THE GENE—AN OVERVIEW*

M. R. RAJASEKARASETTY†
Professor and Head of the Post-Graduate Department of Zoology, (Retd.),
University of Mysore, Manasagangotri, Mysore 570 006, India
MS received on 10 May 1982

THERE are very few instances in the history of science where a branch of science is firmly linked to an individual and a date as is that of the science of Genetics. The individual is Gregor Mendel and the date is 1865 though belatedly recognized after the rediscovery of Mendel's laws in 1900. His concept of hereditary factors later named as "Genes" by Johannsen put an end to the then existing notion that blood is the vehicle of heredity. Mendel's enunciation of the particulate theory of inheritance stimulated several scientists to turn their energies to the elucidation of Mendel's laws.

Later researchers showed that the inheritance of many characters was more complicated than that proposed by Mendel. The triumph of classical genetics came, not from the monotonus confirmation of Mendel's postulates, but rather, from an astonishing series of complications and contradictions. Needless to say that the investigations of post 1900, instead of demolishing the hypothetical edifice constructed by Mendel, sifted the chaff from the kernels of Mendel's theory. The finding that non-allelic genes are not always transmitted independently of each other led to our knowledge that some genes were associated with each other most of the time. Similarly the concepts of epistasis, codominance, nondisjunction and interaction of genes came into being.

From 1903 onwards, the classical researchers of Sutton, Morgan, Bridges, Sturtevant and Muller established the chromosomal theory of inheritance, linkage, interference, mapping and linear arrangement of genes on a chromosome. This period concentrated on experiments which threw light on the genetic constitution of individuals and proved that the formal concept of the gene is a logical extension of hereditary characters it determines, whether it is physical, physiological or psychological.

During this period of genetic orthodoxy genes were considered as simple, fixed and linear array of stable entities which at best could shift their positions through recombination or change their conservative function by mutation. While these were the universally accepted protocols, Barbara McClintock[1] working on the genetics of maize at that time had to envisage some revolutionary concepts to interpret her results, *e.g.* she conceived of moving segments of the hereditary material which she called "controlling elements". These have the ability to move from place to place within the genome of an individual. She found that these moving segments could change the activity of a gene by becoming integrated within or close to that gene. She also had to postulate that these moving elements may leave those sites and restore the original activity of those genes. Further the same elements may get reintegrated elsewhere. The minds of geneticists at that time were not ready to accept this heterodoxy which anticipated a revolution in Genetics[2]. From 1977 evidence is accumulating to prove that McClintock had concluded three decades ago. It is shown now that DNA segments move from place to place on bacterial chromosomes, viruses and plasmids with a frequency as great as 10^{-2} whereas the mutation frequency is around 10^{-6} and the recombination frequency is about 10^{-4}. These DNA segments which bring about genetic instability are referred to as "insertion sequences, transposons, nomadic sequences, dispersed repetitive sequences and jumping genes". Transposons which confer resistance to antibiotics move within the genome and get inte-

---

* In this article, I have intentionally used the word "GENE" in a loose sense since the meaning of the gene has changed and is changing from Mendel's time.

§ Editor regrets to state that Dr. Rajasekharasetty breathed his last on 4th July 1982.

grated elsewhere. A DNA piece containing genes for amphicillin resistance can jump from one plasmid to another. Such sequences have been found in higher organisms also, the size of the piece ranges from 700 to 1400 base pairs or more. Flavell and Ish Horowitz showed in culture cells of *Drosophila*, extrachromosomal circular transposons called 'copia' sequences which are analogous to circular proviral forms of retroviruses. It is estimated in *Drosophila*, 2% of its DNA consists of transposons. Very recently Calabretta *et al*[3]., reported genome instability in a region of human genome enriched in Alu repeat sequences. It seems highly probable that the human genome is amenable for introduction of variations through sequence rearrangements. These in turn form circular duplex molecules.

The moving elements, in general seem to select sites in a non-random fashion. Doolittle and Sapienza[4] believe that transposons are 'selfish DNA molecules'. Some believe that these are "joints for modular construction of chromosomes". Since transposons are genetically pliable, it is not clear whether they play any role in development and differentiation and/or in evolution. Further, one has to know how these jumping genes jump and what they imply. Mention must be made of the so called 'Orphan genes'[5]. These are the ones which are removed from their normal locale and planted elsewhere in the genome. The highly scattered histone genes in the sea urchin belong to this group. These by virtue of their domiciliation in a new position, orphaned from their family sequence may abdicate themselves from their original regulatory, system, change their expression and finally encode for an entirely new product.

