Self-similarity of extinction statistics in the fossil record

Ricard V. Solé[*][†], Susanna C. Manrubia[*],
Michael Benton[‡] & Per Bak[§][†]

[*] *Department of Physics FEN, Universitat Politècnica de Catalunya,
Campus Nord, Mòdul B4, 08034 Barcelona, Spain*
[†] *Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA*
[‡] *Department of Geology, University of Bristol, Bristol BS8 1RJ, UK*
[§] *The Niels Bohr Institute, Biegdamsvej 17, DK-2100 Copenhagen, Denmark*

....

The dynamical processes underlying evolution over geological timescales remain unclear[1,2]. Analyses of time series of the fossil record have highlighted the possible signature of periodicity in mass extinctions[3,4], perhaps owing to external influences such as meteorite impacts. More recently the fluctuations in the evolutionary record have been proposed to result from intrinsic non-linear dynamics for which self-organized criticality provides an appropriate theoretical framework[5–7]. A consequence of this controversial[8] conjecture is that the fluctuations should be self-similar, exhibiting scaling behaviour like that seen in other biological[9] and socioeconomic[10,11] systems. The self-similar character is described by a $1/f$ power spectrum $P(f)$, which measures the contributions of each frequency f to the overall time series. If self-similarity is present, then $P(f) \approx f^{-\beta}$ with $0 < \beta < 2$. This idea has not been sufficiently tested, however, owing to a lack of adequate data. Here we explore the statistical fluctuation structure of several time series obtained from available palaeontological data bases, particularly the new 'Fossil Record 2'[18]. We find that these data indeed show self-similar fluctuations characterized by a $1/f$ spectrum. These findings support the idea that a nonlinear response of the biosphere to perturbations provides the main mechanism for the distribution of extinction events.

The Phanerozoic record of life consists almost entirely of extinct groups of organisms, indicating that extinction is the fate of most lineages and that no biotas are infinitely resilient[13]. Considerable effort has been devoted to explaining the so-called mass extinction events in terms of external physical events[14,15]. The dynamics of the biosphere are reflected, to some extent, in the time series provided by the fossil record. Different theories have been tested by means of analyses of this data set.

A direct inspection of the available time series[16] of fluctuations in the number of Ammonoidea families (Fig. 1) suggests a fractal structure in this fossil record, which is consistent with statistical measures such as the frequency distribution of extinctions and

lifetimes[12,17], which also exhibit power-law decay. Although self-organized criticality (SOC) is not the only mechanism leading to power laws, we should first test whether or not there is evidence for $1/f$ dynamics in real time series. We have tested this idea by using the data sets available from 'The Fossil Record 2' which includes all groups that have a fossil record[18] as well as information from other databases, such as the Sepkoski compilation of Phanerozoic diversity and extinction events for marine genera[19].

Using a given time series $\xi(t)$ $(t = 1, 2, …, N)$, where the scale is typically in millions of years, we compute $P(f)$; this is done by using the discrete Fourier transform[20] through a standard
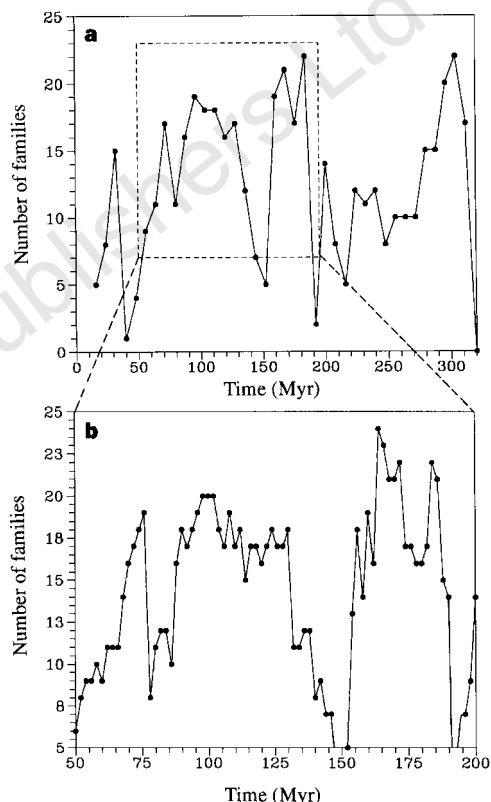

Figure 1 Self-similarity in the fossil record time series (redrawn from House[16]). Here we show the fluctuations in families of Ammonoidea (over a period of 320 Myr, with a time resolution of 2 Myr). In **a** only one of each four original data points is shown, giving a coarse resolution of 8 Myr. In **b** we plot a segment of the previous data, but at full (2-Myr) resolution.

