Gene

In biology, the word **gene** has two meanings. The Mendelian gene is a basic unit of heredity. The molecular gene is a sequence of nucleotides in DNA that is transcribed to produce a functional RNA. There are two types of molecular genes: protein-coding genes and non-coding genes.[1][2][3] During gene expression (the synthesis of RNA or protein from a gene), DNA is first copied into RNA. RNA can be directly functional or be the intermediate template for the synthesis of a protein.

The transmission of genes to an organism's offspring, is the basis of the inheritance of phenotypic traits from one generation to the next. These genes make up different DNA sequences, together called a genotype, that is specific to every given individual, within the gene pool of the population of a given species. The genotype, along with environmental and developmental factors, ultimately determines the phenotype of the individual.

A chromosome and its packaged long strand of DNA unraveled. The DNA's base pairs encode genes, which provide functions. A human DNA can have up to 500 million base pairs with thousands of genes.

Most biological traits occur under the combined influence of polygenes (a set of different genes) and gene–environment interactions. Some genetic traits are instantly visible, such as eye color or the number of limbs, others are not, such as blood type, the risk for specific diseases, or the thousands of basic biochemical processes that constitute life. A gene can acquire mutations in its sequence, leading to different variants, known as alleles, in the population. These alleles encode slightly different versions of a gene, which may cause different phenotypical traits.[4] Genes evolve due to natural selection or survival of the fittest and genetic drift of the alleles.

Definitions

There are many different ways to use the term "gene" based on different aspects of their inheritance, selection, biological function, or molecular structure but most of these definitions fall into two categories, the Mendelian gene or the molecular gene.[1][5][6][7][8]

The Mendelian gene is the classical gene of genetics and it refers to any heritable trait. This is the gene described in *The Selfish Gene*.[9] More thorough discussions of this version of a gene can be found in the articles *Genetics* and *Gene-centered view of evolution*.

The molecular gene definition is more commonly used across biochemistry, molecular biology, and most of genetics—the gene that is described in terms of DNA sequence.[1] There are many different definitions of this gene—some of which are misleading or incorrect.[5][10]
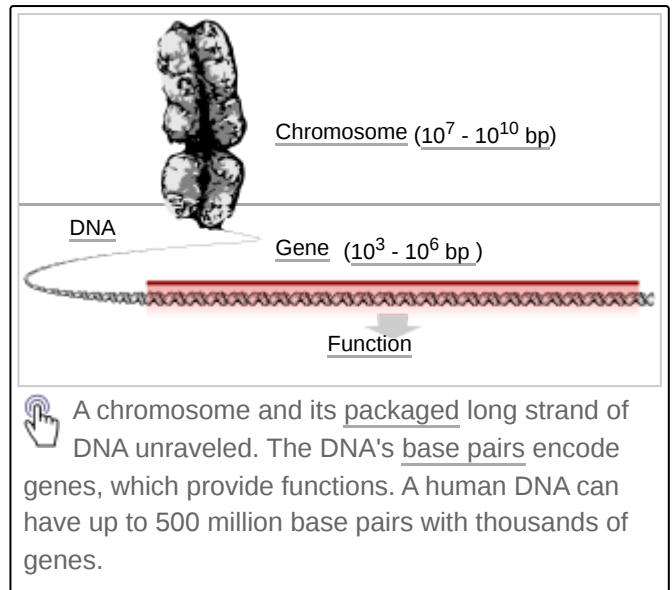
Very early work in the field that became molecular genetics suggested the concept that one gene makes one protein (originally 'one gene – one enzyme').[11][12] However, genes that produce repressor RNAs were proposed in the 1950s[13] and by the 1960s, textbooks were using molecular gene definitions that included those that specified functional RNA molecules such as ribosomal RNA and tRNA (noncoding genes) as well as protein-coding genes.[14]

This idea of two kinds of genes is still part of the definition of a gene in most textbooks. For example,

> The primary function of the genome is to produce RNA molecules. Selected portions of the DNA nucleotide sequence are copied into a corresponding RNA nucleotide sequence, which either encodes a protein (if it is an mRNA) or forms a 'structural' RNA, such as a transfer RNA (tRNA) or ribosomal RNA (rRNA) molecule. Each region of the DNA helix that produces a functional RNA molecule constitutes a gene.[15]

> We define a gene as a DNA sequence that is transcribed. This definition includes genes that do not encode proteins (not all transcripts are messenger RNA). The definition normally excludes regions of the genome that control transcription but are not themselves transcribed. We will encounter some exceptions to our definition of a gene - surprisingly, there is no definition that is entirely satisfactory.[16]

> A gene is a DNA sequence that codes for a diffusible product. This product may be protein (as is the case in the majority of genes) or may be RNA (as is the case of genes that code for tRNA and rRNA). The crucial feature is that the product diffuses away from its site of synthesis to act elsewhere.[17]

The important parts of such definitions are: (1) that a gene corresponds to a transcription unit; (2) that genes produce both mRNA and noncoding RNAs; and (3) regulatory sequences control gene expression but are not part of the gene itself. However, there is one other important part of the definition and it is emphasized in Kostas Kampourakis' book *Making Sense of Genes*.

> Therefore in this book I will consider genes as DNA sequences encoding information for functional products, be it proteins or RNA molecules. With 'encoding information', I mean that the DNA sequence is used as a template for the production of an RNA molecule or a protein that performs some function.[5]

The emphasis on function is essential because there are stretches of DNA that produce non-functional transcripts and they do not qualify as genes. These include obvious examples such as transcribed pseudogenes as well as less obvious examples such as junk RNA produced as noise due to transcription errors. In order to qualify as a true gene, by this definition, one has to prove that the transcript has a biological function.[5]

Early speculations on the size of a typical gene were based on high-resolution genetic mapping and on the size of proteins and RNA molecules. A length of 1500 base pairs seemed reasonable at the time (1965).[14] This was based on the idea that the gene was the DNA that was directly responsible for

production of the functional product. The discovery of introns in the 1970s meant that many eukaryotic genes were much larger than the size of the functional product would imply. Typical mammalian protein-coding genes, for example, are about 62,000 base pairs in length (transcribed region) and since there are about 20,000 of them they occupy about 35–40% of the mammalian genome (including the human genome).[18][19][20]

In spite of the fact that both protein-coding genes and noncoding genes have been known for more than 50 years, there are still a number of textbooks, websites, and scientific publications that define a gene as a DNA sequence that specifies a protein. In other words, the definition is restricted to protein-coding genes. Here is an example from a recent article in American Scientist.

> ... to truly assess the potential significance of de novo genes, we relied on a strict definition of the word "gene" with which nearly every expert can agree. First, in order for a nucleotide sequence to be considered a true gene, an open reading frame (ORF) must be present. The ORF can be thought of as the "gene itself"; it begins with a starting mark common for every gene and ends with one of three possible finish line signals. One of the key enzymes in this process, the RNA polymerase, zips along the strand of DNA like a train on a monorail, transcribing it into its messenger RNA form. This point brings us to our second important criterion: A true gene is one that is both transcribed and translated. That is, a true gene is first used as a template to make transient messenger RNA, which is then translated into a protein.[21]

This restricted definition is so common that it has spawned many recent articles that criticize this "standard definition" and call for a new expanded definition that includes noncoding genes. However, some modern writers still do not acknowledge noncoding genes although this so-called "new" definition has been recognised for more than half a century.[22][23][24]

Although some definitions can be more broadly applicable than others, the fundamental complexity of biology means that no definition of a gene can capture all aspects perfectly. Not all genomes are DNA (e.g. RNA viruses),[25] bacterial operons are multiple protein-coding regions transcribed into single large mRNAs, alternative splicing enables a single genomic region to encode multiple district products and trans-splicing concatenates mRNAs from shorter coding sequence across the genome.[26][27][28] Since molecular definitions exclude elements such as introns, promotors, and other regulatory regions, these are instead thought of as "associated" with the gene and affect its function.

An even broader operational definition is sometimes used to encompass the complexity of these diverse phenomena, where a gene is defined as a union of genomic sequences encoding a coherent set of potentially overlapping functional products.[29] This definition categorizes genes by their functional products (proteins or RNA) rather than their specific DNA loci, with regulatory elements classified as *gene-associated* regions.[29]

History

Discovery of discrete inherited units

The existence of discrete inheritable units was first suggested by Gregor Mendel (1822–1884).[30] From 1857 to 1864, in Brno, Austrian Empire (today's Czech Republic), he studied inheritance patterns in 8000 common edible pea plants, tracking distinct traits from parent to offspring. He described these mathematically as 2^n combinations where n is the number of differing characteristics in the original peas. Although he did not use the term *gene*, he explained his results in terms of discrete inherited units that give rise to observable physical characteristics. This description prefigured Wilhelm Johannsen's distinction between genotype (the genetic material of an organism) and phenotype (the observable traits of that organism). Mendel was also the first to demonstrate independent assortment, the distinction between dominant and recessive traits, the distinction between a heterozygote and homozygote, and the phenomenon of discontinuous inheritance.

Gregor Mendel

Prior to Mendel's work, the dominant theory of heredity was one of blending inheritance,[31] which suggested that each parent contributed fluids to the fertilization process and that the traits of the parents blended and mixed to produce the offspring. Charles Darwin developed a theory of inheritance he termed pangenesis, from Greek pan ("all, whole") and genesis ("birth") / genos ("origin").[32][33] Darwin used the term *gemmule* to describe hypothetical particles that would mix during reproduction.

Mendel's work went largely unnoticed after its first publication in 1866, but was rediscovered in the late 19th century by Hugo de Vries, Carl Correns, and Erich von Tschermak, who (claimed to have) reached similar conclusions in their own research.[34] Specifically, in 1889, Hugo de Vries published his book *Intracellular Pangenesis*,[35] in which he postulated that different characters have individual hereditary carriers and that inheritance of specific traits in organisms comes in particles. De Vries called these units "pangenes" (*Pangens* in German), after Darwin's 1868 pangenesis theory.

Twenty years later, in 1909, Wilhelm Johannsen introduced the term "gene" (inspired by the ancient Greek: γόνος, *gonos*, meaning offspring and procreation)[36] and, in 1906, William Bateson, that of "genetics"[37][29] while Eduard Strasburger, among others, still used the term "pangene" for the fundamental physical and functional unit of heredity.[35]:Translator's preface,viii

Discovery of DNA

Advances in understanding genes and inheritance continued throughout the 20th century. Deoxyribonucleic acid (DNA) was shown to be the molecular repository of genetic information by experiments in the 1940s to 1950s.[38][39] The structure of DNA was studied by Rosalind Franklin and Maurice Wilkins using X-ray crystallography, which led James D. Watson and Francis Crick to publish a model of the double-stranded DNA molecule whose paired nucleotide bases indicated a compelling hypothesis for the mechanism of genetic replication.[40][41]

In the early 1950s the prevailing view was that the genes in a chromosome acted like discrete entities arranged like beads on a string. The experiments of Benzer using mutants defective in the rII region of bacteriophage T4 (1955–1959) showed that individual genes have a simple linear structure and are likely to be equivalent to a linear section of DNA.[42][43]

Collectively, this body of research established the central dogma of molecular biology, which states that proteins are translated from RNA, which is transcribed from DNA. This dogma has since been shown to have exceptions, such as reverse transcription in retroviruses. The modern study of genetics at the level of DNA is known as molecular genetics.

In 1972, Walter Fiers and his team were the first to determine the sequence of a gene: that of bacteriophage MS2 coat protein.[44] The subsequent development of chain-termination DNA sequencing in 1977 by Frederick Sanger improved the efficiency of sequencing and turned it into a routine laboratory tool.[45] An automated version of the Sanger method was used in early phases of the Human Genome Project.[46]

Modern synthesis and its successors

The theories developed in the early 20th century to integrate Mendelian genetics with Darwinian evolution are called the modern synthesis, a term introduced by Julian Huxley.[47]

This view of evolution was emphasized by George C. Williams' gene-centric view of evolution. He proposed that the Mendelian gene is a unit of natural selection with the definition: "that which segregates and recombines with appreciable frequency."[48]:24 Related ideas emphasizing the centrality of Mendelian genes and the importance of natural selection in evolution were popularized by Richard Dawkins.[9][49]
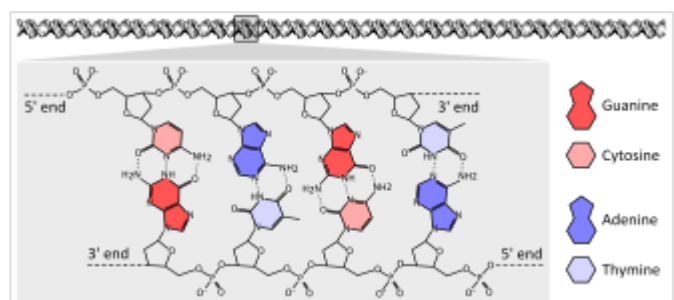
The development of the neutral theory of evolution in the late 1960s led to the recognition that random genetic drift is a major player in evolution and that neutral theory should be the null hypothesis of molecular evolution.[50] This led to the construction of phylogenetic trees and the development of the molecular clock, which is the basis of all dating techniques using DNA sequences. These techniques are not confined to molecular gene sequences but can be used on all DNA segments in the genome.

