diffusive reequilibration), and the disequilibrium between ^{226}Ra and ^{210}Pb. Once ^{210}Pb excess is formed, reestablishing secular equilibrium between ^{226}Ra and ^{210}Pb would require about 100 years; thus, the ^{210}Pb excesses formed by accumulation of gases will persist. Plagioclase grows during decompression (under H_2O-saturated conditions), thus sealing off the melt inclusions shortly after gas is lost. The available experimental data show that Li diffuses more rapidly than all other trace elements in melt (*9*) and plagioclase (*10*), which suggests that Li can be diffusively homogenized in both phases on time scales of hours.

In the context of Mount St. Helens, our results provide a detailed picture of magma and gas movement during 1980. Specifically, the cryptodome followed a two-stage decompression path interrupted at a depth of 4- to 5-km where magma stalled and gas accumulated before further ascent. During magma ascent, continued gas fluxing through a semipermeable magmatic foam buffered the Li concentration of the melt and supplied Rn (Fig. 3). Evidence for breaching of the impermeable cap and gas flow comes from phreatomagmatic eruptions and steam venting in March and April (*2*). Magma erupted during the Plinian phase of the eruption ascended rapidly from a depth of more than about 7 km without stalling (*22*). Before the post–18 May explosive eruptions, magmas stalled at 4- to 5-km before the impermeable cap was breached and magma ascended rapidly. A dome sample from October 1980 also has high Li melt inclusions at $p\mathrm{H_2O}$ < 125 MPa, which suggests that magma ascended more slowly than during the preceding explosive eruptions and, as with the cryptodome, Li concentrations in the melt were buffered. The increase of (^{210}Pb/^{226}Ra) during the summer of 1980 correlates with a decrease in magma ascent rate (*4*, *5*). This suggests that the trend of increasing (^{210}Pb/^{226}Ra) with time reflects prolonged stalling of magma at 4- to 5-km depth as the eruption intensity waned.

Only recently has attention been turned to the complexity of shallow conduit processes and their link to eruption style. Our data shed light on processes occurring on time scales of years to hours before an eruption and, as such, may provide an aid in interpreting observations from established monitoring techniques, for example, gas emissions and seismic surveys. ^{210}Pb excesses are coupled to ^{210}Pb deficits deeper in the system. Thus, magma erupting with ^{210}Pb excess requires the presence of degassing magma at depth.

References and Notes

1. P. W. Lipman, D. R. Mullineaux, Eds., *U.S. Geol. Surv. Prof. Paper* **1250**, 93 (1981).
2. K. V. Cashman, R. P. Hoblitt, *Geology* **32**, 141 (2004).
3. K. Cashman, *Contrib. Mineral. Petrol.* **109**, 431 (1992).
4. C.-H. Geschwind, M. J. Rutherford, *Bull. Volcanol.* **57**, 356 (1995).
5. R. Scandone, S. D. Malone, *J. Volcanol. Geotherm. Res.* **23**, 239 (1985).
6. J. Blundy, K. Cashman, *Contrib. Mineral. Petrol.* **140**, 631 (2001).
7. P.-J. Gauthier, M. Condomines, *Earth Planet. Sci. Lett.* **172**, 111 (1999).
8. J. T. Bennett, S. Krishnaswami, K. K. Turekian, W. G. Melson, C. A. Hopson, *Earth Planet. Sci. Lett.* **60**, 61 (1982).
9. F. M. Richter, A. M. Davis, D. J. DePaolo, E. B. Watson, *Geochim. Cosmochim. Acta* **67**, 3905 (2003).
10. B. J. Giletti, T. M. Shanahan, *Chem. Geol.* **139**, 3 (1997).
11. J. D. Webster, J. R. Holloway, R. L. Hervig, *Econ. Geol.* **84**, 116 (1989).
12. S. Newman, J. B. Lowenstern, *Comput. Geosci.* **28**, 597 (2002).
13. J. Blundy, K. Cashman, in preparation.
14. T. Casadevall *et al.*, *Science* **221**, 1383 (1983).
15. A. C. Rust, K. V. Cashman, P. J. Wallace, *Geology* **32**, 349 (2004).
16. D. S. Stevenson, S. Blake, *Bull. Volcanol.* **60**, 307 (1998).
17. B. E. Taylor, J. C. Eichelberger, H. R. Westrich, *Nature* **306**, 541 (1983).
18. J. C. Eichelberger, C. R. Carrigan, H. R. Westrich, R. H. Price, *Nature* **323**, 598 (1986).
19. C. Klug, K. C. Cashman, *Bull. Volcanol.* **58**, 87 (1996).
20. S. Signorelli, M. R. Carroll, *Geochim. Cosmochim. Acta* **64**, 2851 (2000).
21. H. L. Barnes, Ed., *Geochemistry of Hydrothermal Ore Deposits* (Wiley, New York, ed. 3, 1997).
22. M. J. Rutherford, P. M. Hill, *J. Geophys. Res.* **98**, 19,667 (1993).
23. Methods and data tables are available as supporting material on *Science* Online.
24. During this research, K.B. was supported by a University of Bristol scholarship; J.B. and S.T. acknowledge support from the Royal Society. W. Melson at the Smithsonian Institution provided some of the samples used for this study. Other samples were collected with the generous support of J. Pallister and M. Clynne during field work. We also thank S. Kasemann (Natural Environment Research Council Ion Microprobe Facility, Edinburgh) for assistance with secondary ion mass spectrometry analysis.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1103869/DC1
Materials and Methods
Tables S1 to S3

