Supporting Information

Dini-Andreote et al. 10.1073/pnas.1414261112

SI Materials and Methods

Sampling Location and Data Collection. Soil samples were collected at five different stages of soil development in a salt marsh chronosequence spanning over a century of ecosystem development, in the island of Schiermonnikoog (N53°30′ E6°10′), The Netherlands (for a detailed description of sample collection, see ref. 1). Successional stages were identified and estimated as 0, 5, 35, 65, and 105 years of soil development (referred to as "years") in 2012. For detailed information on chronosequence establishment and calibration, see refs. 1 and 2. Briefly, plots were established at the same base of elevation [position at the initial elevation gradient on the bare sand flats with a base elevation of 1.16 m ± 2.2 cm (mean ± SE) above Dutch Ordinance Level]. Sampling was performed in all sites in four time points: May, July, September, and November 2012. Total soil DNA was extracted using a MoBio PowerSoil DNA extraction kit (MoBio Laboratories), and communities were profiled targeting the V4–V6 region of the bacterial 16S rRNA using a Roche GS-FLX 454 automated pyrosequencer running the Titanium chemistry. Sequences were analyzed in QIIME (3), using PyNast (4) and the Greengenes core set (5) for alignment and FastTree (6) to generate a phylogeny containing all OTUs observed across all samples. OTUs were defined at 97% of nucleotide identity. OTU table was rarified to a depth of 2,600 sequences per sample (the fewest in a single sample), to minimize effects of sampling effort upon analysis. Each individual sample was subjected to measurements of soil physical structure (clay:silt:sand %) and chemical content of total organic matter (OM), nitrate (N-NO_3^-), ammonium (N-NH_4^+), sulfate (S-SO_4), sodium (Na), and pH (for methods, see ref. 1).

Phylogenetic Signal. To test for phylogenetic signal, we used a phylogenetic Mantel correlogram (7–9). The inputs to this analysis are two distance matrices with OTUs as both rows and columns. One matrix contains between-OTU differences in environmental optima, and the other matrix contains between-OTU phylogenetic distances. Differences in environmental optima among OTUs were quantified using all measured physicochemical variables (7, 8). For each OTU we calculated its relative-abundance–weighted mean value for each physicochemical variable; the resulting values estimate the magnitude of each physicochemical variable at which a given OTU is most abundant, which we interpret as a proxy for the level of each physicochemical variable at which a given OTU has its highest fitness (i.e., the OTU's optimal environmental condition with respect to a given physicochemical variable) (7–12). After estimating environmental optima for all OTUs with respect to all physicochemical variables, we generated a matrix containing these estimates, with OTUs as rows and physicochemical variables as columns; estimates for each physicochemical variable were normalized as standard normal deviates. Among-OTU differences in environmental optima were then quantified as Euclidean distances simultaneously using all (normalized) physicochemical axes.

Using a phylogenetic Mantel correlogram (R function mantel.correlog in package vegan), correlation coefficients between differences in environmental optima and phylogenetic distances were quantified for 50 phylogenetic distance bins, and significance of those correlations was evaluated using 1,000 permutations and a progressive Bonferroni correction (13) (Figs. S1 and S2). A positive correlation coefficient indicates that more closely related species are more similar ecologically (and conversely that more distantly related species are more ecologically dissimilar). For each phylogenetic distance bin—phylogenetic distances were normalized to vary between 0 and 1 before analysis—we consider there to be significant phylogenetic signal if the correlation coefficient is positive and significantly larger than the coefficient expected following permutation (7–9).

Turnover in Phylogenetic Community Composition. To characterize the turnover in phylogenetic community composition, we quantified the β-mean nearest taxon distance (βMNTD). This metric was calculated as follows:

$$\beta\text{MNTD} = 0.5\left[\sum_{i_k=1}^{n_k} f_{i_k}\min\left(\Delta_{i_k j_m}\right) + \sum_{i_m=1}^{n_m} f_{i_m}\min\left(\Delta_{i_m j_k}\right)\right],$$

