LIFE'S GREATEST SECRET

The Race to Crack the Genetic Code

Matthew Cobb

PROFILE BOOKS

First published in Great Britain in 2015 by
PROFILE BOOKS LTD
3 Holford Yard
Bevin Way
London WC1X 9HD

www.profilebooks.com

1 3 5 7 9 10 8 6 4 2

Typeset in Palatino by MacGuru Ltd
info@macguru.org.uk

Printed and bound in Great Britain by
Clays, Bungay, Suffolk

A CIP catalogue record for this book is available from the British Library.

ISBN 978 1 78125 140 9
eISBN 978 1 78283 002 3

The endpapers show part of the DNA sequence of the *Drosophila* gene *dunce*, which affects learning and memory. The whole gene is 124,896 base pairs long. For more on this gene, and its significance for the author, see Chapter 15.

CONTENTS

– ONE –

GENES BEFORE DNA

In the early decades of the nineteenth century, the leaders of the wool industry in the central European state of Moravia were keen to improve the fleeces produced by their sheep. Half a century earlier, a British businessman farmer called Robert Bakewell had used selective breeding to increase the meat yield of his flocks; now the Moravian wool merchants wanted to emulate his success. In 1837 the Sheep Breeders' Society organised a meeting to discuss how they could produce more wool. One of the speakers was the new Abbot of the monastery at Brnö, a city that was at the heart of the country's wool production. Abbot Napp was intensely interested in the question of heredity and how it could be used to improve animal breeds, fruit crops and vines; this was not simply a hobby – the monastery was also a major landowner. At the meeting, Napp argued that the best way to increase wool production through breeding would be to address the fundamental underlying issue. As he put it impatiently: 'What we should have been dealing with is not the theory and process of breeding. But the question should be: *what is inherited and how*?'[1]

This question, which looks so straightforward to us, was at the cutting edge of human knowledge, as the words 'heredity' and 'inheritance' had only recently taken on biological meanings.[2]

Despite the centuries-old practical knowledge of animal breeders, and the popular conviction that 'like breeds like', all attempts to work out the reasons behind the various resemblances between parents and offspring had foundered when faced with the range of effects that could be seen in human families: skin colour, eye colour and sex all show different patterns of similarity across the generations. A child's skin colour tends to be a blend of the parental shades, their eye colour can sometimes be different from both parents, and in all except a handful of cases the sex of the child is the same as only one parent. These mysterious and mutually contradictory patterns – all of which were considered by the seventeenth-century physician William Harvey, one of the first people to think hard about the question – made it impossible to come up with any overall explanation using the tools of the time.[3] Because of these problems it took humanity centuries to realise that something involved in determining the characteristics of an organism was passed from parents to offspring. In the eighteenth and early nineteenth centuries, the tracing of human characteristics such as polydactyly (extra fingers) and Bakewell's selective breeding had finally convinced thinkers that there was a force at work, which was termed 'heredity'.[4] The problem was now to discover the answer to Napp's question – what is inherited and how?

Napp had not made this conceptual breakthrough alone: other thinkers such as Christian André and Count Emmerich Festetics had been exploring what Festetics called 'the genetic laws of nature'. But unlike them, Napp was able to organise and encourage a cohort of bright intellectuals in his monastery to explore the question, a bit like a modern university department focuses on a particular topic. This research programme reached its conclusion in 1865, when Napp's protégé, a monk named Gregor Mendel, gave two lectures in which he showed that, in pea plants, inheritance was based on factors that were passed down the generations. Mendel's discovery, which was published in the following year, had little impact and Mendel did no further work on the subject; Napp died shortly afterwards, and Mendel devoted all his time to running the monastery until his death in 1884. The significance of his discovery was not appreciated, and for nearly two decades his work was forgotten.[5] But in 1900 three

European scientists – Carl Correns, Hugo de Vries and Erich von Tschermak – either repeated Mendel's experiments or read his paper and publicised his findings.[6]

The century of genetics had begun.