Molecular basis

DNA

The vast majority of organisms encode their genes in long strands of DNA (deoxyribonucleic acid). DNA consists of a chain made from four types of nucleotide subunits, each composed of: a five-carbon sugar (2-deoxyribose), a phosphate group, and one of the four bases adenine, cytosine, guanine, and thymine.[51]:2.1

The chemical structure of a four base pair fragment of a DNA double helix. The sugar-phosphate backbone chains run in opposite directions with the bases pointing inward, base-pairing A to T and C to G with hydrogen bonds.

Two chains of DNA twist around each other to form a DNA double helix with the phosphate–sugar backbone spiralling around the outside, and the bases pointing inward with adenine base pairing to thymine and guanine to cytosine. The specificity of base pairing occurs because adenine and thymine align to form two hydrogen bonds, whereas cytosine

and guanine form three hydrogen bonds. The two strands in a double helix must, therefore, be complementary, with their sequence of bases matching such that the adenines of one strand are paired with the thymines of the other strand, and so on.[51]:4.1

Due to the chemical composition of the pentose residues of the bases, DNA strands have directionality. One end of a DNA polymer contains an exposed hydroxyl group on the deoxyribose; this is known as the 3' end of the molecule. The other end contains an exposed phosphate group; this is the 5' end. The two strands of a double-helix run in opposite directions. Nucleic acid synthesis, including DNA replication and transcription occurs in the 5'→3' direction, because new nucleotides are added via a dehydration reaction that uses the exposed 3' hydroxyl as a nucleophile.[52]:27.2

The expression of genes encoded in DNA begins by transcribing the gene into RNA, a second type of nucleic acid that is very similar to DNA, but whose monomers contain the sugar ribose rather than deoxyribose. RNA also contains the base uracil in place of thymine. RNA molecules are less stable than DNA and are typically single-stranded. Genes that encode proteins are composed of a series of three-nucleotide sequences called codons, which serve as the "words" in the genetic "language". The genetic code specifies the correspondence during protein translation between codons and amino acids. The genetic code is nearly the same for all known organisms.[51]:4.1

Chromosomes

The total complement of genes in an organism or cell is known as its genome, which may be stored on one or more chromosomes. A chromosome consists of a single, very long DNA helix on which thousands of genes are encoded.[51]:4.2 The region of the chromosome at which a particular gene is located is called its locus. Each locus contains one allele of a gene; however, members of a population may have different alleles at the locus, each with a slightly different gene sequence.

Micrographic karyogram of human male, showing 23 pairs of chromosomes. The largest chromosomes are around 10 times the size of the smallest.[53]

The majority of eukaryotic genes are stored on a set of large, linear chromosomes. The chromosomes are packed within the nucleus in complex with storage proteins called histones to form a unit called a nucleosome. DNA packaged and condensed in this way is called chromatin.[51]:4.2 The manner in which DNA is stored on the histones, as well as chemical modifications of the histone itself, regulate whether a particular region of DNA is accessible for gene expression. In addition to genes, eukaryotic chromosomes contain sequences involved in ensuring that the DNA is copied without degradation of end regions and sorted into daughter cells during cell division: replication origins, telomeres, and the centromere.[51]:4.2 Replication origins are the sequence regions where DNA replication is initiated to make two copies of the chromosome. Telomeres are long stretches of repetitive sequences that cap the ends of the linear chromosomes and prevent degradation of coding and regulatory regions during DNA replication. The length of the telomeres decreases each time the genome is replicated and has been implicated in the aging process.[54] The centromere is required for binding spindle fibres to separate sister chromatids into daughter cells during cell division.[51]:18.2
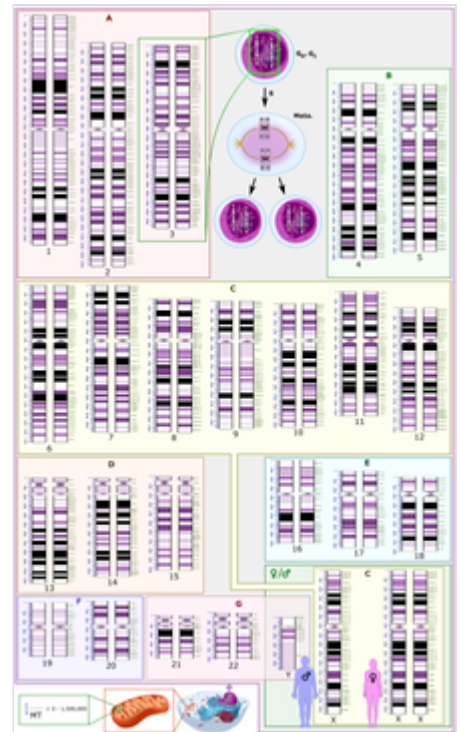
Prokaryotes (bacteria and archaea) typically store their genomes on a single, large, circular chromosome. Similarly, some eukaryotic organelles contain a remnant circular chromosome with a small number of genes.[51]:14.4 Prokaryotes sometimes supplement their chromosome with additional small circles of DNA called plasmids, which usually encode only a few genes and are transferable between individuals. For example, the genes for antibiotic resistance are usually encoded on bacterial plasmids and can be passed between individual cells, even those of different species, via horizontal gene transfer.[55]

Whereas the chromosomes of prokaryotes are relatively gene-dense, those of eukaryotes often contain regions of DNA that serve no obvious function. Simple single-celled eukaryotes have relatively small amounts of such DNA, whereas the genomes of complex multicellular organisms, including humans, contain an absolute majority of DNA without an identified function.[56] This DNA has often been referred to as "junk DNA". However, more recent analyses suggest that, although protein-coding DNA makes up barely 2% of the human genome, about 80% of the bases in the genome may be expressed, so the term "junk DNA" may be a misnomer.[26]
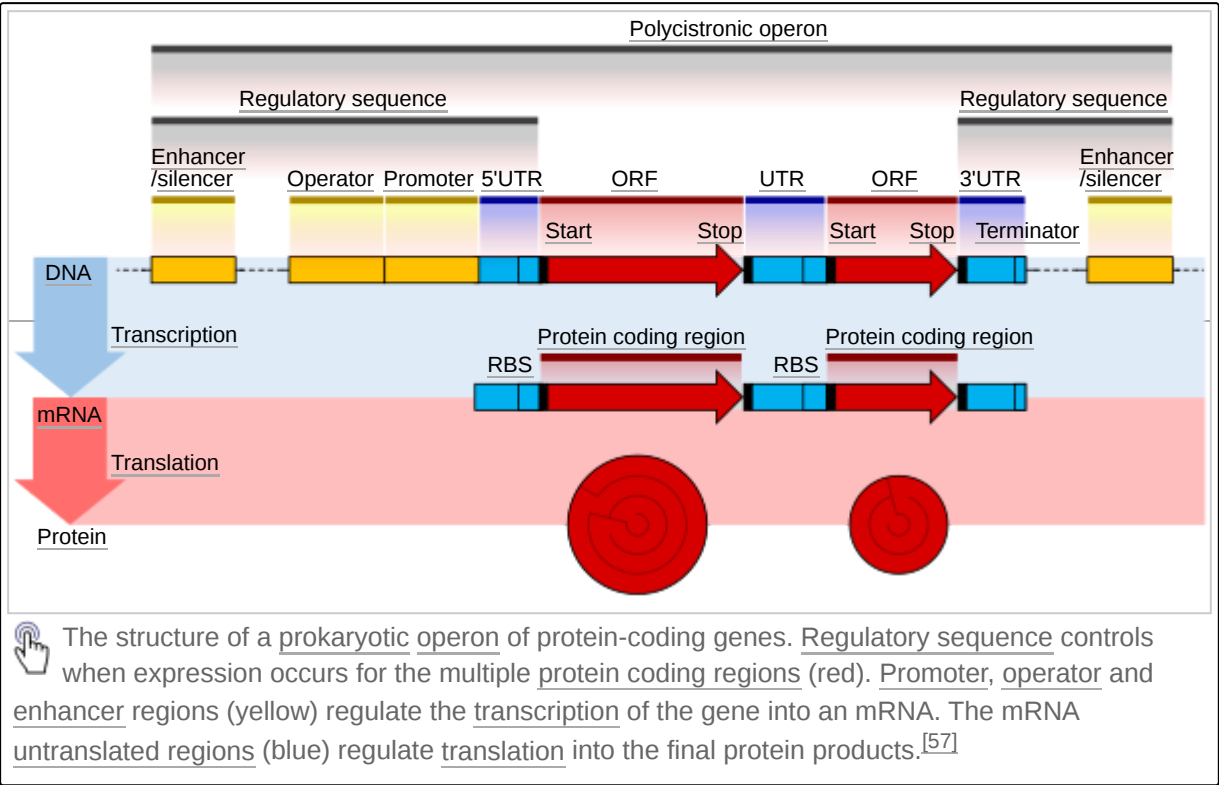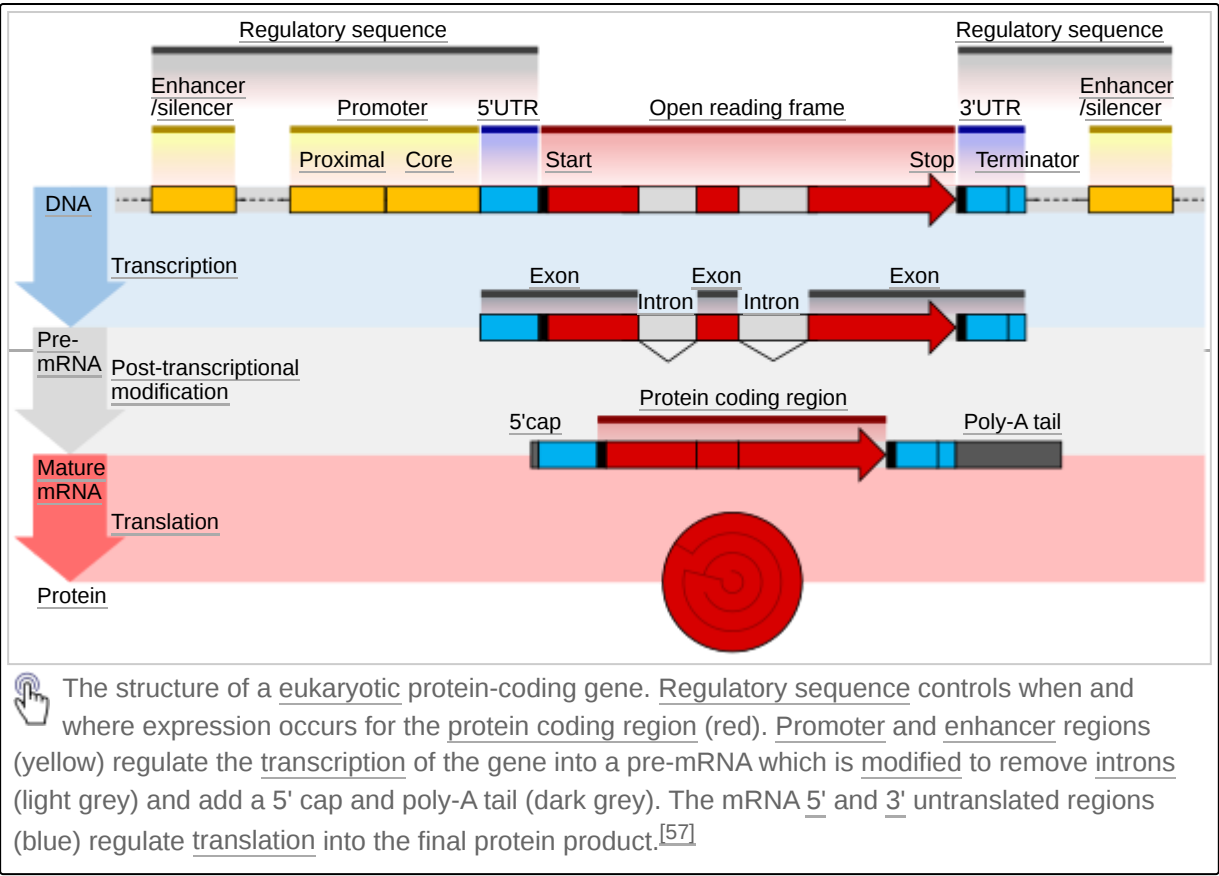

Schematic karyogram of a human, with annotated bands and sub-bands. It shows dark and white regions on G banding. It shows 22 homologous chromosomes, both the male (XY) and female (XX) versions of the sex chromosome (bottom right), as well as the mitochondrial genome (at bottom left).