10 August 2004; accepted 1 October 2004
Published online 14 October 2004;
10.1126/science.1103869
Include this information when citing this paper.

Ventilation of the Glacial Deep Pacific Ocean

Wallace Broecker,[1]* Stephen Barker,[1] Elizabeth Clark,[1]
Irka Hajdas,[2] Georges Bonani,[2] Lowell Stott[3]

Measurements of the age difference between coexisting benthic and planktic foraminifera from western equatorial Pacific deep-sea cores suggest that during peak glacial time the radiocarbon age of water at 2-kilometers depth was no greater than that of today. These results make unlikely suggestions that a slowdown in deep-ocean ventilation was responsible for a sizable fraction of the increase of the ratio of carbon-14 (^{14}C) to carbon in the atmosphere and surface ocean during glacial time. Comparison of ^{14}C ages for coexisting wood and planktic foraminifera from the same site suggests that the atmosphere to surface ocean ^{14}C to C ratio difference was not substantially different from today's.

Hughen *et al.* (*1*) present a strong case that during the last glacial maximum (LGM) (i.e., 22,000 to 16,000 calendar years ago) the ^{14}C to C ratio in the atmosphere and surface ocean was 375 ± 25‰ higher than that for preindustrial time. The obvious explanation for this increase is that Earth's magnetic field was on the average weaker, allowing more cosmic rays to reach our atmosphere. Hughen *et al.* (*1*) contend, however, that although the field was weaker, the consequent increase in ^{14}C production was insufficient to explain the entire observed ^{14}C increase. To explain the remainder, these authors call on a sizable reduction in deep-sea ventilation rate and a possible reduction in shallow marine carbonate deposition (i.e., reef growth). Muscheler *et al.* (*2*) reach a similar conclusion by another route. They make use of the ^{10}Be measurements on the Summit Greenland ice core to reconstruct past production rates of cosmogenic isotopes and conclude that the excess ^{14}C present in the atmosphere and upper ocean cannot be accounted for by higher production alone. They also call on reduced deep-ocean ventilation in an attempt to account for at least part of the remaining radiocarbon increase.

Were the deep sea to have been ventilated at a slower rate during glacial time, then the ^{14}C to C ratio difference between upper-ocean and deep-ocean carbon must have been larger than today's. As shown in Table 1, raising the atmosphere and surface ocean ^{14}C to C ratio by 200 per mil requires a large increase in the difference between the ^{14}C to

---

[1] Lamont-Doherty Earth Observatory of Columbia University, 61 Route 9W/Post Office Box 1000, Palisades, NY 10964, USA. [2] AMS 14C lab, ETH Hoenggerberg HPK H27 and H30, CH-8093 Zurich, Switzerland. [3] University of Southern California, 3651 Trousdale Parkway, Los Angeles, CA 90089, USA.

