Glossary

A

A-DNA It is right handed helical configuration of the DNA with about 11 bp per turn and bases are tilted at an angle to the helix axis. In A-DNA form the bases are relatively inaccessible, the sugar-phosphate backbone being the most prominent part of the molecule. A-DNA occurs *in vitro* at low relative humidity (75%) in the presence of Na^+, K^+ or Cs^+.

Actinomycin D An antibiotic isolated from *streptomyces*. It contains a (red) substituted phenoxazone chromophore linked to two identical pentapeptide lactone rings. All cell types are potentially susceptible, any resistance being due to low permeability of cells to the drug. It specifically inhibits DNA-directed RNA synthesis by binding to B-DNA as intercalating agent. The phenoxazone chromophore intercalates primarily between two adjacent (antiparallel) GC pairs, while the lactone rings fit into the minor groove. The drug dissociates from DNA only very slowly. It blocks the movement of RNA polymerase along its DNA template. Because promoters are usually AT rich, initiation of transcription is not inhibited. The drug does not affect DNA replication because strand separation by the replicative apparatus facilitates dissociation of the antibiotic.

Aerobes Any organism that has the ability to grow in the presence of oxygen.

Agrobacterium tumefaciens A soil bacterium that causes a cancer-like plant disease (crown gall) in dicotyledenous plants (all agricultural crops except cereals). It contains the Ti plasmid. The tumor induction ability of the bacterium spreads to neighboring cells via the plasmid.

α-Helix Common secondary structure of proteins in which the linear sequence of amino acids is folded into a spiral that is stabilized by hydrogen bonds between the carboxyl oxygen of each peptide bond.

AIDS Acquired Immune Deficiency Syndrome is a transmissible, fatal human disease, which affects the immune system. It was first recognized in 1981 and is becoming prevalent among certain high-risk groups of people. AIDS is characterized by a severe defect of cell-mediated immunity. The causal agent of AIDS is an exogenous retrovirus termed Human Immunodeficiency Virus or HIV, which has been isolated from blood, saliva and semen of AIDS patients. HIV infection results in the destruction of the T4 cells. Abnormal function of the T cells is manifest in vivo by a high incidence of neoplastic disease e.g. Kaposi's Sarcoma and by increased susceptibility to diseases caused by various opportunist pathogens. AIDS is transmitted via blood or blood products, sexually and by placenta. Thus, groups at high-risk of infection include promiscuous homosexual males, intravenous drug abusers, haemophiliacs dependent on blood products, patients receiving transfusions of infected blood, etc. AIDS has proven to be 100% fatal.

Alarmone A low molecular weight molecule whose synthesis serves as a trigger or signal for the redirection of cellular metabolism in response to a particular type of stress e.g. the role of ppGpp in stringent control.

Alleles The different forms of a gene. For instance ***Y*** and ***y*** are different alleles of the gene that determines seed color. Alleles occupy the same locus, or position, on chromosomes. As humans have two sets of chromosomes, one from each parent, equivalent genes might be different by single nucleotide (see SNP).

Allolactose [β-D-Galactopyranosyl-(1→6)-D-galactopyranose] is a minor product of β-galactosidase action on lactose [β-D-Galactopyranosyl-(1→4)-D-glucopyranose]. It is the natural inducer of the *lac* operon in *E. coli*.

Allometric growth A heterochronic change can result from a mutation that causes the rate of one cell line of the body to develop at a rate different from that of other cell lines in the body. This can result in allometric growth (from the Greek allo meaning "different" and metr meaning "measure"). In a species that exhibits allometric growth, different cell lines/body parts grow at different rates (relative to an ancestral, isometrically growing form) during development from juvenile to adult.

Allosteric Refers to a change in the properties (usually including shape) of a protein following the binding of another molecule to the protein. This represents one of many controls over enzymatic activity. In fact, enzymatic activity can be either activated or inhibited through noncovalent interaction of the enzyme with small molecules (metabolites) other than the substrate. This mode of regulation was baptized allosteric because the activator or inhibitor binds to the enzyme at a site other than (*allo* means other) the active site. Allosteric regulation is reversible thus making the cellular response instantaneous.

Alternative splicing Describe the mechanism by which multiple forms of mature mRNA are produced from the same gene generating different or slightly different proteins. This increases the coding capacity of the genome.

