foraminifera, which are subject to potentially large systematic errors.

The results show two periods during deglaciation when the bottom water at their site had unexpectedly low $^{14}C/^{12}C$ ratios. The water was so old that it must have been delivered to the site by upwelling from greater depth, presumably from below 2800 m. The oldest waters found by Marchitto *et al.* have an age of ~4000 years. For comparison, the oldest waters in the modern ocean have an age of ~2300 years.

The study provides the strongest evidence to date that the glacial ocean contained some very poorly ventilated water somewhere in its depths. The low-^{14}C periods coincide with the periods when atmospheric radiocarbon decreased and atmospheric CO_2 increased most rapidly during deglaciation. The results are thus a convincing fingerprint of a process that flushed excess carbon from an isolated deep reservoir toward the surface, thereby driving the atmospheric changes.

Today, waters below 2800 m are ventilated by two routes. One involves the sinking of aerated surface waters in the North Atlantic, the other sinking of such waters near Antarctica. During the last glacial period, both routes probably weakened, with the southern route possibly influenced by sea ice or surface freshening (see the figure, top panel). During glacial times, the deep ocean would thus have been less ventilated than it is today.

But how could low-^{14}C waters get to Marchitto *et al.*'s core site during deglaciation? Much of the upwelling of deep water occurs today around Antarctica, resulting in the formation of Antarctic Intermediate Water, a low-salinity water mass that spreads northward at intermediate depths. Marchitto *et al.* hypothesize that a similar process occurred during deglaciation, allowing upwelled water to spread northward to their site (see the figure, bottom panel). However, the evidence for this southern pathway is circumstantial.

The results help to reconcile the reconstructed trends in atmospheric radiocarbon with the estimated trends in the production of radiocarbon by cosmic rays—a comparison that seems to demand an increase in ocean ventilation during deglaciation (*4*). They support theories that attribute the bulk of the glacial-interglacial CO_2 change to changes in ocean ventilation (*5*, *6*).

The study also provides support for a theory for how the glacial ocean differed from today's ocean as a result of the cooling of deep waters to nearly the freezing point. Cooling to this extent is expected to allow the salty brine that is released during sea ice formation to accumulate more easily in the deep ocean (*7*). This idea is supported by sediment pore-water studies (*8*). By blocking the input of fresh water from precipitation, sea ice could also reduce the conversion of upwelled deep water into low-salinity Antarctic Intermediate Water (*7*). A strengthening of intermediate-water formation during deglaciation is consistent with a breakdown of this state caused by warming.

The study nevertheless leaves the skeptics with arrows in their quiver. Marchitto *et al.*'s low-^{14}C waters are so old that they start to stretch credibility, especially considering that the deep reservoir from which the water was drawn must have been even older. (This follows because some mixing with younger water would unavoidably have occurred during upwelling and transit to the site.) How could prior studies have overlooked deep waters this old?

If Marchitto *et al.*'s interpretation is correct, evidence for old water at intermediate depths should be present throughout the South Pacific in sediments of the appropriate age and depth. If subsequent work supports the findings, we may look back at this study as a key turning point in the quest to understand glacial and interglacial CO_2 changes.

References

1. T. M. Marchitto, S. J. Lehman, J. D. Ortiz, J. Flückiger, A. van Geen, *Science* **316**, 1456 (2007); published online 10 May 2007 (10.1126/science. 1138679).
2. W. Broecker *et al.*, *Paleoceanography* **22**, PA2206 (2007).
3. L. D. Keigwin, S. J. Lehman, M. S. Cook, *Eos* **87**, PP44A07 (2006).
4. K. Hughen *et al.*, *Science* **303**, 202 (2004).
5. J. R. Toggweiler, J. L. Russell, S. R. Carson, *Paleoceanography* **21**, PA2005 (2006).
6. B. B. Stephens, R. F. Keeling, *Nature* **404**, 171 (2000).
7. R. F. Keeling, B. B. Stephens, *Paleoceanography* **16**, 112 330 (2001).
8. J. F. Adkins, K. McIntyre, D. P. Schrag, *Science* **298**, 1769 (2002).

Published online 10 May 2007;
10.1126/science.1142326.
Include this information when citing this paper.

MOLECULAR BIOLOGY

Site-Seeing by Sequencing

An advance in DNA sequencing is a crucial component of a rapid, precise, and relatively inexpensive way to identify transcription factor binding sites at a whole-genome level.