*To whom correspondence should be addressed. E-mail: broecker@ldeo.columbia.edu

C ratio for surface and deep-sea inorganic carbon. As this difference is potentially recorded by the ^{14}C age difference between coexisting benthic and planktic foraminifera shells in glacial-age sediments, it is possible to check the prediction by Hughen *et al*. (*1*) and Muscheler *et al*. (*2*) that deep-sea ventilation rates were slower during glacial time.

In a recently published paper (*3*), a rather bleak summary was presented of published attempts to establish the age difference between coexisting benthic and planktic foraminifera in glacial-age sediments from the deep Pacific Ocean. The published results for age difference based on individual benthic-planktic pairs range from as low as a few hundred to as high as 3500 years. New results from the eastern equatorial Pacific presented by Broecker *et al*. (*3*) also yield a wide range of ages for coexisting planktic foraminifera. Further, as acid leaching decreased the age of the residual calcite in these samples, the suggestion is that secondary contaminant calcite must be present. Taken together, the average age difference as a function of water depth for the entire ensemble of benthic-planktic pairs is consistent with that for preindustrial time. However, this average has meaning only if it can be assumed that the biases responsible for the very large range in results are random (i.e., on the average not different for benthics than for planktics). The important point is that only a few of these published age differences are large enough to be consistent with the requirement that the extra atmosphere and upper ocean ^{14}C was the result of a slowdown in deep-sea ventilation.

Having found that foraminifera from glacial-age sediments in the eastern equatorial Pacific appear to be contaminated with secondary calcite, we turned to sediments from the western equatorial Pacific in hopes of avoiding this problem. We obtained 50-g samples from core MD98-2181 from the Morotai Basin at 6°N and 126°E, a location just to the south of the island of Mindanao at a depth of 2.1 km, and from core MD97-2138 from the open Pacific at 1°S and 146°E just to the north of the Admiralty Islands at a depth of 1.9 km (*4*). The planktics in these samples yielded ^{14}C ages that respectively correspond to calendar ages of about 19,000 and 21,500 years. The results are listed in Table 2. Because of the problems previously encountered for the eastern equatorial Pacific samples, the first set of runs was limited to planktics. Excellent agreement was achieved in both cores for four planktic species. Based on this success, we conducted a second run in which one of the planktic species was repeated along with the mixed benthics. Subsequent measurements on two additional samples, one from above and one from below the original one, yielded consistent results.

As a bonus, in the first sample from MD98-2181, we were surprised to find a number of small pieces of wood (*5*). The two largest pieces were included in the second run. The age difference of 560 ± 150 years between the wood and planktic foraminifera is consistent with expectation. Measurements on corals from the western equatorial Pacific indicate that before the onset of nuclear testing the Δ^{14}C for surface waters in that region was $-58 \pm 2‰$ (*6*, *7*), which corresponds to a reservoir age of 480 years (*8*). As pointed out by Bard (*9*), other things being equal, this difference should be inversely related to the atmosphere's CO_2 content. Taking the LGM value to be 195 μatm and the preindustrial to be 280, the prediction would be that during the LGM the age difference would have been 630 years.

Water-column profiles of Δ^{14}C in the western equatorial Pacific yield a value of about $-220‰$ at 2000-m water depth (*10*, *11*). Using the coral-based value of $-58‰$ for preindustrial surface water in that region, this corresponds to an age of 1520 years. In today's ocean below 1500 m, there is little change in Δ^{14}C with water depth.

The benthic-planktic ^{14}C age differences obtained for the Morotai core are 1325, 1170, and 1325 years (Figs. 1 and 2). For the Admiralty core, the differences are, respectively, 1954 and 1640 years. Because atmospheric

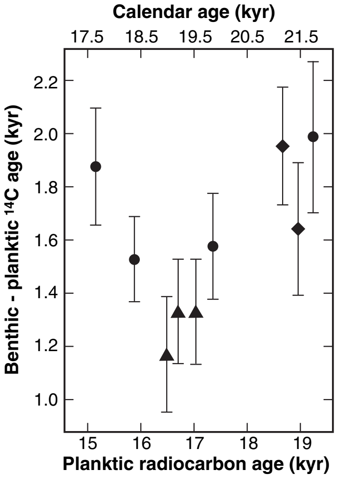

Fig. 1. Benthic-planktic age differences as a function of planktic age for three sites in the western tropical Pacific. The circles represent the results for the South China Sea (*15*), the triangles those for the Morotai Basin, and the diamonds those for Admiralty Island. The upper scale gives the approximate calendar age for the planktics.