Table 1 Scaling exponents for family origination and extinction

	Origination	Total ext.	Per cent ext.	Tot ext. rate	PF ext. rate
Global (min.)	0.94 ± 0.01	0.93 ± 0.01	0.97 ± 0.01	0.83 ± 0.02	0.87 ± 0.02
	0.92 ± 0.03	0.98 ± 0.01	1.00 ± 0.01	0.86 ± 0.02	0.87 ± 0.02
Global (max.)	0.93 ± 0.01	0.92 ± 0.02	0.97 ± 0.01	0.84 ± 0.02	0.88 ± 0.02
	0.91 ± 0.03	0.93 ± 0.01	0.99 ± 0.01	0.88 ± 0.02	0.90 ± 0.02
Terr. (min.)	0.92 ± 0.01	0.92 ± 0.01	0.95 ± 0.01	0.82 ± 0.01	0.88 ± 0.01
	0.90 ± 0.02	0.95 ± 0.01	0.98 ± 0.01	0.91 ± 0.02	0.94 ± 0.01
Terr. (max.)	0.91 ± 0.01	0.89 ± 0.02	0.92 ± 0.02	0.81 ± 0.03	0.87 ± 0.02
	0.90 ± 0.02	0.91 ± 0.01	0.95 ± 0.01	0.87 ± 0.03	0.90 ± 0.02
Marine (min.)	0.97 ± 0.01	0.95 ± 0.01	0.98 ± 0.01	0.88 ± 0.02	0.88 ± 0.02
	0.99 ± 0.01	1.00 ± 0.01	1.00 ± 0.01	0.87 ± 0.02	0.89 ± 0.02
Marine (max.)	0.93 ± 0.01	0.91 ± 0.01	0.97 ± 0.01	0.91 ± 0.02	0.92 ± 0.02
	0.98 ± 0.01	0.96 ± 0.02	1.01 ± 0.01	0.89 ± 0.02	0.92 ± 0.02

Exponents are given for different extinction metrics, involving global, terrestrial and marine time series[2]. Both the exponent for the power spectrum β (first row of each pair of rows) and the r.m.s. exponent α (second row) are given. 'Origination' indicates total number of new families; 'Total ext.' indicates total number of extinct families; 'Per cent ext.' indicates per cent of extinction, measured relative to the number of families in existence; 'Tot. ext. rate' (TER) indicates total extinction rate; 'PF ext. rate' indicates per family extinction rate, measured as the TER divided by taxic diversity. Deviations from the $\beta = 1$ exponent are more frequent when rate-dependent measures are considered. This is expected as more information is involved and deviations from sampling errors are more likely to occur.

fast-Fourier-transform algorithm (FFT). $P(f)$ is

$$P(f) = \frac{1}{\sqrt{N}} \sum_{\tau=1}^{N} C(\tau) \exp\left(\frac{-i2\pi f \tau}{N}\right) \quad (1)$$

where $C(\tau) = \langle \xi(t)\xi(t+\tau)\rangle$ is the so-called autocorrelation function. Then a log–log plot of $P(f)$ on f is used (Fig. 2) and the scaling exponent β is computed from a least-squares fit (a linear regression of log-transformed data is better than nonlinear fit on raw data because the residual error will be distributed as a quadratic and the minimum error is guaranteed). If long-range correlations are present, then a scaling $C(\tau) \approx \tau^{-\gamma}$ will be observed with $\gamma = 1 - \beta$ (refs 10, 21).

Several time series have been used, involving both origination and different measures of extinction for families. First, we consider the extinction sizes in families. Here the available data sets are mainly for stratigraphic stages, which have an average duration of 7 Myr (ref. 18). So in order to get an equally spaced data set with 1-Myr resolution we need to generate more points. We have used a direct, linear interpolation, the results of which are shown in Table 1. Because this method can lead to spurious correlations, a different procedure was also used to test the robustness of our results: a step-function approach to each stage is introduced, together with a square-root transformation[20,21]. Both methods gave basically the same results. Both minimum and maximum measures are available, based on assessments of uncertainty about dates of origin and extinction and other factors[2,22]. In Table 1 we show our results for family origination and extinction by using four different extinction metrics[2]: most of the obtained scaling exponents are close to $\beta \approx 1$. There is some drift towards lower β for time series based on rates of extinction. This could be due to the biases in the estimation of the total number of families per stage[2]. A decrease in β as a consequence of errors introduced by estimated stage duration. These durations are poorly constrained, especially for stages before 100 Myr ago, and their estimates vary considerably among available geological timescales.

