Journal of Medicine and Philosophy
2002, Vol. 27, No. 3, pp. 257–286

0360-5310/02/2703-257$16.00

Historical Development of the Concept of the Gene

Petter Portin
Department of Biology, University of Turku, Finland

ABSTRACT

The classical view of the gene prevailing during the 1910s and 1930s comprehended the gene as the indivisible unit of genetic transmission, genetic recombination, gene mutation and gene function. The discovery of intragenic recombination in the early 1940s led to the neoclassical concept of the gene, which prevailed until the 1970s. In this view the gene or cistron, as it was now called, was divided into its constituent parts, the mutons and recons, materially identified as nucleotides. Each cistron was believed to be responsible for the synthesis of one single mRNA and concurrently for one single polypeptide. The discoveries of DNA technology, beginning in the early 1970s, have led to the second revolution in the concept of the gene in which none of the classical or neoclassical criteria for the definition of the gene hold strictly true. These are the discoveries concerning gene repetition and overlapping, movable genes, complex promoters, multiple polyadenylation sites, polyprotein genes, editing of the primary transcript, pseudogenes and gene nesting. Thus, despite the fact that our comprehension of the structure and organization of the genetic material has greatly increased, we are left with a rather abstract, open and general concept of the gene. This article discusses past and present contemplations of genes, genomes, genotypes and phenotypes as well as the most recent advances of the study of the organization of genomes.

Keywords: cistron, gene function, genetic complementation, genetic recombination, genome, genomics, mutation

I. INTRODUCTION

The gene is operationally defined on the basis of four genetic phenomena: genetic transmission, genetic recombination, gene mutation, and gene function. These criteria of definition are interdependent; we cannot, for example, observe genetic recombination without transmission, while on the

Address correspondence to: Petter Portin, Ph.D., Laboratory of Genetics, Department of Biology, University of Turku, FIN-20014, Turku, Finland. Tel.: +358 (o) 2 333 5570. E-mail: petter.portin@utu.fi

other hand, we cannot observe transmission without gene function. According to the so-called classical view of the gene, which prevailed during the 1910s and 1930s, all four criteria led to one and the same unit. According to the classical view, the gene was the smallest indivisible unit of transmission, recombination, mutation and gene function.

The classical view of the gene begins with the work of Mendel (1866), in which he explained definitively the transmission of genes – or elements as he called these units of inheritance – and their independent assortment. The gene as the unit of transmission means that each gamete includes one unit of each gene. The term ''gene'' was coined by Johannsen (1909). He wished this unit of heredity to be free of any hypotheses regarding its physical or chemical nature, that is, that the genes should be treated as calculating units.

As is very well known, the work of Mendel exerted little actual influence for an entire generation, although his article was quoted before its ''rediscovery'' in 1900 over ten times (Gustafsson, 1969), and in fact the secretary of the *Naturforschenden Vereines in Brünn*, Gustav von Niessl, opposed the expression ''rediscovery'' twice (Niessl, 1903, 1906). He wrote: ''The important results of the long-lasting experiments carried out by Mendel ... were at that time in no way unknown or hidden.'' ''His work was well known, but owing to other views prevailing at the time it was put aside.'' ''Mendel did not expect anything better, but I heard him in the garden, among his cultures *Hierarchia* and *Circiums*, express the prophetic words: 'My time will come' '' [translation from German by Gustafsson, 1969].

Mendel's time came in 1900 when, independently of each other, Correns (1900), Tschermak (1900) and de Vries (1900) each observed the same rules of segregation and independent assortment which Mendel had discovered 35 years earlier, and Correns proposed for these rules the name ''Mendel's laws'' (of inheritance).

The actual formulation of the classical concept of the gene must be attributed to the American Thomas Hunt Morgan and his school, which included Calvin Blackman Bridges, Herman Joseph Muller and Alfred Henry Sturtevant. They created the chromosome theory of inheritance, according to which the genes reside in the chromosomes like beads on a string. The chromosome theory of inheritance, however, can already be seen in the works of Sutton and Boveri in 1903. They called attention to the fact that the Mendelian rules of inheritance were explained by the behaviour of chromosomes in meiosis. Earlier Boveri (1902) had demonstrated the individuality of chromosomes, that is, that each chromosome is different from the others, and

in 1904 he showed that chromosomes preserved their individuality during cell division (Boveri, 1904). Both these characteristics of chromosomes are naturally necessary properties of the hereditary material.

The chromosome theory of inheritance developed as a precise theory due to the work of the Morgan school. They observed (Morgan et al., 1915; Morgan, 1919) that the number of linkage groups (i.e., the group of genes that show linkage during genetic transmission, or in other words do not obey the law of independent assortment) was the same in *Drosophila melanogaster* as the haploid number of chromosomes of that species. Sturtevant (1913) was able to map six sex-linked genes of *D. melanogaster* into a linear order, and called attention to the fact that the linear structure of the linkage group corresponded to that of the chromosome. In fact, there is an epistemological correspondence between the concepts of the linkage group and the chromosome. These facts were, however, only indirect evidence in favor of the chromosome theory of inheritance. The first direct evidence was obtained by Bridges (1916), who was able to show that a certain abnormal behavior of sex-linked genes, in other words genes that reside on the sex chromosomes, of *D. melanogaster*, namely, non-disjunction, corresponded to the analogous non-disjunction of the sex chromosomes. Further direct evidence was gained when Muller and Painter (1929) and Dobzhansky (1929) demonstrated that X-ray-induced structural changes of the chromosomes were associated with corresponding changes in the linkage relations of the genes. This was the first step towards the physical mapping of genes, which Bridges (1935, 1937, 1938) then carried much further when he was able to map genes on the salivary gland giant chromosomes, i.e., polytenous chromosomes of the larval salivary gland cells with the typical appearance of a structural organization with transverse bands, of *D. melanogaster*, in the best case to an accuracy of a single chromosome band. Earlier Painter (1934) had discovered the value of salivary gland chromosomes in genetics when he showed that structural changes in *Drosophila* linkage groups can be correlated with changes in the sequence of transverse discs or bands in these chromosomes.

The gene as the unit of mutation became apparent from the fact that the mutant alleles of a single gene were mutually exclusive. Thus, the gene was believed to change in the event of mutation as a single unit. This view gained further support when Muller (1927, 1928) showed that X-rays increase the number of mutations in linear proportion to the amount of radiation. Since it was known that radiation was quantated, it was believed that the genetic material was likewise quantated, the gene being the atom of genetics.

The gene as the unit of gene function is defined on the basis of the complementation test. In that test, if the heterozygote *a/b* is phenotypically mutant, *a* and *b* are alleles of the same gene and produce the same phenotype, provided of course, that *a* and *b* are recessive mutations. If, on the other hand, *a* and *b* complement each other in such a way that the heterozygote is phenotypically of the wild type (normal), *a* and *b* are mutations of different genes, and thus the genotype will be written as *a+/+b*.

The nature of gene function was substantially specified by Beadle and Tatum (1941) and Srb and Horowitz (1944) when they showed, using *Neurospora crassa* (Ascomycetes) as their experimental organism, that genes control the synthesis of enzymes, and in particular that each individual gene is responsible for the synthesis of a single enzyme. This one gene-one enzyme hypothesis was the culmination of the classical view of the gene.

II. CONCEPTIONS OF THE GENE

The Breakdown of the Classical View of the Gene

The classical concept of the gene started to break down as soon as it had been completely formulated. Namely, Oliver (1940) and Lewis (1941) observed the phenomenon of intragenic recombination in *Drosophila melanogaster*. Mutations, which on the basis of the complementation test were alleles, recombined with a very low frequency. Thus, the atom of genetics was not indivisible. Green and Green (1949) were able to map mutations of the *lozenge* locus of *D. melanogaster* into linear order. Roper (1950) and Pontecorvo (1952) carried the analysis even further. They observed intragenic recombination within genes in the ascomycete *Aspergillus nidulans*, which were known to control the synthesis of one single enzyme. Bonner (1950) and Giles (1952) observed the same in *Neurospora crassa*. Pritchard (1955) was the first to demonstrate with microbic fungi that mutations within a single gene could be mapped into linear order by means of recombination.

The Neoclassical Concept of the Gene

Our comprehension of the nature of the genetic material became more accurate when Avery et al. (1944) demonstrated that the substance causing transformation in bacteria was DNA. Transformation had been discovered by Griffith (1928). He observed that killed bacterial cells injected into mice were able to transform genetically different living bacteria into their own kind.

Dawson and Sia (1931) and Alloway (1932) demonstrated transformation *in vitro* in a cell-free extract. Avery's group (Avery et al., 1944) was able to isolate from this extract the substance which was responsible for transformation. That substance was DNA. Thus, genes consisted of DNA. Further support for the DNA theory of inheritance was gained when Hershey and Chase (1952) demonstrated that DNA is the only component of bacterial virus that enters its bacterial host, and therefore it must be presumed to be the sole bearer of viral genetic information.