Aminoacyl-tRNA synthetase An enzyme that covalently links amino acids to the 3'-ends of their cognate tRNA(s). Aminoacyl-tRNA synthetases belong to one of two classes depending on the amino acid they are responsible for. Class I enzymes are generally monomeric, and attach the carboxyl of their target amino acid to the 2'-OH of adenosine 76 in the tRNA molecule. Class II enzymes are generally dimeric or tetrameric, and attach their amino acid to the 3'-OH of their tRNA, except for phenylalaninyl-tRNA synthetase, which uses the 2'-OH. Note that only 3'-aminoacyl-tRNA molecules are substrates for protein synthesis.

Anaerobe Designates any organism that utilizes energy-rich compounds through oxygen-independent metabolic pathways such as glycolysis and fermentation.

Analogy A similarity due to convergent evolution (common function) but not inheritance from a common ancestor (bat's wings and bird's wings).

Antagonistic pleiotropy The effects of genes that are beneficial early in life (i.e., increasing fitness) but deleterious later in life. Such genes will be maintained by selection, because by the time the gene exerts its damage, its bearers will already have had more offspring than other individuals.

Annotate Identify the locations, the coding as well as the non-coding (or regulatory) regions of genes in a genome and determine their functions.

Antisense RNA Every RNA molecule having a complementary sequence to a region of genetic material (as mRNA) and serving to inhibit gene function. RNA antisense is long though to act at the post-transcriptional level targeting the complementary RNA to degradation.

Antitermination Is a regulatory process during which the RNA polymerase fails to recognize a transcription terminator and continues transcribing the downstream genes. Genes *N* and *Q* both encode antiterminators involved in the regulation of the life cycle of bacteriophage Lambda.

Apoptosis Also called Programmed Cell Death (PCD), is a mechanism that leads to cell death in response to several external and internal stimuli; When activated, the mechanism of PCD leads to the destruction of the cell by a characteristic set of reactions involving several proteases called caspases. The genetically programmed death of cells at specific times during embryonic morphogenesis and development, metamorphosis, and during cell turnover in adults including the maturation of T and B cells of the immune system. Defects in apoptosis are associated with maintenance of the transformed state and cancer. Anti-apoptotic proteins include Bcl-2 and HSP families (see also caspase). Apoptosis is often induced by activation of death receptors (DR) belonging to the tumor necrosis factor receptor (TNFR) family. Examples are Fas (CD95), TNFR-1 and TNFR-related apoptosis-mediated protein (TRAMP). Death signals are conducted through a cytoplasmic motif (death domain - DD) - death-inducing signaling complex (DISC) and caspase-8 that leads to the activation of caspase cascade and eventual death of the cell.

Arabidopsis thaliana A small member of the mustard family (kitchen cress). It has a very small genome (125 Mb), five small chromosomes and contains almost no repetitive DNA. Its genome has been entirely sequenced and published in 2000. It is a plant model system of choice because of the additional advantages of short generation time (about five weeks), high seed production (up to 40,000 seeds per plant) and natural self-pollination (as opposed to natural cross-pollination in maize).

Archaea Form the third domain of life beside Eukarya and Bacteria. They are bacterial looking prokaryotes that harbour many unique genetic and phenotypic properties, testifying for their peculiar evolutionary status. The archaeal ancestor was probably a hyperthermophilic anaerobe. Two major archaeal phyla are presently recognized, the Euryarchaeota and the Crenarchaeota.

Archezoa One of the kingdom level taxa, which consists of the most ancient unicellular eukaryotes with a nucleus and rod shaped chromosome but no mitochondria or plastid, thus believed to be the intermediate stage between prokaryotes and eukaryotes. They are also used as evidence for the evolution of nucleus before the organelles i.e. before endosymbiosis. However, it is now believed that the Archezoa secondarily lost their mitochondria. The intestinal parasite *Giardia lamblia* (a protist) is an example.

Attenuation Is a regulatory mechanism in which gene expression is prevented by the termination of transcription at a rho-independent terminator called the attenuator present in the leader region of the mRNA. This prevents the RNA polymerase from reaching the first structural gene in the operon. At the molecular and structural levels the termination of transcription at the attenuator depends on the growth conditions and on the secondary structures adopted by the leader region. In one conformation the stem-and-loop structure of the rho-independent terminator will form and transcription will be terminated. While in the other conformation the terminator will not form and the RNA polymerase will proceed downstream to transcribe the structural genes.

