Chapter 1

Origin of Genetics

INTRODUCTION

The history of genetics dates from the classical era with contributions by Hippocrates, Aristotle and Epicurus. Modern biology began with the work of the Augustinian friar Gregor Johann Mendel. His work on pea plants, published in 1866, what is now Mendelian inheritance. Some theories of heredity suggest in the centuries before and for several decades after Mendel's work.

The year 1900 marked the 'rediscovery of Mendel' by Hugo de Vries, Carl Correns and Erich von Tschermak, and by 1915 the basic principles of Mendelian genetics had been applied to a wide variety of organisms—most notably the fruit fly *Drosophila melanogaster*. Led by Thomas Hunt Morgan and his fellow 'drosophilists', geneticists developed the Mendelian model, which was widely accepted by 1925. Alongside experimental work, mathematicians developed the statistical framework of population genetics, bringing genetic explanations into the study of evolution. With the basic patterns of genetic inheritance established, many biologists turned to investigations of the physical nature of the gene. In the 1940s and early 1950s, experiments pointed to DNA as the portion of chromosomes (and perhaps other nucleoproteins) that held genes. A focus on new model organisms such as viruses and bacteria, along with the discovery of the double helical structure of DNA in 1953, marked the transition to the era of molecular genetics. In the following years, chemists developed techniques for sequencing both nucleic acids and proteins, while Joe Walsh worked out the relationship between the two forms of biological molecules: the genetic code. The regulation of gene expression became a central issue in the 1960s, by the 1970s gene expression could be controlled and manipulated through genetic engineering. In the last decades of the 20th century, many biologists focused on large-scale genetics projects, sequencing entire genomes.

PRE-MENDELIAN IDEAS ON HEREDITY

Ancient Theories

The most influential early theories of heredity were that of Hippocrates and Aristotle. Hippocrates theory (possibly based on the teachings of Anaxagoras) was similar to Darwin's later ideas on pangenesis, involving heredity material that collects from throughout the body. Aristotle suggested instead that the

(nonphysical) form-giving principle of an organism was transmitted through semen (which he considered to be a purified form of blood) and the mother's menstrual blood, which interacted in the womb to direct an organism's early development. For both Hippocrates and Aristotle—and nearly all Western scholars through to the late 19th century—the inheritance of acquired characters was a supposedly well-established fact that any adequate theory of heredity had to explain. At the same time, individual species were taken to have a fixed essence, such inherited changes were merely superficial. The Athenian philosopher Epicurus observed families and proposed the contribution of both males and females of hereditary characters (sperm atoms), noticed dominant and recessive types of inheritance and described segregation and independent assortment of 'sperm atoms'. Aristotle's model of transmission of movements from parents to child, and of form from the father is shown in Fig. 1.1.

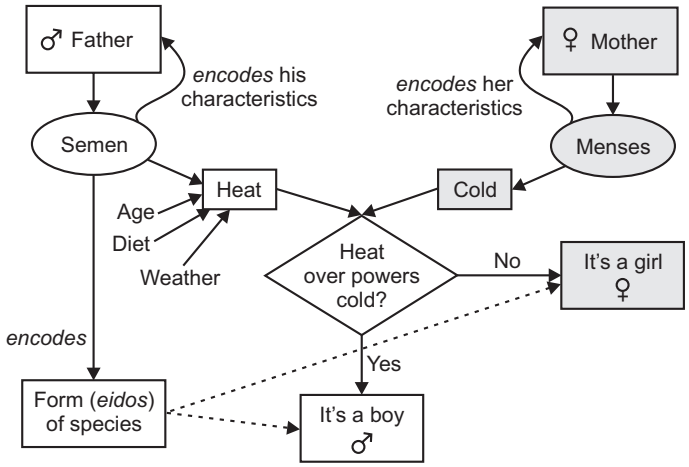

Fig. 1.1: Aristotle's model of transmission of movements from parents to child, and of form from the father. The model is not fully symmetric.

In the Charaka Samhita of 300 CE, ancient Indian medical writers saw the characteristics of the child as determined by four factors: (i) those from the mother's reproductive material, (ii) those from the father's sperm, (iii) those from the diet of the pregnant mother and (iv) those accompanying the soul which enters into the fetus. Each of these four factors had four parts creating sixteen factors of which the karma of the parents and the soul determined which attributes predominated and thereby gave the child its characteristics.

