Newcastle University e-prints

Date deposited: 7th December 2010

Version of file: Author, final

Peer Review Status: Peer Reviewed

Citation for published item:

Ofiteru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, Sloan WT. Combined niche and neutral effects in a microbial wastewater treatment community. *Proceedings of the National Academy of Sciences of the United States of America* 2010,**107** 35 15345-15350.

Further information on publisher website:

http://www.nas.edu/

Publishers copyright statement:

This paper was originally published by National Academy of Sciences, 2010 and can be accessed (with permissions) from the link below:

http://dx.doi.org/10.1073/pnas.1000604107

Always use the definitive version when citing.

Use Policy:

The full-text may be used and/or reproduced and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not for profit purposes provided that:

- A full bibliographic reference is made to the original source
- A link is made to the metadata record in Newcastle E-prints
- The full text is not changed in any way.

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Robinson Library, University of Newcastle upon Tyne, Newcastle upon Tyne. NE1 7RU. Tel. 0191 222 6000

Title: Combined niche and neutral effects in a microbial wastewater treatment community

Authors: Irina Dana Ofițeru[a,b], Mary Lunn[c], Thomas P. Curtis[a], George F. Wells[d], Craig S. Criddle[d], Christopher A. Francis[e] and William T. Sloan[f]

Author affiliation:

[a] School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

[b] Chemical Engineering Department, University Politehnica of Bucharest, RO 011061, Romania

[c] Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK

[d] Department of Civil and Environmental Engineering, Yang & Yamazaki Bld., 473 Via Ortega MC-4020, Stanford University, Stanford, CA 94305, USA

[e] Environmental Earth System Science, Y2E2 Building, 473 Via Ortega, Stanford University, Stanford, CA 94305-4216, USA

[f] Department of Civil Engineering, University of Glasgow, G12 8LT, UK

Corresponding Author:

William T. Sloan

Department of Civil Engineering, University of Glasgow, G12 8LT, UK

Phone: + 44 141 330 4076

Fax: + 44 141 330 4557

Email: Sloan@civil.gla.ac.uk

Manuscript information:

- number of text pages: 20
- number of figures: 3
- number of tables: 2

Abstract

It has long been assumed that differences in the relative abundance of taxa in microbial communities reflect differences in environmental conditions. Here we show that in the economically and environmentally important microbial communities in a wastewater treatment plant, the population dynamics are consistent with neutral community assembly, where chance and random immigration play an important and predictable role in shaping the communities. Using dynamic observations, we demonstrate a straightforward calibration of a purely neutral model and a parsimonious method to incorporate environmental influence on the reproduction (or birth) rate of individual taxa. The calibrated model parameters are biologically plausible, with the population turnover and diversity in the heterotrophic community being higher than for the ammonia oxidising bacteria (AOB) and immigration into AOB community being relatively higher. When environmental factors were incorporated more of the variance in the observations could be explained but immigration and random reproduction and deaths remained the dominant driver in determining the relative abundance of the common taxa. Consequently we suggest that neutral community models (NCMs) should be the foundation of any description of an open biological system.

Introduction

\body

Naturally occurring populations of bacteria and archaea are vital to life on Earth and are of enormous practical significance in medicine (1), engineering (2) and agriculture (3). However the rules governing the formation of such communities are still poorly understood. Typically, microbial community structure is thought to be shaped mainly by deterministic factors such as competition and niche differentiation, where the relationship between taxon traits and the environment dominate (4, 5). However, such theories when applied to macro-organisms struggle to explain very diverse environments where many rare taxa can coexist (6, 7). An alternative neutral theory has emerged (8, 9) that considers birth, death, dispersal and speciation and disregards the differences between species at the same trophic level. Hence in the neutral theory the explicit link between the distribution of relative abundances and the distribution of niches is broken. Despite their apparent simplicity and small number of parameters, neutral models have been remarkably successful at reproducing some of the most widely studied patterns in community ecology, including, species abundance distributions (SADs) and species-area relationships (SARs) in a wide range of communities from tropical trees to bacteria (10-16). However, neutral models are not without their critics. Some argue that alternative phenomenological models fit a particular data set marginally better (e.g. (17, 18)) others that the mechanisms are just plain "too simple" to represent biological reality and yet more that small deviations from neutrality would have large repercussions for the predicted patterns (19, 20). The arguments on the relative importance of niche and neutral forces in shaping community structure are, however, muddied by the inconclusive nature of the most common method for testing neutral theory. In this, the single observed distribution of taxa abundances at one location and at one period of time is compared to a distribution of abundances produced by a neutral model (15). The parameters are calibrated and it has not been possible to validate the models, and hence the underlying mechanisms

(21). There is however a push to move beyond this and use data from multiple sites (11, 22) and explore some of the richer predictions of neutral theory. The capacity of neutral theory to unite SADs and SARs has been demonstrated, which adds strength to the veracity of the underlying assumptions (12). Neutral models are derived from a dynamic stochastic process, so they might gain even more credence if if the dynamics in abundance and the SAD could be explained by the same neutral model (23); until now this has not been achieved .

The initial polarization of “nichists” and “neutralists” caused by Hubbell’s (9) monograph has waned and recognition that neutral models embody mechanisms (birth, death, immigration and sometimes speciation) that are indisputable features of virtually all biological systems (24) has led to calls for, what some call, “reconciliation” (25, 26). To this end a variety of niche models have been extended to include some stochastic elements (27-29). Whilst, these are often elegant expressions of plausible conceptual models they mostly defy calibration. For the microbial communities in which we are interested, where diversity is awe-inspiring (30) and traits are difficult to measure, it is impractical to aim for a model that requires a suite of taxon specific parameters. However, we maintain that a more parsimonious purely statistical approach can be taken to layering the influence of the environment on top of a neutral model when multiple realizations of a community composition exist.

