar et al. (8). General properties of finite-size scaling distributions (as in Eq. 1) are detailed in a study by Banaver et al. (9) and its supplementary information. A brief account of the derivation of the most relevant results, adapted to the case at hand, is given in the following sections.

Normalization. Here, we study the normalization conditions for size distributions of the form:

\[
p_k(m) = \frac{1}{m^{\phi}} F \left( \frac{m}{m_k^\phi} \right),
\]

where \( \phi > 0 \) and, for dimensional reasons:

\[
F \left( \frac{m}{m_k^\phi} \right) = \frac{1}{m_k^{1-\phi}} \Phi \left( \frac{m}{m_k^\phi} \right),
\]

where \( m_0 \) is the minimum mass of an organism or a cutoff in the system. From this point onward, we will measure all masses in units of \( m_0 \); thus, \( F \) and \( F \) coincide and \( \frac{m}{m_k} \to \langle m \rangle_k \) is arbitrarily large.

In order for the distribution in Eq. S1 to be normalized (i.e., \( \int dm p_k(m) = 1 \)), one needs to make the following assumptions:

\( F(x) \) approaches a constant when \( x \ll 1 \). 
\( F(x) \) goes to zero sufficiently fast when \( x \gg 1 \).

With these conditions, one has:

\[
1 = \int_1^\infty dm p_k(m) = \int_1^\infty dm \frac{1}{m^{\phi}} F \left( \frac{m}{m_k^\phi} \right) = \left( m_k^{\phi(1-\phi)} \right) \int_1^\infty dx \frac{1}{x^\phi} F(x) = \langle m \rangle_k^{1-\phi} \phi \int_1^\infty dx x^{-\phi} F(x) = a + b \langle m \rangle_k^{(1-\phi)\phi},
\]

where \( a \) and \( b \) are constants. Now, except for corrections to the scaling [From Eq. S3: \( p_k(m) = m^{-\phi} F \left( \frac{m}{m_k^\phi} \right) \left( a + b \langle m \rangle_k^{(1-\phi)\phi} \right)^{-1} = m^{-\phi} \frac{1}{\phi} F \left( \frac{m}{m_k^\phi} \right) - \frac{b}{\phi} \langle m \rangle_k^{(1-\phi)\phi} F \left( \frac{m}{m_k^\phi} \right) + \ldots \right] \), which adds an additional term to Eq. S1], everything is consistent if \( \alpha = 1, (1 - \Delta) \phi < 0 \), and \( \Delta > 1 (\phi > 0) \) or if \( \Delta = 1 \), in which case one has:

\[
1 = \int_1^\infty dm p_k(m) = \int_1^\infty dm \frac{1}{m} F \left( \frac{m}{m_k^\phi} \right) = \int_1^\infty dx \frac{1}{x} F(x),
\]

and two possibilities arise: (i) \( \int_0^\infty dx \frac{F(x)}{x} = 1, \) with \( F(x) \to 0 \) sufficiently fast for \( x \to 0 \) (which is consistent with our data), and (ii) \( F \) \( x \sim x^{-\phi} (-\ln x)^{-\alpha} \), such that \( \int_0^\infty dx \frac{1}{x} F(x) \sim (\ln m_k)^{1-\omega} \). If \( \alpha > 1 \), one is back to case (i), whereas if \( \alpha < 1 \), one has logarithmic corrections to the scaling. In fact, if \( \alpha < 1 \), one finds \( p_k(m) = \frac{1}{m} (\ln(m_k))^{1-\alpha} F \left( \frac{m}{m_k^{\phi}} \right) \), (i.e., a logarithmic correction to the scaling).

