13. Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26,** 867–878 (1999).
14. Schluter, D. & Ricklefs, R. E. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 1–12 (Univ. Chicago Press, Chicago, 1993).
15. Ettema, C. H. & Wardle, D. A. Spatial soil ecology. *Trends Ecol. Evol.* **17,** 177–183 (2002).
16. Callaway, R. M., Thelen, G. C., Rodriguez, A. & Holben, W. E. Soil biota and exotic plant invasion. *Nature* **427,** 731–733 (2004).
17. Horner-Devine, M. C., Carney, K. M. & Bohannan, B. J. M. An ecological perspective on bacterial biodiversity. *Phil. Trans. R. Soc. Lond. B* **271,** 113–122 (2004).
18. Ward, D. M., Ferris, M. J., Nold, S. C. & Bateson, M. M. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* **62,** 1353–1370 (1998).
19. Poulin, R. The decay of similarity with geographical distance in parasite communities of vertebrate hosts. *J. Biogeogr.* **30,** 1609–1615 (2003).
20. Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier, Boston, 1998).
21. Harte, J., McCarthy, S., Taylor, K., Kinzig, A. & Fischer, M. L. Estimating species-area relationships from plot to landscape scale using species spatial-turnover data. *Oikos* **86,** 45–54 (1999).
22. Ranjard, L. *et al.* Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability. *Appl. Environ. Microbiol.* **67,** 4479–4487 (2001).
23. Kent, A. D. & Triplett, E. W. Microbial communities and their interactions in soil and rhizosphere ecosystems. *Annu. Rev. Microbiol.* **56,** 211–236 (2002).
24. Øvreås, L. Population and community level approaches for analysing microbial diversity in natural environments. *Ecol. Lett.* **3,** 236–251 (2000).
25. Finlay, B. J., Esteban, G. V. & Fenchel, T. Protozoan diversity: converging estimates of the global number of free-living ciliate species. *Protist* **149,** 29–37 (1998).
26. Oliver, I. *et al.* Land systems as surrogates for biodiversity in conservation planning. *Ecol. Appl.* **14,** 485–503 (2004).
27. Yeates, C. & Gillings, M. Rapid purification of DNA from soil for molecular biodiversity analysis. *Lett. Appl. Microbiol.* **27,** 49–53 (1998).
28. Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* (Chapman & Hall/CRC, New York, 1997).
29. Oden, N. L. & Sokal, R. R. An investigation of three-matrix permutation tests. *J. Classif.* **9,** 275–290 (1992).
30. Legendre, P. Comparison of permutation methods for the partial correlation and partial Mantel tests. *J. Stat. Comput. Sim.* **67,** 37–73 (2000).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank F. Ayala, E. Berlow, B. Bohannan, R. Condit, J. Harte, A. Hastings, C. Horner-Devine, J. Hughes, N. Martinez and I. Wright for comments on the manuscript, and U. Malvadkar, M. Holyoak and B. Melbourne for discussions. The authors are grateful to M. Holley and M. Raison for technical assistance with the ARISA analyses. The project was funded by the Australian Research Council and the Resource and Conservation Assessment Council of the NSW Government. J.L.G. acknowledges the NSF Postdoctoral Fellowship Program in Biological Informatics for financial support.

Authors' contributions All authors contributed intellectual input to the project. The original concept, sample design and collections were carried out by A.J.H., M.W., I.O., D.B., M.D., M.G. and A.J.B. The first three authors (J.L.G., A.J.H. and M.W.) took the lead in the analysis and writing up of this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L.G. (jgreen@ucmerced.edu).