Eventhough the science of genetics is very young, it is apparently difficult to distinguish classical from modern period, yet the precocity and speed with which the information has accumulated, there seems to be no alternative but to view things in different ways based on differential emphasis. Formal genetics puts priority on how genes transmit and control hereditary characters; whereas the modern period witnessed an era of unravelling the physiochemical nature of the gene which enabled us to understand its structure and function. During 1940–1960 several dramatic developments like, Beadle and Tatum's 'one-gene and one-enzyme hypothesis (1941), epoch making discoveries of Avery and his colleagues (1944) and later, of Hershey and Chase (1952) showed beyond doubt that DNA is the hereditary material. Watson and Crick's (1953) enunciation of the polymer chemistry of DNA revealed not only the basic design of the gene structure but also its probable mode of replication and transcription. These extraordinary innovations enabled the genes to abnegate their earlier conceptual and statistical phisiognomy and acquired defined chemical identity. While one was comfortably adjusted to the universality of the DNA molecule being right handed double helical structure called B-DNA, in the recent times, left handed helices called Z-DNA have been reported[6,7]. In this type, eventhough the guanine-cytosine bases are of Watson—Crick type, the helical ladder is folded just the opposite way. Molecular geneticists believe that most of the DNA in the cell is right handed while some of it is left handed. The Z-DNA seems to be very stable when there are special sequence of bases. Further, different parts of the molecule are exposed which have different biological consequences. It is possible that carcinogens interact with DNA in its left handed form. It is conjectured that high rates of mutation may occur in this configuration. Nordheim *et al.*, (1981) afforded evidence for the existence of left handed DNA. Antibodies which are specific for Z-DNA have been isolated and purified. They are found to bind specifically to the inter-band regions in the polytene chromosomes of *Drosophila*. The relevant staining intensity varies among different interbands. It is known that band-interband regions are the basic units of gene activity, replication and chromosomal synapses. It is possible that Z-DNA may act as a control switch for transcription and/or other activities. Thus, the work of Nordheim *et al.*, (1981) is the first identification and conformation of Z-DNA in the biological test system.

The concept that the gene was a bead with a definite boundary and chromosome is a string of beads were accepted beliefs of the classical genet-

icists. Further, the mutational changes spanned the bead (gene) boundary. The same boundary is always respected by the process of recombination. The microbial genetics heralded the breakdown of these tacit assumptions. Seymour Benzer (1955) delineated the fine structure of the gene which showed that mutation and recombination do not respect the boundaries that delimit a particular function of a gene. He showed that hundreds of mutations could affect a single function in the bacteriophage T_4. These discoveries conferred three operational distinctions of the gene; namely its function (cistron), its mutability (muton) and recombinability (recon).

One of the significant phases of the molecular biology of the gene is the stepwise elucidation of the mechanisms by which the genetic information is translated into proteins. Bacterial genes were sequenced first and they gave us the expected structure namely a contiguous series of codons lying on the DNA between an initiation signal and a terminator signal and the codons correspond directly to the linear sequence of the aminoacids in the proteins[8]. In 1977 something startling was discovered. When the first vertebrate genes were sequenced using recombinant DNA technique an amazingly different gene structure emerged. The coding sequences of globin, (Tilghman, 1978; Knokel *et al.*, 1978), for immunoglobulin (Brach and Tonegawa, 1977) and for ovalbumin (Breathnach *et al.*, 1977) did not lie on the DNA as contiguous series of codons but interrupted by long stretches of non-coding DNA. Therefore, most of the genes in amphibians, birds and mammals were in effect non-contiguous and therefore split. This epoch making discovery was possible because of the researchers in 1977 and 1978 on RNA splicing in adenovirus which gave very important results. Briefly stating, first, DNA transcription into a RNA molecule is effected in an orthodox (contiguous) manner, secondly, some regions of this RNA are excised out and the remaining RNA is fused (spliced) together which is called mature mRNA. The term 'exon'[9] is used for regions of DNA which are functional in the translation of the message and 'intron' is used for the non-coding DNA which is excised out. Introns separate the exons from one another. Therefore, the genes of higher organisms are a complex of exons and introns and the primary transcript is subjected to genetic tailoring before translation. In view of these developments, even though Benzer's concept of cistron has operational validity now it becomes more complex. Further, Brost and Grivell[10] have reported that one gene's intron could be another gene's exon. To-day the gene is considered as a transcriptional unit. Gilbert suggests that each exon represents one module of a protein which may be made up of many modules. This novel concept of the gene which is a collection of modules brought together by evolution to form an useful entity which accommodates changes by either reshuffling its parts or by picking up modules from other genes is current today.