Structure and function

Structure

The

The structure of a eukaryotic protein-coding gene. Regulatory sequence controls when and where expression occurs for the protein coding region (red). Promoter and enhancer regions (yellow) regulate the transcription of the gene into a pre-mRNA which is modified to remove introns (light grey) and add a 5' cap and poly-A tail (dark grey). The mRNA 5' and 3' untranslated regions (blue) regulate translation into the final protein product.[57]

The structure of a prokaryotic operon of protein-coding genes. Regulatory sequence controls when expression occurs for the multiple protein coding regions (red). Promoter, operator and enhancer regions (yellow) regulate the transcription of the gene into an mRNA. The mRNA untranslated regions (blue) regulate translation into the final protein products.[57]

structure of a protein-coding gene consists of many elements of which the actual protein coding sequence is often only a small part. These include introns and untranslated regions of the mature mRNA. Noncoding genes can also contain introns that are removed during processing to produce the mature functional RNA.

All genes are associated with regulatory sequences that are required for their expression. First, genes require a promoter sequence. The promoter is recognized and bound by transcription factors that recruit and help RNA polymerase bind to the region to initiate transcription.[51]:7.1 The recognition typically occurs as a consensus sequence like the TATA box. A gene can have more than one promoter, resulting in messenger RNAs (mRNA) that differ in how far they extend in the 5' end.[58] Highly transcribed genes have "strong" promoter sequences that form strong associations with transcription factors, thereby initiating transcription at a high rate. Others genes have "weak" promoters that form weak associations with transcription factors and initiate transcription less frequently.[51]:7.2 Eukaryotic promoter regions are much more complex and difficult to identify than prokaryotic promoters.[51]:7.3

Additionally, genes can have regulatory regions many kilobases upstream or downstream of the gene that alter expression. These act by binding to transcription factors which then cause the DNA to loop so that the regulatory sequence (and bound transcription factor) become close to the RNA polymerase binding site.[59] For example, enhancers increase transcription by binding an activator protein which then helps to recruit the RNA polymerase to the promoter; conversely silencers bind repressor proteins and make the DNA less available for RNA polymerase.[60]

The mature messenger RNA produced from protein-coding genes contains untranslated regions at both ends which contain binding sites for ribosomes, RNA-binding proteins, miRNA, as well as terminator, and start and stop codons.[61] In addition, most eukaryotic open reading frames contain untranslated introns, which are removed and exons, which are connected together in a process known as RNA splicing. Finally, the ends of gene transcripts are defined by cleavage and polyadenylation (CPA) sites, where newly produced pre-mRNA gets cleaved and a string of ~200 adenosine monophosphates is added at the 3' end. The poly(A) tail protects mature mRNA from degradation and has other functions, affecting translation, localization, and transport of the transcript from the nucleus. Splicing, followed by CPA, generate the final mature mRNA, which encodes the protein or RNA product.[62]

Many noncoding genes in eukaryotes have different transcription termination mechanisms and they do not have poly(A) tails.

Many prokaryotic genes are organized into operons, with multiple protein-coding sequences that are transcribed as a unit.[63][64] The genes in an operon are transcribed as a continuous messenger RNA, referred to as a polycistronic mRNA. The term cistron in this context is equivalent to gene. The transcription of an operon's mRNA is often controlled by a repressor that can occur in an active or inactive state depending on the presence of specific metabolites.[65] When active, the repressor binds to a DNA sequence at the beginning of the operon, called the operator region, and represses transcription of the operon; when the repressor is inactive transcription of the operon can occur (see e.g. Lac operon). The products of operon genes typically have related functions and are involved in the same regulatory network.[51]:7.3

Complexity

Though many genes have simple structures, as with much of biology, others can be quite complex or represent unusual edge-cases. Eukaryotic genes often have introns that are much larger than their exons,[66][67] and those introns can even have other genes nested inside them.[68] Associated enhancers may be many kilobase away, or even on entirely different chromosomes operating via physical contact between two chromosomes.[69][70] A single gene can encode multiple different functional products by

alternative splicing, and conversely a gene may be split across chromosomes but those transcripts are concatenated back together into a functional sequence by trans-splicing.[71] It is also possible for overlapping genes to share some of their DNA sequence, either on opposite strands or the same strand (in a different reading frame, or even the same reading frame).[72]

Gene expression

In all organisms, two steps are required to read the information encoded in a gene's DNA and produce the protein it specifies. First, the gene's DNA is *transcribed* to messenger RNA (mRNA).[51]:6.1 Second, that mRNA is *translated* to protein.[51]:6.2 RNA-coding genes must still go through the first step, but are not translated into protein.[73] The process of producing a biologically functional molecule of either RNA or protein is called gene expression, and the resulting molecule is called a gene product.

Genetic code

The nucleotide sequence of a gene's DNA specifies the amino acid sequence of a protein through the genetic code. Sets of three nucleotides, known as codons, each correspond to a specific amino acid.[51]:6 The principle that three sequential bases of DNA code for each amino acid was demonstrated in 1961 using frameshift mutations in the rIIB gene of bacteriophage T4[74] (see Crick, Brenner et al. experiment).
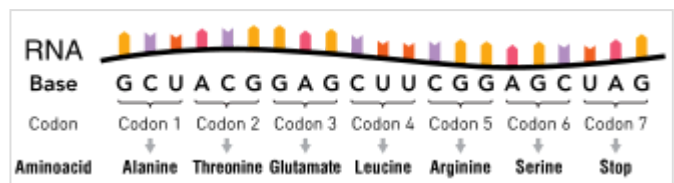
Schematic of a single-stranded RNA molecule illustrating a series of three-base codons. Each three-nucleotide codon corresponds to an amino acid when translated to protein.

Additionally, a "start codon", and three "stop codons" indicate the beginning and end of the protein coding region. There are 64 possible codons (four possible nucleotides at each of three positions, hence 4^3 possible codons) and only 20 standard amino acids; hence the code is redundant and multiple codons can specify the same amino acid. The correspondence between codons and amino acids is nearly universal among all known living organisms.[75]
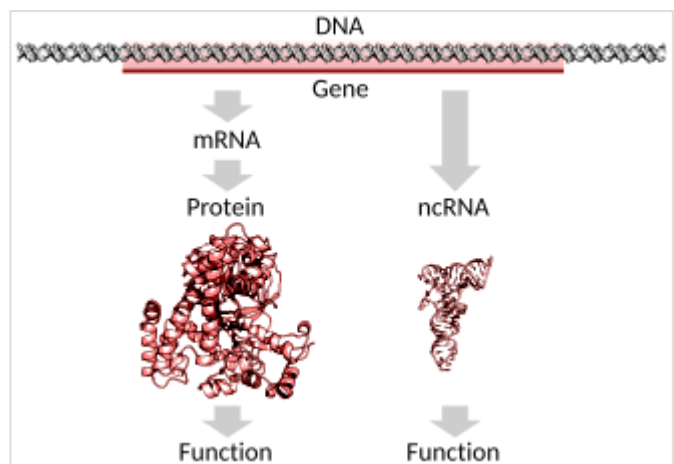
Transcription

Transcription produces a single-stranded RNA molecule known as messenger RNA, whose nucleotide sequence is complementary to the DNA from which it was transcribed.[51]:6.1 The mRNA acts as an intermediate between the DNA gene and its final protein product. The gene's DNA is used as a template to generate a complementary mRNA. The mRNA matches the sequence of the gene's DNA coding strand because it is synthesised as the complement of the template strand. Transcription is performed by an enzyme called an RNA polymerase, which reads the template strand in the 3' to 5' direction and synthesizes the RNA from 5' to 3'. To initiate transcription, the polymerase first recognizes and binds a promoter region of the gene. Thus, a major mechanism of gene regulation is the blocking or sequestering the promoter region, either by tight binding by repressor molecules that physically block the polymerase or by organizing the DNA so that the promoter region is not accessible.[51]:7

In prokaryotes, transcription occurs in the cytoplasm; for very long transcripts, translation may begin at the 5' end of the RNA while the 3' end is still being transcribed. In eukaryotes, transcription occurs in the nucleus, where the cell's DNA is stored. The RNA molecule produced by the polymerase is known as the primary transcript and undergoes post-transcriptional modifications before being exported to the cytoplasm for translation. One of the modifications performed is the splicing of introns which are sequences in the transcribed region that do not encode a protein. Alternative splicing mechanisms can result in mature transcripts from the same gene having different sequences and thus coding for different proteins. This is a major form of regulation in eukaryotic cells and also occurs in some prokaryotes.[51]:7.5[76]

Translation

Translation is the process by which a mature mRNA molecule is used as a template for synthesizing a new protein.[51]:6.2 Translation is carried out by ribosomes, large complexes of RNA and protein responsible for carrying out the chemical reactions to add new amino acids to a growing polypeptide chain by the formation of peptide bonds. The genetic code is read three nucleotides at a time, in units called codons, via interactions with specialized RNA molecules called transfer RNA (tRNA). Each tRNA has three unpaired bases known as the anticodon that are complementary to the codon it reads on the mRNA. The tRNA is also covalently attached to the amino acid specified by the complementary codon. When the tRNA binds to its complementary codon in an mRNA strand, the ribosome attaches its amino acid cargo to the

Protein coding genes are transcribed to an mRNA intermediate, then translated to a functional protein. RNA-coding genes are transcribed to a functional non-coding RNA (PDB: 3BSE (https://www.rcsb.org/structur e/3BSE), 1OBB (https://www.rcsb.org/structure/1OBB), 3TRA (https://www.rcsb.org/structure/3TRA)).

new polypeptide chain, which is synthesized from amino terminus to carboxyl terminus. During and after synthesis, most new proteins must fold to their active three-dimensional structure before they can carry out their cellular functions.[51]:3

Regulation

Genes are regulated so that they are expressed only when the product is needed, since expression draws on limited resources.[51]:7 A cell regulates its gene expression depending on its external environment (e.g. available nutrients, temperature and other stresses), its internal environment (e.g. cell division cycle, metabolism, infection status), and its specific role if in a multicellular organism. Gene expression can be regulated at any step: from transcriptional initiation, to RNA processing, to post-translational modification of the protein. The regulation of lactose metabolism genes in *E. coli* (*lac* operon) was the first such mechanism to be described in 1961.[77]

RNA genes

A typical protein-coding gene is first copied into RNA as an intermediate in the manufacture of the final protein product.[51]:6.1 In other cases, the RNA molecules are the actual functional products, as in the synthesis of ribosomal RNA and transfer RNA. Some RNAs known as ribozymes are capable of enzymatic function, while others such as microRNAs and riboswitches have regulatory roles. The DNA sequences from which such RNAs are transcribed are known as non-coding RNA genes.[73]

Some viruses store their entire genomes in the form of RNA, and contain no DNA at all.[78][79] Because they use RNA to store genes, their cellular hosts may synthesize their proteins as soon as they are infected and without the delay in waiting for transcription.[80] On the other hand, RNA retroviruses, such as HIV, require the reverse transcription of their genome from RNA into DNA before their proteins can be synthesized.
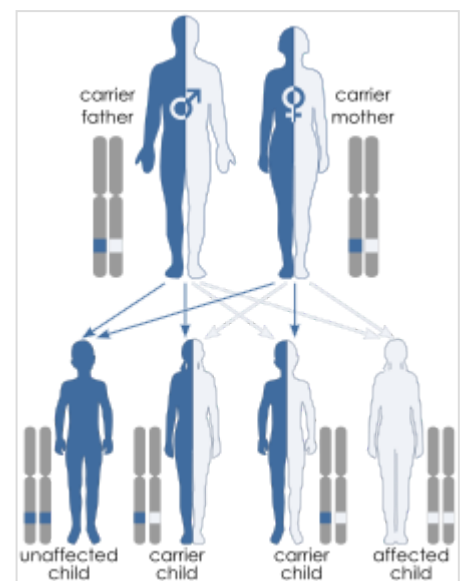
Inheritance

Organisms inherit their genes from their parents. Asexual organisms simply inherit a complete copy of their parent's genome. Sexual organisms have two copies of each chromosome because they inherit one complete set from each parent.[51]:1

Mendelian inheritance

According to Mendelian inheritance, variations in an organism's phenotype (observable physical and behavioral characteristics) are due in part to variations in its genotype (particular set of genes). Each gene specifies a particular trait with a different sequence of a gene (alleles) giving rise to different phenotypes. Most eukaryotic organisms (such as the pea plants Mendel worked on) have two alleles for each trait, one inherited from each parent.[51]:20

Inheritance of a gene that has two different alleles (blue and white). The gene is located on an autosomal chromosome. The white allele is recessive to the blue allele. The probability of each outcome in the children's generation is one quarter, or 25 percent.