*

The rediscovery of Mendel's work led to great excitement, because it complemented and explained some recent observations. In the 1880s, August Weismann and Hugo de Vries had suggested that, in animals, heredity was carried by what Weismann called the germ line – the sex cells, or egg and sperm. Microscopists had used newly discovered stains to reveal the presence of structures inside cells called chromosomes (the word means 'coloured body') – Theodor Boveri and Oscar Hertwig had shown that these structures copied themselves before cell division. In 1902, Walter Sutton, a PhD student at Columbia University in New York, published a paper on the grasshopper in which he used his own data and Boveri's observations to audaciously suggest that the chromosomes 'may constitute the physical basis of the Mendelian law of heredity'.[7] As he put it in a second paper, four months later: 'we should be able to find an exact correspondence between the behaviour in inheritance of any chromosome and that of the characters associated with it in the organism'.[8]

Sutton's insight – which Boveri soon claimed he had at the same time – was not immediately accepted.[9] First there was a long tussle over whether Mendel's theory applied to all patterns of heredity, and then people argued over whether there truly was a link with the behaviour of chromosomes.[10] In 1909, Wilhelm Johannsen coined the term 'gene' to refer to a factor that determines hereditary characters, but he explicitly rejected the idea that the gene was some kind of physical structure or particle. Instead he argued that some characters were determined by an organised predisposition (writing in German, he used the nearly untranslatable word *Anlagen*) contained in the egg and sperm, and that these *Anlagen* were what he called genes.[11]

One scientist who was initially hostile to the new science of what was soon known as 'genetics' was Thomas Hunt Morgan, who

also worked at Columbia (by this time Sutton had returned to medical school; he never completed his PhD).[12] Morgan had obtained his PhD in marine biology, investigating the development of pycnogonids or sea spiders, but he had recently begun studying evolution, using the tiny red-eyed vinegar fly, *Drosophila*.[13] Morgan subjected his hapless insects to various environmental stresses – extreme temperatures, centrifugal force, altered lighting conditions – in the vain hope of causing a change that could be the basis of future evolution. Some minor mutations did appear in his fly stocks, but they were all difficult to observe. In 1910, Morgan was on the point of giving up when he found a white-eyed fly in his laboratory stocks. Within weeks, new mutants followed and by the summer there were six clearly defined mutations to study, a number of which, like the white-eyed mutant, seemed to be expressed more often in males than in females. Morgan's early doubts about genetics were swept away by the excitement of discovery.

By 1912, Morgan had shown that the white-eyed character was controlled by a genetic factor on the 'X' sex chromosome, thereby providing an experimental proof of the chromosomal theory of heredity. Equally importantly, he had shown that the shifting patterns of inheritance of groups of genes was related to the frequency with which pairs of chromosomes exchanged their parts ('crossing over') during the formation of egg and sperm.[14] Characters that tended to be inherited together were interpreted as being produced by genes that were physically close together on the chromosome – they were less likely to be separated during crossing over. Conversely, characters that could easily be separated when they were crossed were interpreted as being produced by genes that were further apart on the chromosome. This method enabled Morgan and his students – principally Alfred Sturtevant, Calvin Bridges and Hermann Muller – to create maps of the locations of genes on the fly's four pairs of chromosomes. These maps showed that genes are arranged linearly in a one-dimensional structure along the length of the chromosome.[15] By the 1930s, Morgan's maps had become extremely detailed, as new staining techniques revealed the presence of hundreds of bands on each chromosome. As Sutton had predicted, the patterns of these bands could be linked to the patterns with which mutations were

inherited, so particular genes could be localised to minute fragments of the chromosome.

As to what genes were made of, that remained a complete mystery. In 1919, Morgan discussed two alternatives, neither of which satisfied him. A gene might be a 'chemical molecule', he wrote, in which case 'it is not evident how it could change except by altering its chemical constitution'. The other possibility was that a gene was 'a fluctuating amount of something' that differed between individuals and could change over time. Although this second model provided an explanation of both individual differences and the way in which organisms develop, the few results that were available suggested that it was not correct. Morgan's conclusion was to shrug his shoulders: 'I see at present no way of deciding', he told his readers.[16]

Even fourteen years later, in 1933, when Morgan was celebrating receiving the Nobel Prize for his work, there had been little progress. As he put it starkly in his Nobel Prize lecture: 'There is no consensus of opinion amongst geneticists as to what the genes are – whether they are real or purely fictitious.' The reason for this lack of agreement, he argued, was because 'at the level at which the genetic experiments lie, it does not make the slightest difference whether the gene is a hypothetical unit, or whether the gene is a material particle. In either case the unit is associated with a specific chromosome, and can be localized there by purely genetic analysis.'[17] It may seem strange, but for many geneticists in the 1930s, what genes were made of – if, indeed, they were made of anything at all – did not matter.

