

So much for genes

Adrian Woolfson

THE CENTURY OF THE GENE BY EVELYN FOX KELLER
Harvard, 186 pp, £15.95, October 2000, ISBN 0 674 00372 1

FROM Peckham Library to the Taj Mahal, the spines of a porcupine to the money bands of City traders, the flailings of a woodlouse emerging from a burning log or the whirlpool generated spontaneously in a bathtub, the Earth is graced with a multiplicity of structures. A great many of them are transient: snowflakes melting on a pair of ski boots, or the decaying remains of a Mayan temple. Others have the capacity to persist, replicate and transform themselves, sometimes into forms quite different from those of their parent structures. We would be greatly surprised if clusters of miniature Nelson's Columns were one day to appear in Trafalgar Square, emerging from the mother column like fledgling mushrooms. We are less surprised, however, when we observe a modest stick insect replicating its structure, almost exactly, to produce a collection of tiny, near-identical young. Indeed, this is the marvel of nature, the characteristic that has, hitherto, been used to divide animate from inanimate forms.

Modern biological thought has focused on four principal questions. What is the nature of the mechanism that enables the 'information' required for the replication of biological structures to be transmitted

across generations? How is a single cell able to differentiate into a complex organism? How are the structures of living things transformed across time? How did life originate in the first place? Any unifying theory must provide a plausible answer to all these questions and in so doing, offer tentative answers to two further ones: what is life? And what will it be like in the future? In spite of recent advances, the answers to the last three questions remain unclear. The concept of genes and the notion that biological transformation is effected by a Darwinian process of evolution by natural selection have, however, brought us much closer to answering all these questions, especially the first three.

The intellectual foundations, which placed genes at the centre of biological explanation, were established in 1900, with the simultaneous and independent rediscovery of Mendel's principles by Hugo de Vries, Carl Correns and Erich von Tschermak. But the chemical nature of genes remained an enigma for some time, the consensus being that the information they carried was encoded in proteins. Eventually, experiments published in 1944 demonstrated that it is DNA and not proteins that encodes the information on which heredi-

ty depends. The jewel in the crown of gene theory, however, came in 1953, when, with the help of X-ray photographs taken by Rosalind Franklin and Maurice Wilkins, Watson and Crick elucidated the three-dimensional structure of DNA.

The DNA-based paradigm of heredity with its accompanying 'genespeak' was strengthened by experimental studies employing a wide range of techniques enabling DNA to be cut, pasted, amplified, modified and manipulated in a great number of ways. Fred Sanger and Walter Gilbert's discovery in 1977 of methods for reading the sequence of DNA molecules raised the enticing possibility that the genetic 'texts' of living things might one day be read in their entirety. Indeed, in the same year, Sanger published the entire genomic sequence of the virus phi-X174. We now know the entire genetic specification for a human being; similar data are available for a host of bacteria, viruses and yeasts, as well as for two multicellular eukaryotes, the nematode worm *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*.

In *The Century of the Gene*, Evelyn Fox Keller states early on that her aim is to explore the impact the human genome project (which encompasses the sequencing of genomes other than the human) has had on both biological thought and practice. It is consequently a great surprise that she fails to address even a small amount of the extensive body of theoretical and experimental work that has issued directly from the monumental achievements of this international enterprise. In much the same

way she raises the interesting question of what the biology of the 21st century will look like, but having done so, makes no attempt to answer it. The century of the gene is, in Keller's view, neatly demarcated by the rediscovery of Mendel's principles of heredity in 1900 and the release of the rough draft of the human genome in 2000. She argues that, paradoxically and somewhat unexpectedly, the very success of the genome project has undermined the conceptual foundations on which it was predicated. Thus, rather than revealing the inner workings of human evolution, the deciphering of the DNA sequence has opened up a Pandora's box of complexity which had previously not been appreciated.