From the point of view of the conceptual framework of genetics, the experiments of Benzer (1955, 1957, 1959, 1961) involving the genetic fine structure of the bacteriophage T4 rII-region turned out to be of fundamental importance. With the aid of a selective technique, he was able to map hundreds of mutations of that region into a linear order. The gene as the unit of function was not indivisible; within the gene, independently mutating mutation sites existed, which could be separated from one another by means of genetic recombination. Benzer created a new terminology. He called the unit of genetic function the "cistron". The cistron is operationally defined with the aid of the *cis-trans* test (Fig. 1). In the *cis-trans* test, *cis-* and *trans-*

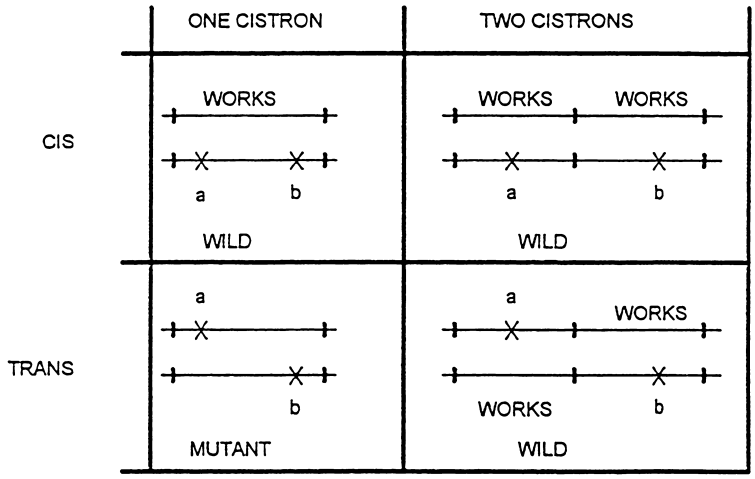

Fig. 1. The principle of the *cis-trans* test. If mutations *a* and *b* belong to the same cistron, the phenotypes of the *cis-* and *trans*-heterozygotes are different. If, however, the mutations belong to different cistrons the *cis-* and *trans*-heterozygotes are phenotypically similar. The notion "works" on the figure means that the cistron is able to produce a functional polypeptide. Mutations *a* and *b* are recessive mutations which produce the same phenotype.

heterozygotes are compared. In the *cis*-heterozygote the mutations are in the same chromosome, but in the *trans*-heterozygote they are in homologous chromosomes. Thus, the genotype of the *cis*-heterozygote is designated as *a b/++,* and that of the *trans*-heterozygote as *a+/+b.* If the *cis*-heterozygote is of a wild type phenotype and the *trans*-heterozygote is mutant, *a* and *b* are mutations of the same cistron. If, however, both *cis*- and *trans*-heterozygotes are phenotypically of a wild type, *a* and *b* are mutations of different cistrons. (Mutations *a* and *b* are recessive mutations producing the same phenotype, and the symbols + refer to their normal counterparts). The cistron is a synonym for the gene, but this term should be used only when it is based on the *cis-trans* test or biochemical evidence, which will be dealt with later.

The smallest unit of recombination Benzer called the ''recon'', and the smallest unit of mutation the ''muton''. The recon is the smallest unit of genetic material which can be separated from other such units by means of genetic recombination but which cannot be divided further. A muton is the smallest unit of genetic material a change in which is sufficient to cause a mutant phenotype. Benzer also called attention to the correspondence between the linear internal structure of the cistron and the linear structure of the DNA molecule referring to the famous model of DNA structure by Watson and Crick (1953). The linear structure of the DNA molecule was demonstrated already in the late 1930s by William T. Astbury and Thorbjörn Caspersson together with Florence Bell by using X-ray structural analysis (see Jahn, 1998).

Dounce (1952) and Gamow (1954) independently presented the so-called colinearity hypothesis, according to which the linear structure of DNA determines the linear primary structure of a polypeptide. The colinearity hypothesis was shown to be true by Sarabhai et al. (1964) and Yanofsky et al. (1964, 1967), by comparing the genetic map of the T4 phage coat protein gene and the corresponding primary structure of the polypeptide, and also by comparing the genetic map of the tryptophane synthetase gene of *Escherichia coli* with the corresponding primary structure of the polypeptide. In other words, Yanofsky's group demonstrated the co-linearity of mutant nucleotide sites and the corresponding mutated aminoacyl polypeptide sites. Further, Yanofsky's group was able to show that the material counterpart of a cistron was that part of the DNA molecule which coded information for the synthesis of a single polypeptide. This they demonstrated by showing that all the mutations of the *E. coli* tryptophan synthethase A-protein were mutations of the A-cistron, and likewise the mutations of the B-protein of the B-cistron. Finally, by studying the A-protein of the tryptophan synthethase Yanofsky's

group was able to show that the material counterpart of both recon and muton was one single nucleotide pair in the structure of DNA (Crawford & Yanofsky, 1958; Yanofsky & Crawford, 1959). Thus, the cornerstone of the neoclassical view of the gene became the one cistron/one polypeptide hypothesis, which replaced the old one gene/one enzyme hypothesis. (An enzyme molecule can consist of a single polypeptide molecule, but usually it consists of several identical [homomultimer] or non-identical [heteromultimer] polypeptides). Thus, the problem was how the information encoded in DNA for the assembly of specific amino acid sequences of the polypeptides was in fact constructed.

The idea was proposed, therefore, that the heterocatalytic function of DNA is a two-stage process. In the first stage of this process, each DNA gene would serve as a template for the synthesis of RNA molecules onto which the precise nucleotide sequence making up that gene, and hence its encoded amino acid sequence information, is *transcribed* into RNA molecules. After their transcription on the DNA template, these RNA molecules were then imagined to migrate to the cytoplasm, where in the second stage of the heterocatalytic function their nucleotide sequences would be *translated* into polypeptide chains of predetermined primary structure (see Stent, 1971; Watson, 1963 for narrative reviews of this part of the history of the concept of the gene).

According to an early version of the theory of the information structure of protein synthesis, the RNA transcript was thought to provide the RNA moieties for newly formed ribosomes. Hence, each gene was imagined to give rise to the formation of one *specialized* kind of ribosome, which in turn would direct the synthesis of one and only one kind of protein – a scheme that Brenner, Jacob, and Meselsohn (1961) epitomized as the ''one gene/one ribosome/one protein'' hypothesis.

Under the ''one gene-one ribosome-one protein'' hypothesis, one would have expected a burst of ribosomal RNA synthesis following phage infection of bacterial cells, while the phage-infected cell is renovated for future production of the polypeptide chains encoded in the phage DNA. But contrary to that expectation, Cohen (1948) had found that upon infection of *Escherichia coli* with T2 phage, net synthesis of RNA, and hence of ribosomes, not only does not accelerate but comes to a stop, indicating that the synthesis of new kinds of ribosomes is not a precondition for the synthesis of new kinds of proteins. This observation was subsequently confirmed by Manson (1953). Hershey (1953) and Hershey et al. (1953), however, were able to detect a *small* amount of RNA synthesis in the bacterial host cell of virus-infected bacteria by using isotopically labeled precursors. Extending this finding, Volkin and

Astrachan (1956a, 1956b) measured phosphorus incorporation into the RNA of different subcellular fractions of *E. coli* after phage infection. The remarkable finding was that the purine-pyrimidine base composition of the RNA of bacterial cells after T7 phage infection is significantly different than that of *E. coli* ribosomal RNA and instead resembles more closely that of the T7 phage DNA, indicating that the RNA was virus specific (Volkin et al., 1958). Subsequently, Volkin (1960) measured a rapid turnover of RNA formed after infection, it having a short half-life of the order of a few minutes, and found that phage production would not occur in the absence of this fraction. Davern and Meselson (1960) and Riley et al. (1960) obtained similar results, i.e., that the RNA involved in phage infection was unstable, unlike the ribosomal RNA, which was stable. Thus, the hypothesis developed that protein synthesis was concluded by the unstable virus-specific RNA and not by the ribosomal RNA.

The time was thus ripe to present direct evidence of the role of RNA in protein synthesis. This was done by Brenner et al. (1961), when they showed with the aid of heavy carbon isotopes and ultra-centrifugation that in phage infection it was indeed the RNA of the phage and not the bacterium which was responsible for the synthesis of the phage coat protein. Jacob and Monod (1961) had proposed the name messenger RNA (mRNA) for this protein-synthesis-conducting RNA. Thus, the neoclassical view of the gene culminated in a theory according to which one gene or cistron controls the synthesis of one messenger RNA molecule, which in turn controls the synthesis of one polypeptide.

The Breakdown of the Neoclassical Gene Concept and the Modern Concept of the Gene

The breakdown of the neoclassical concept of the gene started in the beginning of the 1970s, with the new discoveries of gene technology and molecular biology. These were the discoveries of repeated genes, interrupted genes and alternative splicing, the special case of immunoglobulin genes, overlapping genes, movable genes, complex promoters, multiple polyadenylation sites, polyprotein genes, the editing of messenger RNA and nested genes. Such observations have led to a situation where none of the classical or the neo-classical criteria of the definition of the gene hold strictly true. We, therefore, have to adopt a new, open, general and abstract concept of the gene, despite the fact that our comprehension of the nature and organization of the genetic material has greatly increased.

Linn and Arber (1968) and Meselson and Yuan (1968) found specific restriction endonucleases, i.e., enzymes that cut DNA in bacteria, which act

when the latter defend themselves against the attack of bacteriophages; thus, these enzymes restrict the host range of the bacteriophages. Smith and Wilcox (1970) were able to purify these enzymes, and Kelly and Smith (1970), Danna and Nathans (1971), and Sharp et al. (1973) determined their mode of action. These enzymes cut DNA molecules each at a specific site. These observations made it possible to isolate genes, to clone them and analyze their biochemical structure in great detail. Following the action of restriction endonucleases, there often arise so-called cohesive ends in the DNA molecules (Merz & David, 1972), which tend to join together. By this means it is possible, for example, to join together DNA from any eukaryotic organism and that from bacterial plasmids. Such recombinant DNA molecules were first constructed by Jackson et al. (1972), Lobban and Kaiser (1973) and Cohen et al. (1973). Cloned DNA molecules can be physically mapped, using the cutting points of the restriction endonucleases as markers, (Southern, 1975) and sequenced by means of sophisticated biochemical methods (Maxam & Gilbert, 1977; Sanger et al., 1977).