Automorphy Unique derived characteristic; a trait present in only one member of a lineage or in only one lineage among many.

Autophagy Is the degradation or digestion of the eukaryotic cell component by the cell itself, under the condition of starvation.

Autoradiography A technique that uses X-ray film to visualize radioactively labeled molecules or fragments of molecules; used in analyzing length and number of DNA fragments after they are separated by gel electrophoresis.

Autosomal A locus on any chromosome but a sex chromosome. Not sex-linked.

Autotroph Refers to any organism that is able to synthesize the nutritive substances it requires from the inorganic substance present in its environment. Also refers to any organism capable of surviving on CO_2 as its principle carbon source.

B

BAC clone Bacterial artificial chromosome vector carrying a genomic DNA insert, typically 100–200 kb. Based on naturally occurring F-factor plasmid found in the bacterium *E. coli*. Most of the large-insert clones sequenced in the human project were BAC clones.

Bacillus thuringiensis This bacterium is pathogenic to insects and the gene for its toxin is used to create transgenic plants with their own insecticide.

Bacterial Artificial Chromosomes (BAC) Vectors that are widely used in large-scale DNA sequencing project. They are much more stable than YAC but have a lower insert capacity, which ranges from 100 to 200kb.

Bacteriorhodopsin A protein pigment (20kD) resembling rhodopsin identified in the membranes of some halophilic Archaea, such as *Halobacterium salinarium*. Bacteriorhodopsin acts as a light-driven proton pump, forming protein clusters known as "purple patches" in the membrane.

Baculoviridae A family of covalently closed circular double-stranded DNA (ccc dsDNA) containing viruses that infect arthropods (particularly insects of the Diptera, Hymenoptera and Lepidoptera). Baculovirus virions are structurally complex, consisting of one or more nucleocapsids within an envelope. Virions containing one nucleocapsid per envelope are rod-shaped of about 300 x 50 nm. The genome is monopartite circular supercoiled dsDNA.

Balancing selection Selection involving opposing forces in which selective advantages and disadvantages cancel each other out. Heterozygote advantage (or overdominant selection) is an example in which an allele selected against in the homozygous state is retained because of the superiority of heterozygotes. Other balanced states may occur including when: an allele is favored at one developmental stage and is selected against at another (antagonistic pleiotropy); an allele is favored in one sex and selected against in another (sexual antagonism); an allele is favored when it is rare and selected against when it is common (negative frequency dependent selection).

Base pair (bp) Two nitrogenous bases (adenine and thymine or guanine and cytosine) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs.

Bath culture (closed culture) A form of culture in which a given volume of liquid medium is inoculated with cells (e.g. bacteria) capable of growth in that medium, and the inoculated medium is incubated for an appropriate period of time; cells growing under these conditions are exposed to a continually changing environment due e.g. to the gradual consumption of nutrients and the accumulation of metabolic wastes. The growth curve obtained by monitoring a batch culture commonly exhibits a sequence of four main phases of growth. In the *lag phase* the growth rate (the rate of increase in cell numbers or biomass) is initially minimal but subsequently rises to a value dictated by the prevailing conditions (temperature, concentration of nutrients, etc). At the end of the lag phase the cells enter the *exponential (logarithmic or log) phase* of growth in which the growth rate is both constant and maximal for the particular growth conditions. In this phase there is an exponential increase in the cell

numbers. In the *stationary phase* the growth rate declines and eventually reaches zero. In the *death phase* the number of viable cells in the culture declines.

B-DNA Under physiological conditions B-DNA is the commonest form of the DNA double helix. It is right handed with an average about 10.4 bp per turn, each turn being 3.4-nm long and 2.4 nm in diameter. The bases are stacked roughly perpendicular to the helix axis. Two spiral grooves are distinguishable in B-DNA: the minor groove, located between the two strands, and the major groove, which runs between the turns of the helix.

Binary fission Mode of reproduction not involving any sex but division of a parent cell into two equally sized offspring.

Bioinformatics Also referred to as "*in silico* analysis" or "biology with computer", is defined as the management and analysis of biological information in the databases. Bioinformatics or computational biology research is divided into two main parts: the analysis and interpretation of data and the development of new algorithms and statistics.

Biotechnology A set of biological techniques developed through basic research and now applied to research and product development. In particular, the use by industry of recombinant DNA, cell fusion, and new bioprocessing techniques.

