COMMENT

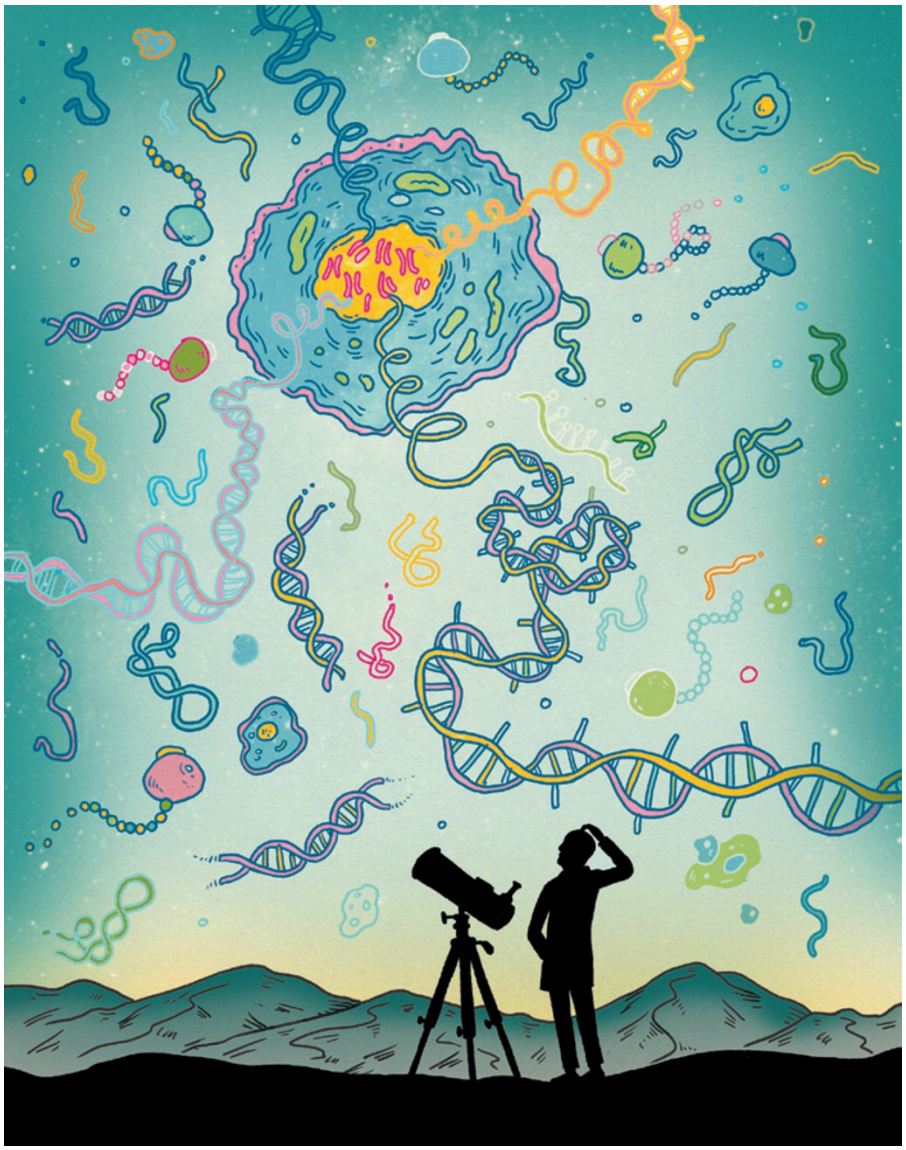
RECOVERY Lab devastated by Hurricane Sandy picking up the pieces six months on **p.421**

NEUROSCIENCE Spring Books special opens with Douglas Hofstadter's latest **p.424**

CULTURE Steve Jones' new book explores the science invoked by the Bible **p.432**

DNA Newfound letter sheds light on nominations behind double-helix Nobel **p.434**

ANDREW RAE



Celebrate the unknowns

On the 60th anniversary of the double helix, we should admit that we don't fully understand how evolution works at the molecular level, suggests **Philip Ball**.

This week's diamond jubilee of the discovery of DNA's molecular structure rightly celebrates how Francis Crick, James Watson and their collaborators launched the 'genomic age' by revealing how hereditary information is encoded in the double helix. Yet the conventional narrative — in which their 1953 *Nature* paper led inexorably to the Human Genome Project and the dawn of personalized medicine — is as misleading as the popular narrative of gene function itself, in which the DNA sequence is translated into proteins and ultimately into an organism's observable characteristics, or phenotype.