Successive Moments Ratios. A test for the validity of a scaling size distribution of the form \( p_k(m) = \frac{1}{m} F \left( \frac{m}{m_k^{\phi}} \right) \) is the proportionality of successive moments ratios \( \langle m^{\phi(j)} \rangle \langle m^{\phi(j-1)} \rangle \) to the first moment \( \langle m \rangle \). In fact, if \( p_k(m) = \frac{1}{m} F \left( \frac{m}{m_k^{\phi}} \right) \), one has:

\[
\frac{\langle m^{\phi(j)} \rangle}{\langle m^{\phi(j-1)} \rangle} = \int dm m^{\phi(j)} F \left( \frac{m}{m_k^{\phi}} \right) = \int dm m^{\phi(j-1)} F \left( \frac{m}{m_k^{\phi}} \right) = \langle m \rangle_k^{\phi(j) - \phi(j-1)} \langle m \rangle_k^{\phi(j-1)} \langle m \rangle_k^{\phi(j-1)}
\]

where \( x = \frac{m}{m_k^{\phi}} \) and \( j > 1 \). In Fig. S1, we plot successive moments ratios calculated from our data [cultures in standard conditions (data are shown in Fig. 2; see also Materials and Methods)] and linear regressions on these data. The slopes of the linear regressions are compatible with the value of 1 (linear regressions
on log-transformed data). Coefficient of determination $R^2$ values for the regressions are $R^2_{\text{m}^2/\text{m}_1}=0.999$, $R^2_{\text{m}^2/\text{m}_2}=0.996$, and $R^2_{\text{m}^2/\text{m}_3}=0.998$. The same is shown for the data of *Chilomonas* sp., *Euglena gracilis*, and *Euplotes aediculatus* in different environmental conditions in Fig. S2 (data are shown in Fig. 3; see also Materials and Methods). The slopes of the linear regressions are compatible with the value of 1 (linear regressions on log-transformed data). Coefficient of determination $R^2$ values for the regressions are $R^2_{\text{m}^2/\text{m}_1}=0.998$, $R^2_{\text{m}^2/\text{m}_2}=0.987$, and $R^2_{\text{m}^2/\text{m}_3}=0.962$.

**Modeling Cell Growth and Division.** To model growth and cell division, we studied the scaling properties of a simple model for these two processes (10), focusing our attention on unicellular vision, we studied the scaling properties of a simple model for

\[
\int_0^M dm \, b(m) = \infty. \tag{S6}
\]

Considering the balance of growth and division in an infinitesimal time interval $dt$ and in the size interval $[m_1, m_2]$, expanding at first order in $dt$, one has:

\[
\int_{m_1}^{m_2} dm \left( \frac{dN}{dt}(m,t) + \frac{\partial j}{\partial m}N(m,t) \right) + b(m)N(m,t) - 4b(2m)N(2m,t) = 0. \tag{S7}
\]

The equation governing the balance of growth and cell division is then:

\[
\frac{d}{dt}N(m,t) + \frac{\partial j}{\partial m}N(m,t) + b(m)N(m,t) - 4b(2m)N(2m,t) = 0. \tag{S8}
\]

Let us assume that $N(m,t) = \lambda(m)e^{\lambda t}$ at stationarity, where $\lambda(m)$ is proportional to the stationary cell size distribution. Introducing $N(m,t) = \lambda(m)e^{\lambda t}$ in Eq. S8, one finds for $\lambda(m)$:

\[
\int_0^M dm \lambda(m) + \mu \int_0^M \frac{d[m\lambda(m)]}{dm} = \int_0^M dm b(m)\lambda(m). \tag{S10}
\]

Eq. S10 imposes $\int_0^M dm b(m)\lambda(m) < \infty$; therefore, in Eq. S6, one has $\lim_{m \to M} \lambda(m) = 0$, which implies $\int_0^M \frac{d[m\lambda(m)]}{dm} = 0$, and as a result:

\[
\int_0^M \frac{d[m\lambda(m)]}{dm} = \mu \int_0^M dm b(m)\lambda(m). \tag{S11}
\]

The scale invariance of $\lambda(m)$ can be deduced directly from Eq. S9 as follows. The value of $\lambda$ depends on $m$ and $M$ [i.e., $\lambda = \lambda(m,M)$]. We assume that the total mass present at $t = 0$ is equal to 1 (Eq. S9 is linear in $\lambda$; therefore, if $\lambda$ is a solution, so is $C\lambda$ with $C$ as an arbitrary constant). Because $M$ is the only scale in the problem, we assume that $b(m,M) = b\left(\frac{m}{M}\right)$ and rewrite Eq. S9 with $x = m/M$ as follows:

\[
\mu \frac{d[x\lambda(Mx,M)]}{dx} = \left[k + b\left(\frac{x}{M}\right)\right]x\lambda(Mx,M) + 4b(2x)x\lambda(2Mx,M), \tag{S12}
\]

where $x \in [0, 1]$. Therefore, one has the solution $\lambda(Mx,M) = \lambda(x)\lambda(M)$ [i.e., $\lambda(m,M) = \frac{1}{M} \lambda(m,M)$], which satisfies:

\[
\mu \frac{d[\lambda(x)]}{dx} = \left[k + b(x)\right]x\lambda(x) + 4b(2x)x\lambda(2x) \tag{S13}
\]

in $x \in [0, 1]$, with $\int_0^1 dx \lambda(x) = 1$ and $\lambda(x) = 0 \forall x > 1$. In particular, the size distribution $p(m)$ can be written as:

\[
p(m) = \frac{1}{m^2} \int_0^1 \frac{dy}{G(y)} G\left(\frac{M}{M}\right). \tag{S14}
\]

Computing the average mass $\langle m \rangle$, one finds it is proportional to $M$:

\[
\langle m \rangle = \int_0^M dm m p(m) = M \int_0^1 \frac{dy}{G(y)} G\left(\frac{m}{m}\right), \tag{S15}
\]

such that the stationary size distribution is of the form:

\[
p(m) = \frac{1}{m} F\left(\frac{m}{M}\right), \tag{S16}
\]

which is precisely the scaling ansatz proposed and observed in the data.

The solution to Eq. S13 can be written as $\lambda(x) = \lambda_0(x)$ with $2^{-n} \leq x \leq 2^{-n+1}$ for $n = 1, 2, \ldots$ and

\[
\lambda_0(x) = e^{-\int_0^x \frac{dy}{h(y)}} C_n + 4 \mu \int_{2^{-n+1}}^{2^{-n}} \frac{dy}{y} h(\lambda_{n-1}(y)), \tag{S17}
\]

where $\lambda_0(y) = 1, x \in [2^{-n}, 2^{-n+1}], h(y) = |k + b(y) + \mu|/y$, and $C_0$ depends on the normalization condition [i.e., $\int_0^1 dy \lambda(x) = 1$]. The $C_n$s ($n \geq 1$) are determined recursively imposing the continuity $\lambda_0(2^{-n+1}) = \lambda_{n-1}(2^{-n+1})$ for $n \geq 2$. For instance, if $n = 2$:

\[
\lambda_1(x) = C_0 e^{-\int_0^x \frac{dy}{h(y)}} \Rightarrow \lambda_1\left(\frac{1}{2}\right) = C_0 \tag{S18}
\]

(note that $\lambda_1(1) = 0$ due to the singularity in Eq. S6):

Giometto et al. www.pnas.org/cgi/content/short/1301552110
the phenotype. A phenotypic character like body size, therefore, (12), has been identi
cal location, which, in the context of phenotypic evolution
portional to the mean $k$ the characteristic mass of the reference species
imply
crease) of the mass corresponds to a shift to the right (left) of the
walk in the variable $x$
where $\lambda$ is the natural logarithm (i.e., the logarithm to the base $e$).

Analytical Form of the Universal Size Distribution. A fitting procedure suggests the viability of an analytical log-normal form for the
universal size distribution; that is,

$$p(m) = \frac{1}{m \sqrt{2\pi \sigma^2}} e^{-\left(\frac{m - \mu}{\sigma}\right)^2}.$$ \[S20\]

where $\sigma$ and $\mu$ are constants, $(m)$ depends on the species, and $\ln$ is the natural logarithm (i.e., the logarithm to the base $e$).