where f_{i_k} is the relative abundance of OTU i in community k, n_k is the number of OTUs in k, and $\min(\Delta_{i_k j_m})$ is the minimum phylogenetic distance between OTU i in community k and all OTUs j in community $m\min(\Delta_{i_k j_k})$. βMNTD was calculated using the R function comdistnt (abundance.weighted = TRUE; package picante). The value of βMNTD provides a raw estimate of phylogenetic turnover between assemblages; by itself the level of βMNTD does not indicate whether turnover between a pair of communities is the result of stochastic or deterministic processes. To infer the ecological process primarily responsible for observed turnover between a given pair of assemblages, we used a null modeling approach that generates an expected level of βMNTD given a dominance of stochastic ecological factors. Each iteration of the null model randomly shuffles OTUs across the tips of the phylogeny, and βMNTD is thereafter recalculated; this provides one null value for βMNTD (12). Going through 999 iterations of the null model provides a null distribution of βMNTD values. To quantify the magnitude and direction of deviation between an observed βMNTD value and the null βMNTD distribution, we used the β-nearest taxon index (βNTI), calculated as follows:

$$\beta\text{NTI} = \left(\beta\text{MNTD}_{\text{obs}} - \overline{\beta\text{MNTD}_{\text{null}}}\right) \Big/ \text{sd}(\beta\text{MNTD}_{\text{null}}),$$

where $\beta\text{MNTD}_{\text{obs}}$ is observed βMNTD, $\beta\text{MNTD}_{\text{null}}$ are null values of βMNTD, and sd indicates the standard deviation of the $\beta\text{MNTD}_{\text{null}}$ distribution. We quantified βNTI for all pairwise comparisons, using a separate null model for each comparison. Detailed interpretation of βNTI values is provided in *Results*.

Simulation Model.

Regional species pool evolution. We aim to develop a model containing the minimal set of elements necessary to evolve a regional species pool that includes information on species evolutionary relationships and species environmental optima. The purpose is to provide a regional species pool that can be used to assemble local communities under defined ecological scenarios that relate to our conceptual interpretation of the βNTI patterns observed in the empirical data. The use of βNTI to make ecological inferences requires significant phylogenetic signal in species' environmental optima; phylogenetic signal occurs when more closely related species are more ecologically similar (14). The evaluation of phylogenetic signal is discussed below, after the description of the regional species pool simulation model.

To simulate regional species pool evolution we begin with one ancestor species that is defined by the environmental conditions that lead to its maximal growth. We use an arbitrary environmental

gradient that takes on values from 0 to 1. In the model, new species arise asexually through mutations according to the environmental optimum of the ancestor such that new species (like the ancestor) are defined simply by their environmental optimum. For our purposes, there is no reason for fitness to vary across the environmental gradient such that we set fitness to be constant, leading effectively to Brownian evolutionary dynamics in species environmental optima. To facilitate comparisons between simulated and empirical patterns we constrained the number of simulated species to be similar to the number retained following rarefaction of the empirical data. To control the number of species in the regional pool, we imposed the following constraints, which are similar to those in Hurlbert and Stegen (15) that ensure equilibrium species richness: (*i*) We constrained the total number of individuals summed across all species' populations to be 2 million such that population sizes decline as species richness increases; this is the zero-sum assumption (15). (*ii*) We assume that the probability of a species going (stochastically) extinct increases as its population size decreases; this assumption was implemented such that extinction probability decreased following a negative exponential function [population extinction probability $\propto \exp(-0.001 \cdot$ population size)].

As a consequence of these constraints, as species richness increases, population sizes decline, which causes extinction rates to increase. The increase in extinction rates with increasing species richness ultimately leads to a balance between speciation and extinction and thus an equilibrium number of species. Parameters were adjusted so that equilibrium species richness arising in the simulation model (4,748) was very close to the empirically observed species richness (4,746).

In the model, mutation occurs probabilistically and increases with population size. When a mutation occurs, a descendant is produced, and the descendant's environmental optimum deviates from its ancestor's by a quantity randomly sampled from a Gaussian distribution ($\mu = 0$, $\sigma = 0.15$). After a new species arises, all population sizes are adjusted so that all population sizes are equal and so that the total number of individuals is 2 million. Population sizes are equal because there are no fitness differences across the environmental axis. Mutation and extinction occur (or not), and population sizes are adjusted within each time step. The simulation model was run for 250 time steps, which was sufficient to reach equilibrium.

To evaluate phylogenetic signal within the simulated regional species pool, we used a phylogenetic Mantel correlogram as described above. The large number of species within the regional species pool made it computationally unfeasible to evaluate statistical significance within the Mantel correlogram. For this reason we randomly selected 2,500 species from the regional pool for evaluation of statistical significance.

We were specifically concerned with the presence/absence of significant phylogenetic signal across relatively short phylogenetic distances because in natural microbial systems, phylogenetic signal is consistently observed but only across relatively short phylogenetic distances (7–9, 12). Significant phylogenetic signal across short phylogenetic distances is also the pattern observed within our study system (Fig. S1). In the simulated regional species pool we likewise find significant phylogenetic signal across relatively short phylogenetic distances (Fig. S2).