This information, Keller argues, has in some respects transformed our expectations, rather than fulfilling them. So, whereas the focus of the 20th century was on the gene and structural genomics, the post-genomic 21st century is centred on what is called functional genomics, the science of deciphering the DNA sequence that we are in the process of generating, and attributing biological meaning to the otherwise meaningless archive of chemical symbols. Indeed, Keller asserts, perhaps prematurely, that at the very moment when genes have reached their apotheosis, the inherent complexity of genomics has undermined their position as the central explanatory concept of modern biology. New concepts, she hopes, will loosen the 'powerful grip that genes have recently come to have on the popular imagination'.

Having said which, Keller acknowledges that gene talk has not yet outlived its use-

fulness and raises the question of what it is actually for. The 'gene paradigm', in its seductive simplicity, plays an important rhetorical role, genespeak having long become a powerful political tool that has captured the imaginations of a wide public, the biotechnology market and government funding agencies. But has our flirtation with the terminology exacted its own price? Will we spend much of this century backtracking from our original premises and attempting in vain to reprogramme the minds of a scientific community and public intoxicated with genespeak? And at what point, if ever, would science benefit from abandoning the now familiar lexicon?

Molecular genomics has, Keller says, taught us four principal lessons. First, that genes play a dynamic role in the evolutionary process: they can adjust the rate of their own evolution by striking a careful balance between the stability of the information they carry and their ability to mutate. This stability is not something intrinsic to DNA, but is rather achieved through the agency of enzymes. Second, that research findings have undermined our notion of what a gene is, and brought the concept to the verge of collapse. If it has any value at all, then it is only as a functional entity, which may not be localisable within a chromosome. But once genes are redefined in this way, it is meaningless to consider them independently from the metabolic processes that play a critical role in determining their function. Third, that the process of development by which a single cell differentiates into a multicellular organism cannot be understood simply in terms of the implemen-

tation of a genetic programme. The results of cloning experiments mean that it is no longer clear that genes cause development. Fourth, that the focus on an exclusively computational paradigm for the analysis of genomes has been at the expense of a cybernetic approach, which, by contrast, emphasises the importance of the overall organisation of biological networks and the emergent properties, like redundancy, that they exhibit. These are important insights and Keller is correct to highlight them. She is incorrect, however, in asserting that they derive from the human genome project: most of the studies they have resulted from predated the genome project or have only a very tangential relationship to it.

The historical notion of a gene as an atomic unit of heredity provided the constancy needed to explain how physical traits such as eye colour and height could be transmitted unchanged across generations. It was also intellectually satisfying, as it obviously fitted in with the parallel sciences of physics and chemistry. Indeed, de Vries wrote that 'just as physics and chemistry are based on the molecules and atoms, even so the biological sciences must penetrate to these units in order to explain by their combinations the phenomena of the living world.' Even Darwin hypothesised the existence of elemental units of heredity, which he called 'gemmules'. August Weisman, one of the greatest zoologists of the late 19th century, took things much further by speculating that the elements of heredity were stored in the substance 'of a definite chemical, and above all, molecular composition' which he called 'germ-

plasm'; and that the germ-plasm stored in germ cells was 'held in trust for coming generations'.

The elucidation of the three-dimensional structure of DNA molecules provided a new basis for understanding how stability is maintained across generations. As DNA consists of two complementary strands, one of them can function as a master copy that can be used as a template for the manufacture and repair of the other. But DNA alone is functionally inept: an insufficient cog in a complex, multi-component and high-fidelity inheritance machine. The normal functioning of DNA is critically dependent on the cell it is housed within and on an ongoing interaction with that cell's metabolic networks. As the stability of the information encoded within DNA is achieved by this dynamic means, it should not come as a surprise that the information's mutability can be regulated. Bacteria, for example, can make use of an emergency, rapid but error-prone mode of repair, which is activated when they are exposed to life-threatening conditions such as ultraviolet light.

It is likely that evolution has selected for genetic systems which are able to adjust their mutation rates to meet the demands of particular situations, and thus selected for 'evolvability': that is, the ability of a species to generate the raw genetic variations on which natural selection can act. The significance of seeing genetic stability as resulting from biological organisation, rather than from the structure of the gene itself, is that the emphasis is transferred from the storehouse (the gene) to the dynamic

mechanism (metabolic process) by which that information is computed and expressed.