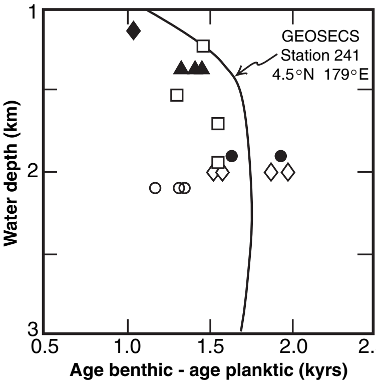

Fig. 2. Summary of LGM ventilation times based on benthic-planktic age differences for cores from the western Pacific. The open squares are for cores from the Sea of Okhotsk (*16*), the solid triangles are for a core from off Japan (*17*), the open diamonds are from a core at a depth of 2.7 km in the South China Sea (sill depth 2.0 km) (*15*), the solid circles are for the Admiralty Island core, and the open circles are for the Morotai Basin core. The solid diamond is for a ^{230}Th-dated benthic coral from the Drake Passage (*18*). Not shown are results obtained on cores stored dry in the Lamont-Doherty repository (*21*) and those from off New Zealand (*22*). The solid line is based on measurements of water samples collected as part of the Geochemical Ocean Sections Study (GEOSECS) survey.

Table 1. Calculation based on a simple box ocean model of the decrease in the ^{14}C to C ratio (relative to that for the preindustrial atmosphere) for deep inorganic carbon required to raise the upper ocean and atmosphere Δ^{14}C with no change in ^{14}C production. Also given is the increase in the radiocarbon age difference between coexisting benthic and planktic foraminifera. Case 1: Two-box ocean; distribution of inorganic carbon between atmosphere, surface, and deep reservoirs = 0.75:1:38. Surface ocean reservoir age is held at 500 years. Case 2: Three-box ocean; as before, except the deep ocean is now divided into intermediate and deep boxes. The ^{14}C age difference between the surface and intermediate boxes is held at 1500 years.

Atmospheric ^{14}C:C	Surface ^{14}C:C	Intermediate ^{14}C:C	Deep ^{14}C:C	Total ^{14}C:C	Benthic-planktic ^{14}C age (years)
Case 1. Two-box ocean; surface:deep = 1:38					
1.00	0.940		0.780	0.788	1500
1.10	1.034		0.775	0.788	2312
1.20	1.128		0.771	0.788	3057
Case 2. Three-box ocean; surface:intermediate:deep = 1:19:19					
1.00	0.940	0.780	0.780	0.788	1500
1.10	1.034	0.858	0.693	0.788	3214
1.20	1.128	0.936	0.606	0.788	4989

and surface ocean $\Delta^{14}C$ was relatively stable during the period 25,000 to 19,000 calendar years ago, it is not necessary to account for changes in atmospheric $\Delta^{14}C$ when interpreting these benthic-planktic offsets (*12*). Further, as the deep sea is continuously ventilated, the ^{14}C to C ratio for the deep sea reflects an average over a time interval of on the order of a millennium rather than a single point in time. The results suggest that waters at 2000 m during the LGM in the western equatorial Pacific were no older with respect to surface waters than today and could actually have been slightly better ventilated at around 19,000 calendar years ago. As these results represent only the mid-depth of the Pacific Ocean, it is possible that the deep high-salinity layer documented for glacial time by Adkins *et al*. (*13*) had a much lower ^{14}C to C ratio. However, as more than half the ocean lies above 2000 m, if the discrepancy pointed out by Hughen *et al*. (*1*) and Muscheler *et al*. (*2*) is to be explained by a slower deep-sea ventilation rate during glacial time, the age of this deep carbon in deeper waters would have to have been at least 4000 years.

Although no reliable ^{14}C data are available for these deeper waters, ^{13}C measurements of glacial-age benthic foraminifera from equatorial zone cores from 1.6 to 4.5 km depth show no particular trend (*14*). If the deep waters were ventilated much more slowly, one would expect that the benthic ^{13}C would have been more negative than that at 2 km. If anything, the data suggest slightly better ventilated waters below 3 km in the glacial Pacific.