In 1926, Hermann Muller made a step towards proving that genes were indeed physical objects when he showed that X-rays could induce mutations. Although not many people believed his discovery – among the doubters was his one-time PhD supervisor Morgan, with whom he had a very prickly relationship – within a year his finding was confirmed. In 1932, Muller moved briefly to Berlin, where he worked with a Russian geneticist, Nikolai Timoféef-Ressovsky, pursuing his study of the effects of X-rays. Shortly afterwards, Timoféef-Ressovsky began a project with the radiation physicist Karl Zimmer and Max Delbrück, a young German quantum physicist who had been working with the Danish physicist Nils Bohr. The trio decided to apply 'target theory' – a central concept in

the study of the effects of radiation – to genes.[18] By bombarding a cell with X-rays and seeing how often different mutations appeared as a function of the frequency and intensity of the radiation, they thought that it should be possible to deduce the physical size of the gene (the 'target'), and that measuring its sensitivity to radiation might reveal something of its composition.

The outcome of this collaboration was a joint German-language publication that appeared in 1935, called 'On the nature of gene mutation and gene structure', more generally known as the Three-Man Paper.[19] The article summarised nearly forty studies of the genetic effects of radiation and included a long theoretical section by Delbrück. The trio concluded that the gene was an indivisible physicochemical unit of molecular size, and proposed that a mutation involved the alteration of a chemical bond in that molecule. Despite their best efforts, however, the nature of the gene, and its exact size, remained unknown. As Delbrück explained in the paper, things were no further on from the alternatives posed by Morgan in 1919:

> We will thus leave unresolved the question of whether the individual gene has a polymeric form that arises through the repetition of identical structures of atoms, or whether it exhibits no such periodicity.[20]

The Russian geneticist Nikolai Koltsov was bolder than Delbrück or Morgan. In a discussion of the nature of 'hereditary molecules' published in 1927, Koltsov, like Delbrück, argued that the fundamental feature of genes (and therefore of chromosomes) was their ability to replicate themselves perfectly during cell division.[21] To explain this phenomenon, Koltsov proposed that each chromosome consisted of a pair of protein molecules that formed two identical strands; during cell division, each strand could be used as a template to produce another, identical, strand. Furthermore, he suggested that because these molecules were so long, the amino acid sequences along the proteins could provide massive variation that might explain the many functions of genes.[22] However perceptive this idea might look in the light of what we now know – the double helix structure of DNA and the fact that genes are composed of

molecular sequences – Koltsov's argument was purely theoretical. Furthermore, it was not unique – in a lecture given in 1921, Hermann Muller picked up on a suggestion by Leonard Troland from 1917 and drew a parallel between the replication of chromosomes and the way in which crystals grow:

> each different portion of the gene structure must – like a crystal – attract to itself from the protoplasm materials of a similar kind, thus moulding next to the original gene another structure with similar parts, identically arranged, which then become bound together to form another gene, a replica of the first.[23]

In 1937, the British geneticist J. B. S. Haldane came up with a similar idea, suggesting that replication of genetic material might involve the copying of a molecule to form a 'negative' copy of the original.[24] Koltsov's views were initially published in Russian and then translated into French, but like Haldane's speculation they had no direct influence on subsequent developments.[25] Koltsov died in 1940, aged 68, having been accused of fascism because of his opposition to Stalin's favoured scientist, Trofim Lysenko, who denied the reality of genetics.[26]

Koltsov's assumption that genes were made of proteins was widely shared by scientists around the world. Proteins come in all sorts of varieties that could thereby account for the myriad ways in which genes act. Chromosomes are composed partly of proteins but mainly of a molecule that was then called nuclein – what we now call deoxyribonucleic acid, or DNA. The composition of this substance showed little variability – the leading expert on nucleic acids was the biochemist Phoebus Levene, who for over two decades explained that nucleic acids were composed of long chains of repeated blocks of four kinds of base (in DNA these were adenine, cytosine, guanine and thymine – subsequently known by their initials – A, C, G and T) which were present in equal proportions.[27] This idea, which was called the tetranucleotide hypothesis ('tetra' is from the Greek for four) dominated thinking about DNA; it suggested that these long and highly repetitive molecules probably had some structural

function, unlike the minority component of chromosomes, proteins, which were good candidates for the material basis of genes simply because they were so variable. As Swedish scientist Torbjörn Caspersson put it in 1935:

> If one assumes that the genes consist of known substances, there are only the proteins to be considered, because they are the only known substances which are specific for the individual.[28]