Biotin Also called vitamin H or coenzyme R, is a vitamin that acts as a cofactor in carboxylation reactions, acting as a carrier of CO_2. Biotin is bound via its carboxyl group to the ε-amino group of a lysine residue in the apoenzyme. Biotin is required as a growth factor by many fungi and bacteria. Some organisms can synthesize biotin from cysteine, pimelic acid and carbamoyl phosphate.

BLAST Basic Local Alignment Search Tool. A heuristic sequence comparison developed by researchers at the National Center for Biotechnology Information (NCBI) and others, which is used to search sequence databases for optimal local alignments to a query.

β-pleated sheet A planar secondary structure element of proteins. It is created by hydrogen bonding between the backbone atoms in two different polypeptide chains or segments of a single folded chain.

C

Cap A methylated guanine residue (GTP) which is added to the 5' end of eukaryotic mRNAs in a post-transcriptional reaction. It protects the mRNA against 5'-exonuclease.

Catabolite repression The repression of certain inducible metabolic pathways of alternative energy sources by the presence of a preferred carbon source e.g. glucose in the enterobacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae.* For instance, in *E. coli,* various operons concerned with the degradation of substrates other than glucose (*ara* operon and *lac* operon) are repressed in the presence of glucose, even if their respective inducers are available. Thus, when *E. coli* is growing in a medium containing both glucose and lactose, the *lac* operon remains repressed until the glucose supply is used up, at which point the intracellular level of cyclic AMP rises and the *lac* operon becomes derepressed (activated).

Caenorhabditis elegans A normally self-fertilizing hermaphrodite soil nematode whose developmental genetics has been extensively studied. It is no more than 1 mm long. Loss of an X chromosome by meiotic disjunction leads to the production of males. The genetic basis of apoptosis was first shown in *C. elegans in 1986*. It has five equally sized chromosomes and it is the first animal whose whole genome has been sequenced (in 1998). The 97 Mbp

genome contains 19,000 genes on 6 chromosomes. About 74% of human genes have their homologues in the *C. elegans* genome.

Cap A methylated and inverted guanine residue (GTP) which is added to the 5' end of eukaryotic mRNAs in a post-transcriptional reaction. It protects the mRNA against 5'-exonuclease.

Cap site The initiation site of transcription in a eukaryotic gene. The initiation of translation of most eukaryotic mRNAs involves recognition of the cap followed by either the first downstream AUG or by a 5' proximal AUG with a consensus sequence surrounding it CCACC (in which the A at position –3 is highly conserved; it is found in ~80% of cases).

Cellular slime molds A great swarm resulting from the aggregation single-celled amoeboid protists. There are two groups of cellular slime molds, the Dictyostelida and the Acrasida, which may not be closely related to each other.

Centromere It is a constricted region where sister chromatids are attached in mitotic chromosomes. The centromere is generally flanked by repetitive DNA sequences and it is late to replicate. The centromere is an A-T region of about 130 bp. It binds several proteins with high affinity to form the kinetochore which is the anchor for the microtubules of the mitotic spindle.

Centimorgan (1) A unit of genetic distance equivalent of 1% of recombination. It determines how frequently two genes or loci located on the same chromosome tend to be inherited together. Centimorgan (cM): (2) A unit of measure of recombination frequency. One centimorgan is equal to a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing over in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1 million base pairs.

Chaperons Are proteins that mediate correct assembly of proteins. They cause a target protein to acquire one possible conformation instead of others. This is accomplished by binding to reactive surfaces in the target protein that are exposed to during the assembly process, and preventing those surfaces from interacting with other regions of the protein to form an incorrect conformation. It refers also to any cellular protein that binds to an unfolded or partially folded target protein to prevent misfolding, aggregation, and/or degradation of it. Chaperones also facilitate the target protein's proper folding, translocation and assembly within cells, preventing inappropriate interactions with other proteins.

Chemoautotroph An autotroph that utilizes energy stored in inorganic molecules (such as ammonia, hydrogen sulfide, nitrites, etc.) to convert CO_2 into organic compounds.

Chemolithoautotroph Refers to any organism that gets energy and its electrons from an inorganic chemical and its carbon from carbon dioxide. Processes include: nitrification, sulfur oxidation, iron oxidation and hydrogen oxidation.