An additional, complementary measure is obtained by means of the use of 'random walk' methods able to detect long-range correlations[10,21,23–25]. Starting from the time series $\xi(t)$, the running sum

$$y(\tau) \equiv \sum_{i=t}^{\tau} \xi(i) \quad (2)$$

Figure 2 a, Pattern of family extinctions through time of all organisms in terms of the number of families. Maximum and minimum curves[2] are shown (inset) as well as the corresponding power spectra (main figure). **b**, Fluctuation in the marine families (inset) and the corresponding power spectrum (main figure). Both series are in a timescale of Myr and in both cases the spectrum is not like white, uncorrelated noise ($\beta = 0$) but almost like $1/f$-spectra, characteristic of long-range correlations in scale-invariant systems.
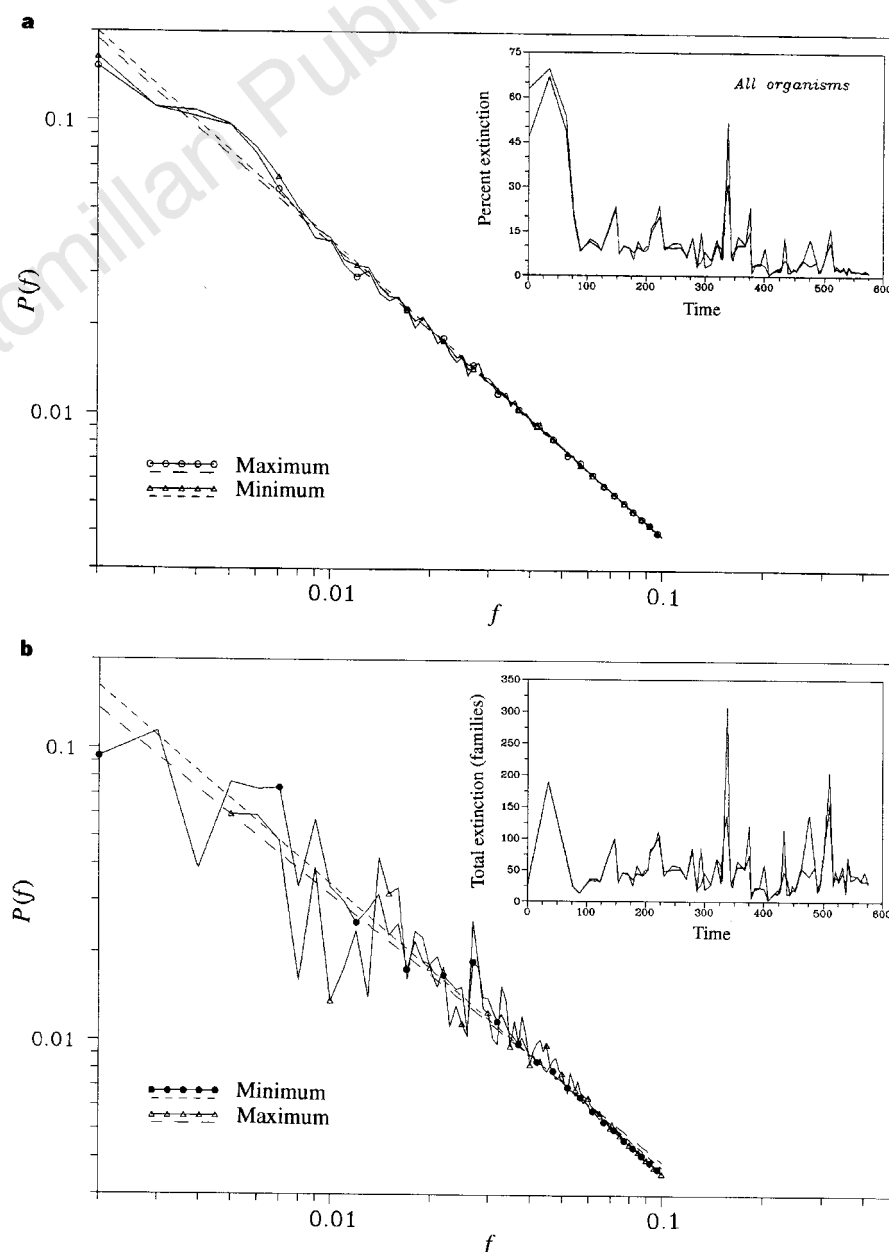
is computed, and the root-mean-square (r.m.s.) fluctuation $F(\tau)$ is obtained from

