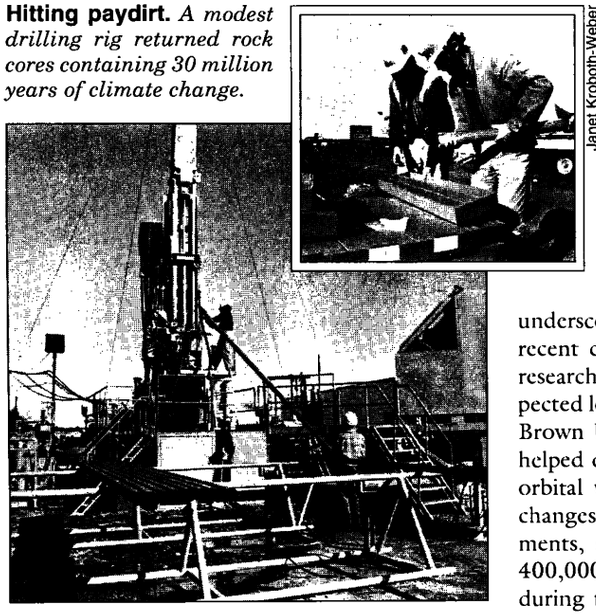

Janet Kroboth-Weber

**Hitting paydirt.** *A modest drilling rig returned rock cores containing 30 million years of climate change.*

The sediments accumulating on the lake bottom changed from black or gray organic-rich ooze when the water was deep to reddish cracked muds patterned with dinosaur footprints when the lake dried up.

These markers of the lake's repeated emptying and filling were laid down one on top of another like a deck of cards 7 kilometers thick, but tectonic forces have since tilted the deposit. Instead of needing a 7-kilometer hole to retrieve a complete sample of the deposit, Olsen and Kent could make do with a string of relatively shallow—and therefore inexpensive—holes, spread out so that each hole pierced a different part of the deposit (*Science*, 11 January, p. 158).

The drilling—six 1-kilometer holes in central New Jersey—was completed at the end of last month. This week Olsen, Kent, and Bruce Cornet of Lamont reported at the meeting of the American Geophysical Union in Baltimore that their cores showed climate periodicities aplenty operating between 200 million and 230 million years ago. The familiar 20,000- and 100,000-year cycles stood out in the first 4 million years of record they have analyzed. Even more prominent, though, was a 400,000-year cycle, which was evident in a pronounced intensification of the dry extremes of the 100,000-year cycle.

Olsen has not analyzed enough core to be very quantitative about even longer periodicities, but "it's clear by visual examination" of the cores that lake levels also varied with a period of about 1.6 million to 2 million years, he says. That's close to an orbital variation having a period of 2.3 million years. And about every 4 million to 6 million years the Newark Basin record shows a million-year period of extreme dryness. Andre Berger of the Catholic University of Louvain, Belgium, who calculates orbital variations, thinks that shorter orbital periods might interact to drive this longest climate cycle.

The appearance of all these stately climate cycles in the ancient record may be less surprising than what it underscores: their apparent absence in recent climatic history. It's not that researchers haven't looked for the expected longer periods. John Imbrie of Brown University, who in the 1970s helped demonstrate the link between orbital variations and recent climate changes recorded in deep-sea sediments, sees little evidence that the 400,000-year variation had any effect during the past 2 million years. If it did, he says, the effect was subtle, while "the 100,000-year period jumps out and shouts at you."

Imbrie suggests that the shorter period's recent dominance probably reflects the other dramatic differences between our world and the one Olsen and Kent are probing—in particular the ice sheets that now cover Antarctica and Greenland. Climatologists have assumed that the ice has somehow acted to amplify the 100,000-year period at the expense of the 400,000-year one—although so far all the suggested ice-related mechanisms act to amplify both periods equally. Imbrie calls this "the 400,000-year problem."

Olsen has a different view of the problem. "I think there may be more of a problem with the climate recorder than the climate system," he says. Rather than worrying about how the climate system could amplify one period and not the other, he says, paleoceanographers should be considering whether their marine sediments have faithfully recorded the 400,000-year cycle.