Chi-like Sequence An octamer nucleotide sequence (consensus GCTGGTGG) that creates a recombinational hotspot in the genome (originally discovered in coliphage lambda). MHC class I transmembrane domain length variation, gene conversions within the MHC class II genes in mice and humans, many oncogene translocations (BCL2 for example) are attributed to chi-like sequences at the breakpoint region. It acts like a restriction site for recombinase.

Chlamydomonas The unicellular green alga that is probably the closest living organism to the ancestor of green plants. It reproduces both asexually and sexually (two mating types).

When reproduces sexually, the mitochondria are inherited from the (-) mating type and chloroplasts from the (+) mating type.

Chromista[1] Means "colored", and although some chromists, like mildews, are colorless, most are photosynthetic. Chromists are not at all closely related to plants, or even to other algae. Unlike plants, the Chromista have chlorophyll c, and do not store their energy in the form of starch. Also, photosynthetic chromists often carry various pigments in addition to chlorophyll, which are not found in plants. It is these pigments which give them their characteristic brown or golden color. In the great kingdom-level taxon Chromista it may seem hard to believe that microscopic diatoms, with their delicate silica skeletons only forty millionths of a meter long, can be related to the giant kelps, which may grow as long as fifty meters, or that either one is related to the downy mildew that nearly destroyed the French wine industry. Chromista have genetic material derived from four different ancestors (purple bacteria, cyanobacteria, red alga and green alga).

Chromomere[2] (1) Also known as an idiomere, is one of the serially aligned beads or granules of an eukaryotic chromosome, resulting from local coiling of a continuous DNA thread. It is visible on a chromosome during the prophase of meiosis and mitosis. (2) Chromomere is a small, bead-like structure that is visible in a chromosome during prophase of meiosis and mitosis, when it is relatively uncoiled (in particular at the leptotene and zygotene stages of meiosis). In polytene chromosomes, the chromomeres lie in parallel, giving the chromosome its banded appearance. Chromomeres in corresponding positions on homologous chromosomes pair during meiosis in many organisms .... (3) A chromomere is a serially aligned beadlike granule of concentrated chromatin that constitutes a chromosome during the early phases of cell division.

Chromosomes The self- replicating genetic structures of cells containing the cellular DNA that bears in its nucleotide sequence the linear array of genes. In prokaryotes, chromosomal DNA is circular, and the entire genome is carried on one chromosome. Eukaryotic genomes consist of a number of chromosomes whose DNA is associated with different kinds of proteins.

Cistron A DNA segment coding for a specific polypeptide, and includes its own start and stop codons. When an mRNA encodes two or more proteins, it is called polycistronic.

Chromatid Each of a pair of identical DNA molecules after DNA replication, joined at the centromere.

Chromatin Protein/DNA complex making the chromosome.

---

[1] Note that Photosynthetic chromists are some of the most important organisms in aquatic ecosystems. The cool and temperate coasts of continents are lined with kelp forests, where many commercially important fish and shellfish feed and reproduce, and diatoms are frequently the primary source of food for both marine and fresh-water organisms. In additional to their roles as producers for marine animals, chromists provide many products for industry. Alginates are viscous chemicals extracted from kelp; these are used in paper production, toothpaste, and in ice cream, where the alginate helps to improve texture and ensure uniform freezing and melting. The skeletons of dead chromists accumulate on the floor of lakes and oceans, where they may become thick deposits of silica or calcium carbonate. These deposits are useful for interpreting ancient climate, and in searching for oil.

[2] Giant banded (Polytene) chromosomes resulting from the replication of the chromosomes and the synapsis of homologs without cell division is a process called endomitosis. These chromosomes consist of more than 1000 copies of the same chromatid that are aligned and produce alternating dark and light bands when stained. The dark bands are the chromomere. The chromomeres are present during leptotene phase of prophase I during meiosis. During zygotene phase of prophase I, the chromomeres of homologs align with each other to form homologous rough pairing (homology searching). These chromomeres helps provide a unique identity for each homologous pairs. There are more than 2000 chromomeres on 20 chromosomes of maize.

Chromosomes Molecules of DNA complexed with specific proteins. They are responsible in eukaryotes for storage and transmission of genetic information.

Clade Refers to all descendants of any given species. It is also a single whole branch of a phylogeny i.e. a group of species or biological taxa, which derive from a common ancestor. When applied to DNA, clade is a group of monophyletic DNA sequences that comprises all the sequences included in the analysis that are descended from a particular common ancestral sequence.

