NEWS & VIEWS

HUMAN GENETICS

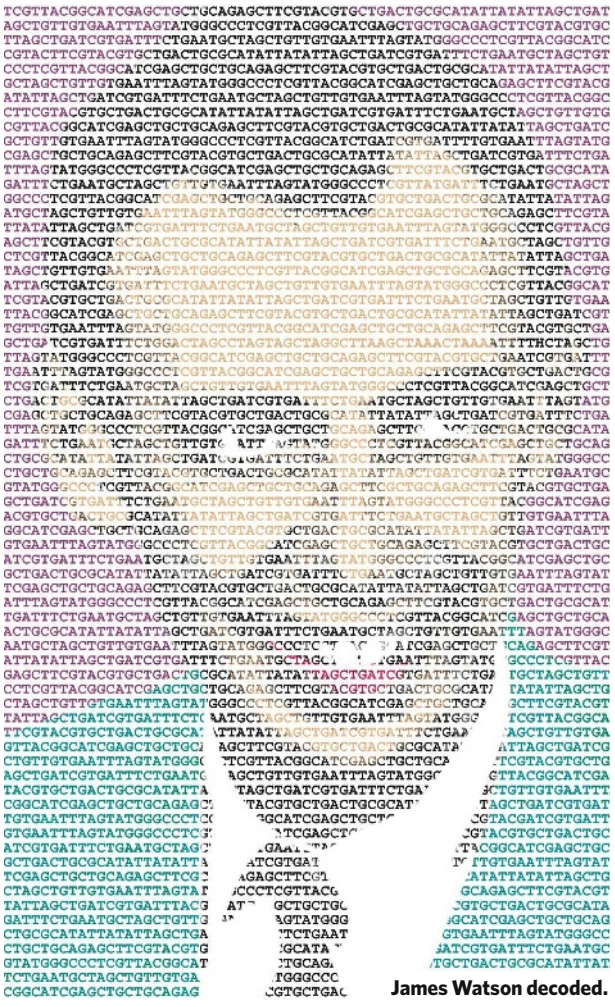
Dr Watson's base pairs

Maynard V. Olson

The application of new technology to sequence the genome of an individual yields few biological insights. Nonetheless, the feat heralds an era of 'personal genomics' based on cheap sequencing.

This issue of *Nature* contains a paper that is, in a curious way, a sequel to one published 55 years ago — the description by James Watson and Francis Crick[1] of the double-helical structure of DNA. At the information-carrying core of this beautiful structure, with its far-reaching implications for biology and medicine, are the base pairs that Watson discovered by fitting together cardboard cut-outs of the bases adenine, thymine, guanine and cytosine. Now, on page 872, Wheeler *et al.*[2] describe the use of massively parallel DNA sequencing to determine the order of the base pairs in Watson's own genome. This achievement is a technical *tour de force* that points towards routine use of whole-genome sequencing as a research tool in human genetics. Given the choice of James Watson as an identified research subject, the paper is also a conspicuous effort to publicize the arrival of the era of personal genomics and the willingness of a famous geneticist to put his genome sequence in the public domain.

Technically, the paper's interest stems from its reliance on a DNA-sequencing platform that differs greatly from the one used during the first great era of genome sequencing, which culminated in the Human Genome Project (HGP). In the HGP platform, each kilobase-pair fragment of genomic DNA was captured as a bacterial 'clone' using recombinant-DNA techniques and processed in its own microlitre-scale well in a microtitre plate. Following a series of biochemical steps, each sample was analysed electrophoretically in a dedicated, metre-long glass capillary. To achieve the required redundancy in sequence coverage, investigators in the HGP processed tens of millions of individual samples en route to determining the order of the 3 billion base pairs present in a single, composite instance of the human genome. This achievement required the development of industrial-scale genome centres that looked more like manufacturing plants than laboratories. The data-production costs alone were hundreds of millions of dollars.



James Watson decoded.