A taxa–area relationship for bacteria

M. Claire Horner-Devine[1]*, Melissa Lage[2]*, Jennifer B. Hughes[2] & Brendan J. M. Bohannan[1]

[1]*Department of Biological Sciences, Stanford University, Stanford, California 94305, USA*
[2]*Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA*

* These authors contributed equally to this work

A positive power-law relationship between the number of species in an area and the size of that area has been observed repeatedly in plant and animal communities[1]. This species–area relationship, thought to be one of the few laws in ecology[2], is fundamental to our understanding of the distribution of global biodiversity. However, such a relationship has not been reported for bacteria, and little is known regarding the spatial distribution of bacteria, relative to what is known of plants and animals[3]. Here we describe a taxa–area relationship for bacteria over a scale of centimetres to hundreds of metres in salt marsh sediments. We found that bacterial communities located close together were more similar in composition than communities located farther apart, and we used the decay of community similarity with distance to show that bacteria can exhibit a taxa–area relationship. This relationship was driven primarily by environmental heterogeneity rather than geographic distance or plant composition.

In the 1920s, the empirical relationship between the number of species and area was generalized[4,5] as a power-law, $S = cA^z$, where S is the number of species, A is the area sampled and c is the intercept in log–log space. The species–area exponent, z, is a measure of the rate of change of the slope with increasing area, that is, the rate of turnover of species across space. Variation in the values for c, and especially for z, is of interest because it may indicate that different processes underlie the species–area relationship at different spatial scales[6,7]. Although not as well studied as species–area relationships, other taxa–area relationships (for example, genera–area and family–area) have been identified for plants and animals; such relationships conform to the same power-law as species–area relationships, although they differ in their values of c and z[8,9].

Bacteria are among the most abundant and diverse groups of organisms on earth[10] and mediate important ecosystem processes, including trace gas emissions, decomposition and nitrogen cycling. Whereas taxa–area relationships have been observed repeatedly for numerous plant and animal taxa regardless of ecosystem type[1], they have not been explicitly examined for bacteria. Unique aspects of bacterial biology may prevent bacteria from exhibiting taxa–area relationships. For example, if most bacteria are not dispersal limited (for example, owing to small size and environmental hardiness)[11] and if they exhibit a high degree of ecological redundancy (for example, if bacteria are flexible in habitat requirements and physiological abilities, or if they can easily obtain traits through horizontal gene transfer that are necessary for survival in a given habitat), then one would not expect to observe a taxa–area relationship[3].

Here we investigated whether bacteria exhibit a taxa–area relationship in a New England salt marsh. We conducted our work in a salt marsh because the spatial ecology of salt marshes is especially well understood[12]. There is an extensive literature regarding the main physical gradients in salt marshes, the spatial distribution of plant species and the ecological processes that underlie this distribution. This information provides an ideal reference point from which to investigate the spatial distribution of bacteria. We sampled 1-cm-diameter sediment cores in a nested manner over a scale of centimetres to hundreds of metres. With the possible exception of the most extreme and depauperate environments[13], the diversity of bacterial communities is too high to be exhaustively sampled. Therefore we used a previously refined distance decay approach[14], which uses data on the spatial turnover of taxa, to determine the taxa–area exponent, z. This approach uses comparisons of community composition rather than richness estimations to describe taxa–area relationships. For comparison, we also estimated the relationship between the number of plant species and area in this ecosystem, using the same distance decay approach.

Because a large proportion of microbes cannot be cultured with current laboratory techniques[15], bacterial taxa are often identified from the sequences of indicator genes extracted from environmental samples[16]. We determined the bacterial community composition of our salt marsh samples by amplifying via the polymerase chain reaction (PCR), cloning and sequencing a region of 16S ribosomal DNA (rDNA), the most commonly used indicator gene for bacterial biodiversity. Because the bacterial diversity of salt marshes is often very high, we used PCR primers targeting a subset of the bacterial

biota (the β-proteobacteria and relatives) to constrain the potential community we sampled. Proteobacteria are commonly found in many different environments, including salt marshes, and are often numerically dominant[17]. Because there is no single best definition of 'species' using this sequencing approach[18], taxa are usually defined as operational taxonomic units (OTUs) based on sequence similarity groupings. We used the three most commonly used groupings—95%, 97% and 99% sequence similarity—to define OTUs in our study. Using multiple OTU definitions is analogous to comparing different taxonomic resolutions (for example, comparing genus, species and sub-species)[16].