$$F(\tau) \equiv [\langle(\delta y(\tau))^2\rangle - \langle(\delta y(\tau))\rangle^2]^{1/2} \tag{3}$$

where $\delta y(\tau) = y(t_0 + \tau) - y(t_0)$. The angle brackets indicate an average over all times t_0. If $\{\xi(t)\}$ is random or a simple Markov process, then we have $F(t) \approx t^{1/2}$. But if no characteristic timescale is involved, then a scaling $F(t) \approx t^\alpha$ is observed with $\alpha \neq 1/2$. The r.m.s. exponent and β are related through $2\alpha = \beta + 1$. We have computed α for all the previous data sets, and the resulting exponents are given in Table 1 (see also Fig. 3). We can see that α is also close to one and that all the pairs (α, β) are consistent with the previous relation.

We also analysed the extinction pattern for other taxonomic levels. The study of the total extinction rate for genera of marine animals[19] gave $\beta = -0.86 \pm 0.03$ and $\alpha = 0.91 \pm 0.03$ using the proportional rate of extinction and $\beta = -0.83 \pm 0.02$, $\alpha = 0.89 \pm 0.03$ for the absolute rate of extinction. We can extend our analysis to other types of data, involving different particular groups of organisms. We have used two particularly well known data sets: fluctuations in Ammonoidea families[16] and the number of planktic foraminiferal species from the Jurassic period to the Holocene epoch[26]. Here no interpolation is needed and again well defined power laws were obtained: $\beta = -0.88 \pm 0.03$ and $\alpha = 0.93 \pm 0.04$ for Ammonoidea and $\beta = -0.86 \pm 0.03$ and $\alpha = 0.91 \pm 0.03$ for Foraminifera.

These results are consistent with a nonlinear, self-similar behaviour of the biosphere. Where do these scale-free dynamics come from? There are several possibilities; one of them is SOC, where fluctuations over all timescales and power laws are a direct consequence of criticality[5–7]. Some authors have shown that power laws do not need to be generated through a SOC phenomenon[17] but can result from stresses (either biotic or abiotic) to which species are subjected by their environment. Additionally, the reported fractal nature of taxonomic systems[27] shows that behind the self-similar time series, a fractal-like organization of the biosphere is also present. Recent SOC models of macroevolution have shown that these scaling relations naturally fit into a self-organized, nonlinear generic process[13,23]. Our results are also in agreement with other studies of palaeontological time series based on other methods from nonlinear dynamics[7,28] as forecasting techniques. In all these studies, data analysis supports the internal biotic organization as the basic component for the response of the biosphere to external perturbations.

Furthermore, these analyses provide evidence that might help resolve two debates about macroevolutionary dynamics. The first is the nature of mass extinctions: are they qualitatively, or merely quantitatively, different from normal (background) levels of extinction? Early statistical tests[29] suggested that there was a clear distinction, and that the five largest (the 'big five'; ref. 14) mass extinctions were qualitatively distinct. This view was supported by the finding that times of mass extinction were associated with selective regimes that were utterly different, and unpredictable, from observations during background times[13]. Comparisons of probabilities and magnitudes of extinction events have subsequently suggested[3] however, that the 'big five' could be the skewed end of a continuous distribution of extinction events of different intensity. This view is supported by the scale-free pattern revealed by our study.