Consistency between simulated and empirically observed phylogenetic signal and species richness in the regional pool suggests that local community assembly from the simulated regional pool will have utility in terms of facilitating conceptual understanding of phylogenetic turnover patterns observed in the natural system. That is, we use local community assembly (next section) from the simulated regional species pool primarily as a device to facilitate understanding; the models are coarse-grain abstractions of the system and are not meant to reflect highly resolved system details.

Local community assembly. We assembled local communities from the regional species pool to evaluate whether our conceptual interpretations of the relationship between βNTI and changes in organic matter concentrations represent reasonable approximations of how ecological selection operates in the natural system. Simulations captured four ecological scenarios that include strong homogeneous selection, weak variable selection, moderate variable selection, and strong variable selection.

For the strong homogeneous selection scenario we assembled 12 communities, which is consistent with the number of communities sampled within each successional stage; homogeneous selection should be strongest within successional stages because organic carbon varies relatively little within (as opposed to between) successional stages. For the other three scenarios we assembled 12 communities within each of two environments meant to represent two successional stages.

The βNTI–organic matter relationship is well approximated by a linear model, and this leads to the hypothesis that the strength of the selective pressure imposed by organic carbon (potentially through both composition and concentration of organic matter) is relatively consistent across organic matter concentrations; nonlinearity in the βNTI–organic matter relationship should arise if the strength of selective pressure imposed by organic matter depends on the absolute value of organic matter concentration. For this reason we set the strength of ecological selection to be consistent across environments.

All communities had 817 species, which is the median OTU richness across samples in the empirical data. To assemble each local community, 10,000 individuals were drawn probabilistically from the regional species pool using ecological rules summarized below. A relatively large number of individuals were used to assemble local communities to generate robust estimates of species' relative abundances; using fewer or more individuals would have generated more or less uncertainty, respectively, in the estimates of species' relative abundances but would not influence species richness.

To assemble each local community in a given environment, 817 species were drawn without replacement from the regional species pool with probabilities proportional to their fitness in that environment; fitness was quantified with a Gaussian function centered on the prevailing environmental condition and with variance of 0.1. To simulate relative abundances, individuals were probabilistically drawn into the selected species until reaching 10,000 individuals; the probability of drawing an individual from a given species into the local community (that is within a given environment) was proportional to the fitness of that species within the environment, as defined by the Gaussian fitness function.

For the scenario of strong homogeneous selection, local communities were assembled within one environment, which took on a value of 0.1; recall that the environmental axis is arbitrary and takes on values from 0 to 1. For the scenarios of weak, moderate, and strong variable selection, local communities were assembled into two environments taking on values of 0.1 and 0.3, 0.1 and 0.5, and 0.1 and 0.6, respectively.

Following community assembly the distributions of βNTI were quantified. For the scenario of strong homogeneous selection, all 12 communities were compared with each other, and the βNTI distribution was generated from the resulting 66 unique pairwise comparisons. For all other scenarios, only between-environment comparisons were used to generate the βNTI distribution, resulting in 144 pairwise comparisons within each βNTI distribution. The distributions were subsequently estimated and visualized using a kernel density function within R (function density within the stats package, using default parameters).

1. Dini-Andreote F, et al. (2014) Dynamics of bacterial community succession in a salt marsh chronosequence: Evidences for temporal niche partitioning. *ISME J* 8(10): 1989–2001.
2. Olff H, De Leeuw J, Bakker JP, Platerink RJ, van Wijnen HJ (1997) Vegetation succession and herbivory in a salt marsh: Changes induced by sea level rise and silt deposition along an elevational gradient. *J Ecol* 85(6):799–814.
3. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
4. Caporaso JG, et al. (2010) PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26(2):266–267.
5. DeSantis TZ, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7):5069–5072.
6. Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26(7):1641–1650.
7. Stegen JC, et al. (2013) Quantifying community assembly processes and identifying features that impose them. *ISME J* 7(11):2069–2079.
8. Wang J, et al. (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: Deterministic versus stochastic processes. *ISME J* 7(7):1310–1321.
9. Lindemann SR, et al. (2013) The epsomitic phototrophic microbial mat of Hot Lake, Washington: Community structural responses to seasonal cycling. *Front Microbiol* 4:323.
10. Andersson AF, Riemann L, Bertilsson S (2010) Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* 4(2):171–181.
11. Pei N, et al. (2011) Exploring tree-habitat associations in a Chinese subtropical forest plot using a molecular phylogeny generated from DNA barcode loci. *PLoS ONE* 6(6): e21273.
12. Stegen JC, Lin X, Konopka AE, Fredrickson JK (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* 6(9):1653–1664.
13. Legendre P, Legendre L (1998) *Numerical Ecology* (Elsevier Science BV, Amsterdam).
14. Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol Lett* 11(10):995–1003.
15. Hurlbert AH, Stegen JC (2014) When should species richness be energy limited, and how would we know? *Ecol Lett* 17(4):401–413.