But if there really is a 400,000-year problem in recent climate history, the Newark Basin record may help researchers resolve it. If they can understand why the longest orbital cycles could drive climate oscillations in the ice-free, single-continent world of 200 million years ago, they may be able to understand why the cycles are absent or inconspicuous in our very different world.

The regularity of the cycles preserved in the Newark Basin drill cores may also make them useful as a timetable for other ancient events. Among them are reversals of Earth's magnetic field, which are recorded in tiny grains of magnetic minerals trapped in the ancient lake sediments. Measuring the duration of the reversals would fill in a major gap in the existing paleomagnetic timetable.

Kent intends to use the fleshed-out paleomagnetic record to date a giant impact that may have scattered debris across the lakes that filled the ancient Newark Basin. He hopes to find out whether the impact coincided with a mass extinction that took place about 205 million years ago (*Science*, 11 January, p. 161). Ironically, climate periodicities and extinction by impact are both subjects that were shunned until recent years as a bit too disreputable. How nice that the two outcasts might now help each other out. ■ RICHARD A. KERR

# Finding DNA Sequencing Errors

The DNA sequence databases are chock full of errors, and investigators should clean up their act. That was the message delivered by molecular biologist Richard Roberts at a recent genome mapping and sequencing meeting at Cold Spring Harbor Laboratory, where he runs his lab. But Roberts did more than chide his colleagues for submitting error-ridden sequences to the databases: He offered a handy computer program to help them spot mistakes.

In doing so, Roberts waded right into a heated debate among human genome project researchers on just how accurate the sequence needs to be. Some have argued that, considering the huge expense of getting the sequence absolutely right, slightly sloppy sequences might be just fine. But Roberts clearly isn't buying that argument. "It is crazy not to get the basic set of genes right," said Roberts, referring to the 50,000 or so genes presumed to exist in the human genome. "It is worth spending extra money on, because it will save money downstream." Since searching the databases for related protein sequences is usually the first step in figuring out what a newly discovered gene does, inaccurate sequence data can mislead investigators or cause them to miss a match entirely, so that they end up not knowing what they have uncovered.

To David Lipman, who is setting up a massive new genome database at the National Library of Medicine, Roberts' call for cleaner data is right on the mark. "There is an incredible number of errors in the database. Real biology is being missed because of errors. What Rich is saying is, 'Wake up and realize there are things you can do that are actually fairly simple to help pick up

errors, so there is no excuse not to.' I think the savvy groups are already doing so."

There are two types of sequencing errors: the substitution of one base for another; and the insertion or deletion of one or more bases. Substitution errors are not particularly bothersome, say numerous sequencers, but insertions and deletions can be catastrophic. That is because of the nature of the genetic code, in which a group of three nucleotides encodes one amino acid. If an extra base is thrown in, it causes a shift in the reading frame used to translate the code, thereby confounding efforts to work out the amino acid sequence of the protein encoded by the gene.

What Roberts has come up with, in collaboration with Janos Posfai of the Institute of Biophysics at the Hungarian Academy of Science, is a simple new program called DETECT that finds these potential frameshift errors. It works by comparing all possible translations of the newly determined DNA sequence to every sequence in a protein database. The program then looks for regions of the new sequence that would resemble a protein already in the database if a base were added or removed.

Finding such a "hit" doesn't necessarily mean that the new sequence is in error, Roberts points out, but rather shows places where the data should be rechecked. Sometimes it is the database sequence that is wrong. And if no error is found, then "something biologically interesting" is likely to be going on, says Roberts. Indeed, he has already found an example of a phenomenon known as ribosomal frameshifting, in which a gene contains extra bases in the coding region that must be skipped over during translation.

Roberts and Posfai have tested their program on sequences in GenBank that flank known bacterial genes, reasoning that these regions were more likely to contain errors than the well-studied genes. They examined a total of 6000 unidentified reading frames, spanning 1.3 million bases—about 4% of the database. "In 156 cases, the program predicted what we consider clear errors," says Roberts. Based on that first run, Roberts suspects that the program might detect several thousand errors in GenBank.