Alleles at a locus may be dominant or recessive; dominant alleles give rise to their corresponding phenotypes when paired with any other allele for the same trait, whereas recessive alleles give rise to their corresponding phenotype only when paired with another copy of the same allele. If you know the genotypes of the organisms, you can determine which alleles are dominant and which are recessive. For example, if the allele specifying tall stems in pea plants is dominant over the allele specifying short stems, then pea plants that inherit one tall allele from one parent and one short allele from the other parent will also have tall stems. Mendel's work demonstrated that alleles assort independently in the production of gametes, or germ cells, ensuring variation in the next generation. Although Mendelian inheritance remains a good model for many traits determined by single genes (including a number of well-known genetic disorders) it does not include the physical processes of DNA replication and cell division.[81][82]

DNA replication and cell division

The growth, development, and reproduction of organisms relies on cell division; the process by which a single cell divides into two usually identical daughter cells. This requires first making a duplicate copy of every gene in the genome in a process called DNA replication.[51]:5.2 The copies are made by specialized enzymes known as DNA polymerases, which "read" one strand of the double-helical DNA, known as the template strand, and synthesize a new complementary strand. Because the DNA double helix is held together by base pairing, the sequence of one strand completely specifies the sequence of its complement; hence only one strand needs to be read by the enzyme to produce a faithful copy. The process of DNA replication is semiconservative; that is, the copy of the genome inherited by each daughter cell contains one original and one newly synthesized strand of DNA.[51]:5.2

The rate of DNA replication in living cells was first measured as the rate of phage T4 DNA elongation in phage-infected *E. coli* and found to be impressively rapid.[83] During the period of exponential DNA increase at 37 °C, the rate of elongation was 749 nucleotides per second.

After DNA replication, the cell must physically separate the two genome copies and divide into two distinct membrane-bound cells.[51]:18.2 In prokaryotes (bacteria and archaea) this usually occurs via a relatively simple process called binary fission, in which each circular genome attaches to the cell membrane and is separated into the daughter cells as the membrane invaginates to split the cytoplasm into two membrane-bound portions. Binary fission is extremely fast compared to the rates of cell division in eukaryotes. Eukaryotic cell division is a more complex process known as the cell cycle; DNA replication occurs during a phase of this cycle known as S phase, whereas the process of segregating chromosomes and splitting the cytoplasm occurs during M phase.[51]:18.1

Molecular inheritance

The duplication and transmission of genetic material from one generation of cells to the next is the basis for molecular inheritance and the link between the classical and molecular pictures of genes. Organisms inherit the characteristics of their parents because the cells of the offspring contain copies of the genes in their parents' cells. In asexually reproducing organisms, the offspring will be a genetic copy or clone of the parent organism. In sexually reproducing organisms, a specialized form of cell division called meiosis produces cells called gametes or germ cells that are haploid, or contain only one copy of each gene.[51]:20.2 The gametes produced by females are called eggs or ova, and those produced by males are called sperm. Two gametes fuse to form a diploid fertilized egg, a single cell that has two sets of genes, with one copy of each gene from the mother and one from the father.[51]:20

During the process of meiotic cell division, an event called genetic recombination or *crossing-over* can sometimes occur, in which a length of DNA on one chromatid is swapped with a length of DNA on the corresponding homologous non-sister chromatid. This can result in reassortment of otherwise linked alleles.[51]:5.5 The Mendelian principle of independent assortment asserts that each of a parent's two genes for each trait will sort independently into gametes; which allele an organism inherits for one trait is unrelated to which allele it inherits for another trait. This is in fact only true for genes that do not reside on the same chromosome or are located very far from one another on the same chromosome. The closer

two genes lie on the same chromosome, the more closely they will be associated in gametes and the more often they will appear together (known as genetic linkage).[84] Genes that are very close are essentially never separated because it is extremely unlikely that a crossover point will occur between them.[84]

Genome

The genome is the total genetic material of an organism and includes both the genes and non-coding sequences.[85] Eukaryotic genes can be annotated using FINDER.[86]

Number of genes

Depiction of numbers of genes for representative plants (green), vertebrates (blue), invertebrates (orange), fungi (yellow), bacteria (purple), and viruses (grey). An inset on the right shows the smaller genomes expanded 100-fold area-wise.[87][88][89][90][91][92][93][94]

The genome size, and the number of genes it encodes varies widely between organisms. The smallest genomes occur in viruses,[95] and viroids (which act as a single non-coding RNA gene).[96] Conversely, plants can have extremely large genomes,[97] with rice containing >46,000 protein-coding genes.[91] The total number of protein-coding genes (the Earth's proteome) is estimated to be 5 million sequences.[98]

Although the number of base-pairs of DNA in the human genome has been known since the 1950s, the estimated number of genes has changed over time as definitions of genes, and methods of detecting them have been refined. Initial theoretical predictions of the number of human genes in the 1960s and 1970s were based on mutation load estimates and the numbers of mRNAs and these estimates tended to be about 30,000 protein-coding genes.[99][100][101] During the 1990s there were guesstimates of up to 100,000 genes and early data on detection of mRNAs (expressed sequence tags) suggested more than the traditional value of 30,000 genes that had been reported in the textbooks during the 1980s.[102]

The initial draft sequences of the human genome confirmed the earlier predictions of about 30,000 protein-coding genes however that estimate has fallen to about 19,000 with the ongoing GENCODE annotation project.[103] The number of noncoding genes is not known with certainty but the latest estimates from Ensembl suggest 26,000 noncoding genes.[104]

Essential genes

Essential genes are the set of genes thought to be critical for an organism's survival.[106] This definition assumes the abundant availability of all relevant nutrients and the absence of environmental stress. Only a small portion of an organism's genes are essential. In bacteria, an estimated 250–400 genes are essential

for *Escherichia coli* and *Bacillus subtilis*, which is less than 10% of their genes.[107][108][109] Half of these genes are orthologs in both organisms and are largely involved in protein synthesis.[109] In the budding yeast *Saccharomyces cerevisiae* the number of essential genes is slightly higher, at 1000 genes (~20% of their genes).[110] Although the number is more difficult to measure in higher eukaryotes, mice and humans are estimated to have around 2000 essential genes (~10% of their genes).[111] The synthetic organism, *Syn 3*, has a minimal genome of 473 essential genes and quasi-essential genes (necessary for fast growth), although 149 have unknown function.[105]
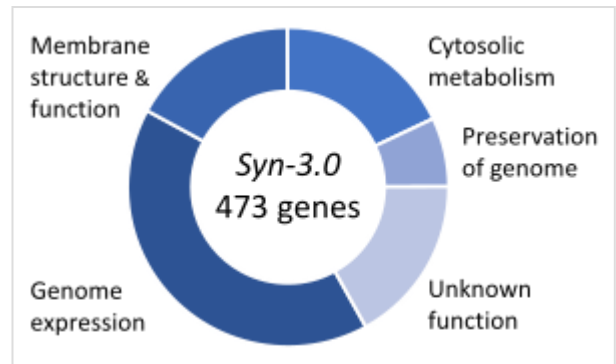
Gene functions in the minimal genome of the synthetic organism, *Syn 3*[105]

Essential genes include housekeeping genes (critical for basic cell functions)[112] as well as genes that are expressed at different times in the organisms development or life cycle.[113] Housekeeping genes are used as experimental controls when analysing gene expression, since they are constitutively expressed at a relatively constant level.

Genetic and genomic nomenclature

Gene nomenclature was established by the HUGO Gene Nomenclature Committee (HGNC), a committee of the Human Genome Organisation, for each known human gene in the form of an approved gene name and symbol (short-form abbreviation), which can be accessed through a database maintained by HGNC. Symbols are chosen to be unique, and each gene has only one symbol (although approved symbols sometimes change). Symbols are preferably kept consistent with other members of a gene family and with homologs in other species, particularly the mouse due to its role as a common model organism.[114]

Genetic engineering

Genetic engineering is the modification of an organism's genome through biotechnology. Since the 1970s, a variety of techniques have been developed to specifically add, remove and edit genes in an organism.[115] Recently developed genome engineering techniques use engineered nuclease enzymes to create targeted DNA repair in a chromosome to either disrupt or edit a gene when the break is repaired.[116][117][118][119] The related term synthetic biology is sometimes used to refer to extensive genetic engineering of an organism.[120]

Genetic engineering is now a routine research tool with model organisms. For example, genes are easily added to bacteria[121] and lineages of knockout mice with a specific gene's function disrupted are used to investigate that gene's function.[122][123] Many organisms have been genetically modified for applications in agriculture, industrial biotechnology, and medicine.

For multicellular organisms, typically the embryo is engineered which grows into the adult genetically modified organism.[124] However, the genomes of cells in an adult organism can be edited using gene therapy techniques to treat genetic diseases.
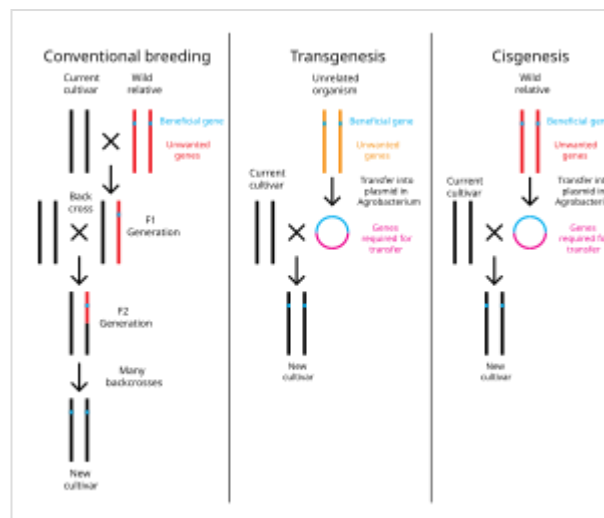

Comparison of conventional plant breeding with transgenic and cisgenic genetic modification

See also

- Biological patent
- Epigenetics
- Gene-centric view of evolution
- Gene dosage
- Gene redundancy
- Gene silencing
- Genetic algorithm
- Haplotype
- List of gene prediction software
- Lists of human genes
- Predictive medicine
- Quantitative trait locus
- Selfish genetic element
- Whole genome sequencing

References

Citations

1. Orgogozo V, Peluffo AE, Morizot B (2016). "The 'Mendelian Gene' and the 'Molecular Gene': Two Relevant Concepts of Genetic Units" (https://hal.archives-ouvertes.fr/hal-01354346/file/Orgogozo2016-gene.pdf) (PDF). *Current Topics in Developmental Biology*. **119**: 1–26. doi:10.1016/bs.ctdb.2016.03.002 (https://doi.org/10.1016%2Fbs.ctdb.2016.03.002). PMID 27282022 (https://pubmed.ncbi.nlm.nih.gov/27282022).
2. "What is a gene?: MedlinePlus Genetics" (https://medlineplus.gov/genetics/understanding/basics/gene/). *MedlinePlus*. 17 September 2020. Archived (https://web.archive.org/web/20200516195548/http://ghr.nlm.nih.gov/handbook/basics/gene) from the original on 16 May 2020. Retrieved 4 January 2021.
3. Hirsch ED (2002). *The new dictionary of cultural literacy*. Boston: Houghton Mifflin. ISBN 0-618-22647-8. OCLC 50166721 (https://search.worldcat.org/oclc/50166721).
4. Elston RC, Satagopan JM, Sun S (2012). "Genetic terminology". *Statistical Human Genetics*. Methods in Molecular Biology. Vol. 850. Humana Press. pp. 1–9. doi:10.1007/978-1-61779-555-8_1 (https://doi.org/10.1007%2F978-1-61779-555-8_1). ISBN 978-1-61779-554-1. PMC 4450815 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4450815). PMID 22307690 (https://pubmed.ncbi.nlm.nih.gov/22307690).
5. Kampourakis K (2017). *Making Sense of Genes*. Cambridge, UK: Cambridge University Press.

6. Gericke N, Hagberg M (5 December 2006). "Definition of historical models of gene function and their relation to students' understanding of genetics". *Science & Education*. **16** (7–8): 849–881. Bibcode:2007Sc&Ed..16..849G (https://ui.adsabs.harvard.edu/abs/2007Sc&Ed..16..849G). doi:10.1007/s11191-006-9064-4 (https://doi.org/10.1007%2Fs11191-006-9064-4).

7. Meunier R (2022). "Stanford Encyclopedia of Philosophy: Gene" (https://plato.stanford.edu/entries/gene/). *Stanford Encyclopedia of Philosophy*. Retrieved 28 February 2023.

8. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. (April 2014). "Defining functional DNA elements in the human genome" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4035993). *Proceedings of the National Academy of Sciences of the United States of America*. **111** (17): 6131–8. Bibcode:2014PNAS..111.6131K (https://ui.adsabs.harvard.edu/abs/2014PNAS..111.6131K). doi:10.1073/pnas.1318948111 (https://doi.org/10.1073%2Fpnas.1318948111). PMC 4035993 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4035993). PMID 24753594 (https://pubmed.ncbi.nlm.nih.gov/24753594).

9. Dawkins R (1976). *The selfish gene*. Oxford, UK: Oxford University Press.

10. Stoltz K, Griffiths P (2004). "Genes: Philosophical Analyses Put to the Test". *History and Philosophy of the Life Sciences*. **26** (1): 5–28. doi:10.1080/03919710412331341621 (https://doi.org/10.1080%2F03919710412331341621). JSTOR 23333378 (https://www.jstor.org/stable/23333378). PMID 15791804 (https://pubmed.ncbi.nlm.nih.gov/15791804).