We sequenced a total of 945 partial 16S rDNA sequences. These sequences grouped into 88 OTUs, using a taxon resolution of 97% sequence similarity. Approximately 34% of the OTUs were singletons ($n = 30$), and 15 OTUs were represented by more than ten sequences. Five-hundred and twenty-three sequences were members of the β-proteobacteria; the remainder were members of the γ-proteobacteria and the δ-proteobacteria.

Bootstrapping of linear regressions between log-transformed values of bacterial similarity and geographic distance revealed significant distance decay curves for all taxonomic resolutions; that is, samples that were located closer in space were significantly more similar in bacterial composition than samples that were located farther apart (Table 1; see Supplementary Methods 2). We then computed the taxa–area *z*-value for bacteria from the slope of these distance decay curves (Table 1; Fig. 1).

We also observed a significant species–area relationship for plants in this same marsh (Fig. 1a). The plant *z*-value was significantly larger than those observed for bacteria but was similar to values estimated for other wetland plant communities[19]. To determine whether taxonomic focus (the particular group of taxa analysed) also affects the taxa–area relationship among different bacterial taxa, we repeated the analyses above for a subgroup of the sequences, restricting our analyses to the β-proteobacteria (Table 1). The *z*-value for the 99% sequence similarity grouping of the β-proteobacteria was significantly lower than that of the entire group of sequences at the 99% resolution (Fig. 1b). This suggests that the turnover in space of β-proteobacteria was lower than that of the other bacteria we sampled.

The *z*-value also varied by taxonomic resolution, increasing with increased taxonomic resolution for bacteria. The estimated *z*-value for all sequences at the 99% resolution was significantly larger than those estimated at the 97% and 95% resolutions. Similarly, the *z*-value at the 99% resolution for β-proteobacteria had a significantly positive *z*-value, whereas the *z*-values at the 97% and 95% resolutions were not significantly different from zero (Table 1; Fig. 1b). A similar increase in *z*-values with increasing taxonomic resolution has been observed for plants[8] and animals[9] (that is, from family to genera to species). We expect that the *z*-value will continue to change with taxonomic resolution at both higher and lower resolutions. We did not present the 100% sequence similarity OTUs because of the possibility that even minor PCR and/or sequencing errors could result in artefactual OTUs at this resolution. The taxonomic resolutions we used probably include a greater diversity of ecological types than contained within an animal or plant species. However, the presence of a taxa–area relationship for bacteria at this relatively coarse resolution suggests that a taxa–area relationship would probably also be present at finer resolutions, resolutions that may more closely correspond to the ecological breadth of animal and plant species.

Environmental heterogeneity often increases with increased area. The increase in heterogeneity with area together with the specificity of taxa for different habitats is the most common explanation of taxa–area patterns[1], especially at scales where dispersal is not limiting to the distribution of taxa. Environmental heterogeneity may also underlie the bacterial taxa–area relationship observed here. When we removed the effect of geographic distance, partial Mantel tests showed that sites that were similar in environmental charac-

Table 1 **Taxa–area *z*-values**

OTU resolution	*z*-value	Regression coefficient	*t*	*P*
All sequences				
99%	0.040	−0.081	−6.53	<0.001*
97%	0.020	−0.039	−4.92	<0.001*
95%	0.019	−0.037	−5.44	<0.001*
β-proteobacteria				
99%	0.019	−0.038	−2.40	0.013*
97%	0.048	−0.096	−1.14	0.396
95%	0.008	−0.015	−1.57	0.177

The *z*-values shown were determined using the distance decay approach. Asterisks denote regressions that were significant at the $P < 0.05$ level. The *z*-value was determined as −(regression coefficient)/2. Percentages refer to the 16S rDNA sequence similarities assumed when defining taxonomic units.