The second debate that is addressed by our study concerns the causes of extinctions. When Raup and Sepkoski[4] proposed that mass extinction events over the past 250 Myr had followed a periodic pattern of occurrence every 26 Myr, many workers focused on extraterrestrial causes. Raup[3] has extended this model to postulate that all extinction events can be explained by extraterrestrial impacts, the magnitude of the event depending broadly on the magnitude of the object striking the Earth (see also Newman[17]). A different (but perhaps complementary) view[5] is that extinction events are triggered by a multiplicity of factors, and that the internal dynamics of the system play an important role. Although the role of nonlinear dynamics in extinction and diversity is not new in palaeobiology[1,26,30,31], previous analysis dealing with fluctuations in the fossil record involved low-dimensional, deterministic approaches based on chaotic dynamics. But it seems clear that the underlying assumptions (such as the low-dimensional character of the dynamics, the lack of stochastic effects or the constancy of parameters) are unlikely to hold. A high-dimensional picture where many species evolve and interact seems more consistent with the process of macroevolution, although the specific mechanisms that could drive the biosphere dynamics are under discussion[32]. Our studies, consistent with scale-free pattern of the fossil record, support this latter view. Future analyses could usefully explore factors such as the stationarity of the previous data sets or the Signor–Lipps effect (the 'backward smearing' of extinctions resulting from incomplete sampling[33]) and their importance in generating correlations. □

1. Chaloner, W. & Hallam, A. (eds) Evolution and extinction. *Phil. Trans. R. Soc. Lond. B* **325,** 239–488 (1989).
2. Benton, M. J. Diversification and extinction in the history of life. *Science* **268,** 52–58 (1995).
3. Raup, D. M. A kill curve for Phanerozoic marine species. *Paleobiology* **17,** 37–48 (1991).
4. Raup, D. M. & Sepkoski, J. J. Jr Periodicity of extinctions in the geologic past. *Proc. Natl Acad. Sci. USA* **81,** 801–805 (1994).
5. Bak, P. & Snepen, K. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.* **59,** 381–384 (1993).
6. Kauffman, S. & Johnsen, J. Coevolution to the edge of chaos: coupled fitness landscapes, poised states and coevolutionary avalanches. *J. Theor. Biol.* **149,** 467–505 (1991).
7. Solé, R. V. & Manrubia, S. C. Extinction and selforganized criticality in a model of large-scale evolution. *Phys. Rev.* **54,** R42–R45 (1996).
8. Maddox, J. Punctuated equilibrium on a computer. *Nature* **371,** 197 (1994).
9. Schroeder, M. *Fractals, Chaos, Power Laws* (Freeman, New York, 1991).
10. Stanley, H. H. *et al.* Scaling and universality in animate and inanimate systems. *Physica A* **A231,** 20–48 (1996).
11. Mandelbrot, B. B. The variation of certain speculative prices. *J. Bus. Univ. Chicago* **36,** 307–317 (1963).
12. Solé, R. V. & Bascompte, J. Are critical phenomena relevant to large-scale evolution? *Proc. R. Soc. Lond. B* **263,** 161–168 (1996).
13. Jablonski, D. Extinctions in the fossil record. *Phil. Trans. R. Soc. Lond. B* **344,** 11–17 (1994).
14. Raup, D. *Extinctions: Bad Genes or Bad Luck?* (Oxford Univ. Press, 1993).
15. Elliott, D. K. (ed.) *Dynamics of Extinction* (Wiley, New York, 1986).
16. House, M. R. Ammonoid extinction events. *Phil. Trans. R. Soc. Lond. B* **325,** 307–326 (1989).
17. Newman, M. E. J. Self-organized criticality, evolution and the fossil extinction record. *Proc. R. Soc. Lond. B* **263,** 1605–1610 (1996).
18. Benton, M. J. (ed.) *The Fossil Record 2* (Chapman & Hall, London, 1993).
19. Sepkoski, J. J. in *Pattern and Process in the History of Life* (eds Ramp, D. & Jablonski, D.) 277–295 (Springer, Berlin, 1986).
20. Davis, J. C. *Statistics and Data Analysis in Geology* (Wiley, New York, 1973).
21. Korvin, G. *Fractal Models in the Earth Sciences* (Elsevier, 1992).
22. Sepkoski, J. J. & Raup, D. in *Dynamics of Extinctions* (ed. Elliott, D. K.) Ch. 2 (Wiley, New York, 1986).
23. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1982).
24. Sugihara, G. & May, R. M. Applications of fractals in ecology. *Trends Ecol. Evol.* **5,** 79–86 (1990).
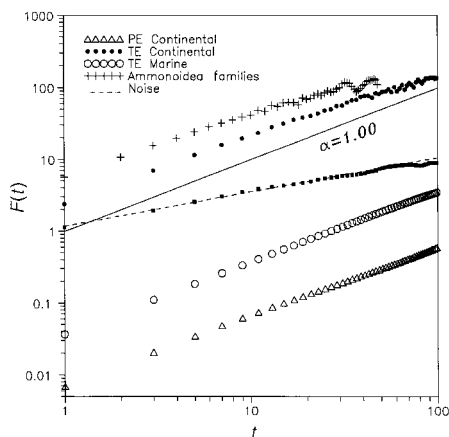
Figure 3 Log-log plot of the root-mean-square $F(t)$ for different (typical) palaeontological time series (see equations (2) and (3) in the text). Here both per cent extinction (PE) and total extinction (TE) for different data sets have been used. We also show the scaling for fluctuations in families of Ammonoidea and the corresponding scaling for a random white-noise process from a $N = 150$ time series. Here $\alpha = 1/2$.