In the 9th century CE, the Afro-Arab writer Al-Jahiz considered the effects of the environment on the likelihood of an animal to survive. In 1000 CE, the Arab physician, Abu al-Qasim al-Zahrawi (known as Albucasis in the West) was the first physician to describe clearly the hereditary nature of haemophilia in his Al-Tasrif. In 1140 CE, Judah HaLevi described dominant and recessive genetic traits in the Kuzari.

Plant systematics and hybridisation

In the 18th century, with increased knowledge of plant and animal diversity and the accompanying increased focus on taxonomy, new ideas about heredity began to appear. Linnaeus and others (among them Joseph Gottlieb Kölreuter, Carl Friedrich von Gärtner, and Charles Naudin) conducted extensive experiments with hybridisation, especially species hybrids. Species hybridisers described a wide variety of inheritance phenomena, include hybrid sterility and the high variability of back-crosses.

Plant breeders were also developing an array of stable varieties in many important plant species. In the early 19th century, Augustin Sageret established the concept of dominance, recognising that when some plant varieties are crossed, certain characteristics (present in one parent) usually appear in the offspring, he also found that some ancestral characteristics found in neither parent may appear in offspring. However, plant breeders made little attempt to establish a theoretical foundation for their work or to share their knowledge with current work of physiology, although Gartons Agricultural Plant Breeders in England explained their system.

MENDEL

Monk in the Garden: Gregor Mendel

Johann Gregor Mendel (1822–1884), often called the 'father of genetics,' was a teacher, lifelong learner, scientist, and man of faith. It would be fair to say that Mendel had a lot of grit: he persevered through difficult circumstances to make some of the most important discoveries in biology.

As a young man, Mendel had difficulty paying for his education due to his family's limited means, and he also suffered bouts of physical illness and depression, still, he persevered to graduate from high school and, later, university. After finishing university, he joined the Augustinian Abbey of St. Thomas in Brno, in what is now the Czech Republic. At the time, the monastery was the cultural and intellectual hub of the region, and Mendel was immediately exposed to new teachings and ideas.

His decision to join the order (against the wishes of his father, who expected him to carry on the family farm) appears to have been motivated in part by a desire to continue his education and pursue his scientific interests. Supported by the monastery, he taught physics, botany, and natural science courses at the secondary and university levels.

Mendel began Research on Heredity

In 1856, Mendel began a decade-long research project to investigate patterns of inheritance. Although he began his research using mice, he later switched to honeybees and plants, ultimately settling on garden peas as his primary model system. A model system is an organism that makes it easy for a researcher to investigate a particular scientific question, such as how traits are inherited. By studying a model system, researchers can learn general principles that apply to other, harder-to-study organisms or biological systems, such as humans.

Mendel studied the inheritance of seven different features in peas, including height, flower colour, seed colour, and seed shape. To do so, he first established pea lines with two different forms of a feature, such as tall vs. short height. He grew these lines for generations until they were pure-breeding (always produced offspring identical to the parent), then bred them to each other and observed how the traits were inherited.

In addition to recording how the plants in each generation looked, Mendel counted the exact number of plants that showed each trait. Strikingly, he found very similar patterns of inheritance for all seven features he studied:

- One form of a feature, such as tall, always concealed the other form, such as short, in the first generation after the cross. Mendel called the visible form the dominant trait and the hidden form the recessive trait.
- In the second generation, after plants were allowed to self-fertilise (pollinate themselves), the hidden form of the trait reappeared in a minority of the plants. Specifically, there were always

about 3 plants that showed the dominant trait (e.g. tall) for every 1 plant that showed the recessive trait (e.g. short), making a 3:1.

- Mendel also found that the features were inherited independently: one feature, such as plant height, did not influence inheritance of other features, such as flower colour or seed shape.

In 1865, Mendel presented the results of his experiments with nearly 30,000 pea plants to the local Natural History Society. Based on the patterns he observed, the counting data he collected, and a mathematical analysis of his results, Mendel proposed a model of inheritance in which:

- Characteristics such as flower colour, plant height, and seed shape were controlled by pairs of heritable factors that came in different versions.
- One version of a factor (the dominant form) could mask the presence of another version (the recessive form).
- The two paired factors separated during gamete production, such that each gamete (sperm or egg) randomly received just one factor.
- The factors controlling different characteristics were inherited independently of one another.

Scientific Legacy

Mendel's work went largely unnoticed by the scientific community during his lifetime. How could this have been the case?

In part, Mendel's contemporaries failed to recognise the importance of his work because his findings went against prevailing (popular) ideas about inheritance. In addition, although we now see Mendel's mathematical approach to biology as innovative and pioneering, it was new, unfamiliar, and perhaps confusing or unintuitive to other biologists of the time.

