An integrative framework for stochastic, size-structured community assembly

J. P. O'Dwyer[a,1], J. K. Lake[b], A. Ostling[b], V. M. Savage[c,d], and J. L. Green[a,e]

[a]Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, OR 97403-5289; [b]Ecology and Evolutionary Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048; [c]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115; [d]Department of Biomathematics, UCLA Medical Center, Box 951766, Los Angeles, CA 90095-1766; and [e]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved February 10, 2009 (received for review December 21, 2008)

We present a theoretical framework to describe stochastic, size-structured community assembly, and use this framework to make community-level ecological predictions. Our model can be thought of as adding biological realism to Neutral Biodiversity Theory by incorporating size variation and growth dynamics, and allowing demographic rates to depend on the sizes of individuals. We find that the species abundance distribution (SAD) is insensitive to the details of the size structure in our model, demonstrating that the SAD is a poor indicator of size-dependent processes. We also derive the species biomass distribution (SBD) and find that the form of the SBD depends on the underlying size structure. This leads to a prescription for testing multiple, intertwined ecological predictions of the model, and provides evidence that alternatives to the traditional SAD are more closely tied to certain ecological processes. Finally, we describe how our framework may be extended to make predictions for more general types of community structure.

neutral theory | allometry | ecological drift | complex systems

The introduction of Neutral Biodiversity Theory (1) (NBT) ignited an ideological debate in community ecology, challenging the viewpoint that deterministic forces play the dominant role in shaping patterns of biodiversity (2–9). NBT proposes that community-level patterns are primarily determined by the effects of demographic stochasticity, and that a detailed knowledge of the traits of and interactions between individuals comprising the community is irrelevant. It has been argued that the predictions of NBT are uninformative of process (10–12), and it has been demonstrated that the assumptions underpinning NBT are often manifestly violated: in the tropical forests where the theory has found striking success (1, 13, 14), there is a huge variation in demographic rates of individuals (15, 16). Although there has been much debate (6, 17, 18) about the importance of neutral, stochastic processes, one common principle has emerged (17–20): the need for a unified, theoretical framework both to quantify the effects of demographic stochasticity relative to other forces, and to generate a broader range of predictions more closely tied to process.

In this article, we integrate the effects of demographic stochasticity with ontogenetic variation in the size of individuals (21), and allow the demographic rates of individuals to explicitly depend on their size. We assume that this variation in demographic rates depends on size alone, and is not linked to species identity, an approach that is closely related to the philosophy of allometric scaling theory (8, 22–24): individuals of a given size play by the same rules, regardless of species identity. Given the strong evidence that demographic rates in nature are correlated with size (8, 16), this synthesis of size variation with demographic stochasticity may be thought of as adding a crucial extra layer of biological realism to NBT.

From these ingredients we derive a functional differential equation to describe ecological communities, and use its analytical solution to answer three key questions. First, are our predictions consistent with deterministic, size-structured models? It is often assumed that demographic stochasticity will have little impact on deterministic results (25), and our model allows us to quantify its effects analytically. Second, are our predictions consistent with NBT? To answer this question we compare the species abundance distributon (SAD) of our size-structured, stochastic model with the SAD predicted by NBT. Finally, can we make macroecological predictions by using our framework? To address this we derive the first theoretical prediction for the species biomass distribution (SBD), a characterization of community structure where abundance is measured in a currency other than number of individuals (26).

Background: Size Structure and Stochasticity

Our model of size-structured, stochastic community assembly may be regarded as a synthesis of two earlier theories, which we review in this section. The first is the Von Foerster equation, a continuity equation that predicts the size spectrum of a population, based on the scaling of individual growth and mortality rates with size. It was introduced in an epidemiological context by McKendrick (27), and later developed by Von Foerster (28) to describe the distribution of ages in cell populations. More recently, equilibrium solutions of the Von Foerster equation have found applications in community ecology, where they have been used to characterize the abundance of trees in tropical forests as a function of their basal diameter (29).

The second ingredient is Hubbell's Neutral Biodiversity Theory (1), which makes general predictions for the distribution of species abundances in a community, based on the effects of demographic stochasticity. At the metacommunity level, NBT describes a community of individuals undergoing stochastic birth, death, and speciation processes, and individuals of all species are taken to have the same demographic rates. Birth rate is taken to be less than mortality rate, so that every extant species is taking a random but inexorable walk toward extinction. We note that this approach of fitness equalization contrasts with many niche-based approaches to species coexistence and community assembly, which emphasize the importance of stabilizing mechanisms (7, 30).

Size Structure and the Von Foerster Equation. The Von Foerster equation predicts the abundance of individuals per unit size, $n(m,t)$, as a function of size, m, and time, t. The equilibrium distribution, $n(m)$, is known in ecology as an individual size distribution or a size spectrum (31), and is a time-independent solution of the following differential equation:

$$\frac{\partial n(m,t)}{\partial t} = -\frac{\partial\left(g(m)n(m,t)\right)}{\partial m} - d(m)n(m,t). \qquad [1]$$

Author contributions: J.P.O., A.O., V.M.S., and J.L.G. designed research; J.P.O., J.K.L., A.O., V.M.S., and J.L.G. performed research; J.P.O. contributed new reagents/analytic tools; and J.P.O., J.K.L., A.O., V.M.S., and J.L.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. E-mail: jodwyer@uoregon.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0813041106/DCSupplemental

The inputs for this equation are the scaling of ontogenetic growth of individuals, $g(m)$ and mortality rate, $d(m)$ with body size, and the boundary condition at the smallest size, m_0, is determined by a balance between birth, growth, and immigration.

The intuitive meaning of the Von Foerster equation is simply that growth into a given body-size class is balanced by growth out, and mortality. However, the form of the resulting size spectrum depends on the precise scaling of growth and mortality as a function of size. These scalings may either be thought of as theoretical predictions, for example, from the metabolic theory of ecology (8, 24), or as empirical inputs.