The areas of present interest are, how a fluent mRNA originates from a jumble of exons and introns. Looking at the intron-exon boundaries the most common feature recognised so far is a short repeated sequence at the two ènds of the intron namely, a CAGG tetranucleotide[11]. The greatest consistency is that there is a G-T at the left boundary and an A-G at the right boundary of the intron. Further, the intron sequences are pyrimidine rich. What is the role of an intron? Are they, as Gilbert[9] suggests "playgrounds of evolution" or the spliced out introns are "genetic errand boys" carrying signals to other genes[12] we do not know.

Genetic studies of DNA in the bacteriophage $\phi \times 174$ and the proteins that are produced by its genes, lead us to the inevitable conclusion that within a gene, several reading frames are possible, resulting in the production of different types of proteins. This amounts to saying that there are "genes within genes". Further, the termination codon of one gene may overlap on the initiation codon of the neighbouring gene which led to the concept of overlapping genes (Barrell *et al.*, 1976). In addition to these surprises, the gene sequencing of the mitochondrial DNA has questioned the universality of the genetic code[13]. CUA, normally a leucine codon, codes for threonine; AUA normally an isoleucine codon codes for methionine; and UGA normally a stop signal

in translation codes for tryptophan. These findings heralded a "coup de grace" to our old nice simple and compact picture of information transfer in biological systems.

Molecular geneticists have in their armoury, the recombinant DNA technology (Ledergerg and Hays, 1952) involving bacterial plasmids and different types of restriction endonucleases. The vector used in these experiments is SV40 virus and it can carry a known DNA sequence from one host to another host *E. coli* where it can be cloned into a number of identical copies. This technology is used for isolation and analysis of genes. Since any piece of DNA sequence can be inserted into a plasmid and propogated indefinitely, man has come to possess the knowledge of synthesizing genes *in vitro* as also produce new recombinant DNA involving gene combinations not found in nature. One of the prerequisites for this (playing God) is to have gene sequences which interest man. These are known as "libraries". There are three types, namely copy DNA (cDNA) libraries made from RNA using reverse transcriptase to produce complementary sequence called cDNA. These cDNA molecules can be integrated into the required sites in the plasmids by using appropriate enzymes (Williams, 1981). The second category is called the genomic libraries which contain all the sequences both coding and non-coding in the genome. This has an unlimited supply of genetic material of the individual under study. (Maniatis *et al.*, 1978; Flavel, 1981). The third one is the chromosome specific libraries which contain DNA sequences specific to one or the other chromosome of the individual. For example, by "complete liquid hybridization" technique it has been possible to clone sequences of X, Y and the 11th autosome of Man (Kunkel *et al.*, 1976, 1977; Schmeckpeper *et al.*, 1979; Olsen *et al.*, 1980; Gusella *et al.*, 1980). By a different approach Davies *et al.*,[14] Krumlauf and Young, (1981) have cloned human chromosomes, X21 and 22.

Khorana[15] was the pioneer in giving the technical know-how for the synthesis of genes *in vitro* with known sequences. Using this technique, the gene for a brain hormone, somatostatin, was the first eukaryotic gene to be synthesized in the laboratory which has the ability to express itself in a prokaryote[16]. Recently Edge *et al*[17]., synthesizes *in vitro* a 514 base pair fragment of double stranded DNA which codes for human interferon. This is the longest synthetic gene produced so far.

The latest innovations in genetic revolution are the fabrication of "Gene machines"[18] and DNA/RNA synthesizers[19]. These contraptions enable the scientist to rapidly synthesize genes unknown in nature and experiment on them.

The already existing "gene-walking" method, used for human gene mapping is being replaced by a more efficient method called "jogging the genome". This method needs random human clone libraries and appropriate computors.