Roberts is quick to point out that DETECT is just a first step in developing computer tools for sequence error detection. The program's greatest limitation is that it works only for those genes that have known relatives already in the database; only about 30% of newly determined sequences now do. But as more and more of the human genome is sequenced, as well as the genomes of other organisms, the chances of finding a match will shoot up quickly.

Reaction at the Cold Spring Harbor meeting was positive, though Sydney Brenner of the Medical Research Council Laboratory in Cambridge noted that Roberts and Posfai have simply automated what many seasoned sequencers have been doing by hand for some time. But it is the amateurs that Roberts is hoping to convert when he makes the program available in a few weeks. "Most of the people who are sequencing don't do it full time, and they make all the errors that amateurs make." What Roberts wants to instill in these occasional sequencers is "the idea that one should try to find errors *before* saying the sequence is finished," which means before publication.

Even if Roberts fails to get that message across, all is not lost, according to new work by David States of the National Library of Medicine and David Botstein of Stanford, which will be published soon in the *Proceedings of the National Academy of Sciences*. Reasoning that some error is inevitable, States and Botstein set out to determine just how flawed a sequence can be and still be useful, in terms of detecting similar proteins in the database. To their surprise, they found that relatively inaccurate data, with up to 5% substitution errors and 1% frameshift errors, could detect similarities, even to distant relatives—provided the researcher knows there are errors in the sequence and ideally knows roughly where they are. But, States and Botstein say, their findings are not a license to be sloppy: "Our view remains that the goal in molecular sequencing is as high an accuracy as can be practically and economically achieved." ■ LESLIE ROBERTS

# Rhino Biology: Keeping Tabs on an Endangered Species

*Since humans began taking over their territory, the rhino populations of Asia and Africa have come near extinction, plummeting from an estimated million at the turn of the century to just 11,000 now. Earlier this month, some 300 scientists and conservationists from around the world gathered at an International Conference on Rhinoceros Biology and Conservation, sponsored by the Zoological Society of San Diego, to compare notes on what can be done to save these endangered mammals.*

## Bursting Bottlenecks

In the 1950s the eradication of malaria in the Ganges River plain of northern India and Nepal enabled farmers and poachers to move into the tall grasslands there. The result: The intruders pushed the Indian or greater one-horned rhinoceros to the brink of extinction. By 1962 fewer than 80 of the creatures remained in what is now the Royal Chitwan National Park in Nepal.

With so few breeding animals left, the Chitwan Park rhinos have long been assumed to be doubly threatened. Like several other near-extinct species—including African rhinos—they were thought to have dramatically low genetic diversity, which would further weaken their survival chances. But when population geneticist Gary McCracken of the University of Tennessee measured genetic variability in the Chitwan Park rhinos, which now number about 400, he got a welcome surprise: It approaches the highest levels ever reported for free-ranging mammals.

"This is good news because it means we needn't expect that all large mammals that have gone through recent bottlenecks should be genetic paupers," says San Diego Zoo geneticist Oliver Ryder. It also leads McCracken and collaborator Eric Dinerstein to suggest that population bottlenecks may have been overemphasized as a cause of low genetic diversity in other threatened species.

Studies performed on more than 200 mammalian species since the early 1970s have shown that, on average, about 4% of an animal's genes are heterozygous, meaning that different variants of the genes were inherited from the mother and father. Genetic heterozygosity in the endangered African rhinos, both blacks and whites, has been measured at less than 2%. Yet similar analyses by McCracken on 23 Chitwan rhinos show an average heterozygosity of almost 10%.

To explain this unexpected finding, McCracken and Dinerstein propose that Indian rhinos acquired high levels of diversity and their African cousins, low levels, before either species suffered its precipitous decline in numbers. Dinerstein, a wildlife ecologist formerly with the Smithsonian/Nepal Terai Ecology Project and now with the World Wildlife Fund, points out that prior to the 15th century, perhaps half a million Indian rhinos ranged in a swath from northwestern Burma, across the floodplain of the Ganges, to northern Pakistan. The creature's high