Stochasticity and Neutral Biodiversity Theory. We use the formulation of Volkov et al. (13), which is essentially equivalent to Hubbell's original zero-sum model (1, 32). The dynamics of the species abundance distribution, $P(N,t)$, are described by a master equation (1, 13, 33, 34). Births and deaths are modeled as density-independent processes, so that births of individuals occur at a constant per capita rate, b, and deaths at a rate, d, such that $b < d$. Speciation is modeled as an immigration process at rate ν from the pool of all possible species:

$$\frac{dP(N,t)}{dt} = b(N-1)P(N-1,t) + d(N+1)P(N+1,t) - (b+d)NP(N,t) + \nu P(N-1,t) - \nu P(N,t). \quad [2]$$

In the limit of small speciation rate, ν, relative to birth rate, b, the NBT prediction for the equilibrium SAD is the following time-independent solution of Eq. **2**:

$$P(N) = \frac{\nu}{bN}\left(\frac{b}{d}\right)^N = \frac{\nu}{bN}\left(\frac{\langle N\rangle}{\langle N\rangle + \frac{\nu}{b}}\right)^N, \quad [3]$$

where the second equation uses the mean abundance per species, $\langle N\rangle$. Hence, NBT predicts the familiar log series species abundance distribution. This distribution was first introduced in an ecological context by Fisher, who took the precisely same limit to obtain the log series as a limit of the negative binomial distribution (35).

Combining Stochasticity with Size-Structure

Our model describes a community of individuals structured by their size, m, and may be thought of as a synthesis of the Von Foerster equation with NBT. We interpret size flexibly, so that m may, for example, represent length, diameter, volume, or mass. Each individual undergoes a continuous, deterministic growth process with size-dependent growth rate, $g(m)$, while birth, death, and speciation processes are stochastic. Individuals die with a size-dependent per capita rate, $d(m)$, and individuals give birth with a size-independent per capita rate, b, to new individuals who enter the lowest size class, m_0. (The case of fixed birth rate, b, is more tractable than the general case of a size-dependent birth rate, which we address in *Discussion*.) Finally, there is speciation at a rate ν, modeled as an immigration from the pool of all species into the lowest size class. The conceptual distinction between our model and NBT is illustrated in Fig. 1.

Master Equation and Partition Function. We now derive the central equation of our model step-by-step, beginning with an approximation based on discrete size classes (33, 36–38). As in NBT, the dynamics of this discrete community are described by a master equation, but in this case the description of a population is given by a set of abundances in discrete size classes, $\{n_i\}$, where each n_i is the abundance in size class i. The probability of finding the population in a given state at time t is $P(n_0, n_1, \ldots, t)$, which we take to satisfy the following discrete master equation:

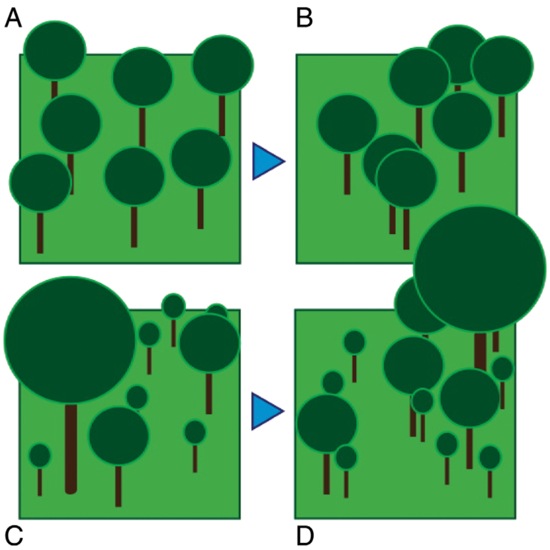

Fig. 1. We illustrate the conceptual differences between NBT and our model by using a single tree species in a forest. (*A* and *B*) Shown is the time evolution of a neutral community through one time step, undergoing stochastic birth and death processes. (*C* and *D*) Shown is the time evolution of a stochastic, size-structured community. Demographic processes are still random, but demographic rates and ontogenetic growth rates depend on size.

$$\begin{aligned}\frac{\partial P(n_0,\ldots,t)}{\partial t} = &\sum_{i=0} d_i(n_i+1)P(n_0,\ldots,n_i+1,\ldots,t)\\ &-\sum_{i=0} d_i n_i P(n_0,\ldots,n_i,\ldots,t)\\ &+\sum_{i=0} b(n_i-\delta_{i0})P(n_0-1,\ldots,n_i,\ldots,t)\\ &-\sum_{i=0} b n_i P(n_0,\ldots,n_i,\ldots,t)\\ &+\nu P(n_0-1,n_1,\ldots,t)-\nu P(n_0,n_1,\ldots,t)\\ &+\sum_{i=0} g_i(n_i+1)P(n_0,\ldots,n_i+1,n_{i+1}-1,\ldots,t)\\ &-\sum_{i=0} g_i n_i P(n_0,\ldots,n_i,\ldots,t) \quad [4]\end{aligned}$$

The mortality terms are precisely analogous to the familiar case of NBT, but instead of a single pair of mortality terms there is now a pair of terms for each size class, labeled by i. The birth terms also take a familiar form, but with the nuance that all individuals are born into the lowest size class, labeled by 0. The third set of demographic terms represent a stochastic speciation/immigration process, with probability ν per unit time of a speciation event occurring, where an individual from a new species is born into the lowest size class. Finally, g_i represents the probability per unit time that an individual in size class i grows into size class $i+1$, a process with no analogue in NBT.

Our aim is to take the biologically relevant limit of Eq. **4** in which discrete size classes become continuous. Taking this limit becomes considerably easier when Eq. **4** is cast in terms of the multivariate generating function, which is defined as:

$$Z(h_0,h_1,\ldots,t) = \sum_{\{n_i\}} P(n_0,n_1,\ldots,t)e^{\sum h_i n_i}. \quad [5]$$

The sum is taken over all possible combinations of abundances, $\{n_i\}$, and the definition of the generating function means that derivatives of $\log Z$, taken at $h_i = 0$, are equal to the moments of the distribution, P. With knowledge of all the moments, one can exactly reconstruct the distribution P, and hence, in principle, address any relevant biological question. Although the use of a generating function may be unfamiliar in this context, we note that one could also describe NBT by using a generating function, $z(h)$, with a single parameter h. Derivatives with respect to h generate the moments of the SAD in the NBT case.

