INVESTIGATING NEUTRAL THEORY IN A BACTERIAL COMMUNITY

by

Sarah M. Clement

June 2012

ABSTRACT

The purpose of this study was to investigate whether a neutral model or non-neutral model dominate the formation of a microbial community. The microbial community that was considered was borrowed from the Werner et al. (13) study in 2011. The data included 16S rRNA sequences for nine anaerobic bioreactors treating brewery waste water. A lognormal distribution was used as the non-neutral model for this study. Using QIIME v1.5 (14) and Microsoft Excel, the best fit lognormal distribution was determined for each of the nine facilities. The R squared values were then calculated. For the nine facilities the R squared values ranged between 0.961 and 0.998. Therefore, the microbial communities in the nine facilities fit the non-neutral model, the lognormal distribution, well.

ACKNOWLEDGMENTS

I would like to thank my advisor Lars Angenent for his guidance and support. In addition, I would also like to thank the other members of the Angenent lab for their support.

TABLE OF CONTENTS

Contents

I. Introduction

Microbial communities are complex systems that have been studied for years. They are an important and necessary component in many fields including medicine, engineering, and agriculture (10). The factors that influence the structure of these microbial communities and other biological communities have been a highly debated and researched topic. Theories such as the trade-off niche and neutral theory have been developed and investigated in an attempt to better understand the rules that govern the formation and development of ecological communities.

The notion of the niche theory was first discussed by Joseph Grinnell (1) and Charles Elton (2) in the 1920s. They were the first to introduce the idea of the ecological niche. However, it wasn't until the 1930s when Gause introduced the basic principle of the niche theory. Now commonly referred to as *Gause's axiom* or *Gause's theorem*, it states that "no two species can occupy the same ecological niche" (3). When applied to a biological community, niche theory argues that it is the deterministic factors, such as competition and niche differentiation that shape the community. The classic niche model is commonly referred to as the trade-off niche model (4). However, the arguments against the trade-off niche theory emphasize the fact that niche theory fails to explain highly diverse environments in which many rare taxa can coexist. In response to this ecologists have recently made attempts to expand niche models to include stochastic processes (4).

The arguments against the niche theory have also led to the development of an alternative theory, the neutral theory. The neutral theory was first introduced by Hubbell (5, 6, 7). Hubbell argued that all species have an equal probability of entering a community or are competitively identical. In his book, The Unified Neutral Theory of Biodiversity and Biogeography, Hubbell states that under the neutral theory "all species are equal in their probabilities of immigrating onto the island, or of going extinct once there" (7). The neutral theory predicts that the species abundance distribution (SAD) will fit the zero-sum multimodal (ZSM) distribution (8). Figure 1 shows a rank abundance plot with the estimated zero-sum multimodal distribution.

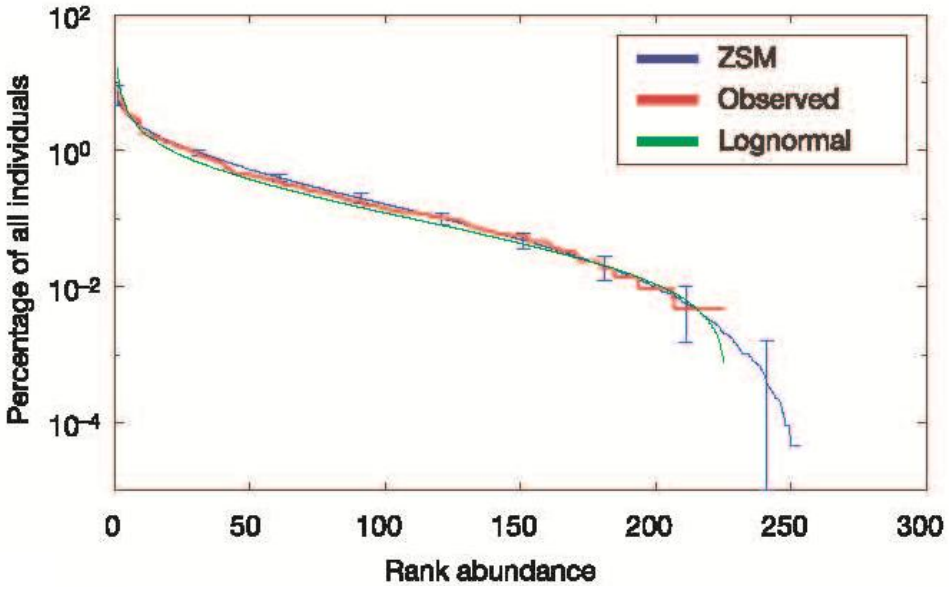

Figure 1. Standard rank abundance lot with the estimated ZSM distribution and the lognormal distribution (9)