11. Beadle GW, Tatum EL (November 1941). "Genetic Control of Biochemical Reactions in Neurospora" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1078370). *Proceedings of the National Academy of Sciences of the United States of America*. **27** (11): 499–506. Bibcode:1941PNAS...27..499B (https://ui.adsabs.harvard.edu/abs/1941PNAS...27..499B). doi:10.1073/pnas.27.11.499 (https://doi.org/10.1073%2Fpnas.27.11.499). PMC 1078370 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1078370). PMID 16588492 (https://pubmed.ncbi.nlm.nih.gov/16588492).

12. Horowitz NH, Berg P, Singer M, Lederberg J, Susman M, Doebley J, Crow JF (January 2004). "A centennial: George W. Beadle, 1903-1989" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470705). *Genetics*. **166** (1): 1–10. doi:10.1534/genetics.166.1.1 (https://doi.org/10.1534%2Fgenetics.166.1.1). PMC 1470705 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470705). PMID 15020400 (https://pubmed.ncbi.nlm.nih.gov/15020400).

13. Judson HF (1996). *The Eight Day of Creation* (Expanded ed.). Plainview, NY (US): Cold Spring Harbor Laboratory Press.

14. Watson JD (1965). *Molecular Biology of the Gene*. New York, NY, US: W.A. Benjamin, Inc.

15. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1994). *Molecular Biology of the Cell: Third Edition*. London, UK: Garland Publishing, Inc. ISBN 0-8153-1619-4.

16. Moran LA, Horton HR, Scrimgeour KG, Perry MD (2012). *Principles of Biochemistry: Fifth Edition*. Upper Saddle River, NJ, US: Pearson.

17. Lewin B (2004). *Genes VIII*. Upper Saddle River, NJ, US: Pearson/Prentice Hall.

18. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, and Vitale L (2019). "On the length, weight and GC content of the human genome" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6391780). *BMC Research Notes*. **12** (1): 106–173. doi:10.1186/s13104-019-4137-z (https://doi.org/10.1186%2Fs13104-019-4137-z). PMC 6391780 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6391780). PMID 30813969 (https://pubmed.ncbi.nlm.nih.gov/30813969).

19. Hubé F, and Francastel C (2015). "Mammalian Introns: When the Junk Generates Molecular Diversity" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394429). *International Journal of Molecular Sciences*. **16** (3): 4429–4452. doi:10.3390/ijms16034429 (https://doi.org/10.3390%2Fijms16034429). PMC 4394429 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394429). PMID 25710723 (https://pubmed.ncbi.nlm.nih.gov/25710723).

20. Francis WR, and Wörheide G (2017). "Similar ratios of introns to intergenic sequence across animal genomes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5534336). *Genome Biology and Evolution*. **9** (6): 1582–1598. doi:10.1093/gbe/evx103 (https://doi.org/10.1093%2Fgbe%2Fevx103). PMC 5534336 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5534336). PMID 28633296 (https://pubmed.ncbi.nlm.nih.gov/28633296).

21. Mortola E, Long M (2021). "Turning Junk into Us: How Genes Are Born" (https://www.americanscientist.org/article/turning-junk-into-us-how-genes-are-born). *American Scientist*. **109**: 174–182. Archived (https://web.archive.org/web/20230227065824/https://www.americanscientist.org/article/turning-junk-into-us-how-genes-are-born) from the original on 27 February 2023. Retrieved 27 February 2023.

22. Hopkin K (2009). "The Evolving Definition of a Gene: With the discovery that nearly all of the genome is transcribed, the definition of a "gene" needs another revision". *BioScience*. **59**: 928–931. doi:10.1525/bio.2009.59.11.3 (https://doi.org/10.1525%2Fbio.2009.59.11.3). S2CID 88157272 (https://api.semanticscholar.org/CorpusID:88157272).

23. Pearson H (2006). "What Is a Gene?" (https://doi.org/10.1038%2F441398a). *Nature*. **441** (7092): 399–401. Bibcode:2006Natur.441..398P (https://ui.adsabs.harvard.edu/abs/2006Natur.441..398P). doi:10.1038/441398a (https://doi.org/10.1038%2F441398a). PMID 16724031 (https://pubmed.ncbi.nlm.nih.gov/16724031). S2CID 4420674 (https://api.semanticscholar.org/CorpusID:4420674).

24. Pennisi E (2007). "DNA study forces rethink of what it means to be a gene" (https://doi.org/10.1126%2Fscience.316.5831.1556). *Science*. **316** (5831): 1556–1557. doi:10.1126/science.316.5831.1556 (https://doi.org/10.1126%2Fscience.316.5831.1556). PMID 17569836 (https://pubmed.ncbi.nlm.nih.gov/17569836). S2CID 36463252 (https://api.semanticscholar.org/CorpusID:36463252).

25. Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, et al. (November 2018). Racaniello VR (ed.). "Origins and Evolution of the Global RNA Virome" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6282212). *mBio*. **9** (6). Eric Delwart, Luis Enjuanes: e02329–18. doi:10.1128/mBio.02329-18 (https://doi.org/10.1128%2FmBio.02329-18). PMC 6282212 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6282212). PMID 30482837 (https://pubmed.ncbi.nlm.nih.gov/30482837).

26. Pennisi E (June 2007). "Genomics. DNA study forces rethink of what it means to be a gene" (https://doi.org/10.1126%2Fscience.316.5831.1556). *Science*. **316** (5831): 1556–7. doi:10.1126/science.316.5831.1556 (https://doi.org/10.1126%2Fscience.316.5831.1556). PMID 17569836 (https://pubmed.ncbi.nlm.nih.gov/17569836). S2CID 36463252 (https://api.semanticscholar.org/CorpusID:36463252).

27. Marande W, Burger G (October 2007). "Mitochondrial DNA as a genomic jigsaw puzzle". *Science*. **318** (5849). AAAS: 415. Bibcode:2007Sci...318..415M (https://ui.adsabs.harvard.edu/abs/2007Sci...318..415M). doi:10.1126/science.1148033 (https://doi.org/10.1126%2Fscience.1148033). PMID 17947575 (https://pubmed.ncbi.nlm.nih.gov/17947575). S2CID 30948765 (https://api.semanticscholar.org/CorpusID:30948765).

28. Parra G, Reymond A, Dabbouseh N, Dermitzakis ET, Castelo R, Thomson TM, et al. (January 2006). "Tandem chimerism as a means to increase protein complexity in the human genome" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1356127). *Genome Research*. **16** (1): 37–44. doi:10.1101/gr.4145906 (https://doi.org/10.1101%2Fgr.4145906). PMC 1356127 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1356127). PMID 16344564 (https://pubmed.ncbi.nlm.nih.gov/16344564).

29. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, et al. (June 2007). "What is a gene, post-ENCODE? History and updated definition" (https://doi.org/10.1101%2Fgr.6339607). *Genome Research*. **17** (6): 669–81. doi:10.1101/gr.6339607 (https://doi.org/10.1101%2Fgr.6339607). PMID 17567988 (https://pubmed.ncbi.nlm.nih.gov/17567988).

30. Noble D (September 2008). "Genes and causation" (https://doi.org/10.1098%2Frsta.2008.0 086). *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*. **366** (1878): 3001–15. Bibcode:2008RSPTA.366.3001N (https://ui.adsabs.harvar d.edu/abs/2008RSPTA.366.3001N). doi:10.1098/rsta.2008.0086 (https://doi.org/10.1098%2 Frsta.2008.0086). PMID 18559318 (https://pubmed.ncbi.nlm.nih.gov/18559318).

31. Magnello, M.E. (2013). "Galton's Law of Ancestral Heredity". *Brenner's Encyclopedia of Genetics*. pp. 143–146. doi:10.1016/B978-0-12-374984-0.00060-7 (https://doi.org/10.1016% 2FB978-0-12-374984-0.00060-7). ISBN 978-0-08-096156-9.

32. "genesis" (https://www.oed.com/search/dictionary/?q=genesis). *Oxford English Dictionary* (Online ed.). Oxford University Press. (Subscription or participating institution membership (http s://www.oed.com/public/login/loggingin#withyourlibrary) required.)

33. Magner LN (2002). *A History of the Life Sciences* (https://books.google.com/books?id=YKJ6 gVYbrGwC) (Third ed.). Marcel Dekker, CRC Press. p. 371. ISBN 978-0-203-91100-6.

34. Henig RM (2000). *The Monk in the Garden: The Lost and Found Genius of Gregor Mendel, the Father of Genetics* (https://archive.org/details/monkingardenlost00heni). Boston: Houghton Mifflin. pp. 1 (https://archive.org/details/monkingardenlost00heni/page/n14)–9. ISBN 978-0395-97765-1.

35. de Vries H (1889). *Intracellulare Pangenese* (http://www.esp.org/books/devries/pangenesis/f acsimile/) [*Intracellular Pangenesis*] (in German). Translated by Gager CS. Jena: Verlag von Gustav Fischer. Archived (https://web.archive.org/web/20230602053529/http://www.esp.org/ books/devries/pangenesis/facsimile/) from the original on 2 June 2023. Retrieved 6 May 2007. Translated in 1908 from German to English by Open Court Publishing Co., Chicago, 1910

36. Johannsen W (1909). *Elemente der exakten Erblichkeitslehre* (https://www.biodiversitylibrar y.org/item/15717#page/134/mode/1up) [*Elements of the exact theory of heredity*] (in German). Jena, Germany: Gustav Fischer. p. 124. Archived (https://web.archive.org/web/20 210624195335/https://www.biodiversitylibrary.org/item/15717#page/134/mode/1up) from the original on 24 June 2021. Retrieved 16 June 2021. From p. 124: *"Dieses "etwas" in den Gameten bezw. in der Zygote, ... – kurz, was wir eben Gene nennen wollen – bedingt sind."* (This "something" in the gametes or in the zygote, which has crucial importance for the character of the organism, is usually called by the quite ambiguous term *Anlagen* [primordium, from the German word *Anlage* for "plan, arrangement; rough sketch"]. Many other terms have been suggested, mostly unfortunately in closer connection with certain hypothetical opinions. The word "pangene", which was introduced by Darwin, is perhaps used most frequently in place of *Anlagen*. However, the word "pangene" was not well chosen, as it is a compound word containing the roots *pan* (the neuter form of Πας all, every) and *gen* (from γί-γ(ε)ν-ομαι, to become). Only the meaning of this latter [i.e., *gen*] comes into consideration here; just the basic idea – [namely,] that a trait in the developing organism can be determined or is influenced by "something" in the gametes – should find expression. No hypothesis about the nature of this "something" should be postulated or supported by it. For that reason it seems simplest to use in isolation the last syllable *gen* from Darwin's well-known word, which alone is of interest to us, in order to replace, with it, the poor, ambiguous word *Anlage*. Thus we will say simply "gene" and "genes" for "pangene" and "pangenes". The word gene is completely free of any hypothesis; it expresses only the established fact that in any case many traits of the organism are determined by specific, separable, and thus independent "conditions", "foundations", "plans" – in short, precisely what we want to call genes.)

37. Bateson W (1906). "The progress of genetic research" (https://www.biodiversitylibrary.org/ite
m/206746#page/129/mode/1up). In Wilks W (ed.). *Report of the Third International
Conference 1906 on Genetics*. London, England: Royal Horticultural Society. pp. 90–97.
Archived (https://web.archive.org/web/20230627221528/https://www.biodiversitylibrary.org/it
em/206746#page/129/mode/1up) from the original on 27 June 2023. Retrieved 17 June
2021. "... the science itself [i.e. the study of the breeding and hybridisation of plants] is still
nameless, and we can only describe our pursuit by cumbrous and often misleading
periphrasis. To meet this difficulty I suggest for the consideration of this Congress the term
Genetics, which sufficiently indicates that our labors are devoted to the elucidation of the
phenomena of heredity and variation: in other words, to the physiology of Descent, with
implied bearing on the theoretical problems of the evolutionist and the systematist, and
application to the practical problems of breeders, whether of animals or plants."

38. Avery OT, Macleod CM, McCarty M (February 1944). "Studies on the Chemical Nature of
the Substance Inducing Transformation of Pneumococcal Types: Induction of
Transformation by a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type III"
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2135445). *The Journal of Experimental
Medicine*. **79** (2): 137–58. doi:10.1084/jem.79.2.137 (https://doi.org/10.1084%2Fjem.79.2.13
7). PMC 2135445 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2135445).
PMID 19871359 (https://pubmed.ncbi.nlm.nih.gov/19871359). Reprint: Avery OT, MacLeod
CM, McCarty M (February 1979). "Studies on the chemical nature of the substance inducing
transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic
acid fraction isolated from pneumococcus type III" (https://www.ncbi.nlm.nih.gov/pmc/article
s/PMC2184805). *The Journal of Experimental Medicine*. **149** (2): 297–326.
doi:10.1084/jem.149.2.297 (https://doi.org/10.1084%2Fjem.149.2.297). PMC 2184805 (http
s://www.ncbi.nlm.nih.gov/pmc/articles/PMC2184805). PMID 33226 (https://pubmed.ncbi.nl
m.nih.gov/33226).