To express the master equation (Eq. **4**) in terms of Z, we multiply each term in Eq. **4** by $e^{\sum h_i n_i}$ and then sum over all combinations of abundances, $\{n_i\}$. This leads to the following transformed master equation:

$$\frac{\partial Z}{\partial t} = \sum_{i=0} d_i \frac{\partial Z}{\partial h_i}(e^{-h_i} - 1) + \sum_{i=0} g_i \frac{\partial Z}{\partial h_i}(e^{h_{i+1}-h_i} - 1) + \left[\nu Z + \sum_{i=0} b \frac{\partial Z}{\partial h_i}\right](e^{h_0} - 1). \qquad [6]$$

Having made this transformation, it is now possible to take the limit of continuous size, as the separation between size classes, $\Delta m \to 0$. We use dimensional analysis to assign the following scalings with Δm:

$$g_i = \frac{g(m_i)}{\Delta m}$$
$$d_i = d(m_i)$$
$$h_i = H(m_i).$$

In the limit of Δm going to zero, the function $Z(h_0, h_1, \ldots)$ becomes a functional, $Z[H]$, which depends on a continuous function $H(m)$ instead of the discrete set of variables, $\{h_i\}$. $Z[H]$ is known as a partition function in statistical physics and quantum field theory (39), and we use this terminology for the remainder of the article.

In the continuum limit, sums over partial derivatives in Eq. **6** become integrals over functional derivatives (which are defined in supporting information (SI) *Appendix*) and the partition function $Z[H]$ satisfies the following functional master equation:

$$\frac{\partial Z}{\partial t} = \int_{m_0}^{\infty} dm\, d(m) \frac{\delta Z}{\delta H(m)}(e^{-H(m)} - 1) + \int_{m_0}^{\infty} dm\, g(m) \frac{\delta Z}{\delta H(m)} \frac{dH}{dm} + \left[b \int_{m_0}^{\infty} dm \frac{\delta Z}{\delta H(m)} + \nu Z\right](e^{H(m_0)} - 1). \qquad [7]$$

The inputs for Eq. **7** are the size-dependent growth and mortality rates of individuals, $g(m)$ and $d(m)$, respectively, the size-independent birth rate, b, and the rate of speciation, ν. The first term on the RHS may be thought of as a size-structured generalization of the mortality terms in the NBT master equation (Eq. **2**). The last term generalizes the birth and speciation terms, while the middle term of Eq. **7** has no analogue in NBT, because it represents the effect of ontogenetic growth on community structure. Our master equation may be thought of as a unification of the Von Foerster equation, Eq. **1**, and the master equation of NBT, Eq. **2**, and solving this equation for the partition function is the central question in our theoretical framework.

Characterizing Community Properties

In this section we present a time-independent solution of our key equation, Eq. **7**, and explore the implications of this solution for the properties of size-structured ecological communities. The solution is as follows:

$$\log Z = -\frac{\nu}{b} \log\left[1 - \frac{b}{\nu}\int_{m_0}^{\infty} dm\, f(m)(e^{H(m)} - 1)\right]. \qquad [8]$$

The input parameters and functions are encoded in the function $f(m)$, which must satisfy the following differential equation and boundary condition:

$$\frac{d}{dm}\big(g(m)\, f(m)\big) + d(m)\, f(m) = 0 \qquad [9]$$

$$f(m_0)\, g(m_0) = \nu + b \int_{m_0}^{\infty} dm\, f(m). \qquad [10]$$

The manipulations necessary to show that Eq. **8** solves Eq. **7** are demonstrated explicitly in *SI Appendix*.

In the remainder of this section, we examine in detail the properties and ecological consequences of this equilibrium solution in three steps. First, we derive the expectation value of the equilibrium size spectrum. This is given by $\langle n(m) \rangle$, the average number of individuals per unit size, per species. We also characterize the fluctuations around this solution and correlations between different sizes, which are analogous to correlation functions in spatial ecology (40, 41). Second, we derive the equilibrium species abundance distribution, integrated across all sizes. Finally, we reinterpret m as individual mass, and derive the corresponding species biomass distribution.

Size Spectrum and Fluctuations. Our partition function is formally defined as the expectation value,

$$Z[H] = \left\langle e^{\int_{m_0}^{\infty} dm\, H(m) n(m)} \right\rangle, \qquad [11]$$

which is the continuum limit of Eq. **5**. This means that we can obtain the expectation value of abundance, $\langle n(m) \rangle$, by taking a single functional derivative of our solution with respect to $H(m)$, and then setting $H = 0$. Taking a single functional derivative of Eq. **8** we find:

$$\langle n(m) \rangle = f(m).$$

This means that the mean abundance per unit size is just equal to $f(m)$, which we know must satisfy Eq. **9**. Comparing Eq. **9** with Eq. **1**, the meaning of this solution and the significance of the function $f(m)$ become quite clear: the expectation value $\langle n(m) \rangle$ is a time-independent solution of the deterministic Von Foerster equation. This means that the effects of demographic stochasticity completely average out, and we have the same result for the mean size spectrum as we would have had without any stochasticity at all.

The Von Foerster equation does not allow for any characterization of fluctuations around the mean size spectrum, because it is an entirely deterministic model. In contrast, our framework characterizes the size and nature of the fluctuations around this mean spectrum, either through the variance of fluctuations within a size class, or more generally the correlation of fluctuations between two different size classes. The two-point correlation in fluctuations between size classes m_1 and m_2 is given by the second functional derivative of Eq. **8** with respect to H, which we find is:

$$\langle n(m_1) \rangle \delta(m_1 - m_2) + \frac{b}{\nu} \langle n(m_1) \rangle \langle n(m_2) \rangle. \qquad [12]$$

If we had found only the first term, the Dirac delta function tells us that this would correspond to a random, Poisson process acting independently in each size class. But the second term of Eq. **12** means that the birth process introduces positive, nonzero correlations in the fluctuations across different size classes.

Much like the mean size spectrum, this positive correlation of fluctuations across different sizes also has an intuitive meaning. Suppose at some point in time that a fluctuation leads to a larger than average abundance of trees in a given diameter class. These trees will then give birth to a higher than expected number of new seedlings in the next generation, which sets up a correlation between the fluctuations in these two size classes. Because this correlation arises from the birth process, the second term in Eq. **12** is naturally proportional to the birth rate, b. Both fluctuations tend to persist as individuals from each of the size classes continue to grow, and so we end up with positive correlations between fluctuations in all size classes.

Species Abundance Distribution. Our next main conclusion is that the species abundance distribution is completely insensitive to size structure, which we now demonstrate. First, we note that we may obtain biologically relevant distributions from the formal definition of the partition function, Eq. **11**, by choosing a specific form for $H(m)$. For example, choosing $H(m)$ to be a constant, h, Eq. **11** reduces to the following function:

$$\begin{aligned} z_{sad}(h) &= \left\langle e^{h \int_{m_0}^{\infty} dm\, n(m)} \right\rangle \\ &= \langle e^{hN} \rangle. \end{aligned} \qquad [13]$$

where N is now the total abundance, integrated across all sizes. This is precisely the definition of the generating function of the SAD.