25. Stanley, H. H. *et al.* Statistical mechanics in biology: how ubiquitous are long-range correlations? *Physica A* **205,** 214–253 (1996).
26. Plotnick, R. E. & McKinney, M. Evidence of self-organization in planktic foraminiferal evolution: implications for interconnectedness of paleoecosystems. *Palaios* **8,** 202–212 (1993).
27. Burlando, B. The fractal geometry of evolution. *J. Theor. Biol.* **163,** 161–172 (1993).
28. Solé, R. V., Bascompte, J. & Manrubia, S. C. Extinctions: Bad genes or weak chaos? *Proc. R. Soc. Lond. B* **263,** 1407–1413 (1996).
29. Raup, D. M. & Sepkoski, J. J. Jr Mass extinctions and the marine fossil record. *Science* **215,** 1501–1503 (1982).
30. Glen, W. (ed.) *The Mass Extinction Debates* (Stanford Univ. Press, 1984).
31. Sepkoski, J. J. Jr A kinetic model of Phanerozoic taxonomic diversity III. Post-Paleozoic families and mass extinctions. *Paleobiology* **10,** 246–267 (1984).
32. Newman, M. E. J. *A Model of Mass Extinction* (Wkg Pap. 97-02-013 Santa Fe Institute, 1997); *J. Theor. Biol.* (in the press).
33. Signor, P. W. & Lipps, J. H. Sampling bias, gradual extinction patterns and catastrophes in the fossil record. *Geol. Soc. Am. Spec. Pap.* **190,** 291–296 (1982).

Genetic tagging of humpback whales

Per J. Palsbøll*, Judith Allen†, Martine Bérubé*‡,
Phillip J. Clapham§, Tonnie P. Feddersen*,
Philip S. Hammond‖, Richard R. Hudson¶,
Hanne Jørgensen*, Steve Katona†, Anja Holm Larsen*,
Finn Larsen#, Jon Lien☆, David K. Mattila§,
Jóhann Sigurjónsson**, Richard Sears††, Tim Smith‡‡,
Renate Sponer*, Peter Stevick† & Nils Øien§§

* *Department of Population Biology, Universitetsparken 15, Copenhagen DK-2100, Denmark*
† *Allied Whale, College of the Atlantic, Bar Harbor, Maine 04609, USA*
‡ *Department of Natural Resources Science, Ste-Anne de Bellevue, Quebec H9X 3V9, Canada*
§ *Center for Coastal Studies, 59 Commercial Street, Provincetown, Massachusetts 02657, USA*
‖ *Sea Mammal Research Unit, University of St Andrews, St Andrews, Fife KY16 8LB, UK*
¶ *Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697, USA*
# *Greenland Institute of Natural Resources, PO Box 570, 3900 Nuuk, Greenland*
☆ *Whale Research Group, Memorial University, Newfoundland A1B 3X9, Canada*
** *Marine Research Institute, Skulagata 4, PO Box 1390, 121 Reykjavik, Iceland*
†† *Mingan Island Cetacean Study Inc., 285 Green Street, Ste-Lambert, Quebec J4P 1T3, Canada*
‡‡ *US National Marine Fisheries Service, Northeast Fisheries Science Center, Woods Hole, Massachusetts 02543, USA*
§§ *Institute of Marine Research, P.O. Box 1870, Nordnes, N1-15024 Bergen, Norway*