In order for the distribution in Eq. S20 to have the scaling form $p(m) = U m F(m/(m))$, one has to impose that $(m) = \int \frac{m}{p} dm m p(m)$, which implies $\mu = -\sigma^2/2$ (i.e., $\mu$ and $\sigma$ are not independent). We thus propose the following analytical form for the universal size distribution, which depends on only one parameter, $\sigma^2$:

$$p(m) = \frac{1}{m \sqrt{2\pi \sigma^2}} e^{-\left(\frac{m - \mu}{\sigma}\right)^2}.$$ \[S21\]

The scaling function $F(x)$ is therefore of the form:

$$F(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\left(\frac{x - \mu}{\sigma}\right)^2}.$$ \[S22\]

as suggested by the fact that a parabola fits the log-transformed data $mp(m)$ vs. $m/(m)$ (least-squares fit on log-transformed data)
well. To have a good estimate of the mean of $F$, we performed the
fit to Eq. S22 in the common support of at least half of the
protoplasmic life of the cell, with only one parameter, $\sigma^2$. The best estimate for the parameter is $\sigma^2 = 0.222(3)$, and the coefficient of
determination is $R^2 = 0.92$. The fit of the scaling function is shown in Fig. S3A, superimposed on the measured single-species size
distributions. Fig. S3B compares the fit of a log-normal scaling
function with the ensemble average of the experimental size
distributions, showing a remarkable overlap.

We now turn to an illustration of how the postulated universal scaling form for $p(m)$ might arise from general dynamical
considerations (9, 12, 13). Consider ecosystem diversity over ecological time scales. Ecological processes governing the
abundances and niche occupancy of species are expected to
to change their characteristic size. One would expect, however, that
offsprings would have a mass proportional to the mass of the
parent organism (13). Thus, fluctuations in size within same
species ought to be measured on the order of percent variations,
and the natural variable is $x = \log(m/\bar{m})$ (12, 13), with $\bar{m}$ being the characteristic mass of the reference species $k$ (e.g.,
proportional to the mean $\bar{m} = a(m)$) (9, 14). In this framework, the results of ecological processes can be represented by a random
walk in the variable $x$, because a fixed percent increase (de-
crease) of the size corresponds to a shift to the right (left) of the
variable $x$ by a constant amount. In the simplest model, the
results of ecological processes could be represented by an Orn-
stein–Uhlenbeck process (15, 16). This process is a modification of a Wiener process, where the walk tends to move toward
a central location, which, in the context of phenotypic evolution
(12), has been identified as the optimum in the adaptive zone for
the phenotype. A phenotypic character like body size, therefore,
is expected to be distributed around a fitness optimum in this
framework. The physical analogy of this process is a noisy re-

$$\frac{\partial q(x,t)}{\partial t} = \frac{D}{2} \frac{\partial^2 q(x,t)}{\partial x^2} + \frac{\partial [kq(x,t)]}{\partial x},$$ \[S23\]

where $k$ is a constant, $k > 0$. The stationary solution is obtained by setting the rate of change of $q(x,t)$ to zero, and it is known to be
Gaussian (15, 16):

$$q(x) = \sqrt{\frac{k}{\pi D}} e^{-\frac{x^2}{2D}}.$$ \[S24\]

For the mass distribution, one then obtains $p(m)$:

$$p(m) = q(x) \frac{dx}{dm} = \frac{1}{m} \sqrt{\frac{k}{2\pi D}} e^{-\frac{(m - \mu)^2}{2D}}.$$ \[S25\]

(i.e., a log-normal distribution of mass). Imposing Eq. S25 to have
mean $(\mu)$, one finds $\alpha = \exp[-D(2k)]$, and the distribution of
size is therefore:

$$p(m) = q(x) \frac{dx}{dm} = \frac{1}{m} \sqrt{\frac{k}{2\pi D}} e^{-\frac{(m - \mu + \frac{\mu}{\alpha})^2}{2D}}.$$ \[S26\]

(i.e., a log-normal distribution of mass with mean $(\mu)$ as in Eq. S21). Therefore, the scaling function $F$ in Fig. 2 and Fig. S3 is $F(x) = \frac{\sqrt{m}}{\sqrt{2\pi D}} \exp \left[ -\frac{1}{2D} \left( \frac{m - \mu}{\sqrt{2D}} \right)^2 \right]$. \[S27\]

One might wonder whether the size distribution obtained in
Eq. S26 is in agreement with our model of cellular growth and
division. The cellular growth and division model, as treated in
the previous section, assumes the existence of a maximum mass
and allows one to study the scaling properties of the stationary
size distribution. It is possible, however, to relax this hypothesis,
allowing the cells to assume all masses in the range $[0, \infty]$ and to
obtain an implicit relation for the stationary size distribution
$p(m)$ (17), which allows one to compute the asymptotic behavior
of the distribution for large mass (i.e., $(m) \Rightarrow \infty$). In the notation
of the previous section, the size distribution for large $m$ satisfies the relation:

$$p(m) \Rightarrow \frac{1}{m} \exp \left[ -\frac{m}{\bar{m}} \frac{k + d(y)}{\mu y} \right],$$ \[S27\]

which behaves as a log-normal distribution if we further assume a division rate $d(y)$ increasing logarithmically with size
[i.e., $d(y) \propto \ln y$].