Setting $H(m) = h$ in our actual solution, Eq. **8**, the result is that for a size-structured community the moments of the SAD are generated by the following function:

$$\log z_{sad}(h) = -\frac{\nu}{b} \log \left[1 - \frac{b}{\nu} \langle N \rangle (e^h - 1) \right]$$

where $\langle N \rangle$ is the expectation value of total abundance per species, integrated over all size classes. The significance of this result is much more obvious when we consider the SAD itself, rather than its generating function (42). In the limit of small $\frac{\nu}{b}$, which is appropriate for comparison with NBT, the SAD is

$$P(N) = \frac{\nu}{bN} \left(\frac{\langle N \rangle}{\langle N \rangle + \frac{\nu}{b}} \right)^N. \qquad [14]$$

This distribution is identical to the Neutral Theory SAD, Eq. **3**, and so whatever the structure of the size variation in our community, we find the same, log series SAD.

Species Biomass Distribution. We have phrased our model in terms of individual size, but we may equally well interpret m as the mass of an individual. Using this interpretation we can derive an expression for the species biomass distribution, $P(M)$, the probability that a species has a total biomass, M, and we find that this distribution is much more closely tied to the size structure of a community than the SAD.

To derive the SBD, we use a method similar to the SAD case. Setting $H(m)$ equal to hm, for constant h, Eq. **11** reduces to the following function:

$$\begin{aligned} z_{sbd}(h) &= \left\langle e^{h \int_{m_0}^{\infty} dm\, m\, n(m)} \right\rangle \\ &= \langle e^{hM} \rangle. \end{aligned} \qquad [15]$$

where M is the total biomass, integrated over all individual masses, rather than total abundance. The SBD corresponding to our solution is generated by:

$$\log z_{sbd}(h) = -\frac{\nu}{b} \log \left[1 - \frac{b}{\nu} \int_{m_0}^{\infty} dm\, f(m) (e^{hm} - 1) \right]. \qquad [16]$$

The form of Eq. **16** means that this model makes no universal prediction for the species biomass distribution. In contrast to the log series result for the SAD, the form of $P(M)$ depends crucially on the details of $g(m)$, $d(m)$, b, and ν.

A Concrete Example

The results of the preceding section are very general, but their implications are not necessarily intuitive. For example, although the generating function given by Eq. **16** determines the moments of the SBD for a community with any given size structure, it is also illuminating to consider a specific case in more detail. In this section we compute and compare the SAD and the SBD for a specific mass-structured community that we term "completely" neutral. This is a very simple case, for which growth and mortality rates are independent of an individual's mass:

$$\begin{aligned} g(m) &= g \\ d(m) &= d. \end{aligned}$$

We also take for simplicity that individuals are born at a very small size, so that $m_0 \approx 0$. We call this community completely neutral because all individuals really are the same, both in terms of their birth and death dynamics, and in terms of their ontogenetic growth. Of course, we may consider more general and more biologically realistic cases by using our framework, but this simple case provides an important insight into the more general differences between the SAD and the SBD.

Completely Neutral Size Spectrum. The size spectrum is given by the solution to Eq. **9** (a time-independent version of the Von Foerster equation, Eq. **1**), with boundary condition (Eq. **10**):

$$\begin{aligned} g\frac{df(m)}{dm} + d \times f(m) &= 0 \\ gf(m_0) &= \nu + b \int_{m_0}^{\infty} dm\, f(m), \end{aligned}$$

which has solution

$$\langle n(m) \rangle = f(m) = \frac{\nu}{g\left(1 - \frac{b}{d}\right)} e^{-\frac{d}{g}m}. \qquad [17]$$

This means that the mean size spectrum, averaged across all species, drops off exponentially with increasing individual size: there are fewer larger individuals simply because more of their cohort are already deceased.

Completely Neutral SAD. We have already concluded that the SAD is always a log series, given by Eq. **14**. To make the comparison with NBT even clearer, this SAD may be further rewritten in terms of the familiar NBT diversity parameter, θ, and the total number of individuals in the metacommunity, J, and for small $\frac{\nu}{b}$:

$$\begin{aligned} \theta &= \frac{S\nu}{b} \\ J &= S\langle N \rangle \end{aligned} \qquad [18]$$

where S is the number of extant species in the metacommunity, and our conventions follow ref. 13. Reexpressing Eq. **14** in terms of θ and J, the absolute species abundance distribution for $N > 0$ is

$$S(N) = SP(N) = \frac{\theta}{N} \left(\frac{J}{J + \theta} \right)^N. \qquad [19]$$

Completely Neutral SBD. We can compute $P(M)$ for this neutral community by taking the inverse Laplace transform of Eq. **16**, with the details presented in *SI Appendix*. The result is:

$$P(M) = \frac{\nu}{bM} \left(e^{-\left(\frac{d-b}{g}\right)M} - e^{-\frac{d}{g}M} \right). \qquad [20]$$

This is a unique analytical prediction of the neutral species biomass distribution, and we note that although it bears some similarity to the species abundance distribution, there are some significant differences. To see the impact of these differences most clearly,