The earlier notion that genes contain a programme able to compute the various activities of a cell is predicated on a model envisaging them as invariant entities. But this is far from being the case. Whereas it was possible in the 1940s for George Beadle and Edward Tatum to equate a gene with a region of DNA that encodes a single enzyme, we now know that things are more complex. The DNA sequence that encodes a protein is itself regulated by many elements which determine when and to what extent the gene is activated. The regions of DNA that encode the amino acids of any given protein are not, furthermore, organised in a linear, contiguous manner, but compartmentalised into regions of protein-encoding information known as exons, themselves separated by intervening sequences called introns. These must be cut in order to bring the exons together and generate a protein-encoding message. The way such editing occurs is not invariant: exons may be spliced together in different ways to generate several alternative forms of the target protein. A single gene can thus, potentially, generate large numbers of different proteins.

Several other tiers of complexity further complicate the way in which genetic information is realised. The DNA sequence may, for example, be read in different configurations, or may contain cryptic 'punctuation marks', or a particular gene may be 'silenced' by chemical modification. Moreover, genes which appear to map to a given

function can often be deleted without any observed loss of the accompanying function. Keller argues that these and other complicating factors ‘threaten to throw the very concept of “the gene” – either as a unit of structure or as a unit of function – into blatant disarray’, and goes on to quote the molecular biologist William Gelbart, who suggests that the gene might be a ‘concept past its time’.

Have we then reached a point when we need to invent a new vocabulary? In 1961, François Jacob and Jacques Monod concluded that ‘the genome contains not only a series of blue-prints, but a co-ordinated programme of protein synthesis and the means of controlling its execution.’ Recent advances in our understanding of developmental processes have, however, served to undermine the notion that the information contained in the programme governing the assembly of a three-dimensional organism resides exclusively within genes. So if genetic information cannot be assimilated to that encoded on the processor of a computer, how are we to understand it? John Gurdon demonstrated in the 1970s that the nucleus of a fully differentiated adult frog cell can be ‘reprogrammed’ by transferring it into a fertilised egg from which the nucleus has been removed: embryonic development follows. The same experiment was replicated in the mammalian cells of Dolly the sheep by Ian Wilmut and his colleagues in 1997. Such reprogramming indicates that the cytoplasm, which is the protoplasm of a cell surrounding the nucleus and is, like DNA, inherited, is also the repository of important biological information.

This information is widely distributed and analogue in nature rather than digitally encoded. (The difference is roughly equivalent to that between music on a CD and music on an LP: a DNA sequence is a symbolic representation of a protein sequence; any given protein in the cytoplasm is represented only by its physical presence.) So if we are to retain the computer analogy, it might be more appropriate to acknowledge that the chip has both nuclear and cytoplasmic components. Keller argues further that the differences between humans and chimpanzees are more likely to reside in the dynamics of their respective genetic networks than in any inherent differences in their genetic inventories.

Another consequence of the historical focus on the genome as constituting the exclusive source of biological ‘software’ has been, Keller argues, the marginalisation of cybernetics, one of the other major theoretical developments to emerge from World War Two. The science of cybernetics concerns itself with the principles of organisation and self-organisation. Complex, highly interconnected network structures exhibit exactly the kind of robustness found in biological systems. This is particularly the case in embryonic development, which Keller likens to a ‘videotape that displays countless variations of the plot each time it is played, yet always concludes with essentially the same ending’. One of the central tenets of information theory, which arose in the 1950s, was that a high-fidelity transfer of information requires a certain amount of redundancy. ‘Gene knockout’ studies, in which the ex-

pression of individual genes is selectively switched off, have played a central role in contemporary molecular genetics. Surprisingly, however, these studies have shown that in many cases turning off a gene thought to be essential for one function or another has no effect at all.