Table 2 **The influence of geographic distance and habitat heterogeneity on bacterial community composition**

Effect of: / Controlling for:	Env. simil. / Geog. dist.		Geog. dist. / Env. simil.		Plant simil. / Env. simil.	
	r	*P*	*r*	*P*	*r*	*P*
All sequences						
99%	−0.36	0.0002*	−0.15	0.11	−0.069	0.17
97%	−0.37	0.0001*	−0.0089	0.46	−0.10	0.080
95%	−0.36	0.0002*	−0.056	0.27	−0.085	0.13
β-proteobacteria						
99%	−0.32	0.0004*	0.012	0.47	−0.043	0.27
97%	−0.29	0.0004*	0.11	0.11	−0.022	0.35
95%	−0.26	0.0009*	0.082	0.17	0.0014	0.51

Partial Mantel tests showed that environmental heterogeneity may underlie the bacterial taxa–area relationship. Sites that were similar in environmental characteristics (env. simil.) were similar in bacterial composition when we removed the effect of geographic distance (geog. dist.) and plant community composition (plant simil.). *r* represents the Mantel statistic, and asterisks denote significance at the $P < 0.05$ level (9,999 permutations).

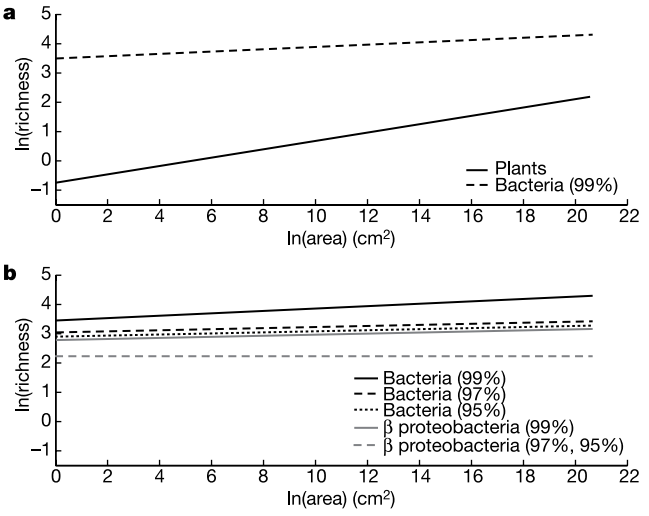

Figure 1 The taxa–area relationship for salt marsh organisms varied with taxonomic focus (**a**) and taxonomic resolution (**b**). **a**, Plants and bacteria both exhibited significant taxa–area relationships, but the *z*-value for plants ($z = 0.103$) was significantly greater than the *z*-value for bacteria ($z = 0.040$; coeff. $= 0.0198$, $t = 159.51$, $P = 0.001$). **b**, Both taxonomic resolution and taxonomic focus influenced the taxa–area relationship for salt marsh bacteria. Lower resolution OTUs (that is, those defined using the criteria of 97% or 95% sequence similarity) had a lower slope ($z = 0.020$ and $z = 0.019$, respectively; coeff. $= -0.041$, $t = -3.32$, $P = 0.0008$) than the higher resolution OTUs (99% sequence similarity). The β-proteobacteria only exhibited a significant taxa–area relationship for the 99% resolution (grey solid line, $z = 0.019$).

teristics were also similar in bacterial composition (Table 2). In contrast, there was no effect of geographic distance on community similarity when we removed the effect of environmental similarity (Table 2). Although there was a significant relationship between environmental similarity and plant community composition ($r = 0.36$, $P < 0.001$), there was no relationship between plant similarity and bacterial similarity independent of environmental similarity (Table 2).