39. Hershey AD, Chase M (May 1952). "Independent functions of viral protein and nucleic acid
in growth of bacteriophage" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2147348). *The
Journal of General Physiology*. **36** (1): 39–56. doi:10.1085/jgp.36.1.39 (https://doi.org/10.10
85%2Fjgp.36.1.39). PMC 2147348 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC214734
8). PMID 12981234 (https://pubmed.ncbi.nlm.nih.gov/12981234).

40. Judson H (1979). *The Eighth Day of Creation: Makers of the Revolution in Biology*. Cold
Spring Harbor Laboratory Press. pp. 51–169. ISBN 978-0-87969-477-7.

41. Watson JD, Crick FH (April 1953). "Molecular Structure of Nucleic Acids: A Structure for
Deoxyribose Nucleic Acid". *Nature*. **171** (4356): 737–8. Bibcode:1953Natur.171..737W (http
s://ui.adsabs.harvard.edu/abs/1953Natur.171..737W). doi:10.1038/171737a0 (https://doi.or
g/10.1038%2F171737a0). PMID 13054692 (https://pubmed.ncbi.nlm.nih.gov/13054692).

42. Benzer S (June 1955). "Fine Structure of a Genetic Region in Bacteriophage" (https://www.n
cbi.nlm.nih.gov/pmc/articles/PMC528093). *Proceedings of the National Academy of
Sciences of the United States of America*. **41** (6): 344–54. Bibcode:1955PNAS...41..344B (h
ttps://ui.adsabs.harvard.edu/abs/1955PNAS...41..344B). doi:10.1073/pnas.41.6.344 (https://
doi.org/10.1073%2Fpnas.41.6.344). PMC 528093 (https://www.ncbi.nlm.nih.gov/pmc/article
s/PMC528093). PMID 16589677 (https://pubmed.ncbi.nlm.nih.gov/16589677).

43. Benzer S (November 1959). "On the Topology of the Genetic Fine Structure" (https://www.nc
bi.nlm.nih.gov/pmc/articles/PMC222769). *Proceedings of the National Academy of Sciences
of the United States of America*. **45** (11): 1607–20. Bibcode:1959PNAS...45.1607B (https://u
i.adsabs.harvard.edu/abs/1959PNAS...45.1607B). doi:10.1073/pnas.45.11.1607 (https://doi.
org/10.1073%2Fpnas.45.11.1607). PMC 222769 (https://www.ncbi.nlm.nih.gov/pmc/articles/
PMC222769). PMID 16590553 (https://pubmed.ncbi.nlm.nih.gov/16590553).

44. Min Jou W, Haegeman G, Ysebaert M, Fiers W (May 1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein". *Nature*. **237** (5350): 82–8. Bibcode:1972Natur.237...82J (https://ui.adsabs.harvard.edu/abs/1972Natur.237...82J). doi:10.1038/237082a0 (https://doi.org/10.1038%2F237082a0). PMID 4555447 (https://pubmed.ncbi.nlm.nih.gov/4555447). S2CID 4153893 (https://api.semanticscholar.org/CorpusID:4153893).

45. Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765). *Proceedings of the National Academy of Sciences of the United States of America*. **74** (12): 5463–7. Bibcode:1977PNAS...74.5463S (https://ui.adsabs.harvard.edu/abs/1977PNAS...74.5463S). doi:10.1073/pnas.74.12.5463 (https://doi.org/10.1073%2Fpnas.74.12.5463). PMC 431765 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765). PMID 271968 (https://pubmed.ncbi.nlm.nih.gov/271968).

46. Adams JU (2008). "DNA Sequencing Technologies" (https://www.nature.com/scitable/topicpage/dna-sequencing-technologies-690/). *Nature Education Knowledge*. SciTable. **1** (1). Nature Publishing Group: 193. Archived (https://web.archive.org/web/20081019192918/https://www.nature.com/scitable/topicpage/dna-sequencing-technologies-690/) from the original on 19 October 2008. Retrieved 2 May 2024.

47. Huxley J (1942). *Evolution: the Modern Synthesis*. Cambridge, Massachusetts: MIT Press. ISBN 978-0262513661.

48. Williams GC (2001). *Adaptation and Natural Selection a Critique of Some Current Evolutionary Thought* (Online ed.). Princeton: Princeton University Press. ISBN 9781400820108.

49. Dawkins R (1989). *The extended phenotype* (Paperback ed.). Oxford: Oxford University Press. ISBN 978-0-19-286088-0.

50. Duret L (2008). "Neutral Theory: The Null Hypothesis of Molecular Evolution" (https://www.nature.com/scitable/topicpage/neutral-theory-the-null-hypothesis-of-molecular-839/). *Nature Education*. **1**: 218.

51. Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter (2002). *Molecular Biology of the Cell* (https://www.ncbi.nlm.nih.gov/books/NBK21054/) (4th ed.). Garland Science. ISBN 978-0-8153-3218-3.

52. Stryer L, Berg JM, Tymoczko JL (2002). *Biochemistry* (https://www.ncbi.nlm.nih.gov/books/NBK21154/) (5th ed.). San Francisco: W.H. Freeman. ISBN 978-0-7167-4955-4. Archived (https://web.archive.org/web/20140311010455/http://www.ncbi.nlm.nih.gov/books/NBK21154/) from the original on 11 March 2014. Retrieved 31 August 2017.

53. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, et al. (May 2005). "Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1084335). *PLOS Biology*. **3** (5): e157. doi:10.1371/journal.pbio.0030157 (https://doi.org/10.1371%2Fjournal.pbio.0030157). PMC 1084335 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1084335). PMID 15839726 (https://pubmed.ncbi.nlm.nih.gov/15839726).

54. Braig M, Schmitt CA (March 2006). "Oncogene-induced senescence: putting the brakes on tumor development" (https://doi.org/10.1158%2F0008-5472.CAN-05-4006). *Cancer Research*. **66** (6): 2881–4. doi:10.1158/0008-5472.CAN-05-4006 (https://doi.org/10.1158%2F0008-5472.CAN-05-4006). PMID 16540631 (https://pubmed.ncbi.nlm.nih.gov/16540631).

55. Bennett PM (March 2008). "Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2268074). *British Journal of Pharmacology*. **153** (Suppl 1): S347-57. doi:10.1038/sj.bjp.0707607 (https://doi.org/10.1038%2Fsj.bjp.0707607). PMC 2268074 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2268074). PMID 18193080 (https://pubmed.ncbi.nlm.nih.gov/18193080).

56. International Human Genome Sequencing Consortium (October 2004). "Finishing the euchromatic sequence of the human genome" (https://doi.org/10.1038%2Fnature03001). *Nature*. **431** (7011): 931–45. Bibcode:2004Natur.431..931H (https://ui.adsabs.harvard.edu/abs/2004Natur.431..931H). doi:10.1038/nature03001 (https://doi.org/10.1038%2Fnature03001). PMID 15496913 (https://pubmed.ncbi.nlm.nih.gov/15496913).

57. Shafee, Thomas; Lowe, Rohan (2017). "Eukaryotic and prokaryotic gene structure" (https://doi.org/10.15347%2Fwjm%2F2017.002). *WikiJournal of Medicine*. **4** (1). doi:10.15347/wjm/2017.002 (https://doi.org/10.15347%2Fwjm%2F2017.002). ISSN 2002-4436 (https://search.worldcat.org/issn/2002-4436).

58. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (July 2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq". *Nature Methods*. **5** (7): 621–8. doi:10.1038/nmeth.1226 (https://doi.org/10.1038%2Fnmeth.1226). PMID 18516045 (https://pubmed.ncbi.nlm.nih.gov/18516045). S2CID 205418589 (https://api.semanticscholar.org/CorpusID:205418589).

59. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G (April 2013). "Enhancers: five essential questions" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445073). *Nature Reviews. Genetics*. **14** (4): 288–95. doi:10.1038/nrg3458 (https://doi.org/10.1038%2Fnrg3458). PMC 4445073 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445073). PMID 23503198 (https://pubmed.ncbi.nlm.nih.gov/23503198).

60. Maston GA, Evans SK, Green MR (2006). "Transcriptional regulatory elements in the human genome" (https://doi.org/10.1146%2Fannurev.genom.7.080505.115623). *Annual Review of Genomics and Human Genetics*. **7**: 29–59. doi:10.1146/annurev.genom.7.080505.115623 (https://doi.org/10.1146%2Fannurev.genom.7.080505.115623). PMID 16719718 (https://pubmed.ncbi.nlm.nih.gov/16719718).

61. Mignone F, Gissi C, Liuni S, Pesole G (28 February 2002). "Untranslated regions of mRNAs" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139023). *Genome Biology*. **3** (3): REVIEWS0004. doi:10.1186/gb-2002-3-3-reviews0004 (https://doi.org/10.1186%2Fgb-2002-3-3-reviews0004). PMC 139023 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC139023). PMID 11897027 (https://pubmed.ncbi.nlm.nih.gov/11897027).

62. Bicknell AA, Cenik C, Chua HN, Roth FP, Moore MJ (December 2012). "Introns in UTRs: why we should stop ignoring them" (https://doi.org/10.1002%2Fbies.201200073). *BioEssays*. **34** (12): 1025–34. doi:10.1002/bies.201200073 (https://doi.org/10.1002%2Fbies.201200073). PMID 23108796 (https://pubmed.ncbi.nlm.nih.gov/23108796). S2CID 5808466 (https://api.semanticscholar.org/CorpusID:5808466).

63. Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J (June 2000). "Operons in Escherichia coli: genomic analyses and predictions" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC18690). *Proceedings of the National Academy of Sciences of the United States of America*. **97** (12): 6652–7. Bibcode:2000PNAS...97.6652S (https://ui.adsabs.harvard.edu/abs/2000PNAS...97.6652S). doi:10.1073/pnas.110147297 (https://doi.org/10.1073%2Fpnas.110147297). PMC 18690 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC18690). PMID 10823905 (https://pubmed.ncbi.nlm.nih.gov/10823905).

64. Blumenthal T (November 2004). "Operons in eukaryotes" (https://doi.org/10.1093%2Fbfgp%2F3.3.199). *Briefings in Functional Genomics & Proteomics*. **3** (3): 199–211. doi:10.1093/bfgp/3.3.199 (https://doi.org/10.1093%2Fbfgp%2F3.3.199). PMID 15642184 (https://pubmed.ncbi.nlm.nih.gov/15642184).

65. Jacob F, Monod J (June 1961). "Genetic regulatory mechanisms in the synthesis of proteins". *Journal of Molecular Biology*. **3** (3): 318–56. doi:10.1016/S0022-2836(61)80072-7 (https://doi.org/10.1016%2FS0022-2836%2861%2980072-7). PMID 13718526 (https://pubmed.ncbi.nlm.nih.gov/13718526). S2CID 19804795 (https://api.semanticscholar.org/CorpusID:19804795).

66. Pozzoli U, Menozzi G, Comi GP, Cagliani R, Bresolin N, Sironi M (January 2007). "Intron size in mammals: complexity comes to terms with economy". *Trends in Genetics*. **23** (1): 20–24. doi:10.1016/j.tig.2006.10.003 (https://doi.org/10.1016%2Fj.tig.2006.10.003). PMID 17070957 (https://pubmed.ncbi.nlm.nih.gov/17070957).

67. Marais G, Nouvellet P, Keightley PD, Charlesworth B (May 2005). "Intron size and exon evolution in Drosophila" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1449718). *Genetics*. **170** (1): 481–485. doi:10.1534/genetics.104.037333 (https://doi.org/10.1534%2Fgenetics.104.037333). PMC 1449718 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1449718). PMID 15781704 (https://pubmed.ncbi.nlm.nih.gov/15781704).

68. Kumar A (September 2009). "An overview of nested genes in eukaryotic genomes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2747821). *Eukaryotic Cell*. **8** (9): 1321–1329. doi:10.1128/EC.00143-09 (https://doi.org/10.1128%2FEC.00143-09). PMC 2747821 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2747821). PMID 19542305 (https://pubmed.ncbi.nlm.nih.gov/19542305)..

69. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA (June 2005). "Interchromosomal associations between alternatively expressed loci". *Nature*. **435** (7042): 637–645. Bibcode:2005Natur.435..637S (https://ui.adsabs.harvard.edu/abs/2005Natur.435..637S). doi:10.1038/nature03574 (https://doi.org/10.1038%2Fnature03574). PMID 15880101 (https://pubmed.ncbi.nlm.nih.gov/15880101). S2CID 1755326 (https://api.semanticscholar.org/CorpusID:1755326).