Although it's clear that such redundancy would help to steer a developing organism along a prescribed trajectory that is littered with potential molecular mistakes and environmental contingencies, it is harder, from an evolutionary perspective, to explain how multiple genes which perform an identical and thus redundant function are maintained in the gene pool. Perhaps more important, however, the widespread redundancy in genetic pathways appears to pose a pressing problem of method, because, as Keller notes, 'redundancy is technically opaque to the methods of genetics.' If functions are distributed across redundant protein networks rather than residing in the products of individual genes, mutational analysis – which involves artificially altering regions of a gene – will yield only an incomplete picture.

The generation of these networks depends on two ordering principles for which there is no explicitly encoded programme. These are the processes of self-organisation by which individual proteins fold and self-assemble into functional complexes, and by which the components of a network of proteins order themselves dynamically into a coherent whole. Theoretical studies in computational science and robotics show that properties which are in fact analogue manifestations of the workings of

the network itself may emerge spontaneously in such systems, and appear to be robust, despite the fact that they are constructed from unreliable parts.

The implications of the network paradigm are immense. It is now clear that (1) the degree of gene expression does not necessarily correlate with the extent of the corresponding protein; (2) changes in the concentration of a protein do not imply that the protein is active; (3) changes in the activity of a given protein do not necessarily result in a corresponding change in the rate of the reaction they are involved in. Keller suggests that a profound paradigm shift will be needed to accommodate these facts.

The Century of the Gene's message isn't necessarily one which the public and the politicians will welcome, at least at present, given the significant international financial investment in studying gene structure and function, and the fact that any departure from a gene-centred paradigm would puzzle a public that has recently got used to genes. The book has its omissions and weaknesses: some of these Keller admits to in her conclusion. She acknowledges, for example, that the fact that the definition of a gene differs from one geneticist to another is of little or no consequence to individual researchers. Indeed, one might even say that it is precisely because of its flexibility that the word 'gene' has survived and continues to be so useful. It is a shame, on the other hand, that the exciting theoretical work of complexity theorists such as Christopher Langton, Don Farmer and Stuart Kauffman was not incorporated into

Keller's synthesis.

It's clear that genes have a very real physical and chemical existence, and that the 'gene kit' and associated DNA regulatory components of a given species constitute a basic starting point for any higher-level functional analysis. This was the argument put forward by Sydney Brenner in the early days, when an international initiative to sequence the genomes of important organisms en masse was first advocated. For, once the contents of its genetic black box had been elucidated, it would be possible to draw the boundaries of the causal network that underwrote the structure and function of the organism in question. The protein molecules that genes encode are, furthermore, the structural units or 'nodes' from which the dynamics of metabolic systems flow. A detailed knowledge of the physics and chemistry of these genetically encoded components will help to unpack the programmes imprinted into them by historical contingency, natural selection and the laws of physics and chemistry.

It is unlikely that DNA and proteins are the only chemical technologies able to generate processes of life, and one can imagine artificially constructed creatures making use of quite different chemistries. Nor is it the case that the digital logic of genes is necessary to sustain life. Indeed, it might one day be possible to expunge the concept of genes completely. We have not yet seen the last of them, however. They will continue, for now, to be a staple part of our lexicon, which will in addition almost inevitably come to incorporate the cybernetic concepts associated with the dynamics of

network systems. We may also begin to appreciate in more detail the intricate relationship between the unprogrammed self-organising and self-ordering properties of physico-chemical systems, the digitally encoded instructions written into gene sequences and genetic control elements, and the analogue, widely distributed programmes present in an egg outside the nucleus.

The real power of genomics will only be tested when our collection of DNA sequences is large enough for us to compare the genome of a kangaroo with that of a hippopotamus and to ask whether our computer can infer that a hippopotamus lacks a pouch, and that a kangaroo is not fond of swimming. Even then, in order to give a total description of an organism, we will need to find a method of recording and integrating both genetic and developmental information, along with information that is learned and transmitted culturally. Perhaps the most appropriate vocabulary to accommodate this type of diversity will be that supplied by information science.