Genetic Discoveries

Repeated Genes

Waring and Britten (1966) and Britten and Kohne (1967, 1968a, 1968b) were the first to observe repeated DNA sequences in many organisms by means of the reassociation kinetics of DNA, the organization of single-stranded DNA molecules experimentally into double stranded molecules. The first observations of repeated structural genes, i.e., genes that encode for RNA or proteins, concerned amphibian ribosomal RNA genes (genes that encode for the RNA structural part of the cytoplasmic particles called ribosomes on which protein synthesis occurs) (Miller & Beatly, 1969a, 1969b, 1969c) and sea urchin histone genes (Weinberg et al., 1972).

The genes of ribosomal RNA are repeated in several tandem copies. Each one consists of one transcription unit, but the gene cluster is usually transmitted from one generation to the next as a single unit. Thus, the units of transmission and transcription are not always the same. Likewise, the histone genes have been observed to be repeated in such tandem repeats in many higher eukaryotic organisms (Lewin, 1980a).

Interrupted Genes and Alternative Splicing

The first observations of interrupted (split) genes, i.e., genes in which there are noncoding *intron* sequences between the coding *exon* sequences, were made

in animal viruses in 1977 (Berget et al., 1978; Broker et al., 1978; Westphal & Lai, 1978). These observations were based on physical mapping of the inner structure of the genes by means of so-called Southern blotting (Southern, 1975). In this method, DNA fragments created by restriction endonucleases are separated according to their size in electrophoresis.

Almost at the same time as the virus observations, split genes were also found in eukaryotic organisms (Bell et al., 1980; Breathnach et al., 1977; Jeffreys & Flavel, 1977; Konkel et al., 1978; Lomedico et al., 1979; Tilghman et al., 1978; Wozney et al., 1981a, 1981b; Yamada et al., 1980). Since then, split genes have been found in all eukaryotic organisms investigated and their viruses, and it can be said that interrupted genes are the rule rather than the exception in the organization of the genes of multicellular eukaryotic organisms. In unicellular eukaryotic organisms, they are found to a lesser extent. However, there are exceptions among multicellular organisms too, of which the evolutionarily old histone genes are worth mentioning (Kedes, 1979).

In terms of the concept of the gene, interrupted genes constitute an interesting case in two respects. Firstly, the existence of introns between the coding exons shows that there is no one-to-one colinear relation between the gene and the polypeptide; coding sequences are interrupted by noncoding DNA. Secondly, tissue and stage-specific alternative splicing occurs in certain genes when interrupted genes produce messenger RNA. The interrupted gene produces a primary transcription product, a heterogenous nuclear RNA molecule, in which both exons and introns are represented. Introns, however, are removed from the primary transcript during the processing of messenger RNA in specific splicing reactions. Splicing is usually constitutive, which means that all exons are joined together in the order in which they occur in the heterogenous nuclear RNA. In many genes, however, alternative splicing has also been observed, in which the exons may be combined in some other way (Fig. 2). For example, some exon or exons may be skipped in the splicing reaction. The primary order of the exons is not, however, altered even in alternative splicing. Thus, alternative splicing makes it possible for a single gene to produce more than one messenger RNA molecule, which contradicts the basic conceptual framework of the neoclassical view of the gene.

Alternative splicing was first observed in animal viruses (Berk & Sharp, 1978a; Canaani et al., 1979; Chost et al., 1978a, 1978b; Crawford et al., 1978; Flavell et al., 1979; Horowitz et al., 1978; Lai et al., 1978; Paucha et al., 1978). The first observations of alternative splicing in the genes of eukaryotes

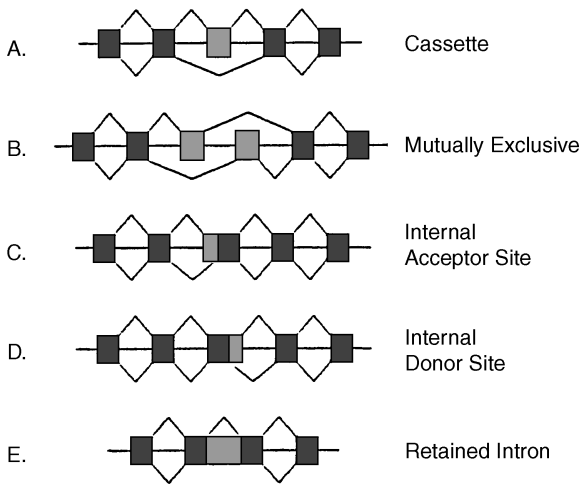

Fig. 2. Patterns of alternative splicing. Constitutive exons (black), alternative exons (dotted), and introns (heavy solid lines) are spliced according to different pathways (fine solid lines). A. A casette is defined as an exon that can be individually excluded from the mature mRNA. B. Mutually exclusive exons are pairs of consecutive casette exons that are included in the mRNA in a mutually exclusive manner. C. and D. Internal acceptor and donor sites are splicing sites that lie entirely within potential RNA coding sequence which may or may not incorporate a casette exon into mRNA. E. A retained intron can be viewed as an 'optional' intron contained within a larger exon, and may or may not be incorporated into mRNA depending on whether the donor-acceptor pair will be failed to be spliced or to be spliced (Andreadis et al., 1987).

concerned murine immunoglobulin genes (Early et al., 1979, 1980; Moore et al., 1981; Sakano et al. 1979a). Since then, alternative splicing has been observed in hundreds of genes in various eukaryotic organisms, man included (see Lewin, 1980b for review).

The tissue specificity of alternative splicing was first shown in the fibrinogene genes of rat and man (the gene which encodes for the fibrinogene protein, the precursor for fibrin) (Crabtree & Kant, 1983). The first observations of developmental stage specificity concerned the alcohol dehydrogenase gene of *Drosophila melanogaster* (the gene that encodes for the alcohol dehydrogenase enzyme, an enzyme of alcohol metabolism) (Benyajati et al., 1983). The first demonstration that alternative splicing was both tissue and stage-specific concerned the trompomyosin gene of *D. melanogaster* and rat (Basi et al., 1984; Medford et al., 1984). The tissue and stage specificity of

alternative splicing naturally constitutes a previously unknown and effective mechanism of gene regulation.

Immunoglobulin Genes

The enormous versatility of antibodies was for a long time a difficult problem in genetics. How was it possible that in the genome there was room for the codes of millions of different antibodies? The matter was solved when it became clear that the functional genes of immunoglobulins mature by means of somatic recombination from a few units in the germ line during the maturation of immune cells.

Each antibody molecule is a tetramere, consisting of two identical light chains and two identical heavy chains. Each chain consists of a constant and a variable region. In the genome of the germ line, there are many genes for the variable region and a few genes for the constant region. In somatic recombination, these can be combined during the maturation of the functional antibody gene into several thousands of different combinations whereby millions of different antibodies are formed. This phenomenon was first demonstrated by Hozumi and Tonegawa (1976; see also Tonegawa et al., 1978). The observation was subsequently confirmed in several laboratories using different methods. This prevailing theory of antibody formation was preceded by the selective theory of antibody formation by Jerne (1955) and Burnet (1959). According to Jerne, all the antibodies are preformed (constitutive) in the immune system, and will be combined with the antigens, as a result of which immune complexes will be formed. These complexes are then introduced into the macrophage cells where the antigen is cleaved from the complex, and the information of the antibody is presented to the lymphocytes, which then begin to form the antibodies in huge quantities. Burnet on his part assumed that natural antibodies do not react with the antigens, but rather that specific clones of immunocompetent cells are selected. Further, Burnet assumed that during embryogenesis somatic mutations cause differences in the antigen receptors of the cells which will then be selected. The immunoglobulin genes, which can be called assembled genes (Dillon, 1987), do not fit any classical or neoclassical definition of the gene, since the genetic unit in the germ line and in the mature immune cell is completely different.

Overlapping Genes

The first observations of overlapping genes were made in the bacteriophages ΦX 174 (Sanger et al., 1977) and G4 (Shaw et al., 1978), and in the animal

virus SV40 (Contreas et al., 1977; Fiers et al., 1978). In the bacteriophage ΦX 174, several genes overlapped, encoding different proteins read from the same DNA strand but in different reading frames. In the G4-phage, the situation was even more complicated. In that phage the same DNA strand encoded as many as three different proteins, the messenger RNAs of which were transcribed overlappingly in all three possible reading frames. (A reading frame is a sequence of the nucleotides of DNA or RNA read in triplets; one of the three possible ways of reading a nucleotide sequence as a series of triplets. The reading frame is open when it starts with an initiator codon, which is followed by a sequence of codons coding for amino acids. An open reading frame ends with a terminator codon). In these phages, however, the genes overlapped for only a few codons. In the SV40 virus, on the other hand, the genes overlapped for as many as 122 nucleotides. Since these early findings, overlapping genes have also been found in eukaryotic multicellular organisms such as *Drosophila melanogaster* (Spencer et al., 1986), mouse (Williams & Fried, 1986), and rat (Eveleth & Marsh, 1987). Overlapping genes can be located on the same DNA strand or on opposite strands.

A particular case of overlapping genes are the so-called hidden genes, hidden frames or mini-cistrons found in the poliovirus (Pierangeli et al., 1998). These are short open reading frames, i.e., sequences initiated by the alternative translation initiation codons ACG, AUA, and GUG in the 5′-terminal extra-cistronic region of poliovirus RNA. Mutations in all but one of the ten found of these mini-cistrons had no effect on the infectivity of full-length cDNA clones, except when they modified a ‘‘hidden frame’’ spanning between nucleotides 157–192 with a starting triplet ACG. The possibility, therefore, had to be entertained that a short ‘‘hidden’’ frame starting with the alternative initiation codon ACG was indeed endowed with coding capacity, and that its suppression entailed the observed loss of infectivity of poliovirus cDNA.