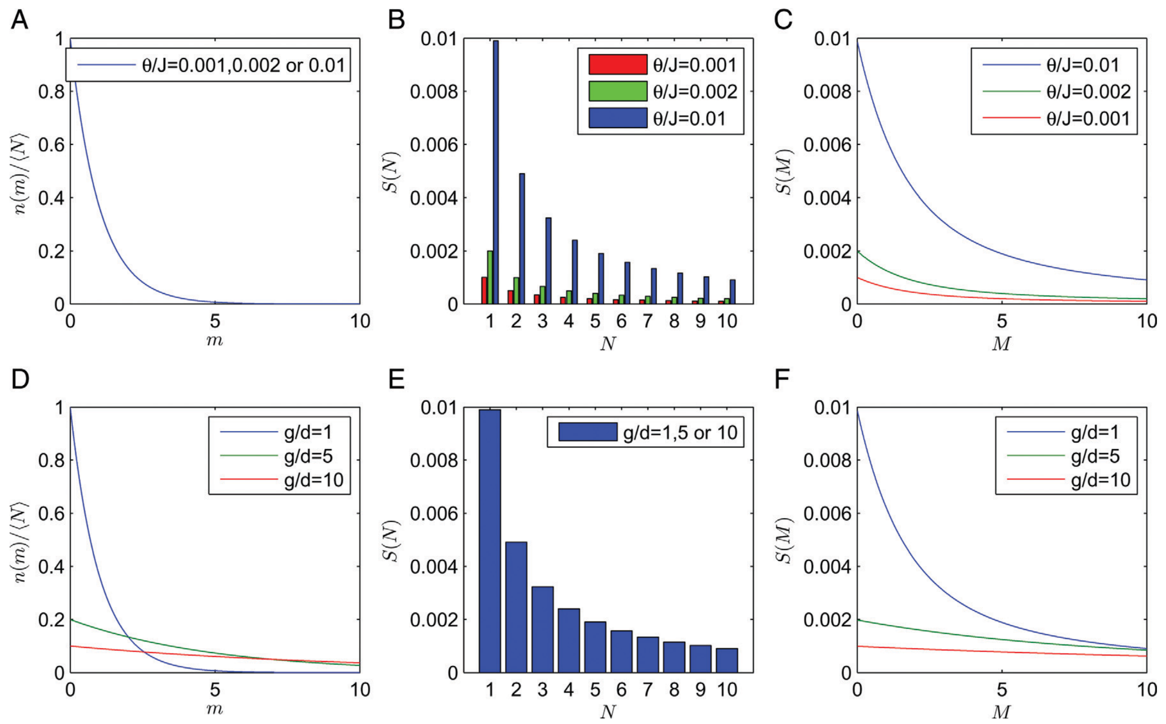

Fig. 2. The species biomass distribution is sensitive both to species richness and to community size-structure. For the completely neutral community defined in the text, *A*, *B*, and *C* show the respective variations of the size spectrum, the SAD, and the SBD with varying species richness. Richness is parametrized by θ, which is the NBT biodiversity parameter (1, 13). We find that the shape of the size spectrum is insensitive to overall species richness, while both the SAD and SBD show increasing numbers of rare or low-biomass species with increasing θ. In constrast, *D*, *E*, and *F* show the variation of these three distributions with varying size structure, parametrized by the ratio of growth rate to mortality rate, $\frac{g}{d}$. With increasing $\frac{g}{d}$, the size spectrum shows increasing numbers of large individuals relative to small individuals, reflecting that more individuals have the chance to grow large before they die. The SBD shows decreasing numbers of low-biomass species relative to abundant species with increasing $\frac{g}{d}$, but the SAD is entirely insensitive to changes in size structure.

we again rewrite this distribution in terms of the NBT diversity parameter and total metacommunity abundance, so that:

$$S(M) = \frac{\theta}{M}\left(e^{-\left(\frac{\theta}{\theta+J}\frac{d}{g}\right)M} - e^{-\frac{d}{g}M}\right). \qquad [21]$$

Like the SAD, the biomass distribution depends on the species richness, through θ, but it also depends on the underlying structure of this community, through the mass-scale $\frac{g}{d}$. We illustrate this central message of our concrete example in Fig. 2: the SBD is sensitive to the size or mass structure of our community, whereas the SAD is not.

Discussion

We have presented an integrated mathematical framework to describe communities structured by size, and subject to demographic stochasticity. This framework provides a language with which we can quantify both the impact of demographic stochasticity and the impact of individual variation on community structure, and by focusing on the importance of size variation (8), we address the call by Hubbell (1) to add biological realism to Neutral Biodiversity Theory (NBT).

Key Results. Using our framework, we have derived a number of predictions for stochastic, size-structured communities. We predict the mean size spectrum, which we find must satisfy the Von Foerster equation, and we make predictions for the size and correlation of fluctuations around this mean spectrum.

We find that the species abundance distribution (SAD) takes the form of a log series, and is completely independent of the underlying variation in individual sizes and demographic rates. The irrelevance of size structure in determining the metacommunity SAD provides a theoretical underpinning for the success of NBT in predicting dispersal-limited, local SADs in tropical forests, where variation in demographic rates is strongly correlated with body size (16).

While the robustness of the neutral SAD provides some explanation for the validity of NBT predictions, this result also underlines the inadequacy of the SAD to distinguish between different underlying processes. Our species biomass distribution (SBD) is a unique kind of prediction, and in contrast to the SAD we find that it depends crucially on the details of the variation in individual sizes. For given scalings of demographic rates, we derive Eq. **16**, the generating function of the SBD, and demonstrate that the SBD itself may be computed via inverse Laplace transform. This derivation provides evidence that characterizing biodiversity in terms of appropriate alternative currencies will allow ecologists to probe community structure more deeply than with the SAD (26).

Applications. Any empirical evaluation of our model based on the SAD will produce the same results, negative or positive, as a test of neutral theory. However, with the addition of size structure, we may generate a much broader suite of predictions, as advocated by recent work on "alternative currency" distributions (26). We have explored what we regard as the three most important and accessible of those predicted distributions: the size spectrum, the SAD, and the SBD.

One prescription for testing the model would be to measure the size spectrum empirically, and use this to compute the SBD by taking the inverse Laplace transform of Eq. **16**. This should then be compared with the empirical SBD, which in practice must be computed by binning individuals into discrete size classes. Whether our framework can reproduce multiple, interrelated distributions in this way constitutes a far more stringent empirical evaluation than simply testing the SAD.

Generalizations. There are three important generalizations of our specific model. First, whereas we have allowed for size-dependent growth and mortality rates, we solved for the case of a size-independent birth rate, b. This approximation is not necessarily realistic, and so an important generalization of Eq. **7** is size-dependent fecundity, $b(m)$:

$$\begin{aligned}\frac{\partial Z}{\partial t} &= \int_{m_0}^{\infty} d(m)\frac{\delta Z}{\delta H(m)}(e^{-H(m)}-1)\\ &+ \int_{m_0}^{\infty} g(m)\frac{\delta Z}{\delta H(m)}\frac{dH}{dm}\\ &+ \left[\int_{m_0}^{\infty} b(m)\frac{\delta Z}{\delta H(m)} + \nu Z\right](e^{H(m_0)}-1). \qquad [22]\end{aligned}$$

Although it is straightforward to generalize the equation itself, finding the equilibrium solution of $Z[H]$ in this case represents important future research. It remains to be seen how sensitive the SAD and SBD are to variation in fecundity.