70. Williams A, Spilianakis CG, Flavell RA (April 2010). "Interchromosomal association and gene regulation in trans" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2865229). *Trends in Genetics*. **26** (4): 188–197. doi:10.1016/j.tig.2010.01.007 (https://doi.org/10.1016%2Fj.tig.2010.01.007). PMC 2865229 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2865229). PMID 20236724 (https://pubmed.ncbi.nlm.nih.gov/20236724).

71. Lei Q, Li C, Zuo Z, Huang C, Cheng H, Zhou R (March 2016). "Evolutionary Insights into RNA trans-Splicing in Vertebrates" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824033). *Genome Biology and Evolution*. **8** (3): 562–577. doi:10.1093/gbe/evw025 (https://doi.org/10.1093%2Fgbe%2Fevw025). PMC 4824033 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824033). PMID 26966239 (https://pubmed.ncbi.nlm.nih.gov/26966239).

72. Wright BW, Molloy MP, Jaschke PR (March 2022). "Overlapping genes in natural and engineered genomes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8490965). *Nature Reviews. Genetics*. **23** (3): 154–168. doi:10.1038/s41576-021-00417-w (https://doi.org/10.1038%2Fs41576-021-00417-w). PMC 8490965 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8490965). PMID 34611352 (https://pubmed.ncbi.nlm.nih.gov/34611352).

73. Eddy SR (December 2001). "Non-coding RNA genes and the modern RNA world". *Nature Reviews. Genetics*. **2** (12): 919–29. doi:10.1038/35103511 (https://doi.org/10.1038%2F35103511). PMID 11733745 (https://pubmed.ncbi.nlm.nih.gov/11733745). S2CID 18347629 (https://api.semanticscholar.org/CorpusID:18347629).

74. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ (December 1961). "General nature of the genetic code for proteins". *Nature*. **192** (4809): 1227–32. Bibcode:1961Natur.192.1227C (https://ui.adsabs.harvard.edu/abs/1961Natur.192.1227C). doi:10.1038/1921227a0 (https://doi.org/10.1038%2F1921227a0). PMID 13882203 (https://pubmed.ncbi.nlm.nih.gov/13882203). S2CID 4276146 (https://api.semanticscholar.org/CorpusID:4276146).

75. Crick FH (October 1962). "The genetic code". *Scientific American*. **207** (4). WH Freeman and Company: 66–74. Bibcode:1962SciAm.207d..66C (https://ui.adsabs.harvard.edu/abs/1962SciAm.207d..66C). doi:10.1038/scientificamerican1062-66 (https://doi.org/10.1038%2Fscientificamerican1062-66). PMID 13882204 (https://pubmed.ncbi.nlm.nih.gov/13882204).

76. Woodson SA (May 1998). "Ironing out the kinks: splicing and translation in bacteria" (https://doi.org/10.1101%2Fgad.12.9.1243). *Genes & Development*. **12** (9): 1243–7. doi:10.1101/gad.12.9.1243 (https://doi.org/10.1101%2Fgad.12.9.1243). PMID 9573040 (https://pubmed.ncbi.nlm.nih.gov/9573040).

77. Jacob F, Monod J (June 1961). "Genetic regulatory mechanisms in the synthesis of proteins". *Journal of Molecular Biology*. **3** (3): 318–56. doi:10.1016/S0022-2836(61)80072-7 (https://doi.org/10.1016%2FS0022-2836%2861%2980072-7). PMID 13718526 (https://pubmed.ncbi.nlm.nih.gov/13718526). S2CID 19804795 (https://api.semanticscholar.org/CorpusID:19804795).

78. Koonin EV, Dolja VV (January 1993). "Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences". *Critical Reviews in Biochemistry and Molecular Biology*. **28** (5): 375–430. doi:10.3109/10409239309078440 (https://doi.org/10.3109%2F10409239309078440). PMID 8269709 (https://pubmed.ncbi.nlm.nih.gov/8269709).

79. Domingo E (2001). "RNA Virus Genomes". *eLS*. doi:10.1002/9780470015902.a0001488.pub2 (https://doi.org/10.1002%2F9780470015902.a0001488.pub2). ISBN 978-0470016176.

80. Domingo E, Escarmís C, Sevilla N, Moya A, Elena SF, Quer J, et al. (June 1996). "Basic concepts in RNA virus evolution" (https://doi.org/10.1096%2Ffasebj.10.8.8666162). *FASEB Journal*. **10** (8): 859–64. doi:10.1096/fasebj.10.8.8666162 (https://doi.org/10.1096%2Ffasebj.10.8.8666162). PMID 8666162 (https://pubmed.ncbi.nlm.nih.gov/8666162). S2CID 20865732 (https://api.semanticscholar.org/CorpusID:20865732).

81. Miko I (2008). "Gregor Mendel and the Principles of Inheritance" (https://www.nature.com/scitable/topicpage/gregor-mendel-and-the-principles-of-inheritance-593/). *Nature Education Knowledge*. SciTable. **1** (1). Nature Publishing Group: 134. Archived (https://web.archive.org/web/20190719224056/http://www.nature.com/scitable/topicpage/gregor-mendel-and-the-principles-of-inheritance-593) from the original on 19 July 2019. Retrieved 2 May 2024.

82. Chial H (2008). "Mendelian Genetics: Patterns of Inheritance and Single-Gene Disorders" (https://www.nature.com/scitable/topicpage/mendelian-genetics-patterns-of-inheritance-and-single-966/). *Nature Education Knowledge*. SciTable. **1** (1). Nature Publishing Group: 63. Archived (https://web.archive.org/web/20190331174639/http://www.nature.com/scitable/topicpage/mendelian-genetics-patterns-of-inheritance-and-single-966) from the original on 31 March 2019. Retrieved 2 May 2024.

83. McCarthy D, Minner C, Bernstein H, Bernstein C (October 1976). "DNA elongation rates and growing point distributions of wild-type phage T4 and a DNA-delay amber mutant". *Journal of Molecular Biology*. **106** (4): 963–81. doi:10.1016/0022-2836(76)90346-6 (https://doi.org/10.1016%2F0022-2836%2876%2990346-6). PMID 789903 (https://pubmed.ncbi.nlm.nih.gov/789903).

84. Lobo I, Shaw K (2008). "Discovery and Types of Genetic Linkage" (https://www.nature.com/scitable/topicpage/discovery-and-types-of-genetic-linkage-500/). *Nature Education Knowledge*. SciTable. **1** (1). Nature Publishing Group: 139. Archived (https://web.archive.org/web/20200512071347/https://www.nature.com/scitable/topicpage/discovery-and-types-of-genetic-linkage-500/) from the original on 12 May 2020. Retrieved 2 May 2024.

85. Ridley, M. (2006). *Genome*. New York, NY: Harper Perennial. ISBN 0-06-019497-9

86. Banerjee S, Bhandary P, Woodhouse M, Sen TZ, Wise RP, Andorf CM (April 2021). "FINDER: an automated software package to annotate eukaryotic genes from RNA-Seq data and associated protein sequences" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8056616). *BMC Bioinformatics*. **44** (9): e89. doi:10.1186/s12859-021-04120-9 (https://doi.org/10.1186%2Fs12859-021-04120-9). PMC 8056616 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8056616). PMID 33879057 (https://pubmed.ncbi.nlm.nih.gov/33879057).

87. Watson, JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. (2004). "Ch9-10", Molecular Biology of the Gene, 5th ed., Peason Benjamin Cummings; CSHL Press.

88. "Integr8 – A.thaliana Genome Statistics" (http://www.ebi.ac.uk/integr8/OrganismStatsAction.do;jsessionid=08E9058B5B688A4F7FF7D161CB9E36A4?orgProteomeId=3). Archived (https://web.archive.org/web/20130523210340/http://www.ebi.ac.uk/integr8/OrganismStatsAction.do;jsessionid=08E9058B5B688A4F7FF7D161CB9E36A4?orgProteomeId=3) from the original on 23 May 2013. Retrieved 26 April 2015.

89. "Understanding the Basics" (http://web.ornl.gov/sci/techresources/Human_Genome/project/info.shtml). *The Human Genome Project*. Retrieved 26 April 2015.

90. "WS227 Release Letter" (https://archive.today/20131128221823/http://www.wormbase.org/wiki/index.php/WS227). WormBase. 10 August 2011. Archived from the original (http://www.wormbase.org/wiki/index.php/WS227) on 28 November 2013. Retrieved 19 November 2013.

91. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. (April 2002). "A draft sequence of the rice genome (Oryza sativa L. ssp. indica)". *Science*. **296** (5565): 79–92. Bibcode:2002Sci...296...79Y (https://ui.adsabs.harvard.edu/abs/2002Sci...296...79Y). doi:10.1126/science.1068037 (https://doi.org/10.1126%2Fscience.1068037). PMID 11935017 (https://pubmed.ncbi.nlm.nih.gov/11935017). S2CID 208529258 (https://api.semanticscholar.org/CorpusID:208529258).

92. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. (April 1981). "Sequence and organization of the human mitochondrial genome". *Nature*. **290** (5806): 457–65. Bibcode:1981Natur.290..457A (https://ui.adsabs.harvard.edu/abs/1981Natur.290..457A). doi:10.1038/290457a0 (https://doi.org/10.1038%2F290457a0). PMID 7219534 (https://pubmed.ncbi.nlm.nih.gov/7219534). S2CID 4355527 (https://api.semanticscholar.org/CorpusID:4355527).

93. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. (March 2000). "The genome sequence of Drosophila melanogaster". *Science*. **287** (5461): 2185–95. Bibcode:2000Sci...287.2185. (https://ui.adsabs.harvard.edu/abs/2000Sci...287.2185.). CiteSeerX 10.1.1.549.8639 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.549.8639). doi:10.1126/science.287.5461.2185 (https://doi.org/10.1126%2Fscience.287.5461.2185). PMID 10731132 (https://pubmed.ncbi.nlm.nih.gov/10731132).

94. Pertea M, Salzberg SL (2010). "Between a chicken and a grape: estimating the number of human genes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2898077). *Genome Biology*. **11** (5): 206. doi:10.1186/gb-2010-11-5-206 (https://doi.org/10.1186%2Fgb-2010-11-5-206). PMC 2898077 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2898077). PMID 20441615 (https://pubmed.ncbi.nlm.nih.gov/20441615).

95. Belyi VA, Levine AJ, Skalka AM (December 2010). "Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976387). *Journal of Virology*. **84** (23): 12458–62. doi:10.1128/JVI.01789-10 (https://doi.org/10.1128%2FJVI.01789-10). PMC 2976387 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976387). PMID 20861255 (https://pubmed.ncbi.nlm.nih.gov/20861255).

96. Flores R, Di Serio F, Hernández C (February 1997). "Viroids: The Noncoding Genomes". *Seminars in Virology*. **8** (1): 65–73. doi:10.1006/smvy.1997.0107 (https://doi.org/10.1006%2Fsmvy.1997.0107).

97. Zonneveld BJ (2010). "New Record Holders for Maximum Genome Size in Eudicots and Monocots" (https://doi.org/10.1155%2F2010%2F527357). *Journal of Botany*. **2010**: 1–4. doi:10.1155/2010/527357 (https://doi.org/10.1155%2F2010%2F527357).

98. Perez-Iratxeta C, Palidwor G, Andrade-Navarro MA (December 2007). "Towards completion of the Earth's proteome" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2267224). *EMBO Reports*. **8** (12): 1135–41. doi:10.1038/sj.embor.7401117 (https://doi.org/10.1038%2Fsj.embor.7401117). PMC 2267224 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2267224). PMID 18059312 (https://pubmed.ncbi.nlm.nih.gov/18059312).

99. Muller HJ (1966). "The gene material as the initiator and the organizing basis of life". *American Naturalist*. **100** (915): 493–517. doi:10.1086/282445 (https://doi.org/10.1086%2F282445). JSTOR 2459205 (https://www.jstor.org/stable/2459205).

100. Ohno S (1972). "So much "junk" DNA in our genome". *Brookhaven Symposia in Biology*. **23**: 366–370. PMID 5065367 (https://pubmed.ncbi.nlm.nih.gov/5065367).

101. Hatje K, Mühlhausen S, Simm D, Killmar M (2019). "The Protein-Coding Human Genome: Annotating High-Hanging Fruits" (https://doi.org/10.1002%2Fbies.201900066). *BioEssays*. **41** (11): 1900066. doi:10.1002/bies.201900066 (https://doi.org/10.1002%2Fbies.201900066). PMID 31544971 (https://pubmed.ncbi.nlm.nih.gov/31544971). S2CID 202732556 (https://api.semanticscholar.org/CorpusID:202732556).

102. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, et al. (October 1996). "A gene map of the human genome". *Science*. **274** (5287): 540–6. Bibcode:1996Sci...274..540S (https://ui.adsabs.harvard.edu/abs/1996Sci...274..540S). doi:10.1126/science.274.5287.540 (https://doi.org/10.1126%2Fscience.274.5287.540). PMID 8849440 (https://pubmed.ncbi.nlm.nih.gov/8849440). S2CID 22619 (https://api.semanticscholar.org/CorpusID:22619).