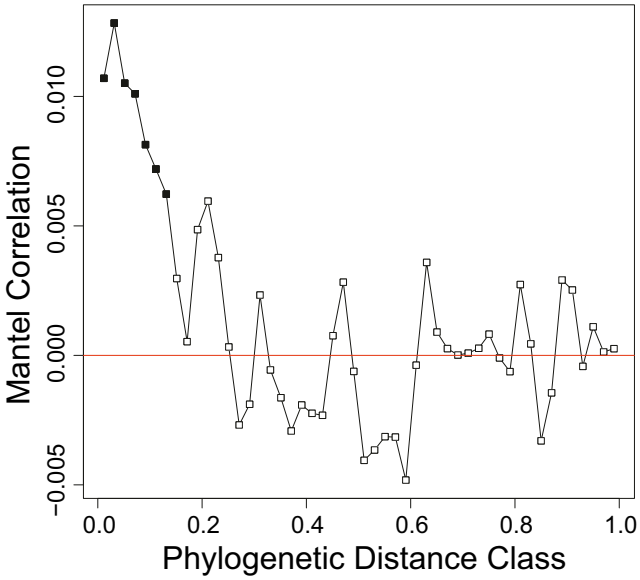

Fig. S1. Phylogenetic Mantel correlogram evaluating phylogenetic signal in the soil microbial communities sampled in this study. The plot relates Pearson correlation coefficients to phylogenetic distances classes. Significant correlations ($P < 0.05$; solid symbols) indicate significant phylogenetic signal in OTU ecological niches but only within the associated phylogenetic distance class. The analysis reveals significant phylogenetic signal but only across relatively short phylogenetic distances.

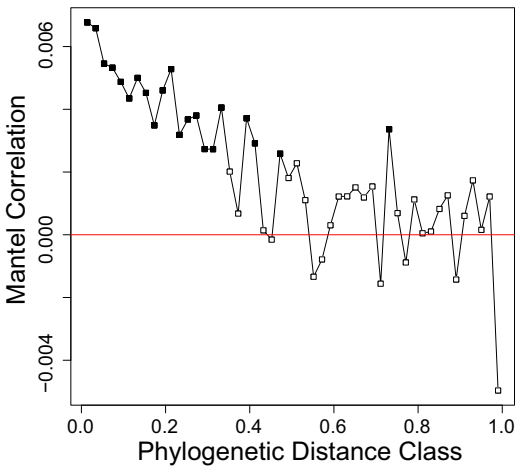

Fig. S2. Phylogenetic Mantel correlogram evaluating phylogenetic signal in the simulated regional species pool. The plot relates Pearson correlation coefficients to phylogenetic distances classes. Significant correlations ($P < 0.05$; solid symbols) indicate significant phylogenetic signal in species ecological niches but only within the associated phylogenetic distance class. The analysis reveals significant phylogenetic signal but primarily across relatively short phylogenetic distances. There is significant phylogenetic signal across somewhat larger phylogenetic distances than the results obtained for the empirical data (Fig. S1), but this difference is not expected to substantially alter inferences drawn from the simulation model.