A second generalization is the introduction of explicit interspecific variation, either through variation in birth size, m_0, or through interspecific variation in the scaling of demographic rates with size (43). This will allow us to make a closer connection with complementary work integrating neutral processes with interspecific size variation (44–46), and to quantify the expected departures from the predictions of the current model. All species in our model share the same underlying dependencies of demographic rates on the size or mass of individuals, and the robustness of the log series SAD may rest precisely on this residual symmetry. The relative size of inter- vs. intraspecific variation seems likely to be a crucial determinant of the validity of our predictions.

Finally, our model may be adapted to integrate variation in heritable traits with demographic stochasticity. The basic framework would follow Eq. **22**, but with variation in trait values from generation to generation alongside ontogenetic variation in size. In all three of these generalizations, the central mathematical tool is the partition function, and to tackle any one of them we must be prepared to cross disciplines, and borrow relevant tools from evolutionary biology or statistical physics (39).

Conclusions. Our model has allowed us to quantify the combined effects of size structure and demographic stochasticity, showing that the SAD is insensitive to size structure, and that the SBD is a more informative characterization of community properties. The model also constitutes progress toward a very general framework for community assembly. Including the variation of heritable traits, in addition to ontogenetic growth, and interspecific variation, in addition to intraspecific variation, will provide us with a powerful framework to unify stochastic and trait-based approaches to community assembly, with the promise of a broad range of stringent empirical tests.

ACKNOWLEDGMENTS. We thank S. P. Hubbell, J. R. Banavar, I. Volkov, and R. S. Etienne for fruitful conversations, M. Westoby, I. J. Wright, B. McGill, and A. Allen for useful comments about trait variation and trait-based community assembly, and S. M. Roper for helpful discussions about inverse Laplace transforms and the SBD. This work was supported by the Gordon and Betty Moore Foundation, the Santa Fe Institute, the National Science Foundation Grant DEB-0739429, and the Center for Ecology and Evolutionary Biology at the University of Oregon. The project was initiated as part of the Unifying Theories of Ecology Working Group, supported by the National Center for Ecological Analysis and Synthesis, a Center funded by National Science Foundation Grant DEB-0553768, the University of California, Santa Barbara, and the State of California.

1. Hubbell SP (2001) *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ Press, Princeton, NJ).
2. MacArthur RH (1957) On the relative abundance of bird species. *Proc Natl Acad Sci USA* 43:293–295.
3. Hutchinson GE (1957) Conculding remarks. *Cold Spring Harbor Symp Quant Biol* 22:415–427.
4. Tilman, D (1982) *Resource Competition and Community Structure*, Monographs in Population Biology (Princeton Univ Press, Princeton, NJ).
5. Weiher E, Keddy PA (1995) Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos* 74:159–164.
6. Westoby M, Wright IJ (2006) Land-plant ecology on the basis of functional traits. *Trends Ecol Evol* 21:261–268.
7. Chase JM, Leibold MA (2003) *Ecological Niches: Linking Classical and Contemporary Approaches* (University of Chicago Press, Chicago).
8. Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a metabolic theory of ecology. *Ecology* 85:1771–1789.
9. Enquist BJ, Niklas KJ (2001) Invariant scaling relations across tree-dominated communities. *Nature* 410:655–660.
10. Chave J, Muller-Landau HC, Levin SP (2002) Comparing classical community models: theoretical consequences for patterns of diversity. *Am Nat* 159:1–23.
11. Purves DW, Pacala SW (2005) Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. *Biotic Interactions in the Tropics*, eds Burslem DFRP, Pinard MA, Hartley SE (Cambridge Univ Press, New York), pp 107–139.
12. Adler PB, Hillerislambers J, Levine JM (2007) A niche for neutrality. *Ecol Lett* 10:95–104.
13. Volkov I, Banavar JR, Hubbell SP, Maritan A (2003) Neutral theory and relative species abundance in ecology. *Nature* 424:1035–1037.
14. Condit R, et al. (2002) Beta-diversity in tropical forest trees. *Science* 295:666–669.
15. Condit R, et al. (2006) The importance of demographic niches to tree diversity. *Science* 313:98–101.
16. Muller-Landau HC, et al. (2006) Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecol Lett* 9:575–588.
17. Hubbell SP (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Funct Ecol* 19:166–172.
18. Alonso D, Etienne RS, McKane AJ (2006) The merits of neutral theory. *Trends Ecol Evol* 21:451–457.
19. Chase JM (2007) Drought mediates the importance of stochastic community assembly. *Proc Natl Acad Sci USA* 104:17430–17434.
20. Chave J (2004) Neutral theory and community ecology. *Ecol Lett* 7:241–253.
21. Werner EE, GIlliam JF (1984) The ontogenetic niche and species interactions in size-structured populations. *Annu Rev Ecol Syst* 15:393–425.
22. West GB, Brown JH, Enquist BJ (1997) A general model for the origin of allometric scaling laws in biology. *Science* 276:122–126.
23. Enquist BJ, Brown JH, West GB (1998) Allometric scaling of plant energetics and population density. *Nature* 395:163–166.
24. West GB, Brown JH, Enquist BJ (2001) A general allometric model of ontogenetic growth. *Nature* 413:628–631.
25. Tilman D (2004) Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proc Natl Acad Sci* 101:10854–10861.
26. McGill BJ, et al. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol Lett 10:* 995:1015.
27. McKendrick AG (1926) Application of mathematics to medical problems. *Proc Edinburgh Math Soc* 44:98–130.
28. Von Foerster H (1959) Some remarks on changing populations. *Kinetics of Cellular Proliferation* (Grune and Stratton, New York), pp 382–399.
29. Muller-Landau HC, et al. (2007) Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecol Lett* 9:589–602.
30. Chesson P (2000) Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 31:343–366.
31. White EP Ernest SKM, Kerkhoff AJ, Enquist BJ (2007) Relationships between body size and abundance in ecology. *Trends Ecol Evol* 22:323–330.
32. Etienne RS, Alonso D, McKane AJ (2007) The zero-sum assumption in neutral biodiversity theory. *J Theor Biol* 248:522–536.
33. Kendall DG (1949) Stochastic processes and population growth. *J Roy Stat Soc B* 230–264.
34. Volkov I, Banavar JR, Hubbell SP, Maritan A (2007) Patterns of relative species abundance in rainforests and coral reefs. *Nature* 450:45–49.
35. Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J Anim Ecol* 12:42–58.
36. Leslie PH (1945) On the use of matrices in certain population mathematics. *Biometrika* 33:183–212.
37. Pollard JH (1966) On the use of the direct matrix product in analysing certain stochastic population models. *Biometrika* 53:397–415.
38. Cohen JE (1979) Comparative statics and stochastic dynamics of age-structured populations. *Theor Popul Biol* 16:159–171.
39. Zinn-Justin J (2002) *Quantum Field Theory and Critical Phenomena* (Oxford Univ Press, New York).
40. Bolker B, Pacala SW (1997) Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor Popul Biol* 52:179–197.
41. Bolker B, Pacala SW (1999) Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *Am Nat* 153:575–602.
42. Spiegel MR (1992) *Theory and Problems of Probability and Statistics* (McGraw–Hill, New York), p 118.
43. Enquist BJ, et al. (2007) A general integrative model for scaling plant growth, carbon flux, and functional trait spectra. *Nature* 449:218–222.
44. Etienne RS, Olff H (2004) How dispersal limitation shapes species-body size distributions in local communities. *Am Nat* 163:69–83.
45. Haegeman B, Etienne RS (2008) Relaxing the zero-sum assumption in neutral biodiversity theory. *J Theor Biol* 252:288–294.
46. Volkov I, Banavar JR, Maritan A (2004) Organization of ecosystems in the vicinity of a novel phase transition. *Phys Rev Lett* 92:2187031–2187034.

