cient to be preserved in the borehole temperature profile.

Practitioners of borehole-temperature paleothermometry cope with the obstacle of limited thermal memory by striking a balance between opposing goals. One goal is fidelity between borehole-temperature observations and the predicted borehole temperature profile based on the inferred surface temperature history. The other goal is simplicity of the inferred surface-temperature history, namely, sufficient simplicity to suppress the wild oscillations that would otherwise be introduced by random error in temperature measurement. These two goals are opposing, because improvements in satisfying one goal (say, simplicity of the temperature history) will detract from the satisfaction of the other (fitting borehole-temperature data).

Cuffey *et al.* strike the balance between the opposing goals in a remarkably adept manner that reflects their aim to challenge the ice-core isotopic paleothermometer (*3*). For the goal of simplicity, Cuffey *et al.* demand that the inferred surface-temperature history be linearly related to only the broadest time-scale patterns of the oxygen-isotope record. For the goal of fidelity, they minimize a least-square measure of misfit between the observed borehole temperature log and the temperature profile predicted by the inferred surface-temperature history. (This prediction involves a sophisticated ice-sheet temperature model, which is one of the technical highlights of their analysis.) The adjustable parameters, which are tuned to achieve this minimization, are simply the slope and intercept of the linear relation defining the ice-core oxygen-isotope paleothermometer. As can be appreciated from figure 2 of the report by Cuffey *et al.*, the twin opposing goals are achieved, and the ice-core oxygen-isotope paleothermometer is both confirmed and calibrated.

This success is qualified, however, by its applicability to only the broadest time-scale feature of the surface-temperature history of the GISP2 site, that is, the glacial to interglacial warm-up. Other interesting climate features implied by the ice-core isotopic record, such as the 10,000-year-old Younger Dryas event or the presence or lack of the controversial Eem/Sangamon cold spells, may never be confirmed. These ancient and short-lived features fall below the resolution threshold of borehole-temperature paleothermometry (*7*).

For scientists who generate and interpret ice-core paleoclimate data, the success of the analysis by Cuffey *et al.* is satisfying because it culminates several decades of effort to use borehole-temperature paleothermometry as a check on other techniques for deriving environmental histories. Of more general interest, their work implies that past estimates of the glacial to interglacial warm-up derived from the ice-core isotope records have been too conservative (that is, were 8°C when they should have been 16°C) (*8*). This deduction is in accord with similar conclusions drawn recently about glacial-period temperatures in the tropics (*9*) and confirms that polar amplification of climate change (for instance, the poles respond with greater amplitude than the tropics) is a central characteristic of Earth's climate.

References and Notes

1. Lord Kelvin, *Trans. R. Soc. Edinburgh* **23**, 157 (1864).
2. F. M. Richter, *J. Geol.* **94**, 395 (1986).
3. K. M. Cuffey *et al.*, *Science* **270**, 455 (1995).
4. R. A. Kerr, *ibid.* **260**, 890 (1993).
5. See exercise 2.06(iii) of R. L. Parker, *Geophysical Inverse Theory* (Princeton Univ. Press, Princeton, NJ, 1994).
6. For example, see D. R. MacAyeal, J. Firestone, E. D. Waddington, *J. Glaciol.* **37**, 326 (1991). See also, D. Dahl-Jensen, S. J. Johnsen, W. S. B. Paterson, C. Ritz, *ibid.* **39**, 421 (1993).
7. J. Firestone, *ibid.* **41**, 39 (1995).
8. This conclusion, that the glacial to interglacial warm-up was much greater than previous interpretations of Greenland ice-core isotopic records indicated, was also achieved in a similar analysis by Cuffey's predecessor at the Geophysics Program of the University of Washington, J. Firestone. Firestone's study involved data from a Greenland ice-core site that was not as well suited for borehole-temperature paleothermometry as the GISP2 site. See (*7*).
9. L. G. Thompson *et al.*, *Science* **269**, 46 (1995). See also W. S. Broecker, *Nature* **376**, 212 (1995).