Neither the classical nor the neoclassical view of the gene encompassed the possibility of overlapping genes, since genes were believed to reside on the chromosome always in tandem. The evolution of overlapping genes is also a difficult problem which, however, will not be dealt with here.

Moveable Genes

Movable genes are DNA elements that can move from one location to another in the genome of an organism. Already in the 1940s, McClintock (1947, 1948) explained certain variegated phenotypes of maize by means of movable genes, which she called ‘‘control elements’’. At the time, however, these elements

appeared so odd that nobody really knew what to think of them. Nowadays movable genes have been found in virtually all organisms and their molecular nature is quite well known (see Berg & Howe, 1989; Cohen & Shapiro, 1980; Fedoroff, 1984 for reviews). Consequently, mobile genetic elements have became one of the most important discoveries of genetics, and Barbara McClintock was awarded the Nobel prize for physiology or medicine in 1983 at the age of 81 years.

In addition to the ability of movable genes to move in the chromosomal complement from one location to another, they can also move from one individual to another and even, to a certain extent, from one species to another at least in the genus *Drosophila* (see Ajioka & Hartl, 1989 and Bi'emont & Cizeron, 1999 for reviews). Movable genes thus constitute an important evolutionary factor. When jumping from one individual or from one species to another, which process is called horizontal transfer, the mobile DNA elements are covered by a proteinaceous envelope (Bi'emont & Cizeron, 1999). The existence of movable genes shows that the hypothesis of a fixed location of the gene in the chromosome, adopted by both the classical and neoclassical view, does not necessarily hold true.

Complex Promoters

Promoters are DNA sequences on the 5′ side of the gene, i.e., in the beginning of the gene, on which the RNA polymerase fastens when transcription begins. In all groups of organisms alternative promoters have been shown for many genes. These alternative promoters have been classified into six classes by Schibler and Sierra (1987) (Fig. 3). Certain types of alternative promoters make it possible for transcription to start from different points of the gene in different cases, and for the transcripts to have initiation codons at different positions of the chromosome. Thus, it is possible for a single gene in this case too to produce more than one type of messenger RNA molecules, encoding more than one polypeptide. This is again against the basic conceptual framework of the neoclassical view of the gene. In higher eukaryotes alternative promoters are typically tissue and/or stage specific, as exemplified by the alcohol dehydrogenase gene of *Drosophila melanogaster* (Benyajati et al., 1983), the murine alpha-amylase gene (Sierra et al., 1986; Young, 1981), and the aldolase gene of rat and man (Joh et al., 1986; Maire et al., 1987; Schweighaffer et al., 1986). Simple and complex transcription units are distinguished according to whether the unit of transcription is controlled by one or several promoters.

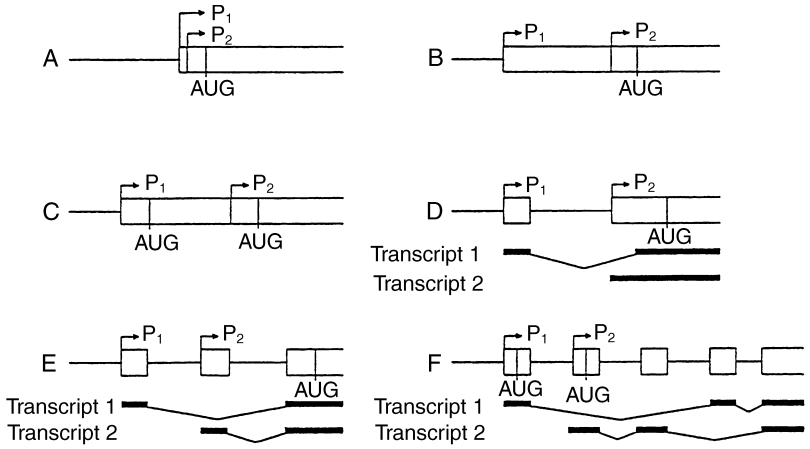

Fig. 3. Various architectures of genes with tandem promoters. Gene regions specifying mature mRNA sequences are shown as open bars. 5′ flanking regions and intervening sequences are drawn as solid lines. Upstream and downstream transcription initiation sites are depicted as bent arrows. The positions of the AUG translation initiation codons (ATG in DNA) are indicated by a vertical line. For interrupted genes, the mode of transcript splicing is indicated. Exons and introns are shown as filled bars and solid bent lines, respectively (Schibler & Sierra, 1987).

Multiple Polyadenylation Sites

During the maturation of messenger RNA, i.e., during the process in which the primary transcript ripens to form messenger RNA, about 200 adenosine nucleotides are added in a polyadenylation reaction at the 3′ end. These are not coded by the corresponding gene. In certain cases, there are multiple alternative polyadenylation sites in the primary transcript. This was first observed in adenoviruses (Klessig, 1977; McGrogan & Raskas, 1978; Nevins & Darnell, 1978; Wilson & Darnell, 1981; Ziff & Fraser, 1980). In cellular genes, many alternative polyadenylation sites have also been found (see Left et al., 1986 for review). Alternative polyadenylation sites usually involve the untranslated trailer sequence in the messenger RNA, but they can also involve translated sequences, and in this case they can affect the structure of the encoded protein. Thus, multiple polyadenylation sites are one mechanism whereby a single gene can control the synthesis of more than one polypeptide.

Polyprotein Genes

In viruses, very many genes encode for one single large polypeptide which, however, after translation, is cleaved enzymatically into smaller subunits. Such polyprotein genes are also known in multicellular eukaryotes. Such are, for example, the neuropeptide genes of mammals (see Douglas et al., 1984; Scheller et al., 1984 for reviews) and the proline-rich proteins of salivary glands (Lyons et al., 1988). Thus, polyprotein genes contradict the hypothesis adopted by the neoclassical view of the gene, that each gene encodes for a single polypeptide.

Editing of Messenger RNA

In trypanosomes, and in certain plant RNA encoded by mitocondrial DNA, a peculiar phenomenon called RNA editing has been observed. In this editing of messenger RNA, uracil nucleotides are removed and cytocine nucleotides are replaced by uracil after transcription. The information for this process comes either from the gene involved or alternatively from outside it. In the latter case, the information source is called guide RNA (see Weissmann et al., 1990 for review). Although probably a rare phenomenon, editing of RNA is contrary to the classical and neoclassical concept of the gene, since in this case messenger RNA can retrieve information from outside the gene by which it was encoded.

Pseudogenes

Pseudogenes are DNA sequences significantly homologous to a functional gene which, however, have been altered so as to prevent any normal function (Rieger et al., 1991). The first pseudogene was found in the 5S DNA gene family of *Xenopus laevis* by Jacq et al. (1977). It was observed that the repeat length of about 700 residues of the oocyte 5S DNA of *X. laevis* could be divided into a gene region coding for the 5S RNA product and a much longer ''spacer'' region separating the coding regions in the adjacent repeats (Brown et al., 1971). Jacq et al. (1977) were able to observe that the gene lies two thirds of the way along the unit length within the G + C-rich region of the repeating gene-spaces unit. Furthermore, a partially duplicated region of the sequence, the pseudogene, lies within the same G + C-rich region on the 3′ side of the gene region. The pseudogene in this case is a structure in which residues 1–101 of the gene are repeated with nine or possibly a few more base changes. Pseudogenes are a common feature of many multigene families in higher eukaryotes. In fact, pseudogenes are ancient genes which have lost their function, and pseudogenes which have in the past been protein coding

genes usually contain many stop codons, i.e., their reading frame has been closed during the course of evolution.

Nested Genes

Nested genes, i.e., a situation in which one gene resides within an intron of another gene, was first demonstrated in the *Gart* locus of *Drosophila melanogaster* by Henikoff et al. (1986). In this particular case the nested genes were on opposite strands of DNA. Chen et al. (1987) in turn demonstrated that in the large intron of the *dunce* locus of *D. melanogaster* there were actually two other genes residing, of which one was the known *Sgs-4* gene. In this case the nested genes were on the same strand of DNA. Levinson et al. (1990) were the first to demonstrate nested genes in man. The 22nd intron of the blood coagulation factor VIII gene included another gene in its opposite strand. The large intron of the human neurofibromatosis gene includes a total of three other transcription units in two opposed orientations (Xu et al., 1990). The existence of nested genes is in contradiction to the central hypothesis adopted by both the classical and neoclassical gene concept, that genes are located in linear order on the chromosome.

Enhancers and Their Relation to Genetic Complementation

Enhancers are located at the 5′ and 3′ end of the gene, which respond to the signals mediated by the proteins regulating the function of the gene. Enhancers can also be located within the introns. The regulative effect of the enhancers is either positive or negative. In the latter case they are often called silencers (for reviews concerning enhancers and silencers, see for example Serfling et al., 1985; Schöler et al., 1988).

The relation of enhancers to genetic complementation is interesting. Denote an enhancer by E and the transcription unit regulated by it by g. In the *cis-trans* test, the $E^- g^-/E^+ g^+$ *cis*-heterozygote is phenotypically wild, whereas the $E^- g^+/E^+ g^-$ *trans-heterozygote* is phenotypically mutant. Thus, the *cis-trans* test gives a positive result. This means that we cannot, on the basis of a genetic test alone, distinguish between an enhancer and the transcription unit regulated by it; biochemical evidence is needed. Thus, by definition, the regulatory elements of a transcription unit, such as enhancers, have to be included in the gene itself. The first enhancers were demonstrated in the SV40-virus (Banerji et al., 1981; Moreau et al., 1981). The first enhancers in nucleated cells were demonstrated in the immunoglobulin heavy chain gene (Banerji et al., 1983; Gillies et al., 1983).