The *z*-values for bacteria reported here are among the lowest *z*-values reported for any organisms (Fig. 2), suggesting that turnover of bacterial taxa at these spatial and taxonomic scales may be lower than that of most other organisms. Bacterial *z*-values may be low, in part, because it is unlikely that bacteria are dispersal limited within a salt marsh. Bacteria can probably disperse easily by air and tidal waters at this scale, and indeed we found no evidence for an effect of geographic distance on bacterial community composition when controlling for environmental heterogeneity and plant community composition. However, it is unlikely that the plants, which have a *z*-value at least two times higher than the bacteria, are dispersal limited at these spatial scales either, suggesting that dispersal alone cannot be responsible for the low bacterial *z*-values. We can think of at least three other potential explanations. First, the levels of taxonomic resolution we examined for bacteria probably reflect a broader ecological breadth than the plant species units, and thus the lower *z*-values for bacteria may be a reflection of taxonomic resolution. In other words, if we could define bacterial OTUs in terms that were equivalent to the ecological breadth of a typical plant species, then perhaps the bacterial species–area curve would have a *z*-value similar to that observed for many macroorganisms. Second, although environmental factors were related to community composition, the bacteria we sampled could have low habitat specificity (relative to the plants), reducing the spatial turnover of taxa relative to the plants in this system. Third, horizontal transfer of ecologically relevant genes could uncouple the relationship between phylotypes (phylogenetically distinct groups) and ecotypes (ecologically distinct groups), resulting in a lower *z*-value[3].

We saw no evidence that the bacterial *z*-value was scale-dependent (that is, the slope of the distance decay curve did not vary with distance). However, it is possible that the *z*-value may increase at larger spatial scales, as has been observed for plants[6], or even at smaller scales, where microscale environmental heterogeneity can be very high[3]. Indeed, recent reports demonstrate that dispersal limitation is possible at large spatial scales for some microbes[20] (although not for all[21]), and this would tend to increase the *z*-value.

For over two-hundred years, ecologists have repeatedly documented that there is a relationship between the number of eukaryotic species found in a given area and the size of the area. Our study demonstrates that both prokaryotes and eukaryotes can exhibit taxa–area relationships, although the quantitative form of the relationship may differ. The taxa–area relationship seems to be a universal law. □

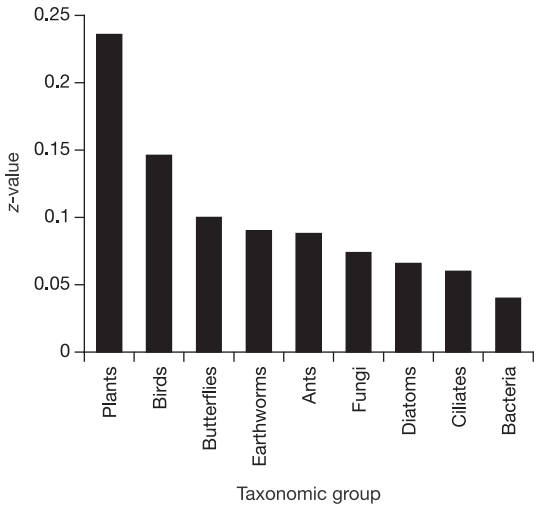

Figure 2 A comparison of *z*-values for both microbial and macrobial taxa in different ecosystems. The bacterial *z*-values observed in this salt marsh system represent some of the lowest values observed so far. Representative *z*-values were selected from a review of taxa–area relationships (see Supplementary Methods 3)

Methods

Sampling

We collected samples from a salt marsh on Prudence Island, Rhode Island in July 2002. Twenty-six sediment samples for bacterial community composition were collected from a nested grid that ranged from 300 × 300 m to 0.03 × 0.03 m (see Supplementary Methods 1).

We sampled plant community composition in two ways. First, we sampled along three transects at the 300-m and 30-m scales, with transects spaced at 100-m and 10-m intervals, respectively. We recorded the presence and per cent cover of each species in 1 × 1 m quadrants every 20 and 10 m, respectively, and used these data to determine the plant species–area relationship. Second, we recorded similar data in a 10 × 10 cm quadrant at each bacterial sampling point and used these data to examine the influence of plant composition on bacterial community composition.