Cladistics A classification approach based on the rule that all members of a group must have shared a common ancestor more recently than they have with any species outside the group. When compared to phenitics, this phylogenetic approach stresses the importance of understanding the evolutionary relevance of the characters that are studied.

Cloning The process of asexually producing a group of cells (clones), all genetically identical, from a single ancestor. In recombinant DNA technology, the use of DNA manipulation procedures to produce multiple copies of a single gene or segment of DNA is referred to as cloning DNA. DNA must first be cloned to produce sufficient amount of it for the sequencing reaction or functional analysis.

Cloning vector DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vectors capacity for self- replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors are often recombinant molecules containing DNA sequences from several sources.

Commensalism A type of symbiosis between two kinds of organisms in which one symbiont (called commensal) derives benefits (e.g. food) from the association, and the other (called host) derives neither benefit nor harm.

Complementary DNA (cDNA) A DNA sequence made from a messenger RNA molecule, using an enzyme called reverse transcriptase. It lacks the introns present in corresponding genomic DNA. cDNAs can be used experimentally to determine the sequence of messenger RNAs after their introns (non-protein-coding sections) have been eliminated. cDNA libraries contain all the mRNA of the cell grown under specific conditions, thus they are context dependent. These libraries are prepared as following: mRNAs are isolated from total RNA extracts, converted to cDNA by reverse transcription, and cloned then into appropriate vector under the control of a convenient (preferentially inducible) promoter.

Complementation Refers to the condition in which two mutations in trans configuration in a heterozygote yield a wild type phenotype. It is due to the ability of two independent genes to generate diffusible products that give wild type phenotype when tested in trans.

Complementation group A set of mutations unable to complement each others when tested in pairwise combinations in trans. Refers to one genetic unit of cistron.

Complementation test Is used to characterize two mutations having the same phenotype and may map close together. This test determine whether the two mutations lie in the same gene or in different genes. First we make a heterozygote for the two mutations (by mating two parents homozygous for each mutation), then we check the resulting phenotype. A mutant phenotype indicates that the mutations lie in the same gene, whereas a wild type phenotype indicates that the mutations lies in two different genes.

Conjugation A sexual process, also termed mating, that occurs in microgorganisms in which the establishment of direct contact between two or more cells is a prerequisite for gene transfer. In bacterial conjugation the male or donor bacterium transfers DNA to the female or recipient bacterium while the cells are in physical contact. A recipient cell that has acquired a DNA molecule from the donor is called transconjugant.

Consensus sequence A sequence that takes account of the base that occurs with the highest frequency at each position, when a series of sequences believed to have common feature are compared. A nucleotide sequence that represents an average of a number of related but not identical sequences.

Conservation Genes that are present in two distinct organisms are said to be conserved. Conservation can be detected by measuring the similarity of the two sequences at the base (RNA or DNA) or amino acid (protein) level. The more similarities there are, the more highly conserved the two sequences.

Contig The result of joining an overlapping collection of sequences or clones.

Cosmids Plasmids into which phage lambda cos sites have been inserted; as a result, the plasmid DNA can be packaged *in vitro* in the phage coat. Cosmids are spatially designed to accept only large DNA inserts (about 45,000bp).

Coverage (or depth) The average number of times a nucleotide is represented by a high quality base in a collection of random raw sequence.

CpG island Repetitive CpG doublets creating a region of DNA greater than 200 bp in length with a G+C content of more than 0.5 and an observed/expected presence of CpG more than 0.6. Usually associated with transcription-initiation regions of (housekeeping) genes transcribed at low rates that do not contain a TATA box. The CpG-rich stretch of 20-50 nucleotides occurs within the first 100-200 bases upstream of the start site region (where promoter-proximal elements reside). A trans-acting transcription factor called SP1 recognizes the CpG islands (see also Htf islands). In vertebrates, many of the non-transcribed genes (and the genes on the inactivated X chromosome) have a 5-methyl group on the C residue in CpG di-nucleotides in transcription-control regions. On the other hand, many genes with restricted expression patterns have (methylated) CpG islands located downstream of transcription initiation; this does not block elongation of the transcript.