PHOTO BETTMANN/CORBIS

Wheeler *et al.*[2] used one of several new DNA-sequencing platforms that can achieve much the same result at perhaps 1% of the cost[3,4]. Note, however, that at their present stage of development none of these methods would allow a simple, bargain-basement replay of the HGP: analysis of the data remains heavily dependent on the high-quality reference sequence produced by that project. The key to the increased efficiency of the new methods lies in massive parallelization of the biochemical and measurement steps. The instruments used by Wheeler *et al.* are marketed by 454 Life Sciences, a component of Roche Diagnostics, which joined forces with the Human Genome Sequencing Center at Baylor College of Medicine in Houston, Texas, to sequence Watson's genome.

The 454 instruments achieve massive parallelization in two different ways[5]. In an initial step, single DNA molecules are attached to synthetic beads and then amplified enzymatically. During amplification, the beads are trapped in tiny water droplets within a water–oil emulsion; hence, more than 100,000 samples can be processed in parallel in a single test tube. In a later step, during which optical measurements are used to collect the actual sequencing data, each bead is confined to a picolitre-scale well etched into the end of a glass fibre within a fibre-optic bundle. Although costs have not yet dropped to the much-ballyhooed target of US$1,000 per genome[6], they are now low enough to make the era of personal genomics a reality rather than a distant dream.

What can we expect to learn from the sequences of individual genomes? The main lesson from the analyses by Wheeler *et al.* is that it will be extremely difficult to extract medically, or even biologically, reliable inferences from individual sequences. Consider the challenge of interpreting Watson's single nucleotide polymorphisms (SNPs — simple substitutions of one base for another at a particular site in the genome). Wheeler *et al.* report about 3,300,000 SNPs in Watson's genome relative to the HGP reference sequence. Of these SNPs, 82% had already been described in other individuals, a result suggesting that public databases are starting to have a good representation of common SNPs. Most of these variants are presumed neutral (that is, of no evolutionary or functional significance).

However, some 11,000 of Watson's SNPs (85% previously known, 15% novel) are predicted to change the amino-acid sequence — and so, perhaps, the function — of a protein. An unknown number of additional SNPs undoubtedly affect the regulation of protein levels. Because there are only around 20,000 protein-coding genes in humans, it seems that if the proteins of any two individuals were compared in detail, a significant fraction would be found to differ. In only a few cases do we have any knowledge of possible biological effects of these differences.

The level and overall pattern of variation in Watson's genome seem to be typical of the variation present in other individuals of predominantly European ancestry. In part, this conclusion rests on the existence of one other example of a personal-genome sequence, that of J. Craig Venter, who was head of Celera Genomics when it competed with the HGP to produce the first human genome sequence. The publication of Venter's sequence last year[7] attracted relatively little attention because the project used HGP-era methods and much of the data had already been reported. Nonetheless, the similarity of the bulk statistics for these two personal-genome sequences helps validate the new methods used by Wheeler and colleagues.

If Watson took his sequence to a genetic counsellor, there would be little to discuss. The sequence seems to show that he is a carrier for a handful of mutations that might catch a counsellor's interest. But these mutations have no known effects on Watson himself, and would confer risk on offspring only in the highly unlikely event of a marriage between two carriers. None of these mutations is ever likely to be considered an appropriate candidate for screening in the general population — of which, for these purposes, Watson is a representative member.

Recognition of the thin clinical value of this sequence may cause some investors in the new sequencing methods to take pause, given that the major capital investments required to commercialize these technologies have been motivated more by their perceived medical potential than by research applications.

Basic scientists are likely to greet the work of Wheeler *et al.*[2], and the era it inaugurates, with more enthusiasm. The central challenge in human genetics now is to learn how to correlate genotype with phenotype (the observable characteristics of an organism), with special attention to disease predisposition and response to therapy. Whole-genome sequencing of case-control populations is an attractive alternative to the surrogates for this procedure on which human geneticists have relied for the past 25 years. These surrogates frequently depend on educated guesses about which genes to sequence in subjects with particular phenotypes. They also rely on efforts to detect co-inheritance of a phenotype — in families or even among individuals who are only evolutionarily related — with 'genetic markers' (neutral DNA variants that can reveal the common ancestry of local segments of the genome). In these indirect 'linkage' or 'association' methods, the observed genetic markers can indicate the approximate genomic position of the unknown functional variants that influence the phenotype. In the era of whole-genome resequencing, all variants, including the functionally important ones, will be observed directly.