Stanley Fields

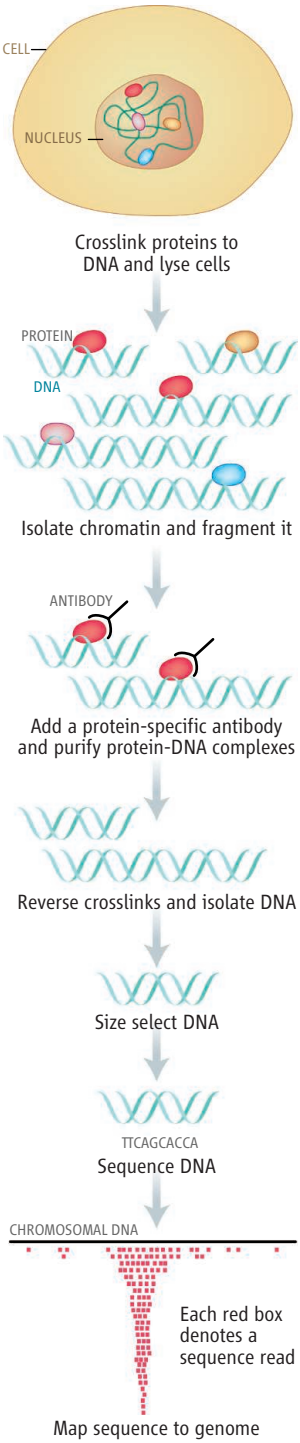
The author is in the Departments of Genome Sciences and Medicine, Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195–5065, USA. E-mail: fields@u.washington.edu

Every few years, a new technology comes along that dramatically changes how fundamental questions in biology are addressed. The impact of the technology is not always appreciated at first—when it is used only by those involved in its development—but becomes clear once the technology begins to spread to the broader scientific community. A well-known example is the DNA microarray, which became widely available to biologists about a decade ago and has since been applied to an ever-expanding set of questions such as determining the profile of genes expressed in a specific cell type. Now it is ultrahigh-throughput DNA sequencing that is making the transition from development to widespread use. Johnson and colleagues are in the vanguard of this movement. On page 1497 of this issue (*1*), they report that an advanced DNA sequencing technology (from Solexa/Illumina) can be used to identify all the locations in the human genome where a specific protein binds. They do this with a speed and precision that goes beyond what has been achieved with previous technologies.

DNA-binding proteins control transcription, replication, DNA repair, and chromosome segregation. Given the importance of these proteins, identifying their binding sites throughout the genome has occupied much attention in recent years. The most common method of locating these sites within a living cell is known as chromatin immunoprecipitation (ChIP). In this approach, cells are treated with a reagent, typically formaldehyde, that crosslinks protein and DNA, and then the cells are lysed. Chromatin (the complex of proteins and DNA in chromosomes) is isolated, the DNA is sheared into small fragments, and an antibody is added to precipitate the protein and its associated DNA. The DNA that is liberated after reversal of the protein-DNA crosslinks is then analyzed. In the initial uses of this method, researchers analyzed the DNA to determine whether single genes were enriched by the immunoprecipitation.

The classic genomic application of this approach—which began about 5 years ago—analyzes not single genes but all precipitated DNA fragments by using them as probes on a DNA microarray, giving rise to the pithy (albeit repetitive) moniker: ChIP-chip. A microarray is an ordered arrangement of defined DNA fragments immobilized on a surface. It is used to identify DNA sequences present in a sample through the hybridization of complementary strands. Pioneered in the yeast *Saccharomyces cerevisiae* (*2*, *3*), ChIP-chip was quickly applied to mammalian cells to identify binding sites for transcription factors (*4*, *5*). More recently, it has found new applications in the analysis of the distribution across the genome of modified histones and histone binding proteins (*6*, *7*). To improve this technique, Johnson *et al.* dispense with the "chip" of ChIP-chip and identify protein-bound DNA fragments by direct DNA sequencing (see the figure), a method they call ChIPSeq.

