5 min, and four sperm counts were performed using a haemocytometer. The average of the four sperm counts was used to estimate the total number of sperm ejaculated by the male (or sperm investment).

Mouse bedding context

Ten male and ten female meadow voles were used in the second part of the study. The ten males were not used as focal or donor males in the first part of the study. Procedures were similar to both the control and RSC contexts, except for the use of 20 g of soiled bedding taken from the cage of a sexually active mouse. Donor mice were individually caged. A different mouse donor was used in each trial.

1. Birkhead, T. R. & Møller, A. P. (eds) *Sperm Competition and Sexual Selection* (Academic, San Diego, 1998).
2. Smith, R. L. (ed.) *Sperm Competition and the Evolution of Animal Mating Systems* (Academic, Orlando, 1984).
3. Wedell, N., Gage, M. J. G. & Parker, G. A. Sperm competition, male prudence and sperm-limited females. *Trends Ecol. Evol.* **17,** 313–320 (2002).
4. Parker, G. A., Ball, M. A., Stockley, P. & Gage, M. J. G. Sperm competition games: a prospective analysis of risk assessment. *Proc. R. Soc. Lond. B* **264,** 1793–1802 (1997).
5. Dewsbury, D. A. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 547–571 (Academic, New York, 1984).
6. Dziuk, P. J. Factors that influence the proportion of offspring sired by a male following heterospermic insemination. *Anim. Reprod. Sci.* **43,** 65–88 (1996).
7. Ginsberg, J. R. & Huck, U. W. Sperm competition in mammals. *Trends Ecol. Evol.* **4,** 74–79 (1989).
8. Schwagmeyer, P. L. & Foltz, D. W. Factors affecting the outcome of sperm competition in thirteen-lined ground squirrels. *Anim. Behav.* **39,** 156–162 (1990).
9. Wedell, N. & Cook, P. A. Butterflies tailor their ejaculate in response to sperm competition risk and intensity. *Proc. R. Soc. Lond. B* **266,** 1033–1039 (1999).
10. Gage, M. J. G. Risk of sperm competition directly affects ejaculate size in the Mediterranean fruit fly. *Anim. Behav.* **42,** 1036–1037 (1991).
11. Pilastro, A., Scaggiante, M. & Rasotto, M. B. Individual adjustment of sperm expenditure accords with sperm competition theory. *Proc. Natl Acad. Sci. USA* **99,** 9913–9915 (2002).
12. Pizzari, T., Cornwallis, C. K., Lovlie, H., Jakobsoon, S. & Birkhead, T. R. Sophisticated sperm allocation in male fowl. *Nature* **426,** 70–74 (2003).
13. Baker, R. R. & Bellis, M. A. *Human Sperm Competition* 203–227 (Chapman & Hall, New York, 1995).
14. Bellis, M. A., Baker, R. R. & Gage, M. J. G. Variation in rat ejaculates consistent with the Kamikaze-Sperm hypothesis. *J. Mamm.* **71,** 479–480 (1990).
15. Lezama, V., Orihuela, A. & Angulo, R. Sexual behavior and semen characteristics of rams exposed to their own semen or semen from a different ram on the vulva of the ewe. *Appl. Anim. Behav. Sci.* **75,** 55–60 (2001).
16. Dewsbury, D. A. Ejaculate cost and male choice. *Am. Nat.* **119,** 601–610 (1982).
17. Johnston, R. E. in *Pheromones and Reproduction in Mammals* (ed. Vandenbergh, J. G.) 3–37 (Academic, New York, 1983).
18. Brown, R. E. & Macdonald, D. W. (eds) *Social Odours in Mammals* 1–36 (Oxford Univ. Press, Oxford, 1985).
19. Ferkin, M. H. & Johnston, R. E. Meadow voles, *Microtus pennsylvanicus*, use multiple sources of scent for sex recognition. *Anim. Behav.* **49,** 37–44 (1995).
20. Boonstra, R., Xia, X. & Pavone, L. Mating system of the meadow vole, *Microtus pennsylvanicus. Behav. Ecol.* **4,** 83–89 (1993).
21. Dewsbury, D. A. Patterns of copulatory behavior in male mammals. *Q. Rev. Biol.* **47,** 1–33 (1972).
22. Dewsbury, D. A. Diversity and adaptation in rodent copulatory behavior. *Science* **190,** 947–954 (1975).
23. Gray, G. D. & Dewsbury, D. A. A quantitative description of the copulation behaviour of meadow voles (*Microtus pennsylvanicus*). *Anim. Behav.* **23,** 261–267 (1975).
24. Pound, N. Effects of morphine on electrically evoked contractions of the vas deferens in two congeneric rodent species differing in sperm competition intensity. *Proc. R. Soc. Lond. B* **266,** 1755–1758 (1999).
25. Parker, G. A. Sperm competition games: raffles and roles. *Proc. R. Soc. Lond. B* **242,** 120–126 (1990).
26. Birkhead, T. R. & Møller, A. P. *Sperm Competition in Birds: Evolutionary Causes and Consequences* (Academic, London, 1992).
27. Simmons, L. W. *Sperm Competition and its Evolutionary Consequences in the Insects* 144–187 (Princeton Univ. Press, Princeton, 2001).
28. Robb, G. W., Amann, R. P. & Killian, G. J. Daily sperm production and epididymal sperm reserves of pubertal and adult rats. *J. Reprod. Fertil.* **54,** 103–107 (1978).
29. Ferkin, M. H., Lee, D. N. & Leonard, S. T. The reproductive state of female voles affects their scent marking behavior and the responses of male conspecifics to such marks. *Ethology* **110,** 257–272 (2004).
30. Dewsbury, D. A. & Baumgardner, D. J. Studies of sperm competition in two species of muroid rodents. *Behav. Ecol. Sociobiol.* **9,** 121–133 (1981).