Since the introduction of Hubbell's neutral theory, many studies have been devoted to testing the neutral theory. The neutral theory has been applied to a number of different types of biological communities, including tree (9), microbial (10), and fish (11) communities. In a study by McGill (9), the zero-sum multimodal distribution and the lognormal, the non-neutral model, are to fit empirical data from a community of trees. The study concludes that the lognormal distribution provides a better fit for the empirical data and that the zero-sum multimodal distribution fails to fit the empirical data a majority of the time. Another study by Ofiteru et al. (10) investigates neutral theory in a microbial community from a wastewater treatment plant. Ofiteru et al. used Hubbell's neutral theory and extended the model into a continuous format that allows the model to include environmental effects. The study demonstrated that it is possible to calibrate a purely neutral model. Ofiteru et al. also provide an economical method for incorporating environmental influence on individual taxa. Finally, a study by Magurran and Henderson (11) focused on an estuarine fish community. The results of the study show that the species abundance distribution for the core species follows a non-neutral model, the lognormal distribution, where as the species in low abundance follow a log series distribution.

II. Objective

This project investigated the microbial community of anaerobic bioreactors that treat waste water from a brewery. The purpose of the study was to determine if a neutral or a non-neutral model fit the species abundance distribution for the microbial community. In order to determine if the microbial communities at each of the facilities fit the neutral or the non-neutral model, the species abundance distributions were fit to a lognormal distribution. The lognormal distribution has long been known as the non-neutral model, as stated in the publication by Etienne and Olff (12).

III. Data and Methodology

A. Data

The data that was used for this project was borrowed from the study by Werner et al. (13). Biomass samples were taken from nine different full-scale methanogenic granular-sludge bioreactor facilities that treated brewery waste water. The biomass samples were taken monthly over a one year period. There were three types of anaerobic digesters used in the study. The study sampled four upflow anaerobic sludge blanket (UASB) facilities, four expanded granule sludge bed (EGSB) facilities, and one internal circulation (IC) facility. The facilities were labeled using abbreviations. The UASB facilities are labeled U1, U2, U3, and U4. The EGSB facilities are labeled E1, E2, E3, and E4. Finally, the IC facility is labeled I1. In addition, the study measured performance and operating conditions of the nine facilities on a daily basis. In the study by Werner et al. it was determined that the taxonomic divisions were dominated by Proteobacteria (mostly Syntrophobacterales and Desulfuromonales), Bacteroidetes, Spirochaetes, Clostridia, Chloroflexi, and Synergistia.

The data that was provided from the study included an operational taxonomic unit (OTU) table and a mapping file. The mapping file included data on environmental conditions as well as digester performance. Using the Quantitative Insights Into Microbial Ecology (QIIME v1.1) the sequences generated from pyrosequencing of bacterial 16S rRNA gene amplicons were processed. The initial processing of the 16S rRNA gene sequences had already been completed in the Werner et al. study. The barcodes had been trimmed, denoised, and then clustered into OTUs at 97% pairwise identity. Next, representative sequences were selected from each OTU and using PyNAST, were aligned to the Greengenes imputed core reference alignment. An OTU table was built and the chimeras were removed from the table. The OTU table and mapping file were both provided in a text file format.

B. Methodology

This project focused on analyzing the relative rank abundance plots in order to determine if the bacterial communities in the nine facilities are dominated by a neutral or a non-neutral model. The non-neutral model, the lognormal distribution, was fit to the relative rank abundance distribution plots generated from each of the nine facilities.

For the analysis of the data both the Quantitative Insights Into Microbial Ecology (QIIME v1.5) (14) software, as well as Microsoft Excel were used. First, QIIME v1.5 was used to separate the data by location. The data was clustered so that the sequences for each month for each of the nine facilities were grouped together, resulting in nine OTU tables. Then using QIIME v1.5 a rank abundance plot was generated for each of the nine facilities over the year time period. For this project, the relative rank abundance was used. Figure 2 is the rank abundance plot for the E1 facility.