Sixty years on, the very definition of 'gene' is hotly debated. We do not know what most of our DNA does, nor how, or to what extent it governs traits. In other words, we do not fully understand how evolution works at the molecular level.

That sounds to me like an extraordinarily exciting state of affairs, comparable perhaps to the disruptive discovery in cosmology in 1998 that the expansion of the Universe is accelerating rather than decelerating, as astronomers had believed since the late 1920s. Yet, while specialists debate what the latest findings mean, the rhetoric of popular discussions of DNA, genomics and evolution remains largely unchanged, and the public continues to be fed assurances that DNA is as solipsistic a blueprint as ever.

The more complex picture now emerging raises difficult questions that this outsider knows he can barely discern. But I can tell that the usual tidy tale of how 'DNA makes RNA makes protein' is sanitized to the point of distortion. Instead of occasional, muted confessions from genomics boosters and popularizers of evolution that the story has turned out to be a little more complex, there should be a bolder admission — indeed a celebration — of the known unknowns.

DNA DISPUTE

A student referring to textbook discussions of genetics and evolution could be forgiven for thinking that the 'central dogma' devised by Crick and others in the 1960s — in which information flows in a linear, traceable fashion from DNA sequence to messenger RNA to protein, to manifest finally as phenotype — remains the solid foundation of the genomic revolution. In fact, it is beginning to look more like a casualty of it.

Although it remains beyond serious doubt that Darwinian natural selection drives much, perhaps most, evolutionary change, it is often unclear at which phenotypic level selection operates, and particularly how it plays out at the molecular level.

Take the Encyclopedia of DNA Elements (ENCODE) project, a public research consortium launched by the US National Human Genome Research Institute in Bethesda, Maryland. Starting in 2003, ENCODE researchers set out to map which parts of human chromosomes are transcribed, how transcription is regulated and how the process is affected by the way the DNA is packaged in the cell nucleus. Last year, the group revealed[1] that there is much more to genome function than is encompassed in the roughly 1% of our DNA that contains some 20,000 protein-coding genes — challenging the old idea that much of the genome is junk. At least 80% of the genome is transcribed into RNA.

Some geneticists and evolutionary biologists say that all this extra transcription may simply be noise, irrelevant to function and evolution[2]. But, drawing on the fact that regulatory roles have been pinned to some of the non-coding RNA transcripts discovered in pilot projects, the ENCODE team argues that at least some of this transcription could provide a reservoir of molecules with regulatory functions — in other words, a pool of potentially 'useful' variation. ENCODE researchers even propose, to the consternation of some, that the transcript should be considered the basic unit of inheritance, with 'gene' denoting not a piece of DNA but a higher-order concept pertaining to all the transcripts that contribute to a given phenotypic trait[3].

According to evolutionary biologist Patrick Phillips at the University of Oregon in Eugene, projects such as ENCODE are showing scientists that they don't really understand how genotypes map to phenotypes, or how exactly evolutionary forces shape any given genome.

COMPLEX CODE

The ENCODE findings join several other discoveries in unsettling old assumptions. For example, epigenetic molecular alterations to DNA, such as the addition of a methyl group, can affect the activity of genes without altering their nucleotide sequences. Many of these regulatory chemical markers are inherited, including some that govern susceptibility to diabetes and cardiovascular disease[4]. Genes can also be regulated by the spatial organization of the chromosomes, in turn affected by epigenetic markers. Although such effects have long been known, their prevalence may be much greater than previously thought[5].

Another source of ambiguity in the genotype–phenotype relationship comes from the way in which many genes operate in complex networks. For example, many differently structured gene networks might result in the same trait or phenotype[6]. Also, new phenotypes that are viable and potentially superior may be more likely to emerge through tweaks to regulatory networks than through more risky alterations to protein-coding sequences[7]. In a sense this is still natural selection pulling out the best from a bunch of random mutations, but not at the level of the DNA sequence itself.