Theory predicts the uneven distribution of genetic diversity within species

Erik M. Rauch[1,2] **& Yaneer Bar-Yam**[2]

[1]*MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA*
[2]*New England Complex Systems Institute, 24 Mount Auburn Street, Cambridge, Massachusetts 02138, USA*

Global efforts to conserve species have been strongly influenced by the heterogeneous distribution of species diversity across the Earth. This is manifest in conservation efforts focused on diversity hotspots[1–3]. The conservation of genetic diversity within an individual species[4,5] is an important factor in its survival in the face of environmental changes and disease[6,7]. Here we show that diversity within species is also distributed unevenly. Using simple genealogical models, we show that genetic distinctiveness has a scale-free power law distribution. This property implies that a disproportionate fraction of the diversity is concentrated in small sub-populations, even when the population is well-mixed. Small groups are of such importance to overall population diversity that even without extrinsic perturbations, there are large fluctuations in diversity owing to extinctions of these small groups. We also show that diversity can be geographically non-uniform— potentially including sharp boundaries between distantly related organisms—without extrinsic causes such as barriers to gene flow or past migration events. We obtained these results by studying the fundamental scaling properties of genealogical trees. Our theoretical results agree with field data from global samples of *Pseudomonas* bacteria. Contrary to previous studies[8], our results imply that diversity loss owing to severe extinction events is high, and focusing conservation efforts on highly distinctive groups can save much of the diversity.

Our approach is to use simulations and analytic studies of the genealogical tree of a population, a method known as coalescent theory[9–12]. We test the robustness of our results in part by varying the model details but primarily by obtaining analytic results (see Box 1). For studies of well-mixed populations, we use the Wright–Fisher model[13,14]: each organism is the offspring of a randomly chosen parent of the previous generation. For spatial populations (Fig. 1), the parent of an organism is either the previous organism at that site, or one of the neighbours on a square lattice with equal probability for all possible parents. Each of the parent–child links is an opportunity for genetic mutation. Thus, the expected genetic distance between two individuals is proportional to the time to their common ancestor (their relatedness), and the expected total diversity is proportional to the number of links traced back from the current population to the common ancestor. (If the rate of mutation is sufficiently high that multiple mutations can occur at a single locus within a genealogical tree, a correction is necessary: the total diversity is $D(B) \approx (\mu/\mu_L)\,(1 - \exp(-\mu_L B))$, where B is the total branch length, μ is the per-genome mutation rate and μ_L is the probability of a particular mutation at a particular locus.) The total diversity has been characterized for well-mixed populations[15]. Here we study the way that the diversity is distributed within populations.