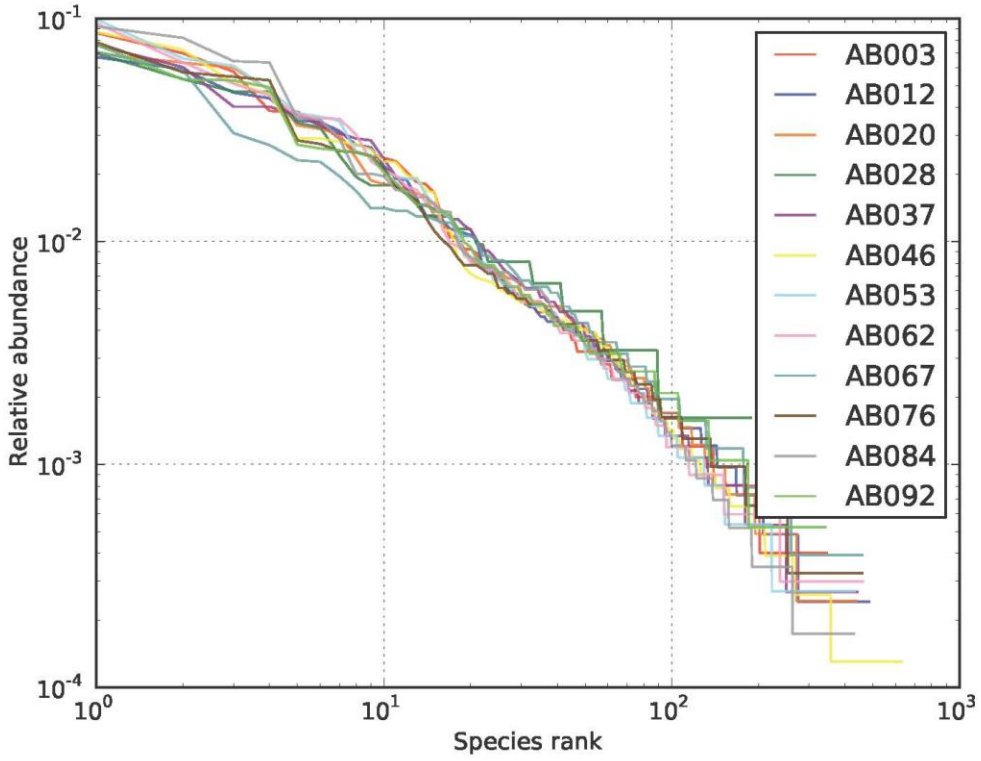

Figure 2. E1 Rank Abundance Plot

Next Microsoft Excel was used to calculate an average relative rank abundance plot for each of the nine facilities. For each facility the relative abundance of each OTU was determined. Then the OTUs were ordered from the most abundant to the least abundant. Next the average relative abundance was calculated for each of the ranks. Figure 3 is the relative rank abundance plot for the E1 facility. The relative rank abundance for each month as well as the average relative rank abundance are shown.