The benthic-planktic age differences presented here are consistent with those published for the South China Sea where concordant ages between multiple planktic species were also obtained (*15*). Taken together, the results from these three locales (representing water depths of 2.0 ± 0.1 km in the western Pacific) suggest a dip in ventilation age centered at 19,000 calendar years (see Fig. 1). However, verification of this dip will require further measurements.

Table 2. Radiocarbon ages for two western equatorial Pacific deep-sea cores. The uncertainty quoted for mean planktic ages is the standard deviation about the mean age.

Depth (cm)	Species	^{14}C age (years)	1 sigma error (years)
	Morotai Basin MD98-2181 6°N 126°E 2.1 km		
1262–1268			
	Pulleniatina obliquiloculata	16,650	110
	Neogloboquadrina dutertrei	16,800	95
	Mix benthics	18,050	130
	Mean planktics	16,725	150
	Δ B-P	1,325	200
1270–1276			
	Wood 1	15,950	120
	Wood 2	15,970	120
	Globigerinoides sacculifer	16,480	120
	P. obliquiloculata	16,330	100
	P. obliquiloculata	16,760	110
	N. dutertrei	16,740	110
	Globorotalia tumida	16,290	110
	Mix benthics	17,690	130
	Mean planktics	16,520	150
	Δ B-P	1,170	220
1279–1285			
	P. obliquiloculata	16,900	130
	N. dutertrei	17,150	130
	Mix benthics	18,350	120
	Mean planktics	17,025	150
	Δ B-P	1,325	200
	Admiralty Island MD97-2138 1°S 146°E 1.9 km		
207–210			
	Globigerinoides sacculifer	18,780	120
	P. obliquiloculata	18,620	120
	N. dutertrei	18,270	120
	Globorotalia tumida	18,890	120
	N. dutertrei	18,620	120
	Mix benthics	20,590	140
	Mean planktics	18,636	150
	Δ B-P	1,954	220
211–215			
	Globigerinoides sacculifer	19,130	130
	P. obliquiloculata	18,970	130
	N. dutertrei	18,730	140
	Globorotalia tumida	18,960	120
	Mix benthics	20,590	150
	Mean planktics	18,950	200
	Δ B-P	1,640	250

LGM benthic–planktic age differences from the Sea of Okhotsk (*16*) and from off Japan (*17*), as well as the estimate based on a ^{230}Th-dated benthic coral from the Drake Passage (*18*), are consistent with those presented here (see Fig. 2). Taken together they suggest that, if anything, LGM ventilation rates were on the average somewhat less than today's.

Clearly, many more measurements will be needed to create an adequate radiocarbon inventory for the LGM ocean. As it is now clear that the formation of secondary calcite is a problem, it will be essential to obtain cross-checks by analyzing several coexisting planktic species. Further, as bioturbation, coupled with opposing gradients in the abundances of planktic and benthic foraminifera can give rise to sizable biases in the reconstructions of surface to deep radiocarbon differences (*19*, *20*), cores with high accumulation rates must be sought. This latter restriction eliminates most open ocean sites and focuses attention on sites located along continental margins. Another problem is that calcite dissolution limits the water depth at which such measurements can be conducted. In the Pacific and Indian Oceans, few calcitic shells are preserved at depths greater than 4400 m.

If it turns out that the reduced storage of ^{14}C in the deep sea is not the answer, then what are the alternatives? The Muscheler *et al*. (*2*) argument rests on the assumption that the ratio of the accumulation of ^{10}Be to accumulation of snow on the Greenland ice cap was not substantially different during the LGM than during the Holocene. Considering that the accumulation rate of ^{10}Be on Greenland is only one-third of the mean global production rate, this assumption rests on shaky ground. If, for example, the accumulation of ^{10}Be relative to snow was lower during glacial time, the discrepancy between the reconstructed ^{14}C and ^{10}Be production rates would disappear.