III. THE PRESENT CONCEPT OF THE GENE

The examples presented above show that none of the classical or neoclassical criteria of the definition of the gene hold strictly true. The current view of the gene is of necessity an abstract, general, and open one, despite the fact that our comprehension of the structure and organization of the genetic material has greatly increased. Simply put, our comprehension has outgrown the classical and neoclassical terminology. Open concepts, with large reference potential, are, however, very useful in science in general, as pointed out by Burian (1985). In fact, it should be stressed that our comprehension of the very concept of gene has always been abstract and open, as indicated already by Johannsen (1909).

Due to the open nature of the concept of the gene, it takes different meanings depending on the context. Singer and Berg (1991) have pointed out that many different definitions of the gene are possible. If we want to adopt a molecular definition, they suggest the following definition: ''A eukaryotic gene is a combination of DNA segments that together constitute an expressible unit. Expression leads to the formation of one or more specific functional gene products that may be either RNA molecules or polypeptides. Each gene includes one or more DNA segments that regulate the transcription of the gene and thus its expression'' (p. 622). Thus the segments of a gene include (1) a transcription unit, which includes the coding sequences, the introns, the flanking sequences – the leader and trailer sequences, and (2) the regulatory sequences, which flank the transcription unit and which are necessary for its specific function. To take a more formal definition of the gene, one can say that a gene consists of elements on the chromosome that give a positive result in the *cis-trans* test. Population geneticists can, on their part, treat the gene as a simple calculation unit segregating in the population.

Past and Present Contemplations on Genes, Genomes, Genotypes and Phenotypes

When Johannsen (1909) coined the term ''gene'', he wished this unit of heredity to be free of any hypotheses regarding its physical or chemical nature, i.e., the genes could be treated as calculating units. Throughout the first half-century of the history of genetics, the concept of the physical structure of genes remained more or less static. The gene was generally regarded as the unit of a genetic system, an indivisible entity in the processes of genetic recombination, self-reproduction, and mutation (Demerec, 1967). The nature

of genes became really more dynamic only with the development of the DNA-theory of inheritance. The gene could now be regarded as a heritable unit of function, a segment of a chromosome, in most instances a segment of a Watson–Crick DNA molecule, which specified the structure of a single polypeptide chain and was made up of a large number of mutational sites, nucleotides by their material nature, biologically separable by recombination. The discoveries of gene technology beginning from the early 1970s and presented in this essay have proven that the nature of genes is even more dynamic than was thought in the late 1960s.

Johannsen (1909) also coined the concepts of genotype and phenotype. He defined the phenotypes as fully concrete realities: the organism as it can be observed and measured. It should be stressed that this observability regards all the levels of the hierarchical biological organization beginning from the lowest molecular level and ending to the highest behavioral or functional level. Johannsen defined the genotype as a total whole of all the genes of an organism.

Woltereck (1909) and Schmalhausen (1949) made an important contribution to the discussion on genotypes and phenotypes by introducing the concept of the norm of reaction. It is the range of potential phenotypes that a single genotype could develop if exposed to a specified range of environmental conditions, i.e., it is the way in which a given genotype reacts to the environmental conditions in which it develops. The genotype determines the norm of reaction.

Mahner and Kary (1997) presented a detailed analysis of the concepts of genomes, genotypes and phenotypes. They, like Brandon (1990) and Lewontin (1992), came to the conclusion that an individual gene or allele is itself a phenotypic character. As precise as the analysis of Mahner and Kary (1997) is, however, they fail by totally neglecting the concept of the norm of reaction.

Lewontin writes: ‘‘At the lowest level the DNA sequence of the genes itself is a phenotype, and the complete description of the DNA sequence is identical with a complete specification of the genotype’’ (Lewontin, p. 143). I would, however, stress a different viewpoint. The complete DNA sequence of an organism, its genotype, is a concrete entity, but the genotype determines the norm of reaction, which, for its part, is an abstract concept. It is the total whole of the *possibilities* residing in the genes.

This is what geneticists have recently written on genes, genotypes and phenotypes. But what about the philosophers of science? Kitcher (1992) stresses that representatives of different fields of biology use the term ‘‘gene’’

in different ways. For example, for population genetics and evolutionary biologists, the word ''gene'' in most instances still means a static unit of calculation, while molecular biologists think in terms of DNA structure and function and do not worry much about the conceptual questions. Therefore, according to Kitcher, the general discussion of genes is problematic but interesting. Finally, he himself takes a pragmatic view: ''A gene is anything a competent biologist has chosen to call a gene'' (p. 131).

Waters, for his part, distinguishes the classical gene concept and the molecular gene concept. The former is, in his terminology, a genetic determinant that is responsible for a given difference in characters, while ''the fundamental concept underlying the application of 'gene' in molecular biology is that of a *gene for a linear sequence in a product at some stage of genetic expression*'' (1994, p. 178; italics original). According to Waters, the domain of the application of the classical and molecular concepts of the gene differs because the classical term applies to regulatory regions such as operators, whereas the molecular one does not. Here he, however, unfortunately fails. According to the *cis-trans* test, which is still a valid operation for the definition of the gene, the regulatory regions such as operators or enhancers are parts of the gene itself as shown in the present essay.

Most Recent Advances in the Study of the Organization of Genomes

In recent years the whole genomes of hundreds of viruses, over a dozen bacteria, and until the summer of 2000 three eukaryotic organisms (viz., baker's yeast (*Saccharomyces cerevisae*), the nematode *Caenorhabditis elegans*, and the fruit fly (*Drosophila melanogaster*)) have been sequenced. The genome of *Saccharomyces cerevisae* appeared to contain 5,885 genes (Goffeau et al., 1996), and that of *Caenorhabditis elegans* 19,099 genes (The *C. elegans* Sequencing Consortium, 1998). The comparison of the yeast and nematode genomes revealed that most of the core biological functions are carried out by orthologous proteins (proteins of different species that can be traced back to a common ancestor) that occur in comparable numbers. The specialized processes of signal transduction and regulatory control that are unique to the multicellular worm appear to use novel proteins, many of which re-use conserved domains. The proteins conserved in yeast and worm are likely to have orthologs throughout eukaryotes; in contrast, the proteins unique to the worm may well define metazoans (Chervitz et al., 1998).

Surprisingly enough, the genome of the fruit fly appeared to contain substantially fewer genes than the nematode, namely, 13,601 genes (Adams

et al., 2000). These encode for 14,113 transcripts since some genes are able to produce more than one transcript through alternative splicing. The 13,601 genes of the fruit fly contain a total of 56,673 exons, i.e., approximately a mean of four exons per one gene. The mean density of genes in the *D. melanogaster* genes is one gene per nine thousand nucleotide pairs of DNA (Adams et al., 2000). Approximately 10% of the genes are nested, i.e., located inside the introns of other genes (Ashburner, 1999a, 1999b), which is an unexpectedly high proportion.

Two human chromosomes (viz., numbers 22 (Dunham et al., 1999) and 21 (Hattori et al., 2000)), have been completely sequenced so far. These chromosomes contain 545 and 225 genes, respectively, and 2% of the DNA content of the whole human genome, giving an estimate of approximately 40,000 genes for the whole human genome, a substantially lower number than the previously anticipated 100,000.

In December 2000, the complete genome of the thale crass *Arabidopsis thaliana* was also decoded, this being the first flowering plant analysed so far. The genome of *A. thaliana* contains 25,498 genes (The Arabidopsis Genome Initiative, 2000). All of the estimates of the number of genes in different species, however, are only best approximations because of the limitations of gene binding programs.

On February 12th, 2001, a working draft of the DNA sequence covering 90% of the entire euchromatic human genome was published by the publicly funded Human Genome Organization project (International Human Genome Sequencing Consortium, 2001) and a commercial project (Venter et al., 2001). According to these reports the human genome consists of approximately 25,000–40,000 (most likely approximately 32,000 genes). Of these, around 15,000 are known genes and 17,000 predictions.

Up to this point (February, 2001), the different sequencing projects, in addition to the work done on the human genome, include the genome sequences of 599 viruses and viroids, 205 naturally occurring plasmids, 185 organelles, 31 eubacteria, seven achaea and four eukaryotic organisms (viz., yeast *Saccharomyce cerevisiseae*, nematode *Caenorhabditis elegans*, fruit fly *Drosophila melanogaster*, and the plant *Arabidopsis thaliana*). Philosophically, the most important result of a preliminary comparative analysis of these genomes is the surprising genetic similarity and uniformity of the whole eukaryotic empire, if not the whole of creation. Less than 1% of homologues of the predicted human proteins are found in prokaryotes only, 21% is found in all prokaryotes and eukaryotes, 32% in all organisms except prokaryotes, 24%

in all animals, and 22% in vertebrates only. Only 1% of the predicted human proteins are unique to human (International Human Genome Sequencing Consortium, 2001). However, it is important to notice that the complexity of the machinery of gene expression regulation increases with the increasing complexity of the organism (Tupler et al., 2001). ''Another notable feature of [the human genome] is the much lower gene tally than anticipated, which indicates that human complexity does not arise solely from the number of genes.'' ''Humans are much more than simply the product of a genome, but in a sense we are, both collectively and individually, defined within the genome. The mapping, sequencing and analysis of the human genome is therefore a fundamental advance in self-knowledge, it will strike a personal chord with many people. And application of this knowledge will, in time, materially benefit almost everyone in the world'' (Danis, Gallagher, & Campbell, 2001).

ACKNOWLEDGEMENT

This work was supported by The Academy of Finland.

REFERENCES

Adams, M.D., et al. (2000). The genome sequence of *Drosophila melanogaster. Science, 287*, 2185–2195.