Life With 482 Genes

André Goffeau

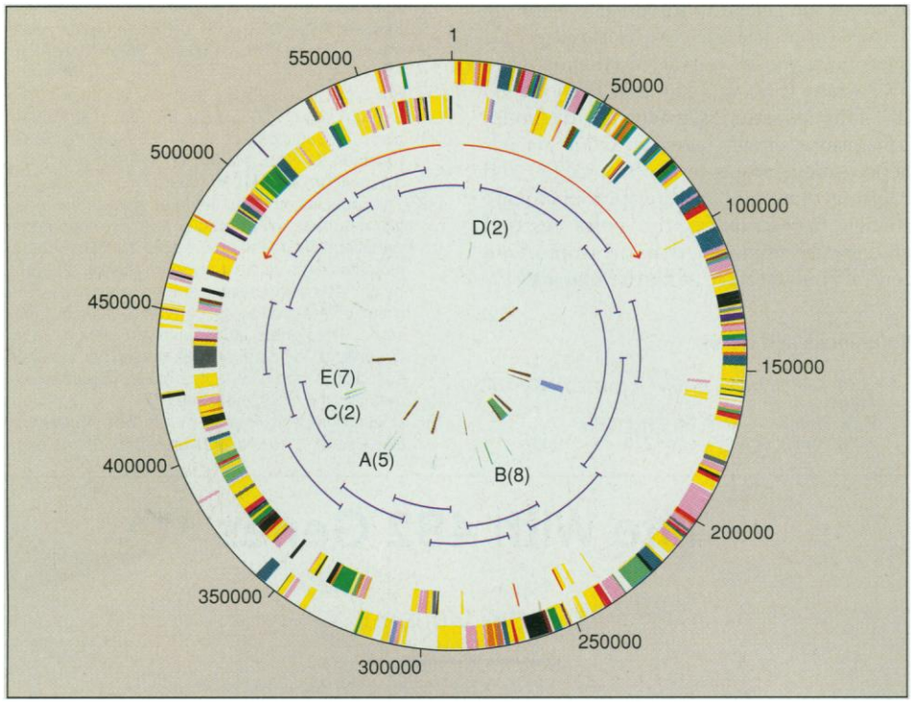
For many years we believed that the first genome to be sequenced entirely would be that of the bacterium *Escherichia coli*, estimated to be 4720 kilobases long (*1*). More recently, the prospect of completing the much longer genome of the yeast *Saccharomyces cerevisiae* (12,500 kilobases) before the end of 1995 has been entertained (*2*). But to everyone's surprise, an outsider won the race for the first complete genome sequence—that of the bacterium *Haemophilus influenzae*, a 1830-kb sequence, recently reported by a team from the Institute for Genomic Research (TIGR) headed by Craig Venter (*3*). This "premiere" was performed on a very small bacterial genome, and not on the long mammalian fragment on which one might have expected TIGR to have concentrated its powerful resources. Now, Venter and his colleagues have focused on an even smaller genome, that of *Mycoplasma genitalium*, the complete sequence of which is reported in the paper by Fraser *et al.* in this issue of *Science* (*4*). This parasite (but not necessarily a pathogen) of human genital and respiratory tracts has a genome of only 580-kb long.

Sequencing of the M. *genitalium* genome was the product of a collaboration among three teams. Foremost among these is TIGR, a nonprofit institute with large-scale, high-throughput DNA sequencing facilities that has given the rights to commercialize its findings to an allied company, Human Genome Sciences, Inc. TIGR and its director Craig Venter are known mainly for their mass production of sequence tags for human genes (ESTs) expressed in different tissues (*5*). The second team is led by Hamilton Smith from the Johns Hopkins University School of Medicine, who is best known for his pioneering work on restriction enzymes in bacteria, work that opened the field of molecular genetics and for which he won the Nobel Prize in 1978. The third team is that of Clyde Hutchison from the University of North Carolina, who is an internationally recognized expert in the study of *Mycoplasma* species.

One of the most impressive features of the sequencing effort for the M. *genitalium* genome is its efficiency, a testament to the power of the TIGR sequencing and informatics facilities. The first DNA extraction from M. *genitalium* was carried out in early January 1995, and the manuscript was submitted on 11 August 1995.