The focus of their study is a protein called neuron-restrictive silencer factor or repressor element-1 silencing transcription factor (indicated here as NRSF), a mammalian transcriptional repressor that silences the expression of neuronal genes in nonneuronal cell types and in neuronal progenitor cells (*8*, *9*). Many NRSF binding sites have been well characterized and comprise a 21–base pair DNA sequence motif containing two nonidentical half-sites of 10 base pairs each. Johnson *et al.* recovered DNA samples from chromatin treated with a monoclonal antibody against NRSF and mapped the DNAs to the human genome by determining their nucleotide sequences (see the figure). Control DNA samples were derived from chromatin not treated with the antibody. For this approach to work, they needed to sequence many fragments, and they did: Two to 5 million sequences were "read" at 25 nucleotides per read. They then used an algorithm they developed to map these reads to the genome and identify regions where reads cluster together. They found all locations in the genome that met two criteria: at least 13 independent sequence reads and an enrichment of at least fivefold relative to the control. The largest cluster contained 6718 reads; in other words, a single NRSF binding site was found 6718 times in the sequence data.



Protein binding, across the genome. The ChIPSeq method described by Johnson *et al.* identifies binding sites across the whole human genome for a specific protein. The control for this procedure omits the antibody step. The final panel shows a cluster of individual sequence reads (red boxes) that map back to the same region of the genome and locate a protein binding site.

What does this study reveal about NRSF binding sites? NRSF binding was detected at nearly all of its canonical motifs in the genome, indicating that all sites are accessible to the protein in the cell type analyzed (a human T cell line). Most sites bound by the factor were also identified, indicating that the sequencing approach is comprehensive. As befits a repressor, binding of NRSF near promoters (DNA regions where transcription factors bind to control their target genes) correlated with low levels of transcription of the associated genes. New binding motifs were also discovered, including those having two half-sites with noncanonical spacing between them, and those composed of only individual half-sites. Finally, genes bound to NRSF are highly enriched for functions involved in synaptic transmission and nervous system development.

What are the advantages of ChIPSeq over ChIP-chip? For one, the whole genome can be assayed by a sequencing approach, rather than only those DNA regions captured on a microarray. However, this advantage diminishes as genome tiling arrays, which present the whole genome in an arrangement of overlapping DNA fragments, approach the resolution of just a few nucleotide bases. ChIPSeq also avoids the complications of array hybridization, such as probes with different optimal temperatures for binding to their complementary strands, probes that hybridize to more than one DNA sequence, and interference of hybridization by DNA secondary structure. Furthermore, ChIPSeq, as currently performed, is only half the cost of human whole-genome tiling arrays. Perhaps most usefully—given the increasing number of genome sequences now available—ChIPSeq can immediately be applied to any of those genomes, rather than only those for which microarrays are available.

With the success of this study, the localization of binding sites for many other proteins—including transcription factors, structural components of chromatin, modified histone proteins, and the enzymes that modify them—will likely be mapped by ultrahigh-throughput DNA sequencing. However, current high-throughput sequencing approaches (Solexa/Illumina and the 454 Life Sciences platforms) are not limited to identifying the sites in the genome visited by DNA binding proteins. They will find considerable use in resequencing genomes; a recent well-publicized example is the genome of James D. Watson. They will be used to discover new genes, such as those encoding small RNAs and microRNAs. And as Johnson *et al.* point out, the large number of individual sequence reads provides a direct count of the sequences present in any sample. Thus, gene expression profiles of cells and tissues, comparative genome hybridization between, for example, DNA from normal versus tumor cells, the messenger RNAs that are present on polysomes and being translated to proteins, and numerous other nucleic acid measurements can be accomplished easily, cheaply, and accurately by a sequencing approach. The technology that is most threatened by the widespread adoption of ultrahigh-throughput sequencing? The DNA microarray.

References

1. D. S. Johnson, A. Mortazavi, R. M. Myers, B. Wold, *Science* **316**, 1497 (2007).
2. B. Ren *et al.*, *Science* **290**, 2306 (2000).
3. V. R. Iyer *et al.*, *Nature* **409**, 533 (2001).
4. C. E. Horak *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2924 (2002).
5. A. S. Weinman *et al.*, *Genes Dev.* **16**, 235 (2002).
6. C. L. Liu *et al.*, *PLoS Biol.* **3**, e328 (2005).
7. D. K. Pokholok *et al.*, *Cell* **122**, 517 (2005).
8. J. A. Chong *et al.*, *Cell* **80**, 949 (1995).
9. C. J. Schoenherr, D. J. Anderson, *Science* **267**, 1360 (1995).

10.1126/science.1144479