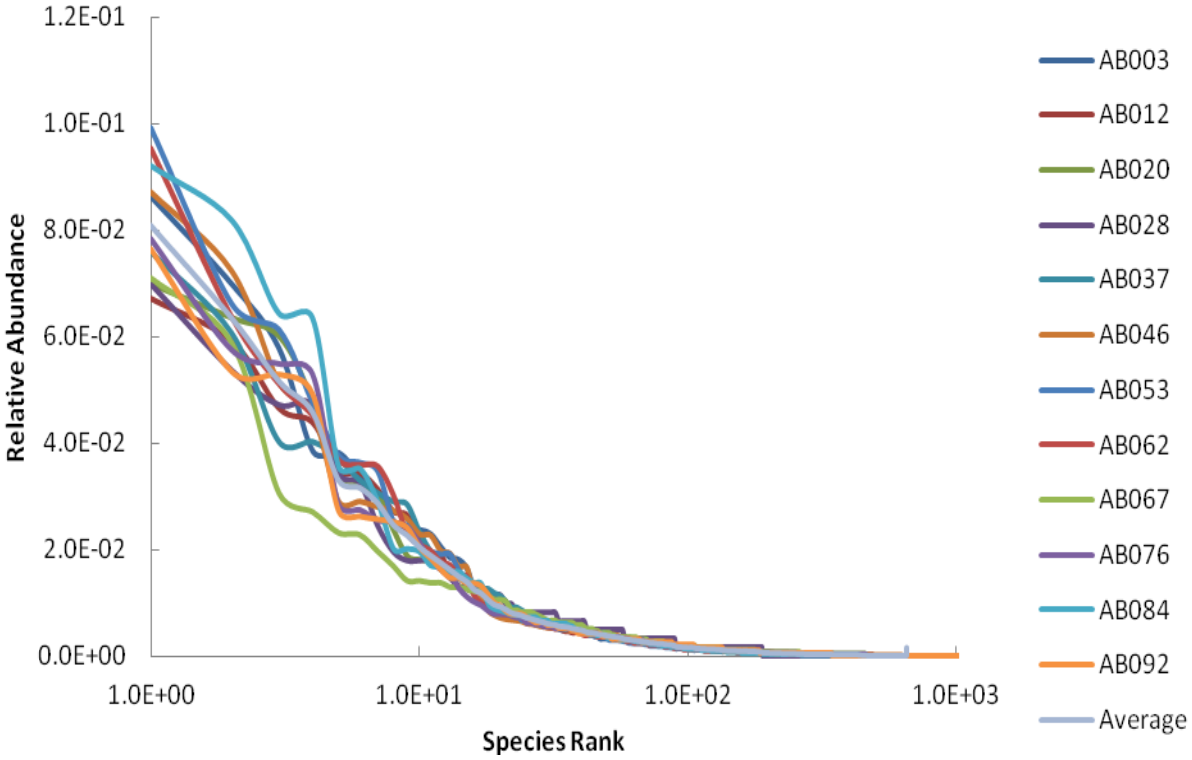

Figure 3. E1 Rank Abundance Plot with Average

The next step also used Excel. Using Excel the lognormal distribution was fit to the average relative rank abundance plot. Equation 1 shows the form of the lognormal distribution function that was used.

$$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \qquad (1)$$

The parameters μ and σ were adjusted to determine the form of the lognormal distribution that best fit the average relative rank abundance for each facility. Using the Solver add-in in Excel, the optimal values of the two parameters were determined to maximize the R squared value for each of the nine facilities. Finally, the R squared values corresponding to each facility were calculated.

IV. Results

Figure 4 shows the average relative rank abundance plot and the optimized lognormal distribution for the E1 facility.