We first consider the genetic uniqueness of an individual or group; that is, the degree to which an individual or group is distinctive from all other individuals in the population, for example, the minimal time to an ancestor common with any other individual. The average of this quantity for well-mixed populations has been calculated[14,16], but its distribution has not been studied. As can be

seen from the genealogy, some of the genetic diversity in the population is found only in particular individuals; other diversity is common to smaller or larger groups that share a common ancestor. We find that uniqueness of individuals has a scale-free (power law) distribution (Fig. 2a). This property is highly robust and is valid for both well-mixed and spatial populations (exponents -2.93 and -2.85 (dashed line), respectively); however, spatial populations contain more highly distinctive individuals and groups, as the tail of the distribution extends farther.

This distribution also applies to the uniqueness of groups composed of genetically related individuals (see Box 1). Sexual reproduction increases the likelihood of moderate values of uniqueness in relation to high values. However, the power law tail remains, so that highly distinctive individuals exist even in a sexually reproducing population (Fig. 2b). The distribution for well-mixed sexual populations is similar to that of spatial sexual populations, although the shorter tail enhances the effect of averaging. Incorporation of spatially uniform selection into the model has little effect unless there are mutations that are advantageous enough to sweep the population in a short time (periodic selection or genetic hitch-hiking). This results in a cutoff of the most ancient part of the tree at the time of the most recent strongly advantageous mutation, thus cutting off the tail of the uniqueness distribution. Otherwise, spatially uniform fitness differences have minimal impact. Selection that varies in space (or multiple local niches) would increase the tail of the uniqueness distribution by providing more highly unique individuals associated with long genealogies in different regions. The power law distribution of diversity suggests that much of the genetic diversity is found in a small portion of the population.

We compared these results with genetic data obtained previously[17] on field samples of *Pseudomonas* bacteria. From the dendrogram of 250 samples, taken from different locations around the world, we obtained the distribution of uniqueness of the

samples, as described in the Supplementary Information, and shown as circles in Fig. 3. The results are long-tailed and can be fitted by a power law distribution. To compare more precisely the genetic data with our theoretical calculations, however, we must include the effect of sampling. Sampling makes the distribution longer-tailed, corresponding to a greater proportion of individuals that are more unique with respect to the sampled population. We directly simulated the ancestral tree of the samples, initializing the simulation by representing organisms at the specific geographical locations where the *Pseudomonas* samples were obtained. The result is shown as a dashed line in Fig. 3. At each step of the simulation, moving backwards in time, a lineage performs a random walk on a 50×25 lattice, staying in place or moving to a neighbouring site. At its destination, with a certain probability, p_c, it coalesces with other lineages at that site. Details are given in the Supplementary Information. p_c, the number of simulation time steps N_T corresponding to one unit T_0 of biological time, and the lattice size are adjustable parameters. The parameters were set by a simple fitting procedure that adjusts the intercepts of $U(T)$ at the U and T axes and accounts for the small effect of sampling on the low-T values of the curve.

The distribution of uniqueness in well-mixed and spatial populations leads to significant temporal variation. It has been noted that the time to the common ancestor can have intermittent large jumps[18]. Here we study the size distribution of fluctuations in diversity and its relationship to the uniqueness distribution. In Fig. 2c we show the diversity, measured by the total branch length of the genealogical tree, over time. The fluctuations are large: diversity grows gradually owing to the accumulation of mutations, but often decreases markedly in a single generation; in the figure, diversity decreases in one generation by up to about two-thirds of the total. The sudden decreases are due to the extinction of lines of descent that have accumulated many unique mutations. The distribution of the losses of diversity (Fig. 2a) has a power law tail with the same exponent as the uniqueness distribution, consistent with the loss of uniqueness of randomly chosen individuals. The flattening of the distribution for small losses reflects averaging due to the loss of multiple lines of descent in a particular generation. Other diversity measures such as nucleotide diversity[19] show similarly large fluctuations.