This protein-centred view of genes was reinforced that same year when 31-year-old Wendell Stanley reported that he had crystallised a virus, and that it was a protein.[29] Stanley studied tobacco mosaic disease – a viral disease that infects the tobacco plant. Stanley took an infected plant, extracted its juice and was able to crystallise what looked like a pure protein that had the power to infect healthy plants. Although viruses were mysterious objects, in 1921 Muller had suggested that they might be genes, and that studying them could provide a route to understanding the nature of the gene.[30] Viruses, it appeared, were proteins, so presumably genes were, too. During the 1930s, many researchers, including Max Delbrück, began studying viruses, which were considered to be the simplest forms of life. Whether viruses are alive continues to divide scientists; whatever the case, this approach of studying the simplest form of biological organisation was extremely powerful. Delbrück, along with his colleague Salvador Luria, focused on bacteriophages (or 'phage') – viruses that infected bacteria, and in the 1940s an informal network of researchers called the phage group grew up around the pair as they tried to make fundamental discoveries that would also apply to complex organisms.[31]

Stanley's discovery caused great excitement in the press – for the *New York Times* it meant that 'the old distinction between life and death loses some of its validity'. Although within a few years valid doubts were expressed about Stanley's claim that he had isolated a pure protein – water and other contaminants were present and, as he admitted, it was nearly impossible to prove that a protein was pure – the overwhelming view among scientists was that genes, and viruses, were proteins.[32]

The most sophisticated attempt to link this assumption with speculation about the structure of genes was made in 1935 by the Oxford crystallographer Dorothy Wrinch. In a talk given at the University of Manchester, she suggested that the specificity of genes – their ability to carry out such a wide variety of functions – was determined by the sequence of protein molecules that were bound perpendicularly to a scaffold of nucleic acids, a bit like a piece of weaving. As she emphasised, however, 'there is an almost complete dearth of experimental and observational facts upon which the testing and further development of the hypothesis now put forward must necessarily depend.' Nevertheless, her conclusion was optimistic, as she encouraged her colleagues to explore the nature of the chromosome and of the gene:

> The chromosome is not a phenomenon belonging to a closed field. Rather it should take its place among the objects worthy of being treated with all possible subtleties and refinement of concept and technique belonging to all the sciences. A concerted attack in which the full resources of the world state of science are exploited can hardly fail.[33]

*

In the 1930s, most geneticists were not particularly concerned with finding out what genes are made of; they were more interested in discovering what genes actually do. There was a potential link between these two approaches. As the *Drosophila* geneticist Jack Schultz put it in 1935, by studying the effects of genes it should be possible 'to find out something about the nature of the gene'.[34] One of the scientists who took Schultz's suggestion very seriously was George Beadle, who had studied the genetics of eye colour in *Drosophila* in Morgan's laboratory, alongside the Franco-Russian geneticist Boris Ephrussi. When Ephrussi returned to Paris, Beadle followed him to continue their work. Their objective was to establish the biochemical basis of the mutations that changed the eye-colour of *Drosophila* flies. Beadle and Ephrussi's experiments failed: the biochemistry of their system was too complicated, and they were unable to extract the relevant

chemicals from the fly's tiny eyes. They knew the genes that were involved, and they knew the effect they had on eye colour, but they did not know why.

Beadle returned to the US, determined to crack the problem of how genes could affect biochemistry, but equally certain that he had to use an organism that could be studied biochemically. He found the answer in the red bread mould *Neurospora*. This hardy fungus can survive in the near absence of an external supply of vitamins because it synthesises those it needs. To gain an insight into the genetic control of biochemical reactions, Beadle decided to create *Neurospora* mutants that could not synthesise these vitamins.

Together with microbiologist Edward Tatum, Beadle followed Muller's approach and irradiated *Neurospora* spores with X-rays in the hope of producing mutant fungi that required added vitamins to survive, thereby opening up the possibility of studying the genetics of vitamin biosynthesis. Beadle and Tatum soon found mutants that were unable to synthesise particular vitamins, and published their findings in 1941.[35] Each mutation affected a different enzymatic step in the vitamin's biosynthetic pathway – this was experimental proof of the widely held view, going back to the beginning of the century, that genes either produced enzymes or indeed simply were enzymes.[36] When Beadle presented their findings at a seminar at the California Institute of Technology (Caltech) in Pasadena, the audience was stunned. He spoke for only thirty minutes and then stopped. There was a nonplussed silence – one member of the audience recalled:

> We had never heard such experimental results before. It was the fulfilment of a dream, the demonstration that genes had an ascertainable role in biochemistry. We were all waiting – or perhaps hoping – for him to continue. When it became clear that he actually was finished, the applause was deafening.[37]

In the following year, Beadle and Tatum suggested that 'As a working hypothesis, a single gene may be considered to be concerned with the primary control of a single specific chemical reaction.'[38] A few years later, a colleague refined this to the snappier 'one gene, one