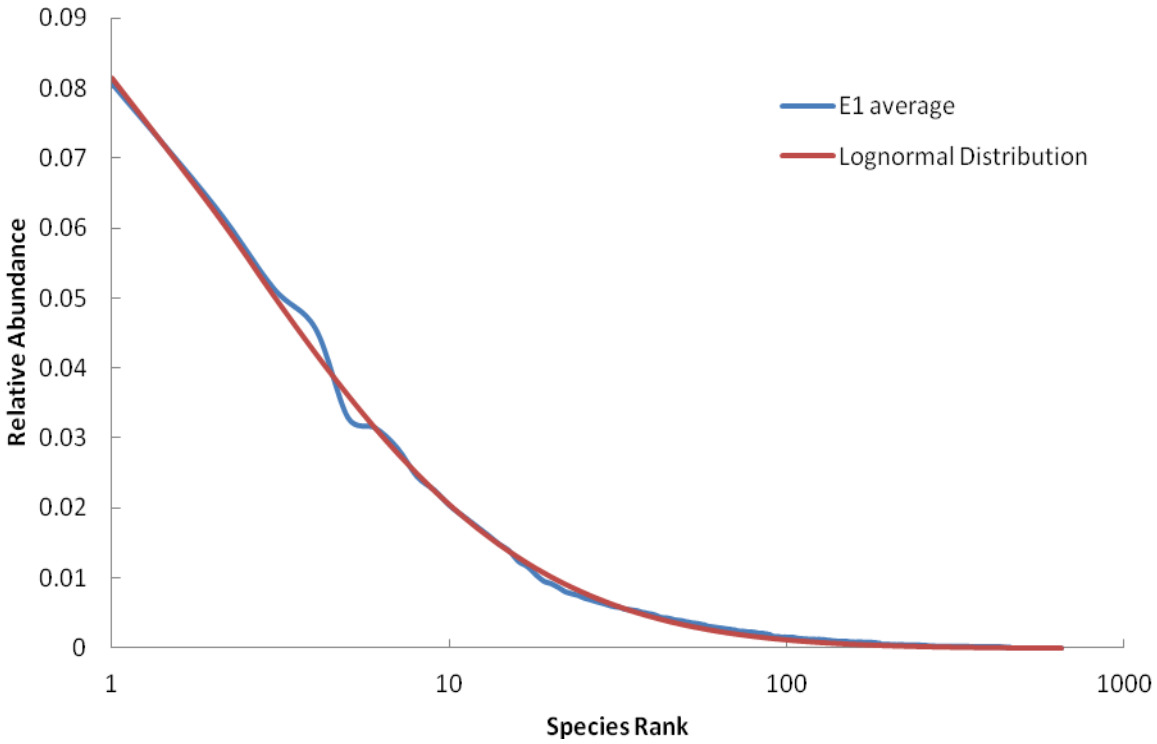

Figure 4. Average relative rank abundance plot and optimized lognormal distribution.

The R squared values that were calculated for each of the nine facilities are summarized in Table 1.

Facility	*R squared Value*
E1	0.9972
E2	0.9944
E3	0.9922
E4	0.9967
U1	0.9876
U2	0.9955
U3	0.9981
U4	0.9609
I1	0.9876

Table 1. The R squared values for each of the nine facilities

For each of the nine facilities the average relative rank abundance was shown to fit the lognormal distribution. All of the R squared values for each of the nine facilities are larger than 0.961; therefore, the average relative rank abundance plots for each of the nine facilities fit the lognormal distribution well. From the results of this project, it can be concluded that the bacterial communities in each of the nine facilities are non-neutral.

V. Conclusion

This project investigated the microbial communities found in nine anaerobic digesters treating brewery waste water. The study used the lognormal distribution as the non-neutral model. Using the data from the study by Werner et al. (13), the species abundance distributions for each of the nine facilities were fit to the lognormal distribution. The results of this analysis suggest that the microbial communities are non-neutral. The R squared values for the nine microbial communities range from 0.961 to 0.998; therefore, the lognormal distribution provides a good fit for the species abundance distributions.

VI. References

1. Grinnell, J. 1928. *Presence and absence of animals*. University of California Chronicle, October 1928.
2. Elton, C. 1 927. *Animal Ecology*. London: Sidgewick & Jackson.
3. Vandermeer, J. 1972. *Niche Theory*. Ann. Rev. Ecol. and System. 3:107-132.
4. Tilman D. 2004. *Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly.* Proc Natl Acad Sci USA 101:10854–10861.
5. Hubbell, S.P. 1979. *Tree dispersion, abundance and diversity in a dry tropical forest*. Science, 203, 1299–1309.
6. Hubbell, S.P. 1997. *A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs*. Coral Reefs, 16, S9–S21.
7. Hubbell, S.P. 2001. A *Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
8. McGill, Brian J., Brian A. Maurer, and Michael D. Weiser. 2006. *Empirical Evaluation of Neutral Theory.* 87:1411–1423.
9. McGill, B. J. 2003. *A test of the unified neutral theory of biodiversity*. Nature **422**:881–885.
10. I. D. Ofiteru et al. 2010. *Combined niche and neutral effects in a microbial wastewater treatment community.* Proc. Natl. Acad. Sci. U.S.A. 107, 15345.
11. Magurran, A.E. & Henderson, P.A. 2003. *Explaining the excess of rare species in natural species abundance distributions*. Nature, 422, 714–716.
12. Etienne, R.S. *&* Olff, H. *2004*. *A novel genealogical approach to neutral biodiversity theory*. Ecol. Lett.*,* 7*,* 170–175.
13. Werner JJ, Knights D, Garcia ML, Scalfone NB, Smith S, Yarasheski K *et al*. 2011. *Bacterial community structures are unique and resilient in full-scale bioenergy systems*. Proc Natl Acad Sci USA 108: 4158–4163.
14. Caporaso JG, et al. 2010 *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods 7:335–336.