Box 1
Analytic result

The distribution of genetic uniqueness, and hence also of fluctuations in the total branch length of the genealogy, can be understood as follows. The probability $P(U > u)$ that an individual has uniqueness greater than u is the probability that its lineage, traced backwards, never exists on a site that has another lineage for all time $T \le u$. In the well-mixed model, the probability that no other lineage jumps to a particular site is $((N-1)/N)^{p(T)N} \approx \exp(-p(T))$, where $p(T)$ is the fraction of individuals living at time T in the past that have descendants in the present. $p(T)$ is approximately $\frac{a}{T}$, with a measured from simulations of $p(T)$ to be 1.95, and expected analytically to be 2 for the well-mixed case[14]. This gives:

$$P(U > u) \approx \prod_{T=1}^{u} \exp\left(-\frac{a}{T}\right) = \exp\left(-a\sum_{T=1}^{u}\frac{1}{T}\right)$$
$$\sim \exp(-a\log(bu)) \sim u^{-a}$$

Thus the exponent depends on the coefficient a of $p(T)$. The probability density is $P(u) = -\mathrm{d}P(U > u)/\mathrm{d}u$, giving

$$P(u) \sim u^{-a-1} \sim u^{-2.95}$$

consistent with both the well-mixed and spatial simulations of $P(u)$.

The scale-free distribution of uniqueness also applies to subgroups defined by a given level of relatedness T_g. For each individual, define the subgroup it belongs to by the identity of its ancestor T_g generations ago. We define the uniqueness u_g of the group to be the uniqueness of the ancestor. The genealogical tree of the ancestors of these groups have the same properties as that of the present population of individuals, only starting with a lower value of $p(T)$. Thus, their uniqueness follows the same power law distribution. For the same reason, the distribution is not affected by the level of resolution in genetic distances (the smallest measurable difference).
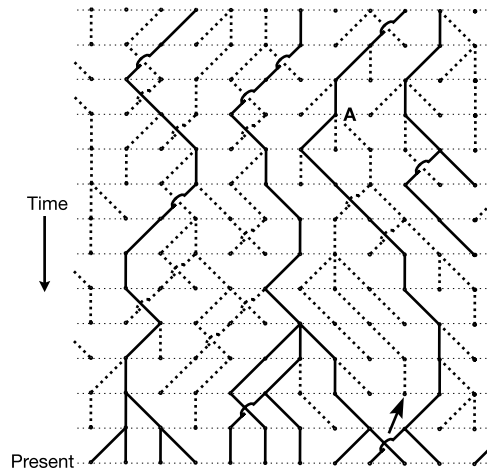

Figure 1 Section of a genealogical tree for a one-dimensional population (part of a much larger tree). Time proceeds down the page, with each horizontal row representing a (non-overlapping) generation and offspring connected to their parent by a line. The tree of the currently living individuals is shown as solid lines; the ancestry of those that have no descendants in the present, and thus do not contribute to diversity, is shown as dashed lines. At the arrow, a lineage goes extinct, causing the loss of accumulated differences on the line of descent from A, the most recent ancestor with descendants in the present.