The ability to recognize individual animals has substantially increased our knowledge of the biology and behaviour of many taxa[1]. However, not all species lend themselves to this approach, either because of insufficient phenotypic variation or because tag attachment is not feasible. The use of genetic markers ('tags') represents a viable alternative to traditional methods of individual recognition, as they are permanent and exist in all individuals. We tested the use of genetic markers as the primary means of identifying individuals in a study of humpback whales in the North Atlantic Ocean. Analysis of six microsatellite loci[2,3] among 3,060 skin samples collected throughout this ocean allowed the unequivocal identification of individuals. Analysis of 692 'recaptures', identified by their genotype, revealed individual local and migratory movements of up to 10,000 km, limited exchange among summer feeding grounds, and mixing in winter breeding areas, and also allowed the first estimates of animal abundance based solely on genotypic data. Our study demonstrates that genetic tagging is not only feasible, but generates data (for example, on sex) that can be valuable when interpreting the results of tagging experiments.

Skin biopsy[4] or sloughed skin[5] samples from free-ranging humpback whales (*Megaptera novaeangliae*) were collected across the North Atlantic between 1988 and 1995. Total-cell DNA was extracted[6], and the sex[7] and genotype at six mendelian inherited[8] microsatellite loci[9] were determined for each sample. From the 3,060 samples analysed, we detected 2,368 unique genotypes. The expected number of samples collected from different individuals with identical genotype arising by chance was estimated at less than one (see Methods). Because of this, and the fact that all samples with identical genotypes were of consistent sex, we believe that the 3,060 samples represented 2,368 individual whales. Of the 692 recaptures observed during the study, 216 occurred on the summer feeding grounds (Fig. 1). Of these, 96% (*n* = 207) occurred within the same feeding area (Fig. 1), confirming previous behavioural[10,11] and genetic[12,13] observations of maternally directed fidelity to specific feeding grounds. The remaining 4% (*n* = 9) of the recaptures on the summer feeding grounds were detected in different but adjacent feeding grounds (Fig. 1). However, significantly more recaptures were detected within these sampling areas than would be expected if the areas constituted one intermixing feeding aggregation and is consistent with the notion of maternally directed site fidelity (see Methods).

Of the 114 individuals recorded on both summer feeding and winter breeding grounds (Fig. 1), two had migrated from the West Indies to either Jan Mayen or Bear Island (in the Barents Sea). These genetic recaptures represent the most extensive one-way movements (6,435 and 7,940 km, respectively) recorded in this study, and support recent findings[13,14] that whales feeding in the Barents Sea share a common breeding ground with other North Atlantic humpback whales. In three other individuals, which were each sampled on three occasions, movements were documented from a feeding ground to the breeding range and back, involving minimum migration distances of up to 10,000 km between the first and last sampling event. No feeding ground was disproportionately represented among the recaptures in the West Indies (Fig. 1), supporting the current view that humpback whales in the North Atlantic constitute a single panmictic population[13,15,16] (*G*-test, $G_{5 d.f.} = 4.68$, $P < 0.46$).

As with traditional identification methods[1], microsatellite data lend themselves to abundance estimation using mark–recapture statistical methods[17], although to our knowledge this has not previously been attempted. Using breeding-ground samples collected during 1992 and 1993, we estimated the North Atlantic humpback whale population at 4,894 (95% confidence interval, 3,374–7,123) males and 2,804 (95% confidence interval, 1,776–4,463) females. This total of 7,698 whales is substantially (albeit not significantly) higher than the most recent previous photographic-based estimate of 5,505 (ref. 10) (95% confidence interval, 2,888–8,122). Preliminary results from new and more reliable photographic estimates are also larger than previous estimates (T.S. *et al.*, manuscript in preparation), and could partly be due to population growth during the intervening decade since the previous estimate[18]. The significantly different estimates for males and females are unexpected given the even sex ratio observed on the feeding grounds[19] (Table 1) and among 198 calves that we sampled in the breeding range (data not shown). The estimates are independent of between-sex sampling biases, and so the observed deficit of females probably reflects within-sex behavioural differences, for example that individual females display a higher degree of preferences with respect to region and/or residence time in the breeding range than do males.

Our results demonstrate that genetic tagging is effective even in a large population of wide-ranging and inaccessible mammals such as cetaceans. Further, the data obtained from genetic tags can be used to address evolutionary[20], demographic[19] and behavioural[21]