Here we examine the microbial communities in a wastewater treatment plant to see if the stationary taxa rank abundance distribution is consistent with neutral theory. From ranked abundance distributions alone we cannot rigourously calibrate the model we can only determine whether or not the neutral model is a candidate; Etienne et al (2008) (31) and Hubbell (2001) (9) demonstrate the insensitivity of the abundance distributions where modest changes in the parameter values are only reflected in the abundance of rare taxa. In microbial surveys using molecular finger printing techniques like T-RFLP we can only observe taxa that exist at a relative abundance above approximately 0.01 and hence the rare taxa are missed. However, in removing the taxon names and considering merely their rank a huge amount of information contained within the time series is lost. Etienne et al (2008) (35) suggest that the uncertainty in parameter values might be reduced by supplementing taxa-abundance distributions with time series data. Therefore, we examine the dynamics of the most abundant taxa to see if they are also consistent with the neutral model, to refine the parameter estimates and to see if adding the influence of environmental covariates allows more of the variance to be explained.

Wastewater treatment plants are inherently open systems that rely on dozens, perhaps hundreds, of different species of bacteria and protozoa coming together to form a microbial community that will transform the waste into biomass, CO_2 or some other, less harmful, substances. Thus a model of the community assembly process could have wide practical application. Wells *et al*'s (32) comprehensive study of bacterial population dynamics of the Palo Alto Regional Water Quality Control Plant (PARWQCP) is used to test our hypotheses with dynamic data for the heterotrophic and ammonia oxidizing communities. They collected samples weekly for one year and profiled the communities (AOB and heterotrophs) using terminal restriction fragment length polymorphism (T-RFLP) analysis. Ten operational taxonomic units (OTUs) were identified for AOB on the basis of *amo*A analysis and 126 of heterotrophic bacterial OTUs were identified from 16S rRNA gene analysis. Wells *et al* (32) managed to encapsulate the patterns of relative abundance of taxa in a reduced set of ordinates that did a good job of preserving a measure of the distance between samples in the original data. They were then able to relate these new co-ordinates of samples to combinations and interactions between a large number of operating/environmental conditions, explaining as much as 30.2% and 25.5% of the variance for the AOB community and heterotrophic bacteria respectively. So the community as a whole, at least partially, responds to the environment with temperature, dissolved oxygen, influent nitrite, and chromium appearing to be important. However, the response of the microbial communities in a wastewater treatment plant to changing operating conditions is unlikely to be immediate. Therefore, it becomes important to analyze serial correlations and characterize the dynamics of individual taxa, preferably in a manner that lends itself to biological interpretation. Furthermore, it is natural to speculate on the 70% variance that is unexplained by the multivariate statistical methods, which when used with many environmental variables offer up the best prospect of explaining the variance in the data. Whilst modeling the fluctuation in the biomass of distinct functional groups of organisms has been successfully achieved in microbial ecology we know of no studies where a significant portion of the dynamics of individual taxa within a functional group has been explained. In our study we assume that the relative height of peaks in T-RFLP plots are estimates of the relative abundance of taxa.

The neutral model we use is that of Hubbell (9) formulated and extended for microbial communities into a continuous format that permits the inclusion of environmental effects(11). Thus, the wastewater treatment communities were assumed to be fed by immigrants from a source community where taxa abundances are distributed according to a logseries distribution with a single parameter θ that determines its shape. High values of θ correspond to diverse source communities and low values to less diverse communities. The distribution of taxa in the local community deviates from that in the source community as a function of the product of a pair of parameters, N_T and m ($N_T m$). N_T is the number of individuals in the neutrally assembled local community and m is the probability that when a member of the local community dies or is removed it is replaced by an individual from the source community rather than through local reproduction. Low migration tends to deplete the local richness of taxa and promote the dominance of common taxa. Advantage or disadvantage is conferred on a particular taxon by a factor α' (11) applied to the probability of birth.

The dynamics of the relative abundance, *X(t)*, of the i^{th} taxon at time t is entirely governed by $N_T m$ and the relative abundance of the taxon in the source community, p say, and can be described by a stochastic differential equation (see supplementary material)

$$dX(t) = \left(N_T m\left(p - X(t)\right) + 2\alpha' X(t)\left(1 - X(t)\right)\right)\frac{1}{a}dt + \frac{1}{\sqrt{a}}\sqrt{2X(t)(1-X(t)}dW_t \qquad \textbf{[1]}$$

where W_t is a Wiener process (standard Brownian motion) and a is an unknown constant that is related to the time between births and deaths. This differential equation is more general than a purely neutral model as the term involving α' confers an advantage $(\alpha' > 0)$ or disadvantage ($\alpha' < 0$) in the birth rate of the i^{th} taxon (33). The advantage coefficient α' is assumed to depend on external factors, thereby breaking the neutrality assumption but consistent with the simplifying assumptions of "mean field" models (34) in that it uses an equation for a given species which does not involve relative abundance of other species. A different α' can be used for each taxon and hence the model allows for differential birth rates but is not specific about the biological mechanisms that convey the advantage; we allow the data to define the advantage. When $\alpha' = 0$ then the differential equation describes purely neutral dynamics. From the observations of abundance for the i^{th} taxon, *X(t)* is known at 52 discrete time points and *dX(t)* can be crudely approximated as the change in relative abundance between successive times. So Eq. **1** maps on to a simple linear model,