We can also characterize how diversity is distributed spatially. In Fig. 1, it can be seen that organisms are often closely related to their neighbours in a spatial population, but that neighbours can also be quite distantly related. Previous work has only considered one-dimensional populations[20]. Using a simulation in two dimensions, Fig. 2d shows that there are specific patch-like areas, which represent groups separated by a large genealogical distance from the rest of the population. The shaded areas are separated by 10,000
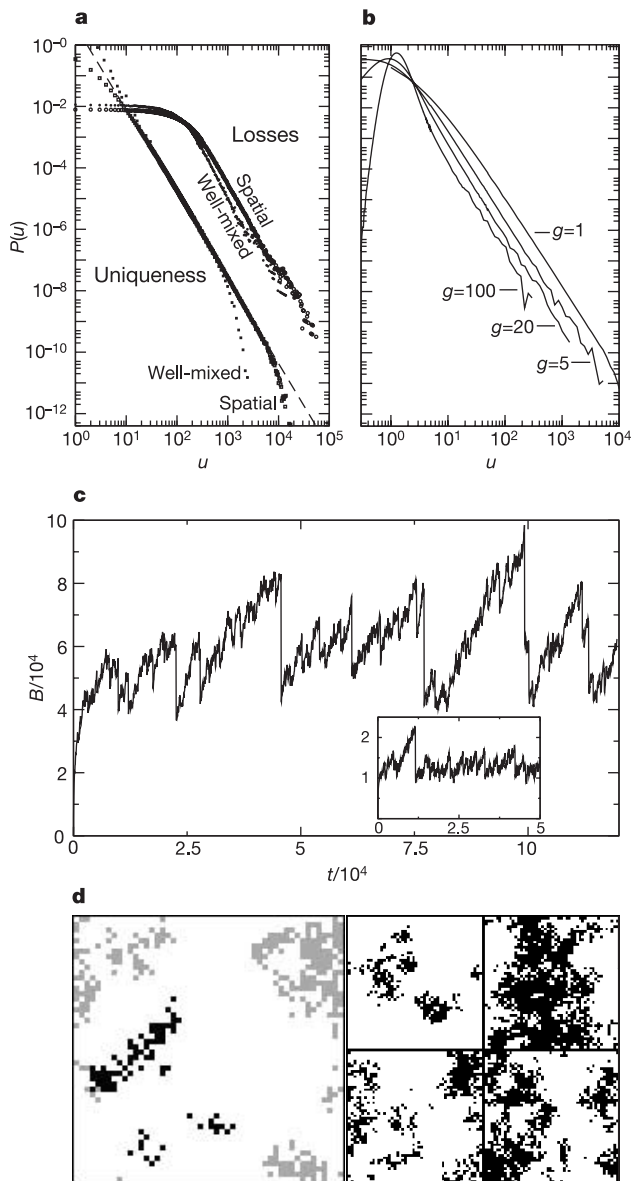
generations from the rest of the population; the grey areas are separated from the black ones by 2,000 generations. Continued simulation shows that specific boundaries move and disappear, but distinctive features exist at any particular time. In this simulation, dispersal of offspring is limited to neighbouring sites. If offspring disperse farther away but still near to their parents, types become interspersed locally and the patches do not have sharp boundaries. Still, as long as the dispersal distance is small compared with the size of the habitat, there remains a spatial patchy structure of genealogically well-separated types. If recombination occurs, there are still such patterns but they are distinct for each independently inherited part of the genome.

Explanations for the occurrences of divergent populations in particular spatial areas, and boundaries between types, are often sought in habitat variation, barriers that prevent organism motion or in a past migration event[21–23]. The model shows that divergent patches can arise from the internal structure of the genealogical tree. Divergent populations are not necessarily confined to a single area; alleles can be geographically widespread in a population even if it is not well-mixed. Thus the spatial patterns of genetic variation in homogeneous habitats must be considered before making inferences about the properties and history of a population. Regarding global microbial populations in particular, the model suggesting that these populations are unaffected by geographic barriers or limited dispersal, and vary spatially only due to selection[24], has been challenged by recent work[17,25]. Here we consider this question only for a specific bacterial species; for others our work does not exclude the cosmopolitan limit of globally mixed populations through which only selection produces local variation. However, our work implies that observations that seem to suggest a globally mixed population, or that variation arises from selection acting on types with otherwise cosmopolitan distributions, must be considered in light of the intrinsic diversity distribution of genealogies. Moreover, even if global dispersal dominates on longer timescales and eliminates biogeographic variation observable in studies with
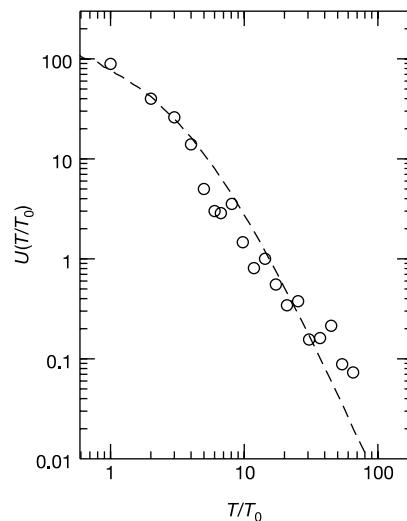

Figure 2 Distribution of diversity and its fluctuation. **a**, Distribution $P(u)$ of genetic uniqueness of individuals, and of losses of diversity in a single generation in well-mixed (small symbols) and spatial populations (larger symbols). Horizontal axis represents uniqueness in generation of divergence. **b**, Distribution of uniqueness in a spatial sexual population for different numbers of independently inherited segments of the genome $g = 1, 5, 20$ and 100. The total genome size is fixed. Linkage would reduce the difference from the $g = 1$ case. **c**, Time series of the diversity of a spatial population. Inset shows a similar well-mixed case. **d**, Simulated pattern of the most divergent groups (grey and black, see text) in a spatial habitat (lattice size $l = 50$). The four panels on the right show the most divergent group in the later evolution of the same population (black), at intervals of 2,000 generations.