Ajioka, J.W., & Hartl, D.L. (1989). Population dynamics of transposable elements. In: D.E. Berg & M.M. Howe (Eds.), *Mobile DNA* (pp. 939–958). Washington, DC: American Society of Microbiology.

Alloway, J.L. (1932). The transformation in vitro of R pneumococci into S forms of different specific types by the use of filtered pneumococcus extracts. *Journal of Experimental Medicine*, *55*, 91–99.

Andreadis, A., et al. (1987). Generation of protein isoform diversity by alternative splicing: Mechanistic and biological implications. *Annual Review of Cell Biology*, *3*, 207–242.

Ashburner, M. (1999a). The *Drosophila* genome project: A progress report. *Abstracts of the 16th European Drosophila Research Conference* (p. 24). Zürich.

Ashburner, M. (1999b). *Drosophila* comes of age – The sequence. *Abstracts of the Genetical Society Autumn Meeting* (p. 9). London.

Avery, O.T., et al. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcal type III. *Journal of Experimental Medicine*, *79*, 137–158.

Banerji, J., et al. (1981). Expression of a globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*, 299–308.

Banerji, J., et al. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, *33*, 729–740.

Basi, G.S., et al. (1984). Alternative splicing of a Drosophila trompomyosin gene generates muscle trompomyosin isoforms with different carboxyterminal ends. *Molecular and Cellular Biology*, *12*, 2828–2836.

Beadle, G.W., & Tatum, E.L. (1941). Genetic control of biochemical reactions in Neurospora. *Proceedings of the National Academy of Science USA*, *27*, 499–506.

Bell, G.I., et al. (1980). Sequence of the human insulin gene. *Nature*, *284*, 26–32.

Benyajati, C., et al. (1983). The messenger RNA for alcohol dehydrogenase in Drosophila melanogaster differs in its 5′ end in different developmental stages. *Cell*, *33*, 125–133.

Benzer, S. (1955). Fine structure of a genetic region in bacteriophage. *Proceedings of the National Academy of Science USA*, *41*, 344–354.

Benzer, S. (1957). The elementary units of heredity. In: W.D. McElroy & B. Glass (Eds.), *The chemical basis of heredity* (pp. 70–93). Baltimore: Johns Hopkins Press.

Benzer, S. (1959). On the topology of the genetic fine structure. *Proceedings of the National Academy of Science USA*, *45*, 1607–1620.

Benzer, S. (1961). On the topography of the genetic fine structure. *Proceedings of the National Academy of Science USA*, *47*, 403–415.

Berg, D.E., & Howe, M.M. (Eds.). (1989). *Mobile DNA*. Washington, DC: American Society of Microbiology.

Berget, S.M., et al. (1978). Spliced segments at the 5′ termini of adenovirus-2 late mRNA: A role for heterogenous nuclear RNA in mammalian cells. *Cold Spring Harbor Symposia on Quantitative Biology*, *42*, 523–529.

Berk, A.J., & Sharp, P.A. (1978a). Splicing early mRNAs of SV40. *Proceedings of the National Academy of Science USA*, *75*, 1274–1278.

Bi'emont, L., & Cizeron, G. (1999). Distribution of transposable elements in Drosophila species. *Genetica*, *105*, 43–62.

Bonner, D.M. (1950). The Q locus of Neurospora. *Genetics*, *35*, 655-656.

Boveri, T. (1902). Über mehrpolige Mitosen als Mittel zur Analyse des Zellkerns.*Verhandlungen physiologischen und medizinischen Gesellschafts Würzburg, N.F.*, *35*, 60–90.

Boveri, T. (1903). Über die Konstitution der chromatischen Kernsubstanz. *Verhandlungen deutschen zoologischen Gesellschafts*, *13*, 10–33.

Boveri, T. (1904). *Ergebnisse über die konstitution der chromatischen substanz des zellkerns* (pp. 1–22). Jena: Gustav Fischer.

Brandon, R.N. (1990). *Adaptation and environment.* Princeton: Princeton University Press.

Breathnach, R., et al. (1977). Ovalbumin gene is split in chicken DNA. *Nature*, *270*, 314–319.

Brenner, S., Jacob, F., & Meselsohn, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, *190*, 576–580.

Bridges, C.B. (1916). Non-disjunction as proof of the chromosome theory of heredity. *Genetics*, *1*, 1–52, 107–163.

Bridges, C.B. (1935). Salivary chromosome maps with a key to the banding of the chromosomes of Drosophila melanogaster. *Journal of Heredity*, *26*, 60–64.

Bridges, C.B. (1937). Correspondence between linkage maps and salivary chromosome structure as illustrated in the tip of chromosome 2R of Drosophila melanogaster. *Cytological Fuiji Jubileum*, 745–755.

Bridges, C.B. (1938). A revised map of the salivary gland X-chromosome of Drosophila melanogaster. *Journal of Heredity*, *29*, 11–13.

Britten, R.J., & Kohne, D.E. (1967). Nucleotide sequence repetition in DNA. *Carnegie Institute Washington Yearbook*, *65*, 78–106.

Britten, R.J., & Kohne, D.E. (1968a). Repeated sequences in DNA. *Science*, *161*, 529–540.

Britten, R.J., & Kohne, D.E. (1968b). Repeated nucleotide sequences. *Carnegie Institute Washington Yearbook*, *66*, 73–88.

Broker, T.R., et al. (1978). Adenovirus-2 messenger – An example of baroque molecular architecture. *Cold Spring Harbor Symposia on Quantitative Biology*, *42*, 531–553.

Brown, D.D., et al. (1971). Purification and some characteristics of 5S DNA from Xenopus laevis. *Proceedings of the National Academy of Science USA*, *68*, 3175–3179.

Burian, R.M. (1985). On conceptual change in biology: The case of the gene. In: D.J. Deprew & B.H. Weber (Eds.), *Evolution at a crossroads* (pp. 21–42). Cambridge, MA, London, England: Bradford Books.

Burnet, F.M. (1959). *The clonal selection theory of acquired immunity.* London.

Canaani, D., et al. (1979). Sequence heterogeneity at the 5′ termini of late SV40 19S and 16S mRNAs. *Proceedings of the National Academy of Science USA*, *76*, 3078–3082.

Chen, C., et al. (1987). At least two genes reside within a large intron of the dunce gene of Drosophila. *Nature*, *329*, 721–724.

Chervitz, S.A., et al. (1998). Comparison of the complete protein sets of worm and yeast: Orthology and divergence. *Science*, *282*, 2022–2028.

Chosh, P.K., et al. (1978a). The 5′ terminal leader sequence of the late 16s mRNA from cells infected with SV40. *Journal of Biological Chemistry*, *253*, 3643–3647.

Chosh, P.K., et al. (1978b). Heterogeneity and 5′ terminal structures of the late RNAs of SV40. *Journal of Molecular Biology*, *126*, 813–846.

Cohen, S.S. (1948). Synthesis of bacterial viruses: Synthesis of nucleic acid and protein in Esherichia coli B infected, with $T2r^{+}$ bacteriophage. *Journal of Biological Chemistry*, *174*, 281–293.

Cohen, S., et al. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Science USA*, *70*, 3240–3244.

Cohen, S.N., & Shapiro, J.A. (1980). Transposable genetic elements. *Scientific American*, *242*, 36–45.

Contreras, R., et al. (1977). Overlapping of the CP2-VP3 gene and the VP1 gene in the SV40 genome. *Cell*, *12*, 529–538.

Correns, C. (1900). G. Mendel's Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. *Berichte der Deutschen Botanischen Gesellschaft*, *18*, 158–168.

Crabtree, G.R., & Kant, J.A. (1983). Organization of the rat gamma-fibrinogen gene: Alternative mRNA splice patterns produce the A and gamma B (gamma') chains of fibrinogen. *Cell*, *31*, 159–166.

Crawford, I.P., & Yanofsky, C. (1958). On the separation of the tryptophan synthetase of Escherichia coli into two protein components. *Proceedings of the National Academy of Science USA*, *44*, 1161–1170.

Crawford, L.V., et al. (1978). Organization and expression of early genes of SV40. *Proceedings of the National Academy of Science USA*, *75*, 1274–1278.

Danna, K., & Nathans, D. (1971). Specific cleavage of simial vitus 40 DNA by restriction endonuclease of Hemophilus influenzae. *Proceedings of the National Academy of Science USA*, *68*, 2913–2917.

Davern, C.I., & Meselson, M. (1960). The molecular conservation of ribonucleic acid during bacterial growth. *Journal of Molecular Biology*, *2*, 153–160.

Dawson, M.H., & Sia, R.H.P. (1931). In vitro transformation of pneumococcal types. I. A technique for inducing transformation of pneumococcal types in vitro. *Journal of Experimental Medicine*, *54*, 681–700.

de Vries, H. (1900). Das Spaltungsgesetz der Bastarde. *Berichte der Deutchen Botanischen Gesellschaft*, *18*, 83–90.

Demerec, M. (1967). Properties of genes. In: R.A. Brink & E.D. Styles (Eds.), *Heritage from Mendel* (pp. 49–61). Milwaukee: University of Wisconsin Press.

Dennis, C., Gallagher, R., & Campbell, P. (2001). Everyone's genome. *Nature*, *409*, 813.

Dillon, L.S. (1987). *The gene: Its structure, function, and evolution* (p. 12). New York: Plenum Press.

Dobzhansky, T. (1929). Genetical and cytological proof of translocations involving the third and fourth chromosome in Drosophila melanogaster. *Biologisches Zentralblatt*, *49*, 408–419.

Douglass, J., et al. (1984). Polyprotein gene expression: Generation of diversity of neuroendocrine peptides. *Annual Review of Biochemistry*, *53*, 665–715.