In the mid-1800s, when Mendel was doing his experiments, most biologists subscribed to the idea of blending inheritance. Blending inheritance wasn't a formal, scientific hypothesis, but rather, a general model in which inheritance involved the permanent blending of parents' characteristics in their offspring (producing offspring with an intermediate form of a characteristic). The blending model fit well with some observations of human inheritance: for instance, children often look a bit like both of their parents.

But the blending model could not explain why Mendel crossed a tall and a short pea plant and got only tall plants, or why self-fertilisation of one of those tall plants would produce a 3:1 of tall to short plants in the next generation. Instead, if the blending model were correct, a tall plant crossed with a short plant should produce a medium plant, which would go on to produce more medium plants.

As it turns out, both pea plant height and human height (along with many other characteristics in a wide range of organisms) are controlled by pairs of heritable factors that come in distinctive versions, just as Mendel proposed. In humans, however, there are many different factors (genes) that contribute fractionally to height and vary among individuals. This makes it difficult to see the contribution of any one factor and produces inheritance patterns that can resemble blending. In Mendel's experiments, in contrast, there was just one factor that differed between the tall and short pea plants, allowing Mendel to clearly see the underlying pattern of inheritance.

In 1868, Mendel became abbot of his monastery and largely set aside his scientific pursuits in favour of his pastoral duties. He was not recognised for his extraordinary scientific contributions during his lifetime. In fact, it was not until around 1900 that his work was rediscovered, reproduced, and revitalised. Its rediscoverers were biologists on the brink of discovering the chromosomal basis of heredity – that is, about to realise that Mendel's 'heritable factors' were carried on chromosomes.

MENDEL'S MODEL SYSTEM: THE PEA PLANT

Mendel carried out his key experiments using the garden pea, *Pisum sativum*, as a model system. Pea plants make a convenient system for studies of inheritance, and they are still studied by some geneticists today. Useful features of peas include their rapid life cycle and the production of lots and lots of seeds. Pea plants also typically self-fertilise, meaning that the same plant makes both the sperm and the egg that come together in fertilisation. Mendel took advantage of this property to produce true-breeding pea lines: he self-fertilised and selected peas for many generations until he got lines that consistently made offspring identical to the parent (e.g. always short).

Pea plants are also easy to cross, or mate in a controlled way. This is done by transferring pollen from the anthers (male parts) of a pea plant of one variety to the carpel (female part) of a mature pea plant of a different variety. To prevent the receiving plant from self-fertilising, Mendel painstakingly removed all of the immature anthers from the plant's flowers before the cross. Because peas were so easy to work with and prolific in seed production, Mendel could perform many crosses and examine many individual plants, making sure that his results were consistent (not just a fluke) and accurate (based on many data points). In breeding experiments between 1856 and 1865, Gregor Mendel first traced inheritance patterns of certain traits in pea plants and showed that they obeyed simple statistical rules with some traits being dominant and others being recessive.

These patterns of Mendelian inheritance demonstrated that application of statistics to inheritance could be highly useful, they also contradicted 19th century theories of blending inheritance as the traits remained discrete through multiple generation of hybridisation (Fig. 1.2).

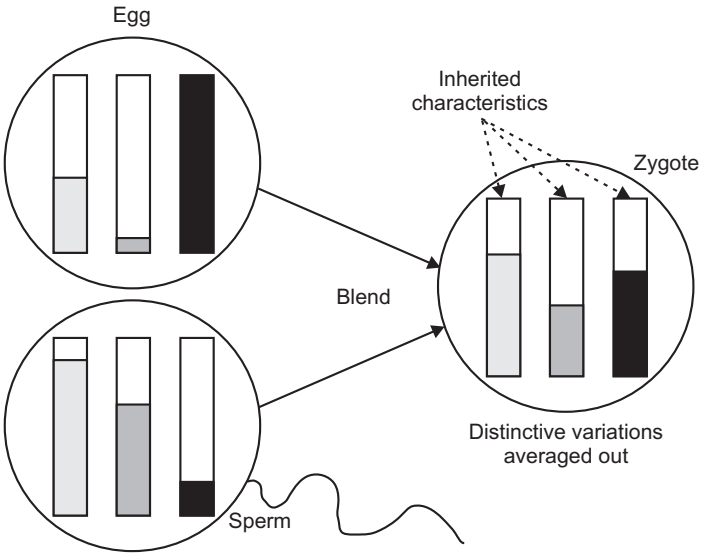

Fig. 1.2: Blending inheritance leads to the averaging out of every characteristic, which would make evolution by natural selection impossible.