Figure 3 Comparison of theoretical values of uniqueness with data from field populations. Circles represent $U(T/T_0)$, the number of samples with a uniqueness of T/T_0, for a sampled *Pseudomonas* population. The dashed line is an average over 1,000 spatial simulations. We normalize T by dividing by T_0, the time to the smallest genetic difference considered. Sampling causes a shallower slope than for the whole population (Fig. 2), and this slope is matched by the simulation. In the simulation, T_0 corresponds to 160 time steps, the lattice size is 50×25, and the coalescence probability, p_c, is 0.15. Details are given in the Supplementary Information.

coarse genetic resolution, the finer-scale differences now becoming observable by genetic fingerprinting are more likely to display biogeography. Extrinsic factors, such as local selection or habitat boundaries, can be distinguished from intrinsic genealogical diversity. For sexually reproducing populations, for extrinsically imposed diversity, spatial boundaries of independently inherited parts of the genome would tend to coincide, whereas for intrinsic genealogical diversity they would tend to be different. For microbial populations, where non-systematic recombination cannot provide the same information, short generation times may allow space–time sampling to study the biogeographical dynamics.

Although the model used here is simple, our analysis shows that the results are robust, and independent of the level of genetic resolution. In summary, our results show that genetic diversity is very unevenly distributed. A small fraction of a population is responsible for a disproportionate fraction of the diversity. Diversity has its own internal dynamics distinct from external influences such as habitat change and species interactions. Increases happen only gradually, but large decreases may occur without an extrinsic perturbation.

It was recently suggested[8], on the basis of the analysis of model phylogenetic trees (similar to our genealogical trees), that extinction of 95% of species would leave 80% of the tree of life (total diversity) retained, and that as a result ecological planning to preserve diversity is not constructive. These results arise because random losses, even when high, are unlikely to remove all individuals belonging to a deep branch of the tree even when it forms a small proportion of the population, thus preserving most of the diversity.
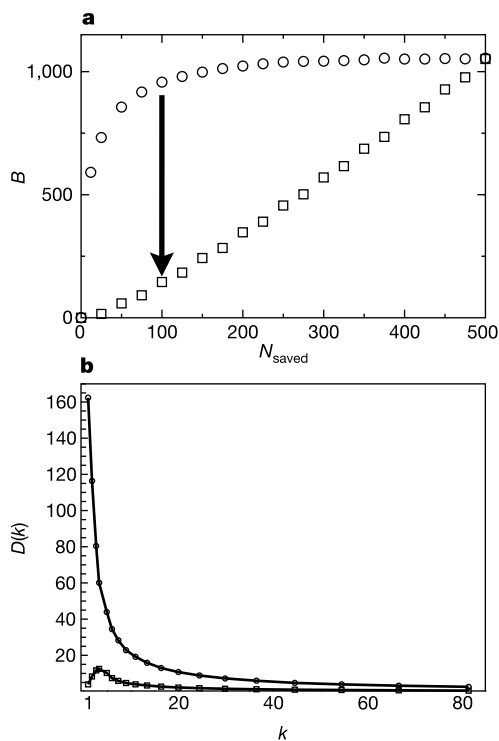
In contrast, our studies suggest that conservation planning is important and can enable substantial improvement of diversity preservation. We note that for spatial populations, the loss of some of the habitat leads to a greater probability of eliminating all members of a divergent group because members of divergent groups are spatially clustered, therefore potentially leading to a much greater loss of diversity. Even for well-mixed populations, however, the small immediate loss (Fig. 4a; circles, similar to Fig. 2b of ref. 8) is followed by a much greater loss over time (Fig. 4a; squares) owing to the vulnerability of residual divergent groups to extinction. Simulations show that most of the subsequent loss occurs within 20 generations. The vulnerability of residual small groups of related organisms can be seen from a plot revealing the multiplicity (redundancy[26]), k, of mutations (Fig. 4b). For the case shown, almost 25% of the total diversity just after the extinction is found in only one or two individuals, and most of the diversity lost is found in less than ten individuals. This shows that ensuring the reproduction of rare types by conservation planning[5,6,27,28] during or even just after an extinction episode can markedly improve diversity retention.   □

1. Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C. & Gibbons, D. W. Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365,** 335–337 (1993).
2. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403,** 853–858 (2000).
3. Gaston, K. J. Global patterns in biodiversity. *Nature* **405,** 220–227 (2000).
4. Ehrlich, P. R. & Wilson, E. O. Biodiversity studies: science and policy. *Science* **253,** 758–761 (1991).
5. Faith, D. P. Genetic diversity and taxonomic priorities for conservation. *Biol. Conserv.* **68,** 69–74 (1994).
6. Amos, A. & Balmford, A. When does conservation genetics matter? *Heredity* **87,** 257–265 (2001).
7. Frankham, R. *et al.* Do population size bottlenecks reduce evolutionary potential? *Anim. Conserv.* **2,** 255–260 (1999).
8. Nee, S. & May, R. M. Extinction and the loss of evolutionary history. *Science* **278,** 692–694 (1997).
9. Hudson, R. R. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7,** 1–44 (1990).
10. Barton, N. H. & Wilson, I. Genealogies and geography. *Phil. Trans. R. Soc. Lond. B* **349,** 49–59 (1995).
11. Notohara, M. The structured coalescent process with weak migration. *J. Appl. Prob.* **38,** 1–17 (2001).
12. Wilkins, J. F. & Wakeley, J. The coalescent in a continuous, finite, linear population. *Genetics* **161,** 873–888 (2002).
13. Wright, S. Isolation by distance. *Genetics* **28,** 114–138 (1943).
14. Fisher, R. A. The distribution of gene ratios for rare mutations. *Proc. R. Soc. Edinb.* **50,** 205–220 (1930).
15. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7,** 256–276 (1975).
16. Wakeley, J. & Takahashi, T. Gene genealogies when the sample size exceeds the effective size of the population. *Mol. Biol. Evol.* **20,** 208–213 (2003).
17. Cho, J. C. & Tiedje, J. M. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66,** 5448–5456 (2000).
18. Watterson, G. A. Mutant substitutions at linked nucleotide sites. *Adv. Appl. Prob.* **14,** 206–224 (1982).
19. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
20. Irwin, D. E. Phylogeographic breaks without geographic barriers to gene flow. *Evolution* **56,** 2383–2394 (2002).
21. Hewitt, G. M. The genetic legacy of the Quaternary ice ages. *Nature* **405,** 907–913 (2000).
22. Willis, K. J. & Whittaker, R. J. The refugial debate. *Science* **287,** 1406–1407 (2000).
23. Cann, R. Genetic clues to dispersal in human populations: retracing the past from the present. *Science* **291,** 1742–1748 (2001).
24. Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296,** 1061–1063 (2002).
25. Whitaker, R. J., Grogan, D. W. & Taylor, J. W. Geographic barriers isolate endemic populations of hyperthermophilic archea. *Science* **301,** 976–978 (2003).
26. Bar-Yam, Y. Multi-scale variety in complex systems. *Complexity* **9,** 37–45 (2004).
27. Crozier, R. H. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annu. Rev. Ecol. Syst.* **28,** 243–268 (1997).
28. Moritz, C. Defining evolutionarily significant units for conservation. *Trends Ecol. Evol.* **9,** 373–375 (1994).

Figure 4 Diversity retained after an extinction episode in a well-mixed population. **a**, The average diversity (total branch length B) remaining after the population is reduced from 500 to the value N_{saved} (horizontal axis). Circles show the values immediately after the extinction and squares the value subsequently reached. The arrow indicates the diversity lost after the initial extinction for a loss of 80% of the population. **b**, $D(k)$, the number of mutations carried by exactly k individuals (multiplicity or redundancy). The area under this curve is the total diversity. $D(k)$ is shown just after the extinction event (upper curve), and in the long term.