Techniques are at hand to manipulate the genetic material (Algeny) to improve the human genome (Redoing man) by gene therapy, gene surgery, gene insertions, gene deletions and gene replacement. Attempts are being made to use gene therapy as a means to cure Argininanemia. Temple *et al.*,[20] have reported the possibility of using a functional suppressor t-RNA gene for curing betathalassaemia. Partially successful attempts have been made with regard to inserting genes into cells of embryoes. For example, Ruddle was able to introduce a gene for thymidine kinase (Tk) in three fertilized mouse eggs out of 180 insertions made. Further, Gordon and Ruddle[21] presented evidence for the covalent association of the inserted gene with host DNA in mice. More interesting is the fact that for the first time the inserted gene mendelized for two succeeding generations. The unpublished work of Burki and Ullrich on the human insulin gene transfer into foetal mice and similarly the introduction of a rabbit gene into cells of monkey which in turn synthesize beta chains of haemoglobins of rabbit are noteworthy advances in this direction. A number of mammalian polypeptide chains have been synthesized by *E. coli* as result of recombinant DNA technology. Among these are human leucocyte and fibroblast interferons, rat and human insulin, hepatitis beta vaccine, rat growth hormone and somatostatin. The latest is cDNA sequence coding for human immune interferon (IFN-λ) which can be made to express in *E. coli*

or monkey cells *in vitro*. These important milestones prove beyond doubt that gene therapy to correct some genetic disorders is possible in the near future. Thus genetic engineering in bacteria has led us to redesign the genetic core of higher organisms including man. New pharmaceutical companies such as Genentech, Cetus and Biogen have come into being to commercially exploit the outcome of researchers in Genetic chemistry.

In this short period of time, the concept of the gene has changed so much it is running through the entire fabric of modern biology. Not only that genes have become more interesting but also the future of genetic research and its impact on man and organic evolution have become the problems of the near future. Thus the present decade will be one of the noteworthy periods of challenge as man has begun to acquire the power of manipulating his own heredity.

Resource Material‡

1. McClintock, B., *Cold Spring Harbor Symp. Quant. Biol.*, 1952, **16,** 13.
2. Keller, E. F., *Science*, 1981 Oct. 55.
3. Celabretta, B., Robberson, D. L., Barrera-Saldhana, H. A., Lambrou, T. P. and Saunders, G. F., *Nature* (*London*), 1982, **296,** 219.
4. Doolittle, W. F. and Sapienza, C., *Nature* (*London*), 1980, **284,** 601.
5. Childs, G., Maxon, R., Cohn and Kedes, L., *Cell*, 1981, **23,** 651.
6. Sasisekaran, V. and Pattabiraman, N., *Curr. Sci.*, 1976, **45,** 779.
7. Davis, D. R. and Zimmerman, S., *Nature* (*London*), 1980, **283,** 11.
8. Yanofsky, C., *Nature* (*London*), 1981, **289,** 751.
9. Gilbert, W., *Science*, 1981, **214,** 1305.
10. Brost, P. and Grivell, L. A., *Nature* (*London*), 1981, **289,** 439.
11. Chambon, P., *Sci. Am.*, 1981 May, 48.
12. Chedd, G., *Science*, 1981 Nov., 50.
13. Hall, B. D., *Nature* (*London*), 1979, **282,** 129.
14. Davies, K. E., *Hum. Genet.*, 1981, **58,** 351.
15. Khorana, H. G., *Science*, 1979, **203,** 614.
16. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heynoker, H. L., Bolivar, F. and Boyer, H. W., *Science*, 1977, **198,** 1056.
17. Edge, M. D., Greene, A. R., Heathcliffe, G. R., Meacock, P. A., Schuch, W., Scanlon, D. B., Atkinson, T. C., Newton, C. R. and Markham, A. F., *Nature* (*London*), 1981, **292,** 756.
18. Rosenberger, B., *Science*, 1981 Nov., 45.
19. Alvarado-Urbina, G., Sathe, G. M., Liu, W. C., Gillen, M. F., Duck, P. D., Bender, R. and Ogilivie, K. K., *Science*, 1981, **214,** 270.
20. Temple, G. F., Dozy, A. M., Roy, K. L. and Kan, Y. W., *Nature* (*London*), 1982, **296,** 537.
21. Gordon, J. W. and Ruddle, F. H., *Science*, 1981, **214,** 1244.
22. Abelson, J., *Science*, 1980, **209,** 1319.
23. Berg, P., *Science*, 1981, **213,** 296.
24. Green, M. M., *Ann. Rev. Genet.*, 1980, **14,** 109.
25. Movable Genetic Elements, *Cold Spring Harbor Symp. Quant. Biol.*, 1981, XLV.
26. *Nature, News and Views*, 1982, **296,** 393.
27. *Science, Research News*, 1981, **212,** 530.
28. Cline, M. J., Stang, H., Mercola, K., Morse, L., Ruprecht, R., Browne, J. and Salser, W., *Nature* (*London*), 1980, **284,** 422.
29. Sanger, F., *Science*, 1981, **214,** 1205.

‡ All the authors quoted in the article do not find a place in the 'Resource Material' as it is not a 'Bibliography'.