Crossing over A recombinational process that results in the exchange of the double-stranded regions between two linear and homologous DNA molecules. It involves breakage and reunion of the strands of the two interacting molecules in such a way that the end of one molecule is joined to the end of the second molecule. Therefore, if the two interacting DNA molecules carry the alleles ***AB*** and ***ab***, crossing over between these two markers results in recombinant DNA molecules ***Ab*** and ***aB***. Crossing over occurs between non-sister chromatids during meiosis. Moreover, recombination between two circular DNA duplexes (see site specific recombination) leads to a larger circular DNA molecule (e.g. insertion or integration of the bacteriophage λ in the *E. coli* chromosome)

Cruciform In a nucleic acid (DNA or RNA) a cruciform is obtained when two hairpins (also designated stem-and-loop structures), opposite to one another, form a cross-shaped structure.

C value Refers to The amount of DNA comprising the haploid genome for a given species (picograms per cell; 2-3 pg in mammals). The C value paradox is the lack of correlation between the C values of species and their evolutionary complexity. For example, some amphibians having 30 times as much DNA as we have are not more complex than humans.

Cyanobacteria Unicellular, photosynthetic (photo-autotroph) prokaryote (a phylum of Eubacteria in the Kingdom Monera), formerly known as the 'blue green algae'. These prokaroytes are the only organisms known to be capable of oxygenic photosynthesis. They contain chlorophyll *a* but not chloroplast. Some species are believed to be the ancestor of the present-day chloroplasts. They reproduce by fission and never sexually.

Cyclin b A regulatory protein whose abundance varies during the cell cycle and which regulates biochemical events in a cell cycle-specific manner.

Cyclophilins Constitute a subgroup of large family of proteins called immunophilins, which also include FKBPs and Parvulins. They are remarkably conserved in all genera, highlighting their pivotal role in important cellular processes. Most cyclophilins display PPIase enzymatic activity, multiplicity, diverse cellular locations and active role in protein folding which render them to be included in the class of diverse set of proteins called molecular chaperones. Due to their distinct PPIase function, besides protein disulfide isomerases and protein foldases, cyclophilins have been deemed necessary for in vivo chaperoning activity. Unlike other cellular chaperones, these proteins are specific in their respective targets. Not all cyclophilin proteins possess PPIase activity, indicating a loss of their PPIase activity during the course of evolution and gain of function independent of their PPIase activity. The PPIase function of cyclophilins is also compensated by their functional homologs, like FKBPs. Multiple cyclophilin members in plants like Arabidopsis and rice have been reported to be associated with diverse functions and regulatory pathways through their foldase, scaffolding, chaperoning or other unknown activities. Although many functions of plant cyclophilins were reported or suggested, the physiological relevance and molecular basis of stress-responsive expression of plant cyclophilins is still largely unknown. However, their wide distribution and ubiquitous nature signifies their fundamental importance in plant survival. Several of these members have also been directly linked to multiple stresses.

D

Dark reaction Also termed light independent reaction, refers, collectively, to those reactions in which photosynthetically derived energy is used for the synthesis of carbohydrates.

Death phase In a batch culture (or closed culture) refers to the phase where the number of viable cells in the culture (maximal in the stationary phase) declines.

Degeneracy A feature of the genetic code that more than one nucleotide triplet or codon codes for the same amino acid. The same applies to the termination signal which is specified by three different stop codons (UAA, UAG and UGA).

Denaturation Reversible disruption of hydrogen bonds between nucleotides converting a double-stranded DNA molecule to single-stranded molecules. Heating or strong alkali treatment result in denaturation of DNA.

Dendritic cells Also called (DC), they are antigen-presenting cells (APCs). They ingest antigen by phagocytosis or pinocytosis, degrade it, and present fragments of the antigen in MHC molecules on their surface. There are two kinds of dendritic cell: (1) DC1 - these are descended from monocytes and (2) DC2 - these appear to be derived from lymphocytes.

Disulfide bond (-S-S-) A covalent linkage between two cysteine residues in different parts of a protein or between two different proteins. Insulin (a small protein having two polypeptide chains) and immunoglobulin molecules, for example, have interchain and intrachain disulfide bonds.

DNA binding motif Common sites on different proteins which facilitate their binding to DNA. Examples are leucine zipper and zinc finger proteins. Any protein containing such a motif is called DNA-binding protein.

DNA chip (DNA microarrays) technology New technology for parallel processing thousands of DNA segments, such as for detecting mutation patterns in genomic DNAs or expression of patterns of mRNAs.