The actual practice of personalized medicine based on genome sequences will have to wait until we can make reliable predictions from the data. At present, we have little ability to do so. To take one seemingly simple example of a highly heritable trait, we could not even make a rough prediction of Watson's height from his genome sequence: the most informative known SNPs that influence height account for only a few millimetres of variation in a trait whose standard deviation is 7 centimetres (ref. 8).

In short, the symbolic significance of Wheeler and colleagues' paper[2] is greater than its immediate contributions to human biology. But it is irresistible to speculate on how biologically revealing the data will ultimately prove to be. James Watson is a brilliant scientist with a remarkable life story. He both laid the deep scientific foundations for genomic biology and devoted much of his life — through his teaching, his leadership and the sheer force of his personality — to building this science to its current productive state. Along the way, he stepped on more than one landmine. Future historians will find him a rich and elusive subject. Perhaps, informed by the advanced genetics of their day, they will scrutinize the data left behind by Wheeler *et al.* for clues to why he was the way he was. However, I suspect that they will have to rely instead, as historians do today, on what Watson wrote, said and did during his lifetime rather than on the order of the base pairs in his genome. ■

Maynard V. Olson is in the Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, Washington 98195-7720, USA.
e-mail: mvo@u.washington.edu

1. Watson, J. D. & Crick, F. H. C. *Nature* **171,** 737–738 (1953).
2. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
3. Bentley, D. R. *Curr. Opin. Genet. Dev.* **16,** 545–552 (2006).
4. Shendure, J., Mitra, R. D., Varma, C. & Church, G. M. *Nature Rev. Genet.* **5,** 335–344 (2004).
5. Margulies, M. *et al. Nature* **437,** 376–380 (2005).
6. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. *Nature* **422,** 835–847 (2003).
7. Levy, S. *et al. PLoS Biol.* **5,** e254 (2007).
8. Weedon, M. N. *et al. Nature Genet.* **39,** 1245–1250 (2007).

PLANETARY SCIENCE

Message from Mercury

H. Jay Melosh

After a 30-year gap, all eyes are back on Mercury as the MESSENGER probe gives us our second glance at the Sun's nearest neighbour. Hints of intriguing results to come are already at hand.

On 10 March, scientists from MESSENGER, NASA's mission to the planet Mercury, reported their first observations of the planet at the annual conference of the Lunar and Planetary Institute*. The probe — its name stands for 'Mercury Surface, Space Environment, Geochemistry, and Ranging'— was launched from Cape Canaveral on 3 August 2004. Its goal is to orbit the Solar System's innermost planet, and thus answer 30-year-old questions raised when the Mariner 10 spacecraft flew past the planet three times during 1974 and 1975.

Mercury sits deep inside the Sun's gravity well, making it more difficult to reach from Earth than even Pluto. No chemically fuelled spacecraft can achieve the velocity necessary to cancel out the motion of Earth's orbit and thus rendezvous directly with Mercury under its own steam. MESSENGER instead flew a tortuous path, using the gravity of Earth once and that of Venus twice to shed angular momentum and so to spiral down into the inner Solar System. Mission destination was finally reached on 14 January of this year, when the probe completed the first of three planned fly-bys at just 200 kilometres above Mercury's surface (S. Solomon, Carnegie Inst.).

Mariner 10 had already revealed the Moon-like nature of this surface, and returned images of a vast impact crater, the Caloris Basin, as well as evidence of a distinctly un-Moon-like uniformity of colour: Mercury's lavas seem to be almost as bright as the rest of its surface, and so do not produce the Man-in-the-Moon contrasts familiar from our satellite. Widespread lobe-shaped scarps, seemingly the result of compressive thrust faulting as the planet cooled in its early history, diversify the planet's surface.

Mariner 10 was never able to image Mercury's entire surface, and its camera systems provided only 1-km resolution over the 45% of the planet imaged, with a few images at a higher resolution of about 300 metres. MESSENGER's more sophisticated instruments have already confirmed and added

* 39th Lunar and Planetary Science Conference, 10–14 March 2008, League City, Texas.