Stochastic, Size-Structured Community Assembly: Supporting Information

1 Full Derivation of the Equilibrium Partition Function

In the main body of the text, we provided an equilibrium solution of our functional Master Equation. First, we set $\frac{\partial Z}{\partial t} = 0$, so that we are looking for an equilibrium solution. Then, rewriting the Eq.[7] in terms of $\log Z$, rather than Z, we need to solve

$$\begin{aligned} 0 = & \int_{m_0}^{\infty} dm\ d(m) \frac{\delta \log Z}{\delta H(m)} \left(e^{-H(m)} - 1 \right) \\ & + \int_{m_0}^{\infty} dm\ g(m) \frac{\delta \log Z}{\delta H(m)} \frac{dH}{dm} \\ & + \left[b \int_{m_0}^{\infty} dm \frac{\delta \log Z}{\delta H(m)} + \nu \right] \left(e^{H(m_0)} - 1 \right). \end{aligned} \tag{1}$$

In the main text, we stated that this equilibrium solution took the form:

$$\log Z = -\frac{\nu}{b} \log \left[1 - \frac{b}{\nu} \int_{m_0}^{\infty} dm f(m) \left(e^{H(m)} - 1 \right) \right]. \tag{2}$$

We may now confirm that (2) solves (1) by direct substitution. To do this, we must make use of the following simple rule for taking functional derivatives:

$$\frac{\delta H(m')}{\delta H(m)} = \delta(m' - m), \tag{3}$$

where $\delta(m' - m)$ is the Dirac delta function. (Informally-speaking, the delta function represents an infinitely sharp peak at $m = m'$, bounding unit area, so that the functional derivative vanishes unless m' is equal to m.)

Applying this rule in conjunction with the chain rule, we may compute the functional derivative of Eq.(2) to obtain

$$\begin{aligned} \frac{\delta \log Z}{\delta H(m)} & = \frac{\frac{\nu}{b} \frac{b}{\nu} \int_{m_0}^{\infty} d\tilde{m} f(\tilde{m}) \frac{\delta H(\tilde{m})}{\delta H(m)} e^{H(\tilde{m})}}{1 - \frac{b}{\nu} \int_{m_0}^{\infty} d\tilde{m} f(\tilde{m}) \left(e^{H(\tilde{m})} - 1 \right)} \\ & = \frac{f(m) e^{H(m)}}{1 - \frac{b}{\nu} \int_{m_0}^{\infty} d\tilde{m} f(\tilde{m}) \left(e^{H(\tilde{m})} - 1 \right)}. \end{aligned} \tag{4}$$

We may now check that Eq.(2) solves Eq.(1) by plugging this result into Eq.(1), and checking that we have a time-independent solution, as claimed. To see this, we show that the right-hand-side of Eq.(1) sums to zero:

$$\begin{aligned}
&\int_{m_0}^{\infty} dm \left[d(m)\frac{\delta \log Z}{\delta H(m)}\left(e^{-H(m)}-1\right)+g(m)\frac{\delta \log Z}{\delta H(m)}\frac{dH}{dm}\right]+\left[\nu+b\int_{m_0}^{\infty} dm\frac{\delta \log Z}{\delta H(m)}\right]\left(e^{H(m_0)}-1\right)\\
&=\frac{1}{1-\frac{b}{\nu}\int_{m_0}^{\infty} d\tilde{m} f(\tilde{m})\left(e^{H(\tilde{m})}-1\right)}\int_{m_0}^{\infty}\left[-d(m)f(m)\left(e^{H(m)}-1\right)+g(m)f(m)\frac{d}{dm}\left(e^{H(m)}-1\right)\right]\\
&+\left[\nu+\frac{b}{1-\frac{b}{\nu}\int_{m_0}^{\infty} d\tilde{m} f(\tilde{m})\left(e^{H(\tilde{m})}-1\right)}\int_{m_0}^{\infty} dm f(m)e^{H(m)}\right]\left(e^{H(m_0)}-1\right)\\
&=\frac{e^{H(m_0)}-1}{1-\frac{b}{\nu}\int_{m_0}^{\infty} d\tilde{m} f(\tilde{m})\left(e^{H(\tilde{m})}-1\right)}\left[-g(m_0)f(m_0)+\nu-\nu\frac{b}{\nu}\int_{m_0}^{\infty} d\tilde{m} f(\tilde{m})\left(e^{H(\tilde{m})}-1\right)\right.\\
&\left.+b\int_{m_0}^{\infty} dm f(m)e^{H(m)}\right]\\
&=\frac{e^{H(m_0)}-1}{1-\frac{b}{\nu}\int_{m_0}^{\infty} d\tilde{m} f(\tilde{m})\left(e^{H(\tilde{m})}-1\right)}\left[-b\int_{m_0}^{\infty} d\tilde{m} f(\tilde{m})e^{H(\tilde{m})}+b\int_{m_0}^{\infty} dm f(m)e^{H(m)}\right]\\
&=0,
\end{aligned}\tag{5}$$

as required. We used the boundary condition for $f(m)$,

$$f(m_0)g(m_0)=\nu+b\int_{m_0}^{\infty} dm\ f(m),\tag{6}$$

to obtain the third equality.