Dounce, A.L. (1952). Duplicating mechanism for peptide chain and nucleic acid synthesis. *Enzymologia*, *15*, 503–507.

Dunham, I., et al. (1999). The DNA sequence of human chromosome 22. *Nature*, *402*, 489–495.

Early, P., et al. (1980). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell*, *20*, 313–319.

Early, P.W., et al. (1979). Immunoglobulin heavy chain gene organization in mice: Analysis of a myeloma genomic clone containing variable and α constant regions. *Proceedings of the National Academy of Science USA*, *76*, 857–861.

Eveleth, D.D., & Marsh, J.L. (1987). Overlapping transcription units in Drosophila: Sequence and structure of the Cs gene. *Molecular and General Genetics*, *209*, 290–298.

Fedoroff, N.V. (1984). Transposable genetic elements in maize. *Scientific American*, *250*, 64–74.

Fiers, W., et al. (1978). Complete nucleotide sequence of SV40 DNA. *Nature*, *273*, 113–210.

Flavell, A.J., et al. (1979). Multiple 5′ terminal cap structures in late polyoma virus RNA. *Cell*, *16*, 357–373.

Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. *Nature*, *173*, 318.

Giles, N.H. (1952). Studies on the mechanism of reversion in biochemical mutants of Neurospora crassa. *Cold Spring Harbor Symposia on Quantitative Biology*, *16*, 283–313.

Gillies, S.D., (1983). A tissue-specific transcription enhancer elements is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, *33*, 717–728.

Goffeau, A., et al. (1996). Life with 6000 genes. *Science*, *274*, 546–567.

Green, M.M., & Green, K.C. (1949). Crossing over between alleles of the lozenge locus in Drosophila melanogaster. *Proceedings of National Academy of Science USA*, *35*, 586–591.

Griffith, F. (1928). Significance of pneumococcal types. *Journal of Hygieny Cambridge*, *27*, 113–159.

Gustafsson, Å. (1969). The life of Gregor Johan Mendel – tragic or not? *Hereditas*, *62*, 239–258.

Hattori, M., et al. (2000). The DNA sequence of human chromosome 21. *Nature*, *405*, 311–319.

Henikoff, S., et al. (1986). Gene within a gene: Nested Drosophila genes encode unrelated proteins on opposite DNA strands. *Cell*, *44*, 33–42.

Hershey, A.D. (1953). Nucleic acid economy in bacteria infected with bacteriophage T2. II. Phage precursor nucleic acid. *Journal of General Physiology*, *37*, 1–23.

Hershey, A.D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology*, *36*, 39–56.

Hershey, A.D., et al. (1953). Nucleic acid economy in bacteria infected with bacteriophage T2. *Journal of General Physiology*, *36*, 777–789.

Horowitz, M., et al. (1978). Polyoma infected cells contain at least three spliced late RNAs. *Nucleic Acids Research*, *5*, 4663–4676.

Hozumi, N., & Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Science USA*, *73*, 3628–3632.

International Human Genome Sequencing Consortion. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*, 860–921.

Jackson, D., et al. (1972). Biochemical method for inserting new genetic information into DNA of simian virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proceedings of the National Academy of Science USA*, *69*, 2904–2909.

Jacob, F., & Monod, J. (1961). Molecular and biological characterization of messenger RNA. *Journal of Molecular Biology*, *3*, 318–356.

Jacq, C., et al. (1977). A pseudogene structure in 5S DNA of Xenopus laevis. *Cell*, *12*, 109–120.

Jahn, I. (1998). *Gesichte der biologie* (p. 645). Jena, Stuttgart, Lübeck, Ulm: Gustav Fisher.

Jeffreys, A.J., & Flavell, R.A. (1977). The rabbit β-globin gene contains a large insert in the coding sequence. *Cell*, *12*, 1097–1108.

Jerne, N.K. (1955). The natural selection theory of antibody formation. *Proceedings of the Academy of Sciences, Paris*, *41*, 849–857.

Joh, K., et al. (1986). Expression of three mRNA species from a single rat aldolase A gene, differing in their 5′ non-coding regions. *Journal of Molecular Biology*, *190*, 401–410.

Johannsen, W. (1909). *Elemente der exakten erblichkeitslehre* (pp. 143-144). Jena: Gustav Fischer.

Kedes, L.H. (1979). Histone genes and histone messengers. *Annual Reviews of Biochemistry*, *48*, 837–870.

Kelly, T.J., Jr., & Smith, H.D. (1970). A restriction enzyme from Hemophilus influenzaea. II. Base sequence of the recognition site. *Journal of Molecular Biology*, *51*, 393–409.

Kitcher, P. (1992). Gene: Current usages. In: E.F. Keller & E.A. Lloyd (Eds.), *Keywords in evolutionary biology* (pp. 128–131). Cambridge: Harvard University Press.

Klessig, D.F. (1977). Two adenovirus mRNAs have a common 5′ terminal leader sequence encoded of least 10 kb upstream from their main coding regions. *Cell*, *12*, 9–21.

Konkel, D., et al. (1978). The sequence of the chromosomal mouse β-globulin major gene: Homologies in capping, splicing and poly (A) sites. *Cell*, *15*, 1125–1132.

Lai, C.-J., et al. (1978). Mapping the spliced and unspliced late lytic SV40 mRNAs. *Cell*, *14*, 971–982.

Leff, S.E., et al. (1986). Complex transcriptional units: Diversity in gene expression by alternative RNA processing. *Annual Review of Biochemistry*, *55*, 1091–1117.

Levinson, B., et al. (1990). A transcribed gene in an intron of the human factor VIII gene. *Genomics*, *7*, 1–11.

Lewin, B. (1980a). *Gene expression, Vol. 2. Eukaryotic chromosomes* (2nd ed.). New York: Wiley.

Lewin, B. (1980b). Alternatives for splicing: Recognizing the ends of introns. *Cell*, *22*, 324–326.

Lewis, E.B. (1941). Another case of unequal crossing over in Drosophila melanogaster. *Proceedings of National Academy of Science USA*, *27*, 31–34.

Lewontin, R.C. (1992). Genotype and phenotype. In: E.F. Keller & E.A. Lloyd (Eds.), *Keywords in evolutionary biology* (pp. 137–144). Cambridge: Harvard University Press.

Linn, S., & Arber, W. (1968). Host specificity of DNA produced by Escherichia coli. X. In vitro restriction of phage fd replicative form. *Proceedings of the National Academy of Science USA*, *59*, 1300–1306.

Lobban, P., & Kaiser, A.D. (1973). Enzymatic end-to-end joining of DNA molecules. *Journal of Molecular Biology*, *78*, 453–471.

Lomedico, P., et al. (1979). The structure and evolution of the two nonallelic rat preproinsulin genes. *Cell*, *18*, 545–558.

Lyons, K.M., et al. (1988). Many protein products from a few loci: Assigment of human salivary proline-rich proteins to specific loci. *Genetics*, *120*, 255–265.

Mahner, M., & Kary, M. (1997). What exactly are genomes, genotypes and phenotypes? And what about phenomes? *Journal of Theoretical Biology*, *186*, 55–63.

Maire, P., et al. (1987). Characterization of 3 optional promoters in the 5′ region of the human aldolase A gene. *Journal of Molecular Biology*, *197*, 425–437.

Manson, L.A. (1953). The metabolism of ribonucleic acid in normal and bacteriophage infected *E. coli. Journal of Bacteriology*, *66*, 703–711.

Maxam, A.M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Science USA*, *74*, 560–564.

McClintock, B. (1947). Cytogenetic studies of maize and Neurospora. *Carnegie Institute Washington Yearbook*, *46*, 146–152.

McClintock, B. (1948). Mutable loci in maize. *Carnegie Institute Washington Yearbook*, *47*, 155–169.

McGrogan, M., & Raskas, J.H. (1978). Two regions of the adenoviru 2 genome specify families of late polysomal RNAs containing common sequences. *Proceedings of the National Academy of Science USA*, *75*, 625–629.

Medford, R.M., et al. (1984). A novel mechanism of alternative RNA splicing for the developmentally regulated generation of troponin T iso-forms from a single gene. *Cell*, *38*, 409–421.

Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verhandlungen naturforschenden Vereins Brünn*, *4*, 3–44.

Merz, J.E., & Davis, R.W. (1972). Cleavage of DNA by RI restriction endonuclease generates cohesive ends. *Proceedings of the National Academy of Science USA*, *69*, 3370–3374.

Meselson, M., & Yuan, R. (1968). DNA restriction enzyme from E. coli. *Nature*, *217*, 1110–1114.

Miller, O.L., Jr., & Beatty, B.R. (1969a). Visualization of nucleolar genes. *Science*, *164*, 955–957.

Miller, O.L., Jr., & Beatty, B.R. (1969b). Extrachromosomal nucleolar genes in amphibian oocytes. *Genetics*, *61*, 133–143.

Miller, O.L., Jr., & Beatty, B.R. (1969c). Portrait of a gene. *Journal of Cell Physiology*, *74* (Suppl. 1), 225–232.

Moore, K.W., et al. (1981). Expression of IgD may use both DNA rearrangement and RNA splicing mechanism. *Proceedings of the National Academy of Science USA*, *78*, 1800–1804.

Moreau, P., et al. (1981). The SV40 72 bp repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Research*, *9*, 6047–6069.

Morgan, T.H. (1919). *The physical basis on heredity.* New Haven: Yale University Press.

Morgan, T.H., et al. (1915). *The mechanism of Mendelian heredity.* New York: Henry Holt.

Muller, H.J. (1927). Artificial transmutation of the gene. *Science*, *66*, 84–87.

Muller, H.J. (1928). The problem of genic modification. *Zeitschrift Vererbunslehre* Suppl. 1, 234–260.