$$dX = m_0 + m_1 Y_1 + m_2 Y_2 + \varepsilon \qquad \textbf{[2]}$$

where $m_0 = \frac{N_T mp}{a}$, $m_1 = -\frac{N_T m}{a}$, $m_2 = \frac{2\alpha'}{a}$, $Y_1 = X$, $Y_2 = X(1 - X)$ and ε is an error term given by $\varepsilon = \frac{1}{\sqrt{a}}\sqrt{2X(t)(1 - X(t))}dW_t$. Thus, while W_t is normally distributed, $N(0,1)$, ε is not. However, equation 2 gives us a straightforward method of calibrating the unknown parameters $N_T m$ and a, under the assumption that $\alpha' = 0$. Performing a weighted least squares regression analysis, using observations of the dependent dX and independent variable in which the weights are $\left(X\ (1-X\ )\right)^{-1}$, gives estimates of the parameters m_0 and m_1. The weighted errors should be normally distributed and hence the standard residual error produced by the least squares analysis should be $\sqrt{\frac{2}{a}}$. Thus, all of the original model parameters in Eq. **1** can be retrieved from a linear least squares analysis (See supplementary material). Furthermore if we allow a non-zero advantage term, α', to be a linear function of n observed covariates, $\{Z_j\}_{j=1}^{n}$, such as temperature or chemical concentration,

$$\alpha' = \alpha_0 + \sum_{j=1}^{n} \alpha_j Z_j, \qquad \textbf{[3]}$$

then incorporating the effects of environment on the birth-death process in the community is achieved by merely extending the linear least-squares analysis to incorporate more independent variables,

$$dX = m_0 + m_1 Y_1 + m_2 Y_2 + m_3 (Y_2 Z_1) + m_4 (Y_2 Z_2) + \ldots + m_{n+2} (Y_2 Z_n) + \varepsilon \qquad \textbf{[4]}$$

where the coefficients are related to the advantage parameters by $m_j = \frac{2\alpha_{j-2}}{a}$ for $j \geq 2$.

Results

One of the predictions of neutral theory is that for a neutrally assembled community the distribution of ranked abundances for the taxa will essentially remain constant within bounds imposed by the natural variability of a stochastic birth-death-immigration process. The relative abundance of the most abundant AOB and heterotrophic bacteria, is very dynamic (Fig. 1). The identity of the top ranked taxon changes many times during the year. However, ignoring the taxon labels and merely ranking their relative abundance for each week (Fig. 2) we see order emerge from what appeared to be a highly complex and dynamic system. It is extremely rare to see a time series of so many ranked abundance distributions from a single site and the prediction that ranked abundances will remain constant even if the individual taxa abundances are highly dynamic has never previously been shown experimentally. We sought the best fit of the neutral model to these data in a least squares sense. It transpires, however, that a very good fit to the data was achievable for both communities for a broad swathe of the parameter space (Fig. 3), which confirms previous findings (31). However, much of the information in the original time series is lost by ignoring the identity of the taxon. Therefore, working with the dynamic representation of the model (Eq. 1), we calibrated a completely neutral model $\alpha' = 0$ using the time series of abundances of the two most abundant organisms in the two communities (Table 1). The estimates of the model parameters are statistically significant at the 99.9% level and the 95% confidence limits of the parameter estimates within each community overlap significantly for both the AOB and the heterotrophs. The R- squared values indicate that approximately a fifth of the variance in the time series of abundance are explained by a purely neutral model.

Examining the dynamics has allowed us to determine the parameter $N_T m$ for a neutral model without any knowledge of the distribution of taxa in the source community; the parameter θ does not appear in the stochastic differential equation **1** for the relative abundance of a single taxon. Armed with this knowledge, we can go back to the ranked abundance distribution which gives an indication of how the log-series distribution of the taxa abundances in the source community is distorted by dispersal limitation into the local wastewater treatment plant, and refine our estimates of θ for both for AOB and the heterotrophs. The best least-squares fit between the observed and simulated ranked abundance distribution was achieved using a θ value of 2.5 for the AOB and 23 for the heterotrophs. This is consistent with the widespread and plausible assumption that the AOB are much less diverse than the heterotrophs.

For our optimal parameter pairs ($N_T m = 55$ and $\theta = 2.5$ for AOB data, $N_T m = 198$ and $\theta = 23$ for heterotrophs) we generated 500 realisations of the wastewater treatment communities from which we sampled 10^6 individuals at random to simulate the physical sampling done in a T-RFLP analysis. From this the average abundances and the 5th and 95th percentile abundances for each rank were calculated (Fig 2). Clearly the vast majority of the observed ranked abundances for each week fall within the 90% confidence limits of the simulated abundance distributions. Thus determining the $N_T m$ from dynamic data constrains our search for the value of data θ using the ranked abundance distributions.

We tested whether more of the variance in the time series data might be explained by incorporate the effects of environment by conveying an advantage on the birth rate of taxa that is linearly related to environmental factors. This was achieved by adding extra terms in the linear model (Eq. **3**). Wells *et al* (32) and Wells *et al* (35) collated times series of twenty environmental variables measured at the same time as the microbiological samples were taken. We tested a suite of linear models (Eq. **3**) that included each of these variables individually and models where combinations of the variables were included but we have only presented the models that explained the most variance over-and-above the purely neutral model as defined by the first two terms in the linear model (Eq. **2**) using statistically significant estimators for the environmental factors included. For the most abundant heterotrophs the model which best met these criteria was achieved by making the advantage term in Eq. **3** a linear function of the dissolved oxygen concentration (Z_1),

$$m_2 = \frac{2\alpha'}{a} = -0.08 \cdot Z_1 \qquad \textbf{[5]}$$

while for the most frequently occurring AOB species it is a function of temperature (Z_1) and chromium (Z_2) concentration

$$m_2 = \frac{2\alpha'}{a} = 0.027 \cdot Z_1 - 0.026 \cdot Z_2 \qquad \textbf{[6]}$$

The complete sets of parameters m for both sets of data are reported in Table 2. α confers a relative advantage on the individual taxon not the community as whole. Thus, whilst the productivity of the whole heterotroph community may increase with an increasing concentration of dissolved oxygen, some taxa within the community will respond relatively less well than others.