We collected porewater samples adjacent to each sediment core to measure salinity, pH and nutrient concentrations (phosphate, ammonium and sulphate). Water was filtered with a 40-micron filter immediately after collection. Sulphate, ammonia and phosphate were measured in the field using Hach portable colorimeters (http://www.hach.com). Salinity was measured with a refractometer and pH with a calibrated probe.

DNA analyses

DNA was extracted from the top 0.5 g of each sediment core using a combination of phenol/chloroform extraction and a MoBio Ultraclean Soil DNA Kit (MoBio Laboratories). We chose primers (βAMOf and βAMOr) that amplify 16S rDNA from a subset of the domain Bacteria, as it is not tractable to sufficiently sample bacteria in our samples using domain level primers[22]. We reduced PCR biases by limiting the PCR cycles to twenty-five and included BSA in the reaction mix[23]. Each cycle consisted of 30 s at 94 °C, 30 s at 57 °C and 90 s at 72 °C. Gel-extracted and purified amplicons were cloned using the TOPO-TA cloning kit for sequencing (Invitrogen). We used an ABI 377 automated DNA sequencer to determine the sequence of the 5′ terminal 600 nucleotides of 945 of the cloned rDNA amplicons.

Phylogenetic analysis

We used the RDP database[24] and ARB software[25] to align the rDNA sequences from our 26 clone libraries. Ambiguously and incorrectly aligned positions were aligned manually on the basis of conserved primary and secondary structure. We identified and excluded potential chimaeras using the Chimera_Check program of the RDP[24] and using ARB to compare trees generated from the 5′-end versus the 3′-end sequences separated at the break point suggested by Chimera_Check. Similarity matrices were generated using 510 unambiguously aligned positions. We grouped sequences into OTUs on the basis of rDNA sequence similarity, using DOTUR[26]. We used the three most commonly used groupings—95%, 97% and 99% sequence similarity—to define OTUs in our study and to examine how the taxa–area relationship varies with taxonomic resolution.

Community similarity

First, we calculated turnover using the Sorensen index, for use in estimating the *z*-values[14]. Second, we calculated the Bray–Curtis similarity coefficient (S_{BC}) for each pair of samples, for use in the Mantel tests[27]. We calculated S_{BC} on square-root-transformed data to decrease the influence of highly dominant sequences, because the most dominant sequences in a clone library may not be the most active or dominant types in the actual community, owing to primer bias[28].

To control for unequal sampling (numbers of sequences) between sediment cores, we used a form of community rarefaction when calculating the pairwise community similarity indices. We randomly sampled the lowest number of sequences found at any point from each core and calculated the similarity values between all cores. We then repeated this randomization 1,000 times to get a rarefied community similarity matrix.

Taxa–area relationships

We used linear regression to examine the relationship between geographic distance between samples and similarity in bacterial composition. Because our data consisted of pairwise comparisons and thus were not independent, we used bootstrapping (10,000 replications) to test if the slope of the regression was significantly different from zero (see Supplementary Methods 2). We estimated the power-law exponent *z* with a distance decay approach[14], using the equation $\log(S_S) = \text{constant} - z\log(D)$, where S_S is pairwise similarity in community composition and D is distance between two samples. Because

some similarity values were equal to zero (that is, no OTUs in common), we coded the similarity data by adding 0.01 before log transforming each value[29]. The approach outlined in ref. 14 allowed us to use relative comparisons of bacterial community composition rather than richness to examine the taxa–area relationship; richness is very difficult to estimate accurately in hyperdiverse communities such as bacterial communities. There is no reason to believe that undersampling and/or PCR biases will co-vary with intersample distance or will result in preferential sampling of those taxa most likely to be shared among samples located close together in space; thus, these factors, although likely to be present, are unlikely to influence the *z*-values we observed. In addition, we can think of no model in which PCR biases and/or undersampling could generate a taxa–area relationship that was completely artefactual.

