A data-intensive assessment of the species abundance distribution.

Abstract

The hollow curve species abundance distribution describes the pattern of large numbers of rare species and a small number of common species in a community. The species abundance distribution is one of the most ubiquitous patterns in nature and many models have been proposed to explain the mechanisms that generate this pattern. While there have been numerous comparisons of species abundance distribution models, most of these comparisons only use only a small subset of available models, focus on a single ecosystem or taxonomic group, and fail to use the most appropriate statistical methods. This makes it difficult to draw general conclusions about which, if any, models provide the best empirical fit to species abundance distributions. I compiled data from the literature to significantly expand the available data for underrepresented taxonomic groups, and combined this with other macroecological datasets to perform comprehensive model comparisons for the species abundance distribution. A multiple model comparison showed that most available models for the species abundance distribution fit the data equivalently well across a diverse array of ecosystems and taxonomic groups. In addition, a targeted comparison of the species abundance distribution predicted by a major ecological theory, the unified neutral theory of biodiversity (neutral theory), against a non-neutral model of species abundance, demonstrates that it is difficult to distinguish between these two classes of theory based on patterns in the species abundance distribution. In concert, these studies call into question the potential for using the species abundance distribution to infer the processes operating in ecological systems.

(XXX pages)

Public abstract

One of the most commonly observed patterns in ecology is that fact that at most locations there are a large number of relatively rare species, composed of only a few individuals per species, and a small number of relatively common species. This pattern of commonness and rarity is quantified by the species abundance distribution.

As one of the most commonly observed patterns in ecology, it has been studied intensively for over 100 years. A major emphasis of this research has been developing models to try to understand the forces that generate such a general pattern. As a result, there are now dozens of different models for how this pattern might be generated, and these models are based on at least five different major categories of forces. Since many of these models appear to match ecological observations reasonably well, it is difficult to tell which model or models are most likely to be correct.

In order to address this issue I compiled data on over 16,000 ecological systems. I then used the best available statistical methods to compare a number of different models to observed data. While there have been numerous comparisons of species abundance distribution models, most of these comparisons only use only a small subset of available models, focus on a single ecosystem or type of species, and fail to use the most appropriate statistical methods. My approach overcomes all of these challenges and thus provides the best opportunity for figuring out which models provide the best description of real data.

Both broad stroke and detailed comparisons of this pattern of commonness and rarity suggest that even when using large amounts of data from across the world and the diversity of life, and the most current and powerful statistical methods, that it is generally not possible to distinguish between many common models of the species abundance distribution. I could exclude some models as clearly poorer descriptions of the pattern than others, but several models provided equivalently good descriptions of the data. This calls into question the potential for using the species abundance distribution to understand what processes are driving ecological systems.

Introduction

Macroecology is a data-intensive approach to studying ecological patterns and processes. As the field has matured, increasingly greater amounts of data have become available to address these questions (Reichman et al. 2011). Although the macroecological research program has matured a great deal since its introduction (McGill and Nekola 2010, Keith et al. 2012), there are still a number of areas in which it can be improved (Beck et al. 2012). One of the major criticisms of macroecology is that it is biased towards analyzing data from terrestrial systems in North America (Beck et al. 2012), a legacy of the academic heritage of the scientists who developed macroecology, as well as the availability of large ecological datasets suitable for testing macroecological questions. Another major criticism has been that the field focuses too much on pattern description, and not enough on the identification of pattern generating mechanisms (Beck et al. 2012).

One of the most significant patterns in macroecology is the species abundance distribution, which describes the commonness and rarity of species in an ecological community. The form of this pattern is very general, with most communities being composed of a small number of common species and a large number of rare species. Interest in this pattern has generated dozens of models attempting to characterize the form of the pattern and the processes underlying it. However, most attempts to determine which of these models provide the best fit to empirical data, and are therefore most likely to describe the processes generating the pattern, have been limited either by a combination of poor statistical methodology and/or restricted scope.