The purely neutral model ($\alpha = 0$) accounts for 0.23 of the variability (measure by R^2) in the time series data for AOB, and 0.20 for heterotroph data. When these taxa were allowed an advantage, α, in the probability of birth which was linearly related to environmental variables, a bigger percentage of the variability could be explained, increasing the coefficient of determination to 0.37 for AOB data, and 0.28 for the heterotrophs respectively.

Discussion

The call for a "reconciliation" of niche and neutral models (25, 26) of community assembly has, until now, been met by adapting what were deterministic niche based models to include stochasticity and immigration (27-29). We have argued in the introduction to this paper that the emergent parameter rich models defy calibration for very diverse microbial communities. However if, as recent studies suggest, neutral dynamics have a significant influence on the community composition (11) then an alternative approach seems logical where neutral dynamics forms the core of the model and environmental effects are layered on top as and when required. This should ultimately lead to a more parsimonious description of the system. It could be argued that conceptually this is a more pleasing approach to modelling the assembly of any open biological community. Births, deaths and immigration are inevitable whilst the relative importance of environmental effects on individual taxa (as opposed to the community as a whole) may vary. However, building a model on a foundation of neutral dynamics, however conceptually pleasing, is only of practical benefit if neutral dynamics do indeed account for a significant proportion of the variance in the observed dynamics and this has not previously been tested. Using both static and dynamic observations, we have demonstrated that a straightforward calibration of a purely neutral model is possible and we give a parsimonious method to incorporate environmental influence on individual taxa.

We have re-iterated the fact that using taxa abundance distributions from one site is a poor test of neutral theory (31). Many parameter pairs will lead to similar shaped abundance distributions (Fig. 2), especially when the distribution is truncated by methodological constraints like the threshold in abundance below which taxa cannot be observed using T-RFLP. Nonetheless, the ranked abundance distributions for each week from the two bacterial communities in the Palo Alto sewage works are consistent with neutral theory and do remain constant through time. So the taxa abundance distributions by themselves give no reason for rejecting neutral theory as the foundation of a mathematical description of community assembly. Woodcock et al (12) demonstrated that it is possible to pin down the parameters of a neutral model using taxa abundance distributions if they come from multiple sites and either the immigration rate or the population sizes change significantly between sites. With data from a single site then the only alternative is to extract more information from the time series of abundance for named taxa, the Palo Alto sewage works time series are a rare example of such data. We were able to explain 23% and 27% of the variance in the time series of abundance for the two ranked AOB taxa using a purely neutral model. For the top two heterotrophs we could explain 20% and 27% of the variance using neutral dynamics. This suggested that neutral dynamics plays a significant role. In addition, there is a large overlap in the confidence limits on the best values of N_Tm for taxa within each functional group. If the taxa were behaving entirely neutrally then this consistency in the estimates for N_Tm calibrated on the dynamics of individual taxa would extend deeper into the community. However, this is difficult to test using the current data because the abundances of all other AOB taxa often drop below the detection limit of the T-RFLP method and for the very low abundances in the heterotrophs measurement noise is relatively large. Nonetheless, the consistency in estimates of the communities $N_T m$ value using the dynamics of the top two most abundant taxa from each group does suggest that migration driven drift is important and consistent within functional groups. In addition, the difference in the estimated parameters between functional groups makes biological sense. We estimated the timescale constant $a = 520$ for the heterotrophs and $a = 139$ for the AOB. This can be interpreted (see methods) as there being 520 replacements in the heterotrophic community for every 139 replacements in the ammonia oxidising community or the turnover in heterotrophic taxa being 3.75 times greater than the AOB. This partly reflects the different community sizes; the total count, N_T, of AOB (36-38) in a wastewater treatment plant is approximately 5 to 10% that of the heterotrophic community. The best value of $N_T m$ for the heterotrophs is 3.6 times greater than the AOB. It is difficult to translate these values into an estimate of the absolute immigration probability because it will depend on our definition of the local community and thus N_T (12). However, given that the total number of AOB is about 10% of the number of heterotrophs the values would indicate that the probability of replacement in the AOB community by an immigrant is actually higher than for the heterotrophs. This may again reflect the relative population sizes since the smaller the community, the higher the probability of a dead individual being replaced by immigration (39)). The parameter p is the relative abundances of the taxon in the source community, which we estimate to be to be 0.06 for the most common heterotroph and 0.39 for the most common AOB. These values are the same orders of magnitude as the average relative abundances displayed in Fig 1. Unfortunately, our lack of knowledge of the abundances when they drop close to or below the T-RFLP detection limit means that we cannot estimate the average abundance for all taxa in community in this way, which would have defined the source community abundance distribution. Therefore, to estimate θ, the parameter that defines the logseries abundance distribution for the source community, we needed to return to the ranked abundance distributions armed with the knowledge of $N_T m$ gained from examining the timeseries of the most abundant organisms. We estimate that θ is 2.5 for the AOB and 39 for the heterotrophs. Hubbell (9) calls θ the fundamental biodiversity number because it is an index to the richness of taxa in the source community. Our values suggest that the AOB are much less diverse than the putative heterotrophs, an observation consistent with prevailing opinion in microbial ecology and the specificity of the PCR primers used in the analysis of each community.