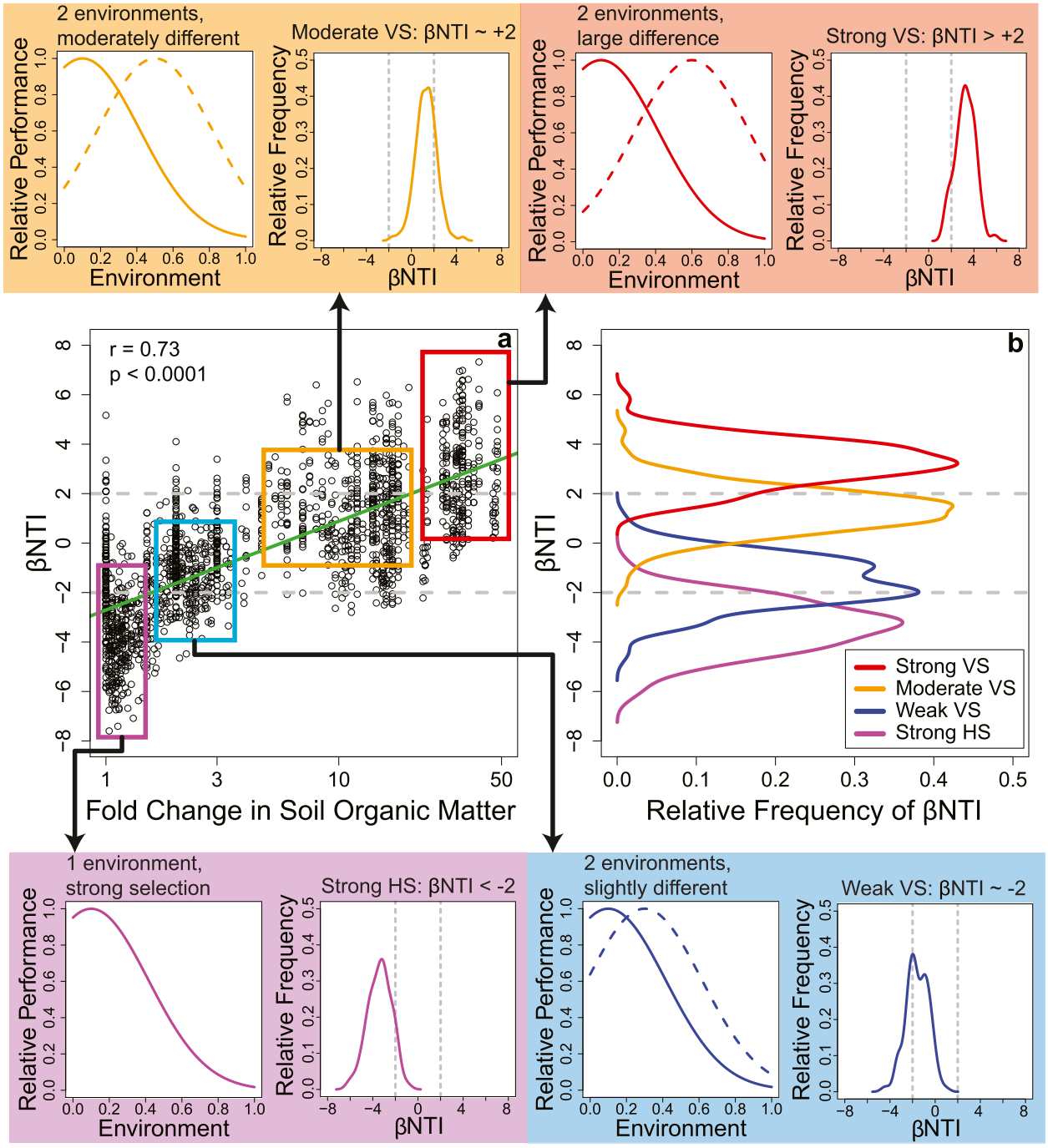

Fig. S3. βNTI patterns from empirical comparisons and simulated ecological scenarios, and linkages to underlying simulation model structure. *A* and *B* are as in Fig. 3. (*B*) The βNTI distribution from simulations of strong HS (purple) aligns with the cluster of empirical βNTI values (*A*) that are centered below −2 and associated with very small changes in SOM, βNTI from weak VS simulations (blue) align with the cluster of empirical βNTI values that are centered near −2 and associated with SOM changes that approach threefold, βNTI from moderate VS simulations (orange) align with the cluster of empirical βNTI values that are centered near +2 and associated with roughly 10- to 20-fold changes in SOM, and βNTI from strong VS simulations (red) align with the cluster of empirical βNTI values centered above +2 and associated with the largest changes in SOM. (*Top* and *Bottom*) Displayed Gaussian functions underlie the ecological simulation model and show the relative performance of an OTU that has a given environmental optimum. In the strong HS scenario (purple), for example, an OTU with an optimal environment near 0.1 will have much higher performance (i.e., higher fitness that translates into higher abundance) relative to an OTU with an environmental optimum near 1.0. Strong HS was achieved with one Gaussian curve such that communities were assembled under a consistent selective pressure. Increasingly strong VS (blue to orange to red) was achieved by increasing the separation between environmental conditions; the increasingly deep valley between performance curves results in increasingly distinct selective environments that select for increasingly distinct ecological communities.