We compared the *z*-values for different taxon definitions by testing if the slopes of the regressions differed (see Supplementary Methods 2). We used the same distance decay approach to determine the *z*-value for plants, using data from the transect quadrants.

Influence of habitat heterogeneity

We used partial Mantel tests (9,999 permutations) to examine the influence of abiotic factors and aboveground plant composition on bacterial community composition, while holding geographic distance constant and vice versa.

We constructed a distance matrix for plant community composition from per cent cover estimates for the four dominant species (*Spartina alterniflora*, *Spartina patens*, *Salicornia virginica* and *Limonium nashii*). We constructed a matrix of environmental distance from the abiotic factors identified as most important to community composition (phosphate and ammonia concentrations), using BIO-ENV[28]. The BIO-ENV procedure selects a subset of available abiotic variables to maximize rank correlation between community similarity and abiotic dissimilarity matrices. We then used these matrices to test for additional distance and plant effects.

Received 10 August; accepted 29 September 2004; doi:10.1038/nature03073.

1. Rosenzweig, M. L. *Species Diversity in Space and Time* 8–48, 190–284 (Cambridge Univ. Press, Cambridge, 1995).
2. Lawton, J. H. Are there general laws in ecology? *Oikos* **84,** 177–192 (1999).
3. Horner-Devine, M. C., Carney, K. M. & Bohannan, B. J. M. An ecological perspective on bacterial biodiversity. *Proc. R. Soc. Biol. Sci. B* **271,** 113–122 (2004).
4. Arrhenius, O. Species and area. *J. Ecol.* **9,** 59–99 (1921).
5. Gleason, H. A. On the relationship between species and area. *Ecology* **3,** 158–162 (1922).
6. Crawley, M. J. & Harral, J. E. Scale dependence in plant biodiversity. *Science* **291,** 864–868 (2001).
7. Losos, J. B. Analysis of an evolutionary species–area relationship. *Nature* **408,** 847–850 (2000).
8. Bennett, J. P. Nested taxa–area curves for eastern United States floras. *Rhodora* **99,** 241–251 (1997).
9. Harcourt, A. H. Biogeographic relationships of primates on South-East Asian islands. *Glob. Ecol. Biogeogr.* **8,** 55–61 (1999).
10. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proc. Natl Acad. Sci. USA* **95,** 6578–6583 (1998).
11. Baas-Becking, L. G. M. *Geologie of Inleiding Tot de Milieukunde* (W. P. Van Stockum & N. V. Zoon, The Hague, The Netherlands, 1934).
12. Bertness, M. D. *The Ecology of Atlantic Shorelines* 313–376 (Sinauer Associates, Sunderland, Massachusetts, 1999).
13. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428,** 37–43 (2004).
14. Harte, J., McCarthy, S., Taylor, K., Kinzig, A. & Fischer, M. L. Estimating species–area relationships from plot to landscape scale using species spatial-turnover data. *Oikos* **86,** 45–54 (1999).
15. Torsvik, V., Goksoyr, J. & Daae, F. L. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56,** 782–787 (1990).
16. Stackebrandt, E. & Goebel, B. M. Taxonomic note: A place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44,** 846–849 (1994).
17. Burke, D., Hamerlynck, E. & Hahn, D. Interactions among plant species and microorganisms in salt marsh sediments. *Appl. Environ. Microbiol.* **68,** 1157–1164 (2002).
18. Rossello-Mora, R. & Amann, R. The species concept for prokayotes. *FEMS Microbiol. Rev.* **25,** 39–67 (2001).
19. Peitinger, M., Bergamini, A. & Schmid, B. Species–area relationships and nestedness of four taxonomic groups in fragmented wetlands. *Basic Appl. Ecol.* **4,** 385–394 (2003).
20. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archea. *Science* **301,** 976–978 (2003).
21. Roberts, M. S. & Cohan, F. M. Recombination and migration rates in natural-populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* **49,** 1081–1094 (1995).
22. McCaig, A. E., Embley, T. M. & Prosser, J. I. Molecular analysis of enrichment cultures of marine ammonia oxidizers. *FEMS Microbiol. Lett.* **120,** 363–367 (1994).
23. Qiu, X. *et al.* Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl. Environ. Microbiol.* **67,** 880–887 (2001).
24. Maidak, B. L. *et al.* The RDP-II (Ribosomal Database Project). *Nucleic Acids Res.* **29,** 173–174 (2001).
25. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
26. Schloss, P. D. & Handelsman, J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. ⟨http://www.plantpath.wisc.edu/fac/joh/dotur.html⟩ *Appl. Environ. Microbiol.* (in the press).
27. Magurran, A. E. *Measuring Biological Diversity* 175–176 (Blackwell, 2004).
28. Clarke, K. R. & Warwick, R. M. *Change in Marine Communities: an Approach to Statistical Analysis and Interpretation* 2.3, 11.8–11.11 (PRIMER-E, Plymouth Marine Laboratory, 2001).
29. Legendre, P. & Legendre, L. *Numerical Ecology* 33–47 (Elsevier Science B. V., Amsterdam, 1998).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We are grateful to C. Anderson, H. P. Horz, A. Martiny, S. Reddy, members of M. Bertness' laboratory at Brown University and K. Nusslein's laboratory at the University of Massachusetts, Amherst for their technical assistance. We thank J. Green, D. Ackerly, P. Ehrlich, D. Relman and D. Petrov for comments on a previous draft of this manuscript. We also thank E. Bathgate, the American Association of University Women, and the National Science Foundation for their support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.H.-D. (mchd@u.washington.edu).