The inclusion of an advantage/disadvantage term which acts on the probability of birth for each taxa means that the core migration and stochastic births and deaths are retained in a model that can also represent niche effects. The birth rates are no longer equivalent and hence the model is no longer neutral, but Sloan et al (33) show that the migration and stochasticity will ensure that biodiversity is maintained. The advantage term was made a linear function of any number of environmental variables and we sought the combination of variables that explained the most of the variance in the time series of abundance. It is gratifying that the same environmental factors (dissolved oxygen, temperature, chromium – Table 2) determined by multivariate statistics to influence the community (32, 35) were also identified by this combined model. There is still substantial unexplained variation in the data which could be attributable to unmeasured environmental factors, or a non-linear relationship between environment and advantage or substantial measurement error. The effect of the environment on the most abundant T-RF could also be weakened by if the T-RF did not comprise an ecologically homogenous group. This could happen because of natural variation within one phylogenetic group or if an unrelated less abundant organism had the same T-RF. However, it should be remembered that the very best quantitative molecular methods have a coefficient of variation of about 20% (38) and so we can expect at least this much "noise". It may be that the model could be improved if the advantage parameter α was allowed to vary non-linearly with environmental factors. However, this is unlikely to be worthwhile until we are able to garner more high resolution and high quality data. In particular our ability to encapsulate the dynamics using a stochastic differential equation model would be enhanced if regular weekly samples were supplement by periods of more frequent sampling.

Sceptics might suggest that the excellent performance of the NCM may occur because the Palo Alto wastewater treatment plant is a carefully managed system in a climate with little seasonal variations. Only high quality, high resolution time series in more variable environments can answer this. However, even if sceptics were right, there are many well controlled stable environments where NCM may find application. The gut, for example, is a plug flow reactor held at a constant temperature. Those wishing to explore or engineer the human or animal microbiome will find NCM invaluable. It could for example be used to rationally design and deploy pro and prebiotics. From an engineers perspective realising that microbial community composition is so dependent on neutral processes and cannot be entirely shaped by environmental conditions could change the way we design sewage works. Bacterial community size correlates with the volume of the sewage works and immigration of new species with the rate at which waste is fed, so changing these two variables could allow us to manipulate the diversity and the timescales over which the population dynamics occur. So, for example, there may be a minimum reactor size and flow rate to ensure that organisms which are rare but important, such as those that can metabolise endocrine disrupting chemicals, are maintained in the system. Or we might be able to predict the frequency with which important taxa are likely drift below critical thresholds. More generally, those seeking to engineer or explore any real microbial environment, and many such systems are under consideration for fuel generation or carbon capture, will benefit from a sound body of theory. We believe that NCM should form the core of that body of theory.

Methods

The two microbial time series examined as part of this study were obtained from the same local environment (the four well-mixed aeration basins of the PARWQCP) and during the same time period (February 2005 - February 2006). The activated sludge samples are 24-h composite (collected every 30 minutes), gathered weekly from the combined outlet of all basins. The first time-series was generated via *β*- proteobacterial-specific *amoA* T-RFLP (digested with the restriction enzyme *TaqI*), and the second was generated via bacterial-specific 16S rDNA T-RFLP (digested with the restriction enzyme *RsaI*). The optimal volume (and hence DNA quantity) applied for fragment sizing was chosen to maximize total fluorescence signal while avoiding detector saturation. Both T-RFLP datasets were binned and normalized, such that individual OTU scores in each sample represent a measure of relative abundance. All peaks below the background noise (<0.5% of the total summed peak heights in any given sample) were neglected, yielding a detection limit of $d = 0.005$.

Acknowledgements

We thank Ian Head and Micol Bellucci for helpful comments. We acknowledge the support of ECOSERV Marie Curie excellence grant, an EPSRC Advanced Fellowship and an EPSRC platform grant.

References

1. Dethlefsen L, Eckburg PB, Bik EM, & Relman DA (2006) Assembly of the human intestinal microbiota *Trends Ecol Evol* **21,** 517-523.
2. Curtis TP, Head IM, & Graham DW (2003) Theoretical ecology for engineering biology *Environ Sci Technol* **37,** 64A-70A.
3. Buckley DH & Schmidt TM (2003) Diversity and dynamics of microbial communities in soils from agro-ecosystems *Environ Microbiol* **5,** 441-452.
4. Ramette A & Tiedje JM (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem *Proc Natl Acad Sci* **104,** 2761-2766.
5. Tilman D (1982) *Resource competition and community structure* (Princeton Univeristy Press, Princeton, New Jersey).
6. Hérault B (2007) Reconciling niche and neutrality through the emergent group approach *Perspect Plant Ecol, Evol and Systematics* **9,** 71-78.
7. Gewin V (2006) Beyond neutrality - ecology finds its niche *PLoS Biol* **4,** 1306-1310.
8. Bell G (2000) The distribution of abundance in neutral communities *Am Nat* **155,** 606-617.
9. Hubbell SP (2001) *The unified neutral theory of biodiversity and biogeography* (Princeton University Press, Princeton).
10. Ulrich W & Zalewski M (2007) Are ground beetles neutral? *Basic Appl Ecol* **8,** 411-420.
11. Sloan WT*, et al.* (2006) Quantifying the roles of immigration and chance in shaping prokaryote community structure *Environ Microbiol* **8,** 732-740.
12. Woodcock S*, et al.* (2007) Neutral assembly of bacterial communities *FEMS Microbiol Ecol* **62,** 171-180.
13. Volkov I, Banavar JR, Hubbell SP, & Maritan A (2007) Patterns of relative species abundance in rainforests and coral reefs *Nature* **450,** 45-49.
14. Muneepeerakul R*, et al.* (2008) Neutral metacommunity models predict fish diversity patterns in mississippi-missouri basin *Nature* **453,** 220-U229.
15. Volkov I, Banavar JR, Hubbell SP, & Maritan A (2003) Neutral theory and relative species abundance in ecology *Nature* **424,** 1035-1037.
16. Babak P & He FL (2008) Species abundance distribution and dynamics in two locally coupled communities *J Theor Biol* **253,** 739-748.
17. McGill BJ (2003) A test of the unified neutral theory of biodiversity *Nature* **422,** 881-885.
18. McGill BJ, Maurer BA, & Weiser MD (2006) Empirical evaluation of neutral theory *Ecology* **87,** 1411-1423.
19. Zhou S-R & Zhang D-Y (2008) A nearly neutral model of biodiversity *Ecology* **89,** 248-258.
20. Fuentes M (2004) Slight differences among individuals and the unified neutral theory of biodiversity *Theor Popul Biol* **66,** 199-203.
21. Gotelli NJ & McGill BJ (2006) Null versus neutral models: What's the difference? *Ecography* **29,** 793-800.
22. Etienne RS & Alonso D (2007) Neutral community theory: How stochasticity and dispersal-limitation can explain species coexistence *J Stat Phys* **128,** 485-510.
23. Vanpeteghem D, Zemb O, & Haegeman B (2008) Dynamics of neutral biodiversity *Math Biosci* **212,** 88-98.
24. Alonso D, Etienne RS, & McKane AJ (2006) The merits of neutral theory *Trends Ecol Evol* **21,** 451-457.
25. Dewar RC & Porte A (2008) Statistical mechanics unifies different ecological patterns *J Theor Biol* **251,** 389-403.
26. Clark JS*, et al.* (2007) Resolving the biodiversity paradox *Ecology Letters* **10,** 647-659.