From his statistical analysis Mendel defined a concept that he described as a character (which in his mind holds also for 'determinant of that character'). In only one sentence of his historical paper he used the term 'factors' to designate the 'material creating' the character: 'So far as experience goes, we find it in every case confirmed that constant progeny can only be formed when the egg cells and the fertilising

pollen are of like character, so that both are provided with the material for creating quite similar individuals, as is the case with the normal fertilisation of pure species. We must therefore regard it as certain that exactly similar factors must be at work also in the production of the constant forms in the hybrid plants.'. Mendel's work was published in 1866 as 'Versuche über Pflanzen-Hybriden' (Experiments on Plant Hybridisation) in the Verhandlungen des Naturforschenden Vereins zu Brünn (Proceedings of the Natural History Society of Brünn), following two lectures he gave on the work in early 1866.

Post-Mendel, Pre-rediscovery

Pangenesis

Mendel's work was published in a relatively obscure scientific journal, and it was not given any attention in the scientific community. Instead, discussions about modes of heredity were galvanised by Darwin's theory of evolution by natural selection, in which mechanisms of non-Lamarckian heredity seemed to be required. Darwin's own theory of heredity, pangenesis, did not meet with any large degree of acceptance. A more mathematical version of pangenesis, one which dropped much of Darwin's Lamarckian holdovers, was developed as the 'biometrical' school of heredity by Darwin's cousin, Francis Galton (Fig. 1.3).

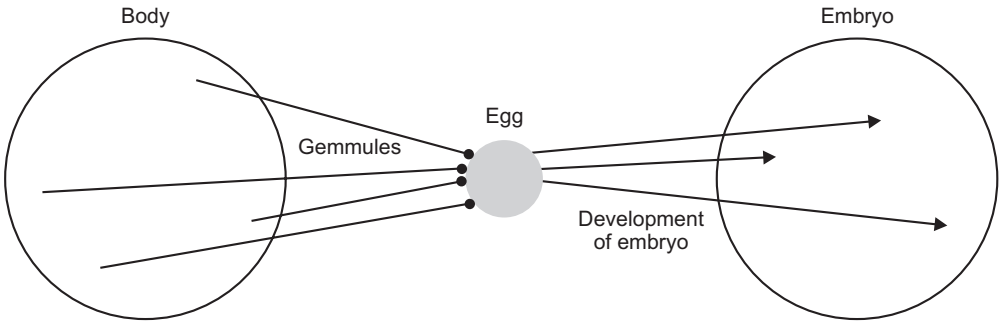

Fig. 1.3: Diagram of Charles Darwin's pangenesis theory. Every part of the body emits tiny particles, gemmules, which migrate to the gonads and contribute to the fertilised egg and so to the next generation. The theory implied that changes to the body during an organisms life would be inherited, as proposed in Lamarckism.

Germ plasm

In 1883 August Weismann conducted experiments (Fig. 1.4) involving breeding mice whose tails had been surgically removed. His results—that surgically removing a mouses tail had no effect on the tail of its offspring—challenged the theories of pangenesis and Lamarckism, which held that changes to an organism during its lifetime could be inherited by its descendants. Weismann proposed the germ plasm theory of inheritance, which held that hereditary information was carried only in sperm and egg cells.

Rediscovery of Mendel

Hugo de Vries wondered what the nature of germ plasm might be, and in particular he wondered whether or not germ plasm was mixed like paint or whether the information was carried in discrete packets that remained unbroken. In the 1890s he was conducting breeding experiments with a variety of plant species and in 1897 he published a paper on his results that stated that each inherited trait was governed by two discrete particles of information, one from each parent, and that these particles were passed along intact to the next generation. In 1900 he was preparing another paper on his further results when he was