On the Goodness of Data Collapses. Data collapse is a tool widely
used in statistical physics to establish scaling laws and extract
information on their exponents (18). Traditionally, the
procedure to produce a data collapse is to rely on the direct vis-
ualization of it and on “eyeballing” the exponent that gives
the best collapse. A less subjective method has been proposed
(19), which introduces a measure (error functional $E$; Figs. 2 and
3, Insets) to quantify the goodness of a collapse. Let $\Lambda$ be the
exponent that we tune to find the best collapse: $E(\Lambda)$ is the
cumulative area enclosed between two pairs of curves that we

Giochetto et al. www.pnas.org/cgi/content/short/1301552110 3 of 7
try to collapse, within their common support, for the value $\Delta$ of the exponent. The value $\Delta^*$ of the exponent, which minimizes $E(\Delta)$, is taken as the best estimate for the exponent (i.e., the smaller the area, the better is the collapse). Errors are associated with the determination of $\Delta^*$ and are obtained from the width of the minimum. Further details can be found in the study by Bhattacharjee and Seno (19).

**Community Size Spectra.** In this section, we show how a power law community-size spectrum arises as a sum of single-species distributions of finite-size scaling form (14). Assume that the size distribution of species $k$ is of the form:

$$p_k(m) = \frac{1}{m} F\left(\frac{m}{m_k}\right). \quad [S28]$$

Let $N_k$ be the stationary abundance of species $k$ in an ecosystem [i.e., $N_k = N_k(t \to \infty)$] and $S$ be the total number of species. The community size spectrum is defined as:

$$f(m) = \sum_{k=1}^{S} N_k p_k(m) / \sum_{k=1}^{S} N_k. \quad [S29]$$

We assume, supported by a number of observations (20), that the population abundance of the $k$th species scales as:

$$N_k \propto (m_k)^{a}.$$

where $a < 0$ implies that the total number of organisms decreases with increasing typical size. From Eqs. 28 and 30, one has that:

$$f(m) \propto \sum_{k=1}^{S} N_k p_k(m) \propto \sum_{k=1}^{S} (m_k)^{a} m^{-1} F\left(\frac{m}{m_k}\right). \quad [S31]$$

Let $g(m)$ be the fraction of species of typical size $m$. The above equation can be rewritten, treating $(m_k)$ as a continuous variable for easiness of computation, as:

$$f(m) \propto \frac{1}{m} \int dm g(m) m^a F\left(\frac{m}{m}\right) \propto \int dm g(m) m^a F\left(\frac{1}{m}\right). \quad [S32]$$

Theoretical predictions from a scaling macroecological framework (9, 14) and considerations on the total number of species on Earth (21, 22) suggest a pure power-law behavior for $g(m)$:

$$g(m) \propto \frac{1}{m^\beta}. \quad [S33]$$

which, because of normalization, is assumed to hold between an upper cutoff and a lower cutoff. One then has for the size spectrum $f(m)$:

$$f(m) \propto m^{a-\beta} \int dx x^{\alpha-\beta} F\left(\frac{1}{x}\right) \propto m^{a-\beta}. \quad [S34]$$

which has the form of a power law. Note, however, that the result still holds for log-normal species abundance distributions (22). In the case of a limited range of sizes, one might argue that the number of species $S$ within the range of sizes investigated could be assumed as constant to first order. This, of course, is the particular case for which $\beta = 0$.

Overall, it is clear that to obtain a scaling community size spectrum (Eq. S29), a necessary condition is adaptive fine-tuning of the specific abundances. This is epitomized by the relation in Eq. S30, which, in turn, implies the thinning relations that are recurrent in the literature of macroecological empirical laws (9, 23).