A set of best practices for testing patterns and models in macroecology has been developed as the discipline has matured (Burnham and Anderson 2002, McGill 2003, McGill et al. 2006, White et al. 2008, 2012). Some of these best practices for testing macroecological theory include:
Testing the generality of patterns with multiple taxonomic groups/ecosystems, both terrestrial and aquatic/marine (White et al. 2012). * Simultaneous testing of multiple models and model predictions (McGill 2003, McGill et al. 2006) * Use of likelihood based methods for comparing distributions (Edwards et al. 2007, White et al. 2008)

Following these best practices allows for a more rigorous assessment of the generality of patterns across taxonomic groups/ecosystems, may provide better insight into identifying pattern generating mechanisms, and improve the ability to make ecological predictions. Here I present research following best practices for comparing species abundance distributions by compiling a unique dataset for addressing macroecological questions for more ecosystems and taxonomic groups, evaluate multiple competing models, and use the best statistical methods available.

First, to address some of the concerns about the lack of data for underrepresented taxonomic groups and ecosystems, I compiled a set of data from the literature. My primary focus for data collection was to collect data for those vertebrate taxa which are not already well represented by publicly available data. I selected these taxa because their taxonomy is fairly well resolved when compared to the majority of invertebrate groups. While my primary focus was on fish, reptiles, and amphibians, I also compiled data on spiders and beetles. In addition, I also collected a small amount of bird data, which could potentially be used in comparision with the large, publicly available bird datasets to determine if the dataset affects the results.

Second, we compete five models from each of four different model families: purely statistical, branching process, population dynamics, and niche partitioning (McGill et al. 2007) with community abundance data for reptiles, amphibians, bony fish, beetles, spiders, birds, trees, mammals, and butterflies to perform the largest test of species abundance distributions to date. We follow the current best practice recommendations for testing species abundance distribution models to determine which models provide the best fit to empirical data. Identifying which, if any, models best describe the data can help determine what pattern generating mechanisms are more likely to have direct influences on the shape of the species abundance distribution.

Third, we expand our exploration of the hollow curve species abundance distribution by performing a detailed analysis evaluating previous work on marine systems by Connolly et al. 2014 to determine if species abundance data is sufficient for identifying two general classes of ecological process, neutral vs. non-neutral processes (Connolly et al. 2014). Connolly et al. found that the majority

of sites were best described by a non-neutral model of species abundance in marine ecosystems; however, this has not yet been tested for terrestrial systems. This work seeks to determine the generality of the non-neutral species abundance distribution method in terrestrial systems, and identify whether there are differences between terrestrial and marine systems relative to species abundance distributions.

Beck, J., L. Ballesteros-Mejia, C. M. Buchmann, J. Dengler, S. A. Fritz, B. Gruber, C. Hof, F. Jansen, S. Knapp, H. Kreft, and others. 2012. What's on the horizon for macroecology? Ecography 35:673–683.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer.

Connolly, S. R., M. A. MacNeil, M. J. Caley, N. Knowlton, E. Cripps, M. Hisano, L. M. Thibaut, B. D. Bhattacharya, L. Benedetti-Cecchi, R. E. Brainard, and others. 2014. Commonness and rarity in the marine biosphere. Proceedings of the National Academy of Sciences:8524–8529.

Edwards, A. M., R. A. Phillips, N. W. Watkins, M. P. Freeman, E. J. Murphy, V. Afanasyev, S. V. Buldyrev, M. G. da Luz, E. P. Raposo, H. E. Stanley, and others. 2007. Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer. Nature 449:1044–1048.

Keith, S. A., T. J. Webb, K. Böhning-Gaese, S. R. Connolly, N. K. Dulvy, F. Eigenbrod, K. E. Jones, T. Price, D. W. Redding, I. P. Owens, and others. 2012. What is macroecology? Biology letters 8:904–906.

McGill, B. J. 2003. A test of the unified neutral theory of biodiversity. Nature 422:881–885.

McGill, B. J., and J. C. Nekola. 2010. Mechanisms in macroecology: aWOL or purloined letter? towards a pragmatic view of mechanism. Oikos 119:591–603.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, and others. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology letters

10:995–1015.

McGill, B. J., B. A. Maurer, and M. D. Weiser. 2006. Empirical evaluation of neutral theory. Ecology 87:1411–1423.

Reichman, O., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. Science 331.

White, E. P., B. J. Enquist, and J. L. Green. 2008. On estimating the exponent of power-law frequency distributions. Ecology 89:905–912.

White, E. P., K. M. Thibault, and X. Xiao. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology 93:1772–1778.