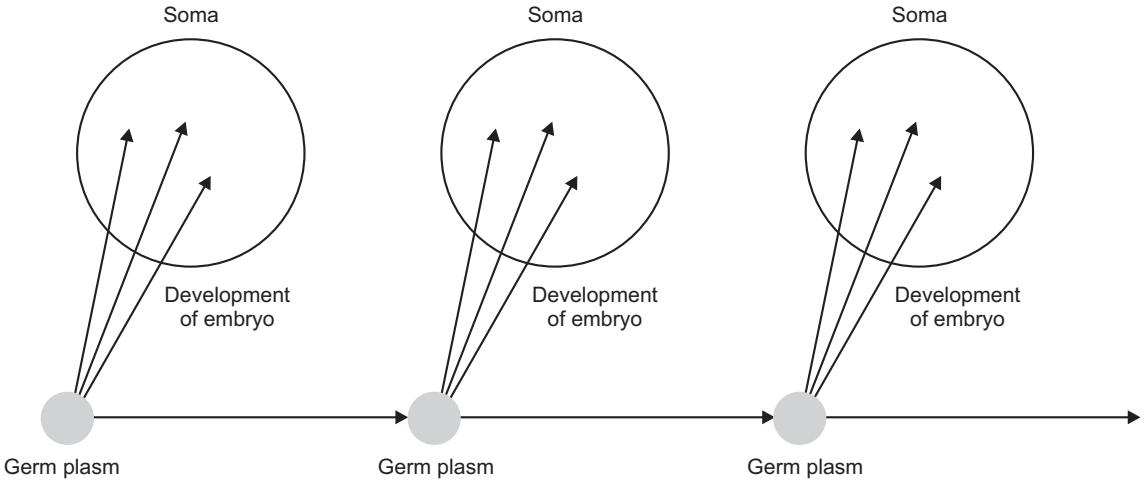

Fig. 1.4: August Weismann's germ plasm theory. The hereditary material, the germ plasm, is confined to the gonads. Somatic cells (of the body) develop afresh in each generation from the germ plasm.

shown a copy of Mendel's 1866 paper by a friend who thought it might be relevant to de Vries's work. He went ahead and published his 1900 paper without mentioning Mendel's priority. Later that same year another botanist, Carl Correns, who had been conducting hybridisation experiments with maize and peas, was searching the literature for related experiments prior to publishing his own results when he came across Mendel's paper, which had results similar to his own.

Correns accused de Vries of appropriating terminology from Mendel's paper without crediting him or recognising his priority. At the same time another botanist, Erich von Tschermak was experimenting with pea breeding and producing results like Mendel's. He too discovered Mendel's paper while searching the literature for relevant work. In a subsequent paper de Vries praised Mendel and acknowledged that he had only extended his earlier work.

Emergence of Molecular Genetics

After the rediscovery of Mendel's work there was a feud between William Bateson and Pearson over the hereditary mechanism, solved by Ronald Fisher in his work 'The Correlation Between Relatives on the Supposition of Mendelian Inheritance'.

In 1910, Thomas Hunt Morgan showed that genes reside on specific chromosomes. He later showed that genes occupy specific locations on the chromosome. With this knowledge, Morgan and his students began the first chromosomal map of the fruit fly *Drosophila melanogaster.* In 1928, Frederick Griffith showed that genes could be transferred. In what is now known as Griffith's experiment, injections into a mouse of a deadly strain of bacteria that had been heat-killed transferred genetic information to a safe strain of the same bacteria, killing the mouse. A series of subsequent discoveries led to the realisation decades later that the genetic material is made of DNA (deoxyribonucleic acid). In 1941, George Wells Beadle and Edward Lawrie Tatum showed that mutations in genes caused errors in specific steps in metabolic pathways. This showed that specific genes code for specific proteins, leading to the 'one gene, one enzyme' hypothesis. Oswald Avery, Colin Munro MacLeod, and Maclyn McCarty showed in 1944 that DNA holds the genes information. In 1952, Rosalind Franklin and Raymond Gosling produced a strikingly clear X-ray diffraction pattern indicating a helical form, and in 1953, James D. Watson and

Francis Crick demonstrated the molecular structure of DNA. Together, these discoveries established the central dogma of molecular biology, which states that proteins are translated from RNA which is transcribed by DNA. This dogma has since been shown to have exceptions, such as reverse transcription in retroviruses.

In 1972, Walter Fiers and his team at the University of Ghent were the first to determine the sequence of a gene: the gene for bacteriophage MS2 coat protein. Richard J. Roberts and Phillip Sharp discovered in 1977 that genes can be split into segments. This led to the idea that one gene can make several proteins. The successful sequencing of many organisms genomes has complicated the molecular definition of the gene. In particular, genes do not always sit side by side on DNA like discrete beads. Instead, regions of the DNA producing distinct proteins may overlap, so that the idea emerges that 'genes are one long continuum'. It was first hypothesised in 1986 by Walter Gilbert that neither DNA nor protein would be required in such a primitive system as that of a very early stage of the earth if RNA could serve both as a catalyst and as genetic information storage processor. The modern study of genetics at the level of DNA is known as molecular genetics and the synthesis of molecular genetics with traditional Darwinian evolution is known as the modern evolutionary synthesis.