**Equal Biomass in Each Size Class.** The case of $f(m) \propto m^{-2}$, which is routinely found in the literature (14, 24), is of special interest because it agrees with the assumption of constant biomass in each size class (25). In this case, the total mass in a range $(m, m + \Delta m)$ with $\Delta m / m \approx 1$ is:

$$\int_{m}^{m + \Delta m} dx f(x) \propto \log \left[1 + \frac{\Delta m}{m}\right] \approx 1. \quad [S35]$$

such that the total biomass in a range between $m$ and $m + \Delta m$, where $\Delta m / m$ is the typical variance at scale $m$, is independent of $m$.

**Bodo saltans Size Distribution.** In Fig. 2, the size distribution of Bodo saltans (BOD) might seem not to collapse as well as that of the other species. This is due to an instrumentation limit and is not inherent to the BOD size distribution. BOD, in fact, has a mean volume of 17 $\mu m^3$ and a mean equivalent diameter of 3.1 $\mu m$. These values lie on the leftmost side of the size spectrum, where the debris peak is dominant and the instrument hardly resolves the peak of the protist culture (Fig. S4, Inset). As an effect of this, it is hard to separate the protist peak from the debris; consequently, the left side of the BOD size distribution is overestimated, causing a deviation from the other collapsing curves in Fig. 2. To show that this is indeed the case, we measured the size distribution of Chlorogonium euchlorum (CHO) with the 150-µm capillary of the Cell Counter and Analyzer System (CASY) model TTC (Roche Applied Science), whereas we used the 60-µm CASY capillary to obtain the data reported in the main text. Using the larger capillary, the protist and the debris peaks appear closer to each other (with there being fewer size channels in the [0, 20]-µm interval; Fig. S5, Inset), and as a result, the size of the CHO size distribution on the left is overestimated. If we plot the corresponding rescaled distribution, together with the other data of Fig. 2, we observe the same kind of deviation observed for BOD (Fig. S5). We argue, therefore, that if the instrumentation could resolve better the small-sized region of the BOD size distribution, it would collapse even better than it does in Fig. 2.


**Fig. S1.** Ratios of successive moments of $m$ are proportional to $\langle m \rangle$. Moments were calculated from size distributions in standard conditions (i.e., calculated from the size distributions shown in Fig. 2 and Materials and Methods).

**Fig. S2.** Ratios of successive moments of $m$ are proportional to $\langle m \rangle$. Moments were calculated from size distributions in nonstandard conditions (i.e., calculated from the size distributions shown in Fig. 3 and Materials and Methods).
Fig. S3. (A) Same collapse of rescaled size distributions as in Fig. 2, with the best log-normal fit superimposed (dashed blue line). The scaling function $F$ is a parabola in a log-log plot; thus, the universal size distribution $p(m)$ is log-normal. Colors are as in Figs. 1 and 2. (B) Average of all curves in A (red curve) shows a remarkable resemblance of the best log-normal fit (dashed blue line). The orange region is the 99.7% confidence interval around the average.

Fig. S4. Size distribution of BOD as a function of the equivalent diameter. (Inset) Output of the CASY instrument for a BOD culture. The protist peak is close to the debris peak on the leftmost side of the spectrum.

Fig. S5. Collapse of rescaled size distributions (as in Fig. 2), where the curve of CHO was measured with a large capillary (150 $\mu$m, black dashed line). The size of the distribution on the left is overestimated and causes the collapse to fail. Colors are as in Figs. 1 and 2. (Inset) CHO size distributions measured with the 60-$\mu$m capillary (red curve) and with the 150-$\mu$m capillary (black dashed line). The 60-$\mu$m capillary correctly resolves the whole spectrum, whereas the 150-$\mu$m capillary overestimates the left side of the distribution.
Fig. S6. Output of a measurement for Chilomonas sp. (CHI) culture. The superimposed exponential decay is the exponential fit of the debris in the region adjacent to the protist peak. (Inset) Decay of the debris to the left of the protist peak is exponential (straight line in a log-linear plot).

Fig. S7. Size distribution of Chilomonas sp. (CHI) as a function of the equivalent diameter. (Inset) Transformed volume size distribution of CHI.