27. Mouquet N & Loreau M (2003) Community patterns in source-sink metacommunities *Am Nat* **162,** 544-557.
28. 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.
29. Scheffer M & van Nes EH (2006) Self-organized similarity, the evolutionary emergence of groups of similar species *Proc Natl Acad Sci* **103,** 6230-6235.
30. Quince C, Curtis TP, & Sloan WT (2008) The rational exploration of microbial diversity *ISME Journal* **2,** 997-1006.
31. Etienne RS, Latimer AM, Silander JA, & Cowling RM (2006) Comment on "Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot" *Science* **311**.
32. Wells GF*, et al.* (2009) Ammonia-oxidizing communities in a highly aerated full-scale activated sludge bioreactor: Betaproteobacterial dynamics and low relative abundance of crenarchaea *Environ Microbiol (Accepted).*
33. Sloan W*, et al.* (2007) Modeling taxa-abundance distributions in microbial communities using environmental sequence data *Microb Ecol* **53,** 443-455.
34. McKane A, Alonso D, & Solé RV (2000) Mean-field stochastic theory for species-rich assembled communities *Phys Rev E* **62,** 8466.
35. Wells GF*, et al.* (2009) Fine-scale bacterial community dynamics in a full-scale nitrifying activated sludge wastewater treatment plant *(In Preparation)*.
36. Harms G*, et al.* (2003) Real-time pcr quantification of nitrifying bacteria in a municipal wastewater treatment plant *Environ Sci Technol* **37,** 343-351.
37. Urakawa H*, et al.* (2006) Abundance and population structure of ammonia-oxidizing bacteria that inhabit canal sediments receiving effluents from municipal wastewater treatment plants *Appl Environ Microbiol* **72,** 6845-6850.
38. Coskuner G*, et al.* (2005) Agreement between theory and measurement in quantification of ammonia-oxidizing bacteria *Appl Environ Microbiol* **71,** 6325-6334.
39. Curtis TP*, et al.* (2006) What is the extent of prokaryotic diversity? *Philos Trans R Soc London Ser B* **361,** 2023-2037.

Figure Legends

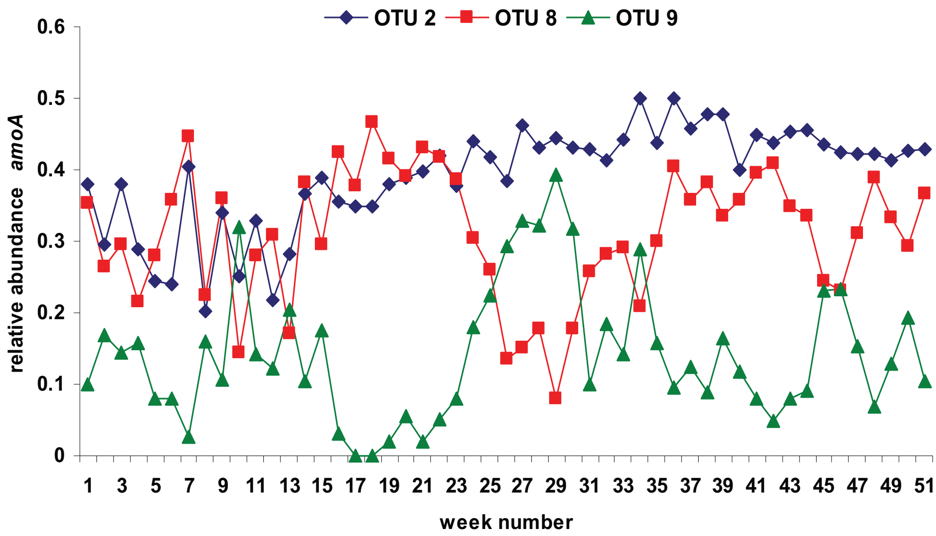
Fig. 1 Relative abundance of **(A)** the three most abundant Ammonia Oxidising Bacteria (AOB) and **(B)** the five most abundant heterotrophic bacterial measure at weekly intervals for one year in the Palo Alto Regional Water Quality Control Plant. The Operational Taxanomic Units (OTUs) and their abundances were using the relative area under peaks in T-RFLP electropherogram of the *amoA* genes for the AOB and the 16S rRNA genes for the heterotrophs. The identity of the most abundant taxon changes over time, both for AOB and heterotrophic OTUs. The abundance of rarer organisms, which are not shown on these graphs, frequently drop below the measurement detection limit.