Muller, H.J., & Painter, T.S. (1929). The cytological expression of changes in gene alignment produced by X-rays in Drosophila. *American Naturalist*, *63*, 193–200.

Nevins, J.R., & Darnell, J.E. (1978). Steps in the processing of Ad2 mRNA: Poly $(A)^+$ nuclear sequences are conserved and poly (A) additions precedes splicing. *Cell*, *15*, 1477–1493.

Niessl, G. von. (1903). Sitzungsberichte Jahresversammlung am 8. Jänner 1902. *Verhandlungen des naturforschenden Vereines in Brünn 1902*, *42*, 20.

Niessl, G. von. (1906). Sitzungsberichte Jahresversammlung am 11. Jänner 1905. *Verhandlungen des naturforschenden Vereines in Brünn 1905*, *44*, 8.

Oliver, C.P. (1940). A reversion to wild type associated with crossing over in Drosophila melanogaster. *Proceedings of National Academy of Science USA*, *26*, 452–454.

Painter, T.S. (1934). A new method for the study of chromosome aberrations and the plotting of chromosome maps in Drosophila melanogaster. *Genetics*, *19*, 175–188.

Paucha, E., et al. (1978). Large and small tumor antigens from SV40 have identical amino termini mapping at 0.65 map units. *Proceedings of the National Academy of Science USA*, *75*, 2165–2169.

Pierangeli, A., et al. (1998). Identification of an alternative open reading frame (''*Hidden Gene*''?) stringently required for infectivity of poliovirus cDNA clones. *Microbiologica*, *21*, 309–320.

Pontecorvo, G. (1952). The genetic formulation of gene structure and action. *Advances in Enzymology*, *13*, 121–149.

Pritchard, R.H. (1955). The linear arrangement of a series of alleles of Aspergillus nidulans. *Heredity*, *9*, 343–371.

Rieger, R., et al. (1991). *Glossary of genetics: Classical and molecular* (5th ed.). Berlin and Heidelberg: Springer-Verlag.

Riley, M., et al. (1960). On the expression of a structural gene. *Journal of Molecular Biology*, *2*, 216–225.

Roper, J.A. (1950). Search for linkage between genes determining a vitamin requirement. *Nature*, *166*, 956.

Sakano, H., et al. (1979a). Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA fragments. *Nature*, *277*, 627–632.

Sanger, F., et al. (1977). Nucleotide sequence of bacteriophage ΦX 174 DNA. *Nature*, *265*, 687–695.

Sanger, F., et al. (1977). DNA sequencing with chain termination inhibitors. *Proceedings of the National Academy of Science USA*, *74*, 5463–5467.

Sarabhai, A.S., et al. (1964). Co-linearity of the gene with the polypeptide chain. *Nature*, *201*, 13–17.

Scheller, R.H., et al. (1984). Neuropeptides: Mediators of behavior in Aplysia. *Science*, *225*, 1300–1308.

Schibler, U., & Sierra, F. (1987). Alternative promoters in developmental gene expression. *Annual Review of Genetics*, *21*, 237–257.

Schmalhausen, I.I. (1949). *Factors of evolution: The theory of stabilizing selection.* Philadelphia: Blakiston.

Schweighaffer, F., et al. (1986). In vivo developmental modifications of the expression of genes encoding muscle-specific enzymes in rat. *Journal of Biological Chemistry*, *261*, 10271–10276.

Schöler, H.J., et al. (1988). Enhancers and trans-acting factors. In: G. Kahl (Ed.), *Architecture of eukaryotic genes* (pp. 89–129). Winheim: VCH Verlagsgesellschaft.

Serfling, E., et al. (1985). Enhancers and eukaryotic gene transcription. *Trends in Genetics*, *1*, 224–230.

Sharp, P.A., et al. (1973). Detection of two restriction endonuclease activities in Hemophilus parainfluenzae using analytical agarose-ethidium bromide electrophoresis. *Biochemistry*, *12*, 3055–3062.

Shaw, D.C., et al. (1978). Gene K, a new overlapping gene in bacteriophage G4; *Nature*, *272*, 510–514.

Sierra, F., Pittet, A.-C., & Schibler, U. (1986). Different tissue-specific expression of the alpha-amylase gene Amy-1 in mouse and rat. *Molecular and Cellular Biology*, *6*, 4067–4076.

Singer, M., & Berg, P. (1991). *Genes and genomes: A changing perspective*. Mill Valley: University Science Books.

Smith, H.O., & Wilcox, K.W. (1970). A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *Journal of Molecular Biology*, *51*, 393–409.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, *98*, 503–517.

Spencer, C.A., et al. (1986). Overlapping transcription units in the dopa decarboxylase region of Drosophila. *Nature*, *322*, 275–279.

Srb, A.M., & Horowitz, N.H. (1944). The ornithine cycle in Neurospora and its genetic control. *Journal of Biological Chemistry*, *154*, 129–139.

Stent, G.S. (1971). *Molecular genetics: An introductory narrative.* San Francisco: Freeman and Company.

Sturtevant, A.H. (1913). The linear arrangement of the six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology*, *14*, 43–59.

Sutton, W.S. (1903). The chromosomes in heredity. *Biological Bulletin*, *4*, 231–251.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thalian. Nature*, *408*, 796–815.

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, *282*, 2012–2018.

Tilghman, S.M., et al. (1978). Intervening sequence of DNA identified in the structural portion of mouse β-globin gene. *Proceedings of the National Academy of Science USA*, *75*, 725–729.

Tonegawa, S., et al. (1978). Organization of immunoglobulin genes. *Cold Spring Harbor Symposia on Quantitative Biology*, *42*, 921–931.

Tschermak, E. (1900). Ueber Künstliche Kreuzung bei Pisum sativum. *Berichte der Deutchen Botanischen Gesellschaft*, *18*, 232–239.

Tupler, R., et al. (2001). Expressing the human genome. *Nature*, *409*, 832–833.

Venter, J.G., et al. (2001). The sequence of the human genome. *Science*, *291*, 1304–1351.

Volkin, E. (1960). The function of RNA in T2-infected bacteria. *Proceedings of the National Academy of Science USA*, *46*, 1336–1348.
Volkin, E., & Astrachan, L. (1956a). Phosphorus incorporation in Escherichia coli ribonucleic acid after infection with bacteriophage T2. *Virology*, *2*, 149–161.
Volkin, E., & Astrachan, L. (1956b). Intracellular distribution of labeled ribonucleic acid after infection of Escherichia coli with bacteriophage T7. *Virology*, *2*, 433–437.
Volkin, E., et al. (1958). Metabolism of RNA phophorus in Escherichia coli infected with bacteriophage T7. *Virology*, *6*, 545–555.
Waring, M., & Britten, R.J. (1966). Nucleotide sequence repetition: A rapidly reassociating fraction of mouse DNA. *Science*, *154*, 791–794.
Waters, C.K. (1994). Genes made molecular. *Philosophy of Science*, *61*, 163–185.
Watson, J.D. (1963). Involvement of RNA in the synthesis of proteins. *Science*, *140*, 17–26.
Watson, J.D., & Crick, F.H.C. (1953). A structure for deoxyribonucleic acid. *Nature*, *171*, 737–738.
Weinberg, E.S., et al. (1972). Genes coding for polysomal 9S RNA of sea urchins: Conservation and divergence. *Nature*, *240*, 225–228.
Weissmann, C., et al. (1990). Sometimes an editor makes sense. *Nature*, *343*, 697–699.
Westphal, H., & Lai, S.-P. (1978). Displacement loops in adenovirus DNA-RNA hybrids. *Cold Spring Harbor Symposia on Quantitative Biology*, *42*, 555–558.
Williams, T., & Fried, M. (1986). A mouse locus at which transcription from both DNA strand produces mRNAs complementary at their 3′ ends. *Nature*, *322*, 275–279.
Wilson, M.C., & Darnell, J.E. (1981). Control of messenger RNA concentration by different cytoplasmic half-life. Adenovirus messenger RNAs from transcription units 1A and 1B. *Journal of Molecular Biology*, *148*, 231–251.
Woltereck, R. (1909). Weitere experimentelle Untersuchungen über Artveränderung, speziell über das Wesen quantitativer, Unterschiede der Daphniden. *Verhandlungen der Deutschen Zoologischen Gesellschaft*, *19*, 110.
Wozney, J., et al. (1981a). Fine structural analysis of the chicken pro-α2 collagen gene. *Proceedings of the National Academy of Science USA*, *78*, 712–716.
Wozney, J., et al. (1981b). Structure of the pro-α2 (I) collagen gene. *Nature*, *294*, 129–135.
Xu, G., et al. (1990). The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell*, *62*, 599–608.
Yamada, Y., et al. (1980). The collagen gene: Evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp. *Cell*, 22, 887–892.
Yanofsky, C., & Crawford, I.P. (1959). The effects of deletions, point mutations on the two components of the tryptophan synthetase of Escherichia coli. *Proceedings of the National Academy of Science USA*, *45*, 1016–1026.
Yanofsky, C., et al. (1964). On the colinearity of gene structure and protein structure. *Proceedings of the National Academy of Science USA*, *51*, 266–272.
Yanofsky, C., et al. (1967). The complete amino acid sequence of the tryptophan synthetase A protein (α subunit) and its colinear relationship with the genetic map of the A gene. *Proceedings of the National Academy of Science USA*, *57*, 296–298.
Young, R.A., et al. (1981). A single mouse alpha-amylase gene specifies two different tissue-specific mRNAs. *Cell*, *23*, 451–458.
Ziff, E., & Fraser, N.W. (1978). Adenovirus type 2 late mRNAs: Structural evidence for 3′ conterminal species. *Journal of Virology*, *25*, 897–906.