The central message of this derivation is that Eq.(2) does indeed solve Eq.(1), our master equation for the partition function $Z[H]$.

2 Obtaining Explicit Expressions for the Species Biomass Distribution

In the main body of the text we derived the following expression for the generating function for the species biomass distribution, $P(M)$:

$$\log z_{sbd}(h)=-\frac{\nu}{b}\log\left[1-\frac{b}{\nu}\int_{m_0}^{\infty} dm f(m)\left(e^{hm}-1\right)\right].\tag{7}$$

The distribution $P(M)$ is defined via the following transform:

$$z_{sbd}(h)=\int dM\ P(M)e^{hM},\tag{8}$$

so that moments of $P(M)$ are equal to appropriate derivatives of $z_{sbd}(h)$ with respect to h, as required. Given this definition, we may compute $P(M)$ explicitly by taking the inverse Laplace transform of $z_{sbd}(-h)$:

$$P(M) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} dh \; z_{sbd}(-h) e^{hM}. \tag{9}$$

The contour is defined so that the real number γ is greater than than the real part of all singularities of $z_{sbd}(-h)$. I.e. the contour is a vertical line in the complex plane, to the right of those singularities.

2.1 The Totally Neutral Community

We gave the example in the main text of a totally neutral community, that is where all individuals have the same birth, death and growth rates, irrespective of species identity *or* size:

$$\begin{aligned} g(m) &= g \\ d(m) &= d. \end{aligned}$$

For this choice of size-structure, we have that

$$f(m) = \frac{\nu}{g\left(1 - \frac{b}{d}\right)} e^{-\frac{d}{g} m} \tag{10}$$

and the problem of finding $P(M)$ reduces to the following contour integral,

$$\begin{aligned} I &= \int_C dh \left[1 - \frac{b}{g\left(1 - \frac{b}{d}\right)} \frac{1}{h + \frac{d}{g}} + \frac{b}{d - b} \right]^{-\frac{\nu}{b}} e^{hM} \\ &= \left(1 + \frac{b}{d - b}\right)^{-\frac{\nu}{b}} \int_C dh \; \frac{\left(h + \frac{d}{g}\right)^{\frac{\nu}{b}}}{\left(h + \frac{d-b}{g}\right)^{\frac{\nu}{b}}} \; e^{hM}, \end{aligned} \tag{11}$$

where the contour C is defined as above. Changing variables to

$$\omega = \frac{gh}{b} + \frac{d}{b}, \tag{12}$$

we have

$$\begin{aligned} I &= \frac{b}{g} e^{-\frac{dM}{g}} \left(1 + \frac{b}{d - b}\right)^{-\frac{\nu}{b}} \int_C dh \; \frac{\left(\frac{b\omega}{g}\right)^{\frac{\nu}{b}}}{\left(\frac{b\omega}{g} - \frac{b}{g}\right)^{\frac{\nu}{b}}} \; e^{\frac{b\omega M}{g}} \\ &= \frac{b}{g} e^{-\frac{dM}{g}} \left(1 + \frac{b}{d - b}\right)^{-\frac{\nu}{b}} \int_C dh \; \frac{(\omega)^{\frac{\nu}{b}}}{(\omega - 1)^{\frac{\nu}{b}}} \; e^{\frac{b\omega M}{g}}. \end{aligned} \tag{13}$$

The function to be integrated has a branch cut between $\omega = 0$ and $\omega = 1$, and so we must take the contour C to be such that $\gamma > 1$.

This contour may be closed with a large semicircle to the left of C in the complex plane, and for $M > 0$ this semicircle gives a vanishing contribution. From Cauchy's integral theorem, this integral may be shrunk to an integral anticlockwise around the branch cut. This contour can be shrunk to be infinitessimally thin, so that all is left is an integral along the top of the branch cut from right to left, $\omega = 1$ to 0, and an integral along the bottom of the branch cut from left to right, $\omega = 0$ to 1. These two integrals are:

$$I_{top} = \frac{b}{g} e^{-\frac{dM}{g}} \left(1 + \frac{b}{d-b}\right)^{-\frac{\nu}{b}} \int_1^0 d\omega \; \frac{(\omega)^{\frac{\nu}{b}}}{(1-\omega)^{\frac{\nu}{b}}} \; e^{\frac{b\omega M}{g}} e^{-\frac{i\pi\nu}{b}} \tag{14}$$

and

$$I_{bottom} = \frac{b}{g} e^{-\frac{dM}{g}} \left(1 + \frac{b}{d-b}\right)^{-\frac{\nu}{b}} \int_0^1 d\omega \; \frac{(\omega)^{\frac{\nu}{b}}}{(1-\omega)^{\frac{\nu}{b}}} \; e^{\frac{b\omega M}{g}} e^{\frac{i\pi\nu}{b}}. \tag{15}$$

Adding both together we obtain

$$I = I_{top} + I_{bottom} = \frac{b}{g} e^{-\frac{dM}{g}} \left(1 + \frac{b}{d-b}\right)^{-\frac{\nu}{b}} 2i \sin\left(\pi\nu/b\right) \int_0^1 d\omega \; \frac{(\omega)^{\frac{\nu}{b}}}{(1-\omega)^{\frac{\nu}{b}}} \; e^{\frac{b\omega M}{g}} \tag{16}$$

and so for $M > 0$ the distribution is

$$P(M) = \frac{b}{g} e^{-\frac{dM}{g}} \left(1 + \frac{b}{d-b}\right)^{-\frac{\nu}{b}} \frac{\sin\left(\pi\nu/b\right)}{\pi} \int_0^1 d\omega \; \frac{(\omega)^{\frac{\nu}{b}}}{(1-\omega)^{\frac{\nu}{b}}} \; e^{\frac{b\omega M}{g}}. \tag{17}$$

Taking the limit of small ν/b we obtain the result relevant for neutral theory:

$$\begin{aligned} P(M) &= \frac{b}{g} e^{-\frac{dM}{g}} \frac{\nu}{b} \int_0^1 d\omega \; e^{\frac{b\omega M}{g}} + O\left(\frac{\nu}{b}\right)^2 \\ &= \frac{\nu}{bM} \left(e^{-\frac{d-b}{g}M} - e^{-\frac{d}{g}M}\right) \end{aligned} \tag{18}$$