Species-typical songs in white-crowned sparrows tutored with only phrase pairs

Gary J. Rose, Franz Goller, Howard J. Gritton, Stephanie L. Plamondon, Alexander T. Baugh & Brenton G. Cooper

Department of Biology, University of Utah, Salt Lake City, Utah 84112-0840, USA

Modern theories of learned vocal behaviours, such as human speech and singing in songbirds[1], posit that acoustic communication signals are reproduced from memory, using auditory feedback[2]. The nature of these memories, however, is unclear. Here we propose and test a model for how complex song structure can emerge from sparse sequence information acquired during tutoring. In this conceptual model, a population of combination-sensitive (phrase-pair) detectors is shaped by early exposure to song and serves as the minimal representation of the template necessary for generating complete song. As predicted by the model, birds that were tutored with only pairs of normally adjacent song phrases were able to assemble full songs in which phrases were placed in the correct order; birds that were tutored with reverse-ordered phrase pairs sang songs with reversed phrase order. Birds that were tutored with all song phrases, but presented singly, failed to produce normal, full songs. These findings provide the first evidence for a minimal requirement of sequence information in the acoustic model that can give rise to correct song structure.

Songbirds must hear conspecific song during a 'sensitive period' early in life to later generate a normal song; birds fail to produce normal song if raised from the egg or the nestling stage, in acoustic isolation[3,4]. From this early acoustic experience, birds form a memory (acquired template) of the song[5,6]. Later, during the sensorimotor phase of song development, birds use auditory feedback to compare their vocalizations with the memorized representation of the tutor song[2]. The neural representation of this acquired template is largely unknown.

A central question with regard to the nature of the acquired template is whether a representation of the full song (Fig. 1a) is required for assembling normal song. Here we propose and test a model of how correct song structure could emerge even if the full song was never experienced. Theoretically, correct phrase order could be deduced if only information about the temporal order of adjacent phrases is represented separately in the tutor model (as phrase pairs, Fig. 1b, c), but not when birds are tutored with phrase types presented in temporal isolation. In support of the latter, Soha and Marler[7] found that white-crowned sparrows (*Zonotrichia leucophrys*) tutored with multiple models, each consisting of repe-