Hughen *et al*.'s (*1*) contention that variations in Earth's magnetic field strength are not large enough to explain the high LGM ^{14}C to C ratios for the atmosphere and surface ocean present a more serious challenge. These authors point out, however, that unexplained differences between simulated ^{14}C and their observations may be the result of uncertainties in the relation between ^{14}C production rate and geomagnetic intensity. Increasing the database of reliable paleo deep-ocean ventilation ages will provide important constraints for such reconstructions in the future (*21*, *22*).

References and Notes

1. K. Hughen *et al.*, *Science* **303**, 202 (2004).
2. R. Muscheler *et al.*, *Earth Planet. Sci. Lett.* **219**, 325 (2004).
3. W. S. Broecker, E. Clark, I. Hajdas, G. Bonani, *Paleoceanography* **19**, doi: 2003PA000974 (2004).
4. L. Beaufort, T. de Garidel-Thoron, A. C. Mix, N. G. Pisias, *Science* **293**, 2440 (2001).
5. Alex Wiedenhoeft of the Forest Products Laboratory in Madison, Wisconsin, advised us that it was not possible to identify milligram-sized fragments of tropical woods; rather, a hand-sized piece of wood would be required. He also commented that, once waterlogged, any wood would be dense enough to sink in seawater.
6. T. P. Guilderson, D. P. Schrag, M. A. Cane, *J. Clim.* **17**, 1147 (2004).
7. T. Guilderson, personal communication, 2004.
8. Because conventional radiocarbon ages are calculated using the Libby half life of 5568 years, we have used this value to calculate all age differences in this study for consistency.
9. E. Bard, *Paleoceanography* **3**, 635 (1988).
10. W. S. Broecker, W. C. Patzert, J. R. Toggweiler, M. Stuiver, *J. Geophys. Res.* **91**, 14345 (1986).
11. R. M. Key *et al.*, *Radiocarbon* **44**, 239 (2002).
12. J. F. Adkins, E. A. Boyle, *Paleoceanography* **12**, 337 (1997).
13. J. F. Adkins, K. McIntyre, D. P. Schrag, *Science* **298**, 1769 (2002).
14. K. Matsumoto, T. Oba, J. Lynch-Stieglitz, H. Yamamoto, *Quat. Sci. Rev.* **21**, 1693 (2002).
15. W. S. Broecker, T.-H. Peng, S. Trumbore, G. Bonani, W. Wolfli, *Global Biogeochem. Cycles* **4**, 103 (1990).
16. L. D. Keigwin, *J. Oceanogr.* **58**, 421 (2002).
17. K. Ohkushi, M. Uchida, N. Ahagon, T. Mishima, T. Kanematsu, *Nucl. Instrum. Methods Phys. Res. Sec. B*, **406**, 223 (2004).
18. S. J. Goldstein, D. W. Lea, S. Chakraborty, M. Kashgarian, M. T. Murrell, *Earth Planet. Sci. Lett.* **193**, 167 (2001).
19. M. Andree, *Radiocarbon* **29**, 169 (1987).
20. W. S. Broecker, K. Matsumoto, E. Clark, I. Hajdas, G. Bonani, *Paleoceanography* **14**, 431 (1999).
21. E. L. Sikes, C. R. Samson, T. P. Guilderson, W. R. Howard, *Nature* **405**, 555 (2000).
22. Four benthic-planktic age differences for LGM samples from cores off New Zealand are reported in (*21*): 2.12 thousand years (ky) (depth 2.07 km, planktic ^{14}C age 24.4 ky); 0.84 ky (depth 2.07 km, planktic ^{14}C age 25.6 ky); 3.48 ky (depth 2.7 km, planktic ^{14}C age 24.1 ky); and 0.75 ky (depth 1.3 km, planktic ^{14}C age 24.8 ky). We choose not to include them because they represent a different age range from those that we studied.

23. We thank L. Beaufort for providing the Admiralty core sample and T. Guilderson for providing a survey of coral-derived prebomb Δ^{14}C values. This paper is funded in part by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. Financial support was provided by NSF grant OCE 02-21979 and by NOAA Consortium on the Ocean's Role in Climate (CORC) grant UCSIO P.O. 10196097-003/NA17RJ1231. This study was funded in part by a Lamont-Doherty postdoctoral fellowship and a fellowship grant (award number CSEF CC3B) from the Comer Science and Education Foundation to S.B., Lamont-Doherty Earth Observatory contribution no. 6663.