103. Chi KR (October 2016). "The dark side of the human genome" (https://doi.org/10.1038%2F538275a). *Nature*. **538** (7624): 275–277. Bibcode:2016Natur.538..275C (https://ui.adsabs.harvard.edu/abs/2016Natur.538..275C). doi:10.1038/538275a (https://doi.org/10.1038%2F538275a). PMID 27734873 (https://pubmed.ncbi.nlm.nih.gov/27734873).

104. "Human assembly and gene annotation" (https://useast.ensembl.org/Homo_sapiens/Info/Annotation). *Ensembl*. 2022. Archived (https://web.archive.org/web/20231103222302/https://useast.ensembl.org/Homo_sapiens/Info/Annotation) from the original on 3 November 2023. Retrieved 28 February 2023.

105. Hutchison CA, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, et al. (March 2016). "Design and synthesis of a minimal bacterial genome" (https://doi.org/10.1126%2Fscience.aad6253). *Science*. **351** (6280): aad6253. Bibcode:2016Sci...351.....H (https://ui.adsabs.harvard.edu/abs/2016Sci...351.....H). doi:10.1126/science.aad6253 (https://doi.org/10.1126%2Fscience.aad6253). PMID 27013737 (https://pubmed.ncbi.nlm.nih.gov/27013737).

106. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, et al. (January 2006). "Essential genes of a minimal bacterium" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1324956). *Proceedings of the National Academy of Sciences of the United States of America*. **103** (2): 425–30. Bibcode:2006PNAS..103..425G (https://ui.adsabs.harvard.edu/abs/2006PNAS..103..425G). doi:10.1073/pnas.0510013103 (https://doi.org/10.1073%2Fpnas.0510013103). PMC 1324956 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1324956). PMID 16407165 (https://pubmed.ncbi.nlm.nih.gov/16407165).

107. Gerdes SY, Scholle MD, Campbell JW, Balázsi G, Ravasz E, Daugherty MD, et al. (October 2003). "Experimental determination and system level analysis of essential genes in Escherichia coli MG1655" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC193955). *Journal of Bacteriology*. **185** (19): 5673–84. doi:10.1128/jb.185.19.5673-5684.2003 (https://doi.org/10.1128%2Fjb.185.19.5673-5684.2003). PMC 193955 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC193955). PMID 13129938 (https://pubmed.ncbi.nlm.nih.gov/13129938).

108. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1681482). *Molecular Systems Biology*. **2**: 2006.0008. doi:10.1038/msb4100050 (https://doi.org/10.1038%2Fmsb4100050). PMC 1681482 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1681482). PMID 16738554 (https://pubmed.ncbi.nlm.nih.gov/16738554).

109. Juhas M, Reuß DR, Zhu B, Commichau FM (November 2014). "Bacillus subtilis and Escherichia coli essential genes and minimal cell factories after one decade of genome engineering" (https://doi.org/10.1099%2Fmic.0.079376-0). *Microbiology*. **160** (Pt 11): 2341–2351. doi:10.1099/mic.0.079376-0 (https://doi.org/10.1099%2Fmic.0.079376-0). PMID 25092907 (https://pubmed.ncbi.nlm.nih.gov/25092907).

110. Tu Z, Wang L, Xu M, Zhou X, Chen T, Sun F (February 2006). "Further understanding human disease genes by comparing with housekeeping genes and other genes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1397819). *BMC Genomics*. **7**: 31. doi:10.1186/1471-2164-7-31 (https://doi.org/10.1186%2F1471-2164-7-31). PMC 1397819 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1397819). PMID 16504025 (https://pubmed.ncbi.nlm.nih.gov/16504025).

111. Georgi B, Voight BF, Bućan M (May 2013). "From mouse to human: evolutionary genomics analysis of human orthologs of essential genes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3649967). *PLOS Genetics*. **9** (5): e1003484. doi:10.1371/journal.pgen.1003484 (https://doi.org/10.1371%2Fjournal.pgen.1003484). PMC 3649967 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3649967). PMID 23675308 (https://pubmed.ncbi.nlm.nih.gov/23675308).

112. Eisenberg E, Levanon EY (October 2013). "Human housekeeping genes, revisited". *Trends in Genetics*. **29** (10): 569–74. doi:10.1016/j.tig.2013.05.010 (https://doi.org/10.1016%2Fj.tig.2013.05.010). PMID 23810203 (https://pubmed.ncbi.nlm.nih.gov/23810203).

113. Amsterdam A, Hopkins N (September 2006). "Mutagenesis strategies in zebrafish for identifying genes involved in development and disease". *Trends in Genetics*. **22** (9): 473–8. doi:10.1016/j.tig.2006.06.011 (https://doi.org/10.1016%2Fj.tig.2006.06.011). PMID 16844256 (https://pubmed.ncbi.nlm.nih.gov/16844256).

114. "About the HGNC" (https://web.archive.org/web/20230326032945/https://www.genenames.org/about/overview). *HGNC Database of Human Gene Names*. HUGO Gene Nomenclature Committee. Archived from the original (https://www.genenames.org/about/overview) on 26 March 2023. Retrieved 14 May 2015.

115. Cohen SN, Chang AC (May 1973). "Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC433482). *Proceedings of the National Academy of Sciences of the United States of America*. **70** (5): 1293–7. Bibcode:1973PNAS...70.1293C (https://ui.adsabs.harvard.edu/abs/1973PNAS...70.1293C). doi:10.1073/pnas.70.5.1293 (https://doi.org/10.1073%2Fpnas.70.5.1293). PMC 433482 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC433482). PMID 4576014 (https://pubmed.ncbi.nlm.nih.gov/4576014).

116. Esvelt KM, Wang HH (2013). "Genome-scale engineering for systems and synthetic biology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3564264). *Molecular Systems Biology*. **9** (1): 641. doi:10.1038/msb.2012.66 (https://doi.org/10.1038%2Fmsb.2012.66). PMC 3564264 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3564264). PMID 23340847 (https://pubmed.ncbi.nlm.nih.gov/23340847).

117. Tan WS, Carlson DF, Walton MW, Fahrenkrug SC, Hackett PB (2012). "Precision editing of large animal genomes". *Advances in Genetics Volume 80*. Vol. 80. pp. 37–97. doi:10.1016/B978-0-12-404742-6.00002-8 (https://doi.org/10.1016%2FB978-0-12-404742-6.00002-8). ISBN 9780124047426. PMC 3683964 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3683964). PMID 23084873 (https://pubmed.ncbi.nlm.nih.gov/23084873).

118. Puchta H, Fauser F (2013). "Gene targeting in plants: 25 years later" (https://doi.org/10.1387%2Fijdb.130194hp). *The International Journal of Developmental Biology*. **57** (6–8): 629–37. doi:10.1387/ijdb.130194hp (https://doi.org/10.1387%2Fijdb.130194hp). PMID 24166445 (https://pubmed.ncbi.nlm.nih.gov/24166445).

119. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (November 2013). "Genome engineering using the CRISPR-Cas9 system" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969860). *Nature Protocols*. **8** (11): 2281–2308. doi:10.1038/nprot.2013.143 (https://doi.org/10.1038%2Fnprot.2013.143). PMC 3969860 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969860). PMID 24157548 (https://pubmed.ncbi.nlm.nih.gov/24157548).

120. Kittleson JT, Wu GC, Anderson JC (August 2012). "Successes and failures in modular genetic engineering". *Current Opinion in Chemical Biology*. **16** (3–4): 329–36. doi:10.1016/j.cbpa.2012.06.009 (https://doi.org/10.1016%2Fj.cbpa.2012.06.009). PMID 22818777 (https://pubmed.ncbi.nlm.nih.gov/22818777).

121. Berg P, Mertz JE (January 2010). "Personal reflections on the origins and emergence of recombinant DNA technology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815933). *Genetics*. **184** (1): 9–17. doi:10.1534/genetics.109.112144 (https://doi.org/10.1534%2Fgenetics.109.112144). PMC 2815933 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815933). PMID 20061565 (https://pubmed.ncbi.nlm.nih.gov/20061565).

122. Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, et al. (September 2004). "The knockout mouse project" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2716027). *Nature Genetics*. **36** (9): 921–4. doi:10.1038/ng0904-921 (https://doi.org/10.1038%2Fng0904-921). PMC 2716027 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2716027). PMID 15340423 (https://pubmed.ncbi.nlm.nih.gov/15340423).

123. Guan C, Ye C, Yang X, Gao J (February 2010). "A review of current large-scale mouse knockout efforts". *Genesis*. **48** (2): 73–85. doi:10.1002/dvg.20594 (https://doi.org/10.1002%2Fdvg.20594). PMID 20095055 (https://pubmed.ncbi.nlm.nih.gov/20095055). S2CID 34470273 (https://api.semanticscholar.org/CorpusID:34470273).

124. Deng C (October 2007). "In celebration of Dr. Mario R. Capecchi's Nobel Prize" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2043165). *International Journal of Biological Sciences*. **3** (7): 417–9. doi:10.7150/ijbs.3.417 (https://doi.org/10.7150%2Fijbs.3.417). PMC 2043165 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2043165). PMID 17998949 (https://pubmed.ncbi.nlm.nih.gov/17998949).

Sources

Main textbook

- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002). *Molecular Biology of the Cell* (https://www.ncbi.nlm.nih.gov/books/NBK21054/) (Fourth ed.). New York: Garland Science. ISBN 978-0-8153-3218-3. – A molecular biology textbook available free online through NCBI Bookshelf.

Referenced chapters of *Molecular Biology of the Cell* (https://www.ncbi.nlm.nih.gov/books/NBK21054/)

Glossary (https://www.ncbi.nlm.nih.gov/books/NBK21052/)

Ch 1: Cells and genomes (https://www.ncbi.nlm.nih.gov/books/NBK26864/)

1.1: The Universal Features of Cells on Earth (https://www.ncbi.nlm.nih.gov/books/NBK26864/)

Ch 2: Cell Chemistry and Biosynthesis (https://www.ncbi.nlm.nih.gov/books/NBK21068/)

2.1: The Chemical Components of a Cell (https://www.ncbi.nlm.nih.gov/books/NBK26883/)

Ch 3: Proteins (https://www.ncbi.nlm.nih.gov/books/NBK26830/)

Ch 4: DNA and Chromosomes (https://www.ncbi.nlm.nih.gov/books/NBK21074/)

4.1: The Structure and Function of DNA (https://www.ncbi.nlm.nih.gov/books/NBK26821/)
4.2: Chromosomal DNA and Its Packaging in the Chromatin Fiber (https://www.ncbi.nlm.nih.gov/books/NBK26834/)

Ch 5: DNA Replication, Repair, and Recombination (https://www.ncbi.nlm.nih.gov/books/NBK21064/)

Further reading

- Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R (2013). *Molecular Biology of the Gene* (7th ed.). Benjamin Cummings. ISBN 978-0-321-90537-6.
- Dawkins R (1990). *The Selfish Gene*. Oxford University Press. ISBN 978-0-19-286092-7.
- Ridley M (1999). *Genome: The Autobiography of a Species in 23 Chapters*. Fourth Estate. ISBN 978-0-00-763573-3.
- Brown T (2002). *Genomes* (https://www.ncbi.nlm.nih.gov/books/NBK21128) (2nd ed.). New York: Wiley-Liss. ISBN 978-0-471-25046-3. PMID 20821850 (https://pubmed.ncbi.nlm.nih.gov/20821850). Archived (https://web.archive.org/web/20201106150314/https://www.ncbi.nlm.nih.gov/books/NBK21128/) from the original on 6 November 2020. Retrieved 31 August 2017.

External links

- Comparative Toxicogenomics Database (https://ctdbase.org/)
- DNA From The Beginning – a primer on genes and DNA (https://www.dnaftb.org/)
- Gene – a searchable database of genes (https://www.ncbi.nlm.nih.gov/gene)

- *Genes* (https://www.mdpi.com/journal/genes) – an Open Access journal
- IDconverter – converts gene IDs between public databases (http://idconverter.bioinfo.cnio.es/) Archived (https://web.archive.org/web/20200929092205/http://idconverter.bioinfo.cnio.es/) 29 September 2020 at the Wayback Machine
- iHOP – Information Hyperlinked over Proteins (https://web.archive.org/web/20051017154959/http://www.ihop-net.org/UniPub/iHOP/)
- TranscriptomeBrowser – Gene expression profile analysis (https://web.archive.org/web/20110720065115/http://tagc.univ-mrs.fr/tbrowser/)
- The Protein Naming Utility, a database to identify and correct deficient gene names (https://archive.today/20121221003541/http://www.jcvi.org/pn-utility)
- IMPC (International Mouse Phenotyping Consortium) (https://www.mousephenotype.org/) – Encyclopedia of mammalian gene function
- Global Genes Project (https://www.globalgenes.org/) – Leading non-profit organization supporting people living with genetic diseases
- Encode threads explorer (https://www.nature.com/collections/aghcdefffg), *Nature*
  - Characterization of intergenic regions and gene definition (https://www.nature.com/articles/nature28172), *Nature*