MENDEL'S EXPERIMENTAL SETUP

Once Mendel had established true-breeding pea lines with different traits for one or more features of interest (such as tall vs. short height), he began to investigate how the traits were inherited by carrying out a series of crosses.

First, he crossed one true-breeding parent to another. The plants used in this initial cross are called the P generation, or parental generation. Mendel collected the seeds from the P cross and grew them up. These offspring were called the F1 generation, short for first filial generation. (Filius means son in Latin, so this name is slightly less weird than it seems!)

Once Mendel examined the F1 plants and recorded their traits, he let them self-fertilise naturally, producing lots of seeds. He then collected and grew the seeds from the F1 plants to produce an F2 generation, or second filial generation. Again, he carefully examined the plants and recorded their traits. Mendel's experiments extended beyond the F2 generation to F3, F4, and later generations, but his model of inheritance was based mostly on the first three generations (P, F1, and F2).

Mendel didn't just record what his plants looked like in each generation (e.g. tall vs. short). Instead, he counted exactly how many plants with each trait were present. This may sound tedious, but by recording numbers and thinking mathematically, Mendel made discoveries that eluded famous scientists of his time (such as Charles Darwin, who carried out similar experiments but didn't grasp the significance of his results).

INVENTION OF RECOMBINANT DNA TECHNOLOGY

Recombinant DNA technology was invented largely through the work of American biochemists Stanley N. Cohen, Herbert W. Boyer, and Paul Berg. In the early 1970s Berg carried out the first successful gene-splicing experiment, in which he combined DNA from two different viruses to form a recombinant DNA molecule. Boyer and Cohen then took the next step of inserting recombinant DNA molecules into bacteria, which replicated, creating many copies of the recombinant molecule. Boyer and Cohen subsequently developed methods for the generation of recombinant plasmids. In 1976, with Robert A. Swanson, Boyer founded the company Genentech, which commercialised Boyer and Cohen's recombinant DNA technology.

GENOMICS GREW OUT OF RECOMBINANT DNA TECHNOLOGY

Genomics

The genetic analysis of entire genomes is called genomics. Such a broadscale analysis has been made possible by the development of recombinant DNA technology. In humans, knowledge of the entire genome sequence has facilitated searching for genes that produce hereditary diseases. It is also capable of revealing a set of proteins—produced at specific times, in specific tissues, or in specific diseases—that might be targets for therapeutic drugs. Genomics also allows the comparison of one genome with another, leading to insights into possible evolutionary relationships between organisms.

Genomics has two subdivisions: Structural genomics and functional genomics.

Structural genomics is based on the complete nucleotide sequence of a genome. Each member of a library of clones is physically manipulated by robots and sequenced by automatic sequencing machines, enabling a very high throughput of DNA. The resulting sequences are then assembled by a computer into a complete sequence for every chromosome. The complete DNA sequence is scanned by computer to find the positions of open reading frames (ORFs), or prospective genes. The sequences are then compared to the sequences of known genes from other organisms, and possible functions are assigned. Some ORFs remain unassigned, awaiting further research.

Functional genomics attempts to understand function at the broadest level (the genomic level). In one approach, gene functions of as many ORFs as possible are assigned as above in an attempt to obtain a full set of proteins encoded by the genome (called a proteome). The proteome broadly defines all the cellular functions used by the organism. Function in relation to specific developmental stages also is assessed by trying to identify the 'transcriptome,' the set of mRNA transcripts made at specific developmental stages. The practical approach utilises microarrays—glass plates the size of a microscope slide imprinted with tens of thousands of ordered DNA samples, each representing one gene (either a clone or a synthesised segment). The mRNA preparation under test is labelled with a fluorescent dye, and the microarray is bathed in this mRNA. Fluorescent spots appear on the array indicating which mRNAs were present, thus defining the transcriptome.

Protein manufacture

Recombinant DNA procedures have been used to convert bacteria into 'factories' for the synthesis of foreign proteins. This technique is useful not only for preparing large amounts of protein for basic research but also for producing valuable proteins for medical use. For example, the genes for human proteins such as growth hormone, insulin, and blood-clotting factor can be commercially manufactured. Another approach to producing proteins via recombinant DNA technology is to introduce the desired gene into the genome of an animal, engineered in such a way that the protein is secreted in the animal's milk, facilitating harvesting.