Fig. 2 Weekly ranked abundance distributions (A) for all the AOB and (B) for the heterotrophic bacteria. Approximately the same ranked abundance patterns are observed each week both for both groups of bacteria even though, as Fig. 1 shows, the identity of the taxon at each rank changes over time. The best fitting model ranked abundance distribution are represented by the green line. The red dashed lines give 90% confidence limits for the modeled abundance distribution derived from 500 realisations. The majority of the data fall within the model confidence limits. It can be seen from figures 3 and 4 that a wide range of parameters give similar fits. However, calibrating the stochastic differential equation representation of the model using the dynamics of the most abundant taxa allows the value of $N_T m$ to be determined independently of the ranked abundance distributions. Thus it is only the value of θ in these plots that has optimized on the basis of the ranked abundance distribution.

Fig. 3 The sum of the square of the errors between the observed ranked abundance distributions (Fig 2A) and the modeled distribution for a wide range of parameter pairs, (**a**) for the AOB and (**b**) for the heterotrophs. The dark blue regions with similarly low sum of squared errors indicates there is a broad swathe of the parameter space, with a good fit to the model and the ranked abundance data. This reinforces the fact that calibrating a neutral model based solely on ranked abundance distributions from one site will yield uncertain parameter estimates. Thus supplementing the ranked abundance distributions with additional data is required to reduce the uncertainty. Times series of the abundance are used to achieve this for the best fitting distributions in Fig 2 .

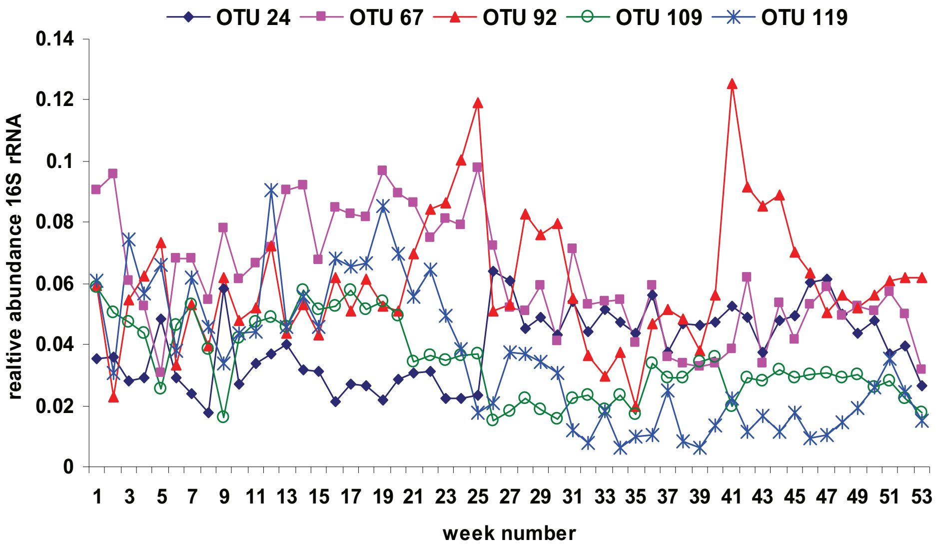
Figure 1

(A)

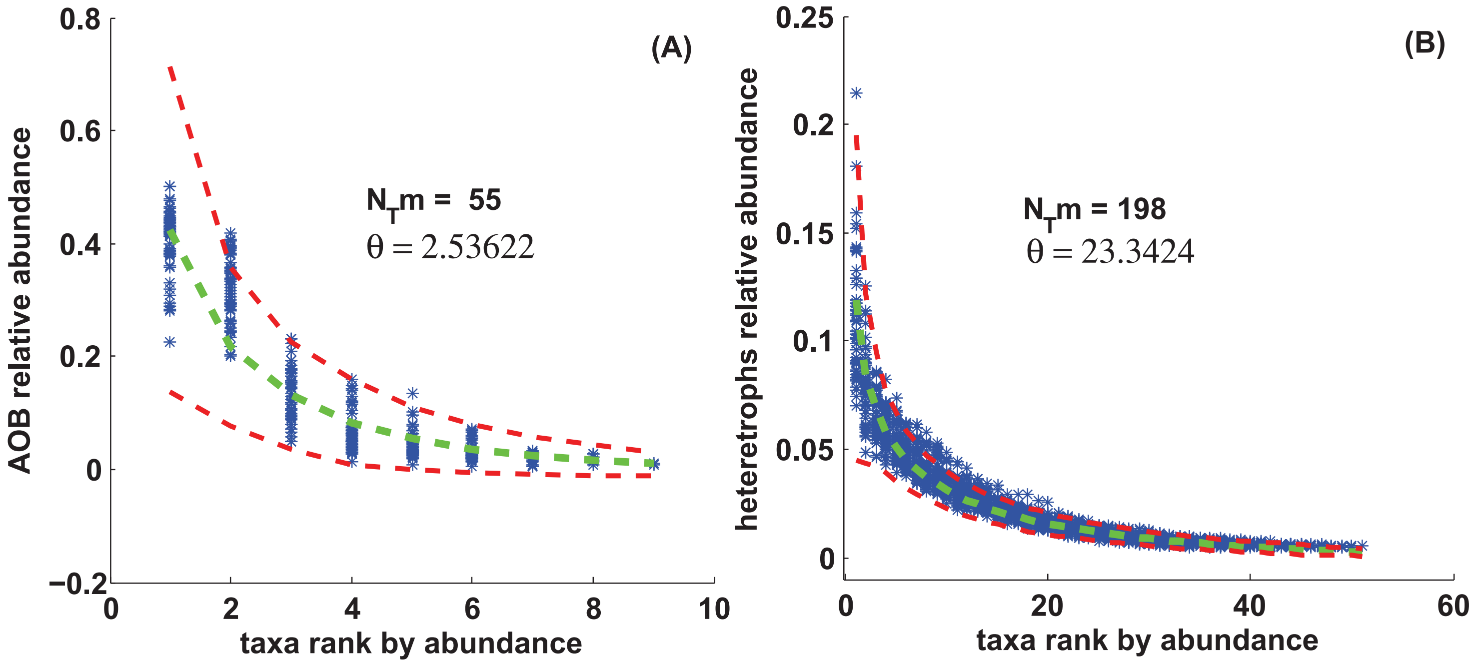


OTU 2
OTU 8
OTU 9
0.6
0.5
0.4
0.3
0.2
0.1
0
relative abundance amoA
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51
week number