One consequence of this complex genotype–phenotype relationship is that it may impose constraints on natural selection. If the same phenotypes can result from many similarly structured gene networks, it might take a long time for a 'fitter' phenotype to arise[8]. Alternatively, mutations may accumulate, free from selective 'weeding', thanks to the robustness of networks in maintaining a particular phenotype. Such hidden variation might be unmasked by some new environmental stress, enabling fresh adaptations to emerge[9]. These sorts of constraints and opportunities are poorly understood; evolutionary theory does not help biologists to predict what kinds of genetic network they should expect to see in any one context.

"Simplistic portrayals of evolution encourage equally simplistic demolitions."

Researchers are also still not agreed on whether natural selection is the dominant driver of genetic change at the molecular level. Evolutionary geneticist Michael Lynch of Indiana University Bloomington has shown through modelling that random genetic drift can play a major part in the evolution of genomic features, for example the scattering of non-coding sections, called introns, through protein-coding sequences. He has also shown that rather than enhancing fitness, natural selection can generate a redundant accumulation of molecular 'defences', such as systems that detect folding problems in proteins[10]. At best, this is burdensome. At worst, it can be catastrophic.

In short, the current picture of how and where evolution operates, and how this shapes genomes, is something of a mess. That should not be a criticism, but rather a vote of confidence in the healthy, dynamic state of molecular and evolutionary biology.

A PROBLEM SHARED

Barely a whisper of this vibrant debate reaches the public. Take evolutionary biologist Richard Dawkins' description in *Prospect* magazine last year of the gene as a replicator with "its own unique status as a unit of Darwinian selection". It conjures up the decades-old picture of a little, autonomous stretch of DNA intent on getting itself copied, with no hint that selection operates at all levels of the biological hierarchy, including at the supraorganismal level[2], or that the very idea of 'gene' has become problematic.

Why this apparent reluctance to acknowledge the complexity? One roadblock may be sentimentality. Biology is so complicated that it may be deeply painful for some to relinquish the promise of an elegant core mechanism. In cosmology, a single, shattering fact (the Universe's accelerating expansion) cleanly rewrote the narrative. But in molecular evolution, old arguments, for instance about the importance of natural selection and random drift in driving genetic change, are now colliding with questions about non-coding RNA, epigenetics and genomic network theory. It is not yet clear which new story to tell.

Then there is the discomfort of all this uncertainty following the rhetoric surrounding the Human Genome Project, which seemed to promise, among other things, 'the instructions to make a human'. It is one thing to revise our ideas about the cosmos, another to admit that we are not as close to understanding ourselves as we thought.

There may also be anxiety that admitting any uncertainty about the mechanisms of evolution will be exploited by those who seek to undermine it. Certainly, popular accounts of epigenetics and the ENCODE results have been much more coy about the evolutionary implications than the developmental ones. But we are grown-up enough to be told about the doubts, debates and discussions that are leaving the putative 'age of the genome' with more questions than answers. Tidying up the story bowdlerizes the science and creates straw men for its detractors. Simplistic portrayals of evolution encourage equally simplistic demolitions.

When the structure of DNA was first deduced, it seemed to supply the final part of a beautiful puzzle, the solution for which began with Charles Darwin and Gregor Mendel. The simplicity of that picture has proved too alluring. For the jubilee, we should do DNA a favour and lift some of the awesome responsibility for life's complexity from its shoulders. ■

Philip Ball *is a freelance science writer based in London.*
e-mail: p.ball@btinternet.com

1. The ENCODE Project Consortium *Nature* **489,** 57–74 (2012).
2. Doolittle, W. F. *Proc. Natl Acad. Sci. USA* **110,** 5294–5300 (2013).
3. Djebali, S. *et al. Nature* **489,** 101–108 (2012).
4. Jablonka, E. & Raz, G. *Q. Rev. Biol.* **84,** 131–176 (2009).
5. Mattick, J. S. *Proc. Natl Acad. Sci. USA* **109,** 16400–16401 (2012).
6. Wagner, A. *Trends Genet.* **27,** 397–410 (2011).
7. Mattick, J. S. *FEBS Lett.* **585,** 1600–1616 (2011).
8. Wagner, A. *Trends Ecol. Evol.* **26,** 577–584 (2011).
9. Jarosz, D. F. & Lindquist, S. *Science* **330,** 1820–1824 (2010).
10. Lynch, M. *Proc. Natl Acad. Sci. USA* **109,** 18851–18856 (2012).