2 July 2004; accepted 8 October 2004

Prospects for Building the Tree of Life from Large Sequence Databases

Amy C. Driskell,[1,2]* Cécile Ané,[1]†‡ J. Gordon Burleigh,[1]†
Michelle M. McMahon,[1]† Brian C. O'Meara,[2]†
Michael J. Sanderson[1]

We assess the phylogenetic potential of ~300,000 protein sequences sampled from Swiss-Prot and GenBank. Although only a small subset of these data was potentially phylogenetically informative, this subset retained a substantial fraction of the original taxonomic diversity. Sampling biases in the databases necessitate building phylogenetic data sets that have large numbers of missing entries. However, an analysis of two "supermatrices" suggests that even data sets with as much as 92% missing data can provide insights into broad sections of the tree of life.

More than 100,000 species—about 6% of all those known to science—have at least one molecular sequence archived in public databases, but what fraction of these sequences is phylogenetically informative? Here, we examine two large samples of proteins and show how the answer depends on the pattern of homology among sequences and the distribution of sequences among taxa. We then parse these databases into phylogenetic supermatrices for metazoans and green plants. Although the databases have sampling biases that cause these matrices to be very sparse, they can still provide useful information for building the tree of life.

We examined the phylogenetic information content of the Swiss-Prot database of 120,000 sequences for nearly 7500 taxa and a "taxonomically enriched" subset of GenBank, which consisted of 185,000 amino acid sequences for more than 16,000 green plant taxa (*1*). Clusters of putative homologs were identified via $N \times N$ BLAST searches (*2*). Clustering procedures involve trade-offs among the reliability of homology assessment, the taxonomic breadth, and the accuracy of tree inference. The trade-offs are controlled by the stringency of homology searches and can be adjusted to maximize the phylogenetic utility of resulting clusters, on the basis of the depth and breadth of the phylogenetic question to be addressed (*3*, *4*). Clusters containing at least four taxa are termed "minimal phylogenetic clusters," because unrooted trees with fewer than four taxa contain no information about relations. Although minimal phylogenetic clusters were a small fraction of all clusters found [6.5% and 2.3% for Swiss-Prot and GenBank, respectively (Table 1)], they retained about one-third of the original sequences and a substantial fraction (74% and 95%) of the taxonomic diversity originally contained in the sample.

We screened minimal phylogenetic clusters for the presence of paralogs with a phylogenetic test of orthology (*5*). A species tree cannot be easily deduced from a cluster containing both orthologs and paralogs, although methods for this have been proposed (*6*–*8*). Screening reduced the candidate minimal clusters to smaller sets of orthologous "single-copy" clusters retaining only 24% and 21% of the original sequences, but still covering 59% and 89% of the original taxa in Swiss-Prot and GenBank, respectively (Table 1). These sequences are very sparsely distributed among taxa as measured by their "densities" (Table 1) (*9*).

Further assessment of the phylogenetic utility of these data requires consideration of how the data should be parsed for phylogenetic analyses. One approach is to build gene trees from individual clusters and to assemble these trees using supertree methods (*3*, *10*). Supertree methods require at least partial taxonomic overlap between trees. A set of trees (each inferred from a cluster) with enough taxonomic overlap to allow supertree construction is a "grove" (*1*). The minimum number of groves in a database is a lower bound on the number of supertrees required to encompass all its sequence data. The single-copy green plant proteins form at least 15 groves (Table 1). The largest of these groves minimally includes trees from 814 clusters and contains more than 14,000 taxa—87% of all the green plant taxa in the GenBank database. Swiss-Prot has at least eight times as many groves, which reflects its greater taxonomic breadth but higher fragmentation (Table 1). Both data sets also contain a small number of "orphans," clusters with no taxonomic overlap with other clusters.

---

[1] Section of Evolution and Ecology, [2] Center for Population Biology, University of California, One Shields Avenue, Davis, CA 95616, USA.

*To whom correspondence should be addressed. E-mail: acdriskell@ucdavis.edu

†These authors contributed equally to this work.

‡Present address: Department of Statistics, University of Wisconsin, Medical Science Center, 1300 University Avenue, Madison, WI 53706, USA.