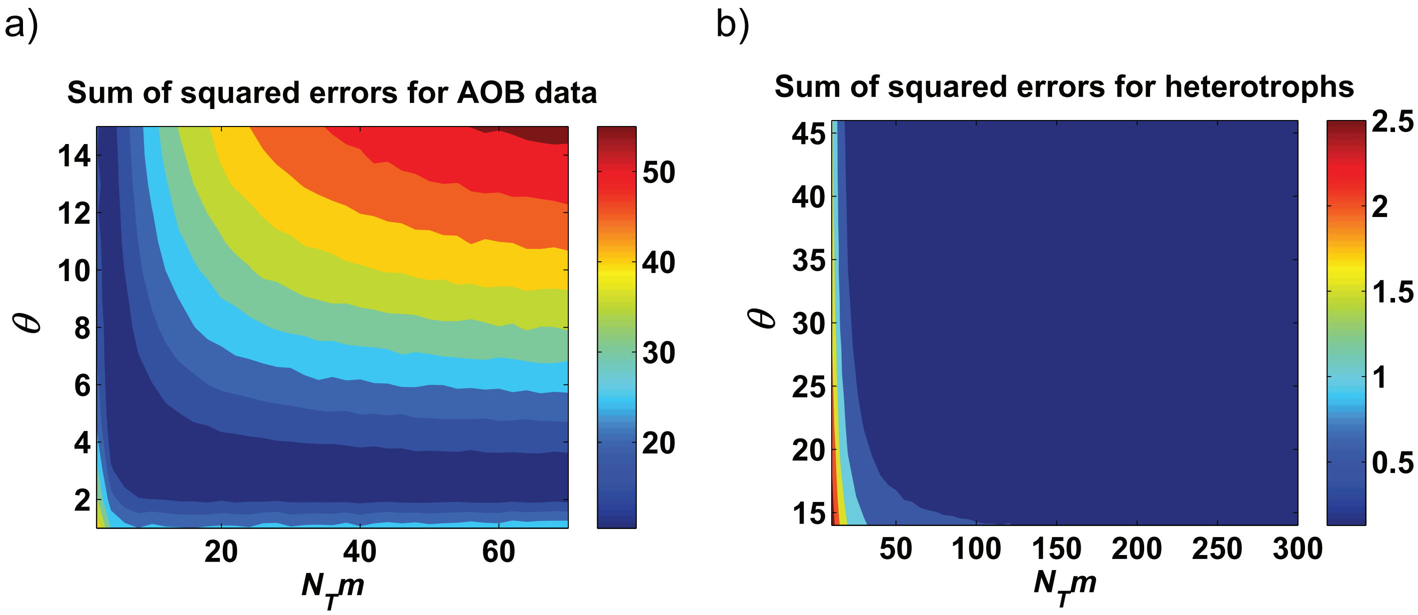
(B)



OTU 24
OTU 67
OTU 92
OTU 109
OTU 119
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
realtive abundance 16S rRNA
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53
week number



(A)
AOB relative abundance
0.8
0.6
0.4
0.2
0
−0.2
0
2
4
6
8
10
taxa rank by abundance
$N_Tm = 55$
$\theta = 2.53622$
(B)
heteretrophs relative abundance
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
taxa rank by abundance
$N_Tm = 198$
$\theta = 23.3424$


a)
Sum of squared errors for AOB data
θ
$N_T m$
14
12
10
8
6
4
2
20
40
60
50
40
30
20
b)
Sum of squared errors for heterotrophs
θ
$N_T m$
45
40
35
30
25
20
15
50
100
150
200
250
300
2.5
2
1.5
1
0.5

		Regression Coefficients										Neutral Model Parameters	
		m_0				m_1				R^2	Standard Error (SE)	From least squares estimates of m_0 and m_1 and SE	
		$Q_{2.5}$	Least Squares Estimate	P-Value	$Q_{97.5}$	$Q_{2.5}$	Least Squares Estimate	P-Value	$Q_{97.5}$			p_i	$N_T m$
Ammonia Oxidising Bacteria	*OTU* 1	0.075	0.157	<0.001	0.239	-0.610	-0.400	<0.001	-0.191	0.23	0.12	0.39	55
	OTU 2	0.077	0.143	<0.001	0.208	-0.678	-0.461	<0.001	-0.245	0.27	0.18	0.31	30
Heterotrophic Bacteria	*OTU* 1	0.010	0.023	<0.001	0.036	-0.605	-0.391	<0.001	-0.176	0.20	0.06	0.06	199
	OTU 2	0.018	0.032	<0.001	0.045	-0.766	-0.526	<0.001	-0.285	0.27	0.05	0.06	170

Table 1. Parameter values for a purely neutral model for the most abundant organisms in the heterotrophic and AOB communities, respectively.

Table 2. Parameter values for the combined model for the most abundant organisms in the heterotrophic and AOB communities, respectively

	COEFFICIENT	**ST-ERROR**	**P-VALUE**
Heterotrophic Bacteria (R^2=0.28)			
m_0	0.03	0.01	< 0.001
m_1	-	-	not sig
m_2 (dissolved oxygen)	-0.08	0.02	<0.0001
Ammonia Oxidising Bacteria (R^2=0.37)			
m_0	0.11	0.05	< 0.05
m_1	-0.86	0.24	< 0.001
m_2 (temperature)	0.027	0.012	0.02
m_2 (chromium)	-0.026	0.011	0.02