On the technical side, the most spectacular aspect of the work is the application of random ("shotgun") sequencing and assembly of the 8650 needed sequencing reactions in a single contig (that is, a collection of overlapping clones that collectively cover the target region). This accomplishment was made possible by the development of a series of highly performing informatics tools already tested during the sequencing of the *H. influenzae* genome. On the scientific front, the originality of the work stems from the fact that the M. *genitalium* genome has one of the smallest known genomes of any free-living organism. It is therefore reasonable to assume that its genome sequence reveals the near-

The author is at the Université Catholique de Louvain, Unité de Biochimie Physiologique, Place Croix du Sud, 2-20, B 1348 Louvain-la-Neuve, Belgium.



A circular representation of the *M. genitalium* chromosome. Outer concentric circle: Coding regions on the plus strand for which a gene was identified. Second concentric circle: Coding regions on the minus strand for which a gene was identified. Third concentric circle: The direction of transcription on each strand of the chromosome starting at the putative origin of replication. Fourth concentric circle: Cosmid and lambda clones (blue). Fifth concentric circle: The locations of the single ribosomal operon (blue) and the 33 tRNAs. The clusters of tRNAs (trnA, trnB, trnC, trnD, and trnE) are indicated by the letters A through E with the number of tRNAs in each cluster listed in parentheses. Sixth concentric circle: Location of the MgPa operon (green) and MgPa repeat fragments (brown). Figure generated with TIGR Genome Display Tool (J. Slagel, unpublished data). [Figure from C. Fraser *et al.*, unpublished data]

minimal set of genes necessary for independent life. Whether this means that each of the predicted coding regions is essential for growth remains to be determined experimentally. Nevertheless, it is likely that the minimal translation machinery requires nearly 90 different proteins to proceed, whereas the complete DNA replication process requires only about 30 proteins. Many other observations can be made on minimalist metabolic or physiological pathways. It is surprising, for example, that this bacterium, which contains only one type of membrane, has devoted 140 (30%) of its 482 genes to encode membrane-inserted proteins. Also unexpected is the observation that up to 4.5% of the genome might be used for evasion of the mammalian host immune response. The scope of our present ignorance is measured by the fact that 117 *M. genitalium* protein sequences (22%) do not match protein sequences from any other organism.

From an evolutionary perspective, sequencing of the *M. genitalium* genome represents the first complete molecular definition of minimal life and will likely become a cornerstone for future comparisons of the genome contents from many species. An example of such an approach is given by the comparison of the gene content of *H. influenzae* and *M. genitalium* sorted by functional category; for instance, *H. influenzae* devotes more than 10 times as many genes to regulatory functions than *M. genitalium* (64 genes compared with 5).

How rapidly will the new information be exploited at the biochemical level? Two facts might hinder such progress, which in principle could result in a definition of the minimal biochemical mechanisms and pathways required for life. One limiting factor might be the relatively small size of the scientific community actively working on molecular aspects of this bacteria (probably a few dozen compared to the thousands of molecular biologists studying yeast or *Escherichia coli*). Another difficulty is that the tools of classical or molecular genetics cannot be applied; no auxotrophic or other mutants are presently available for this parasite, which cannot grow on synthetic media. For further studies to take place, each of the *M. genitalium* genes should be overexpressed in a heterologous host—a procedure that remains cumbersome.

Another question concerns the true error rate of the sequence. The authors admit that about 1% of the total sequence was determined from only one DNA strand. This contrasts with the requirement for 100% coverage of both strands in other systematic sequencing projects, such as those of yeast or *Bacillus subtilis* genomes. Whether full coverage on the second strand is always necessary when the sequence on the opposite strand is known is indeed open to discussion. The authors estimate, rather intuitively, that their error rate is less than 1 in 10,000 bases. In some circles it is believed that this estimate, which corresponds to the highest accepted standards, should have been assessed by independent quality control on the basis of blind resequencing. The debate as to the measurement of exact error frequency should, however, in no way overeshadow this remarkable achievement, which will remain a landmark in contemporary biology.

References

1. Ph. Mordant, see table compiled from the University of Louvain-la-Neuve
2. N. Williams, *Science* **268**, 1560 (1995).
3. R. D. Fleischmann *et al.*, *ibid.* **269**, 496, 1995).
4. C. Fraser *et al.*, *ibid.* **270**, 397 (1995).
5. M. D. Adams *et al.*, *Nature* **377** *Suppl.* 3 (1995).