DNA Polymerase A group of enzymes mainly involved DNA synthesis. They copy a single-stranded DNA molecule to make its complementary strand. Eukaryotic DNA polymerases participate in chromosomal replication, repair, crossing-over and mitochondrial replication. To initiate replication, DNA polymerases require a priming RNA (or chimeric RNA-DNA) molecule. They extend the DNA using deoxyribonucleotide triphosphates (dNTP) as substrates and releasing pyrophosphates (PP_i). The dNMPs are added to the 3' OH end of the growing strand (thus, DNA replication proceeds from 5' to 3' end).

dn/ds ratio In molecular phylogenetic studies, the ratio of the number of non-synonymous nucleotide substitutions (dn) to the number of synonymous nucleotide substitutions (ds). In the case of functionally important (or otherwise constrained) genes, ds is expected to exceed dn (dn/ds <1). Because most amino acid changes will disrupt protein structure and those non-synonymous substitutions (dn) causing them will not be maintained. In a non-functional pseudogene, there will be no discrimination between them and equal numbers of dn and ds are expected (dn/ds=1). When natural selection is acting to favor changes at the amino acid level, it is predicted that dn will exceed ds, hence a high dn/ds ratio.

Domain A portion of a protein that folds independently of the rest of the protein, or at least is assumed to do so.

Dominant trait A trait expressed preferentially over another trait.

Draft clone A large-insert clone for which roughly half-shotgun sequence has been produced. Operationally, the collection of draft clones produced by each center was required to have an average coverage of fourfold for the entire set and a minimum coverage of threefold for each clone.

Drosophila melanogaster The fruit fly, a favorite organism for genetic analysis.

Draft genome sequence The sequence produced by combining the information from the individual sequenced clones (by creating merged sequence contigs and then employing linking information to create scaffolds) and positioning the sequence along the physical map of the chromosomes.

E

Ecogenetics The branch of genetics that studies how (inherited or acquired) genetic factors influence human susceptibility to environmental health risks. It studies the genetic basis of environmental toxicity to develop methods for the detection, prevention and control of environment-related disease. Ecogenetics interacts with ecology, molecular genetics, toxicology, public health medicine and environmental epidemiology.

E. coli Common bacterium that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory.

Electrophoresis A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different

rate, depending on its electrical charge and size. Separation is based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

Electroporation A process using high-voltage current to make cell membranes permeable to allow the introduction of new DNA; commonly used in recombinant DNA technology.

Embden-Meyerhof-Parnas pathway A sequence of reactions in which glucose is broken down to pyruvate.

Emphysema Refers to a condition of the lung being marked by abnormal dilation of its air spaces and distention of its walls and frequently by impairment of heart action. In general, it is a condition characterized by air filled expansion of body tissues.

Endocytosis is a process by which the cell engulfs a drop of the extracellular fluid (ECF) by folding inward a portion of its plasma membrane. The pouch that results is pinched off from the plasma membrane forming a membrane-enclosed bubble-like vesicle called an endosome.

Endonuclease A nuclease which cuts a nucleic acid molecule by cleaving the phosphodiester bonds between two internal residues. Best known examples are restriction endonucleases or enzymes.

Endosymbiosis Refers to an endocytosis-like process during which cells are engulfed but not digested. The endosymbiont and the host cell live together in a mutually benefiting relationship or symbiosis.

Endosymbiont An intracellular organism that contributes to the survival of the host cell and depends on the host for its own persistence.

Endonuclease Is a nuclease that cut the sugar-phosphate backbone of a nucleic acid molecule at internal sites. Cleavage of the phosphodiester bond may occur either on the ***a*** or the ***b***-side (see chapter II).

Endoreplication Also termed endoduplication, is the replication of DNA during the S phase of the cell cycle without the subsequent completion of mitosis and/or cytokinesis.

Enhancer A cis-acting (on either side of a gene and acting in both orientations) promoter element. They are located 10 to 50 kb downstream or upstream of a gene. They may be tissue-specific. The enhancer effect on proximal promoter is mediated through sequence-specific DNA-binding proteins.

Enolase Is an enzyme that catalyzes the formation of *phosphoenolpyruvate* (PEP) from *2-phosphoglycerate* (2-PG) during the glycolysis pathway. The reaction involves the removal of a water molecule –dehydration- to form the enol structure of PEP. The PEP resulting from the enolase reaction drives the phosphorylation of ADP to form ATP in the pyruvate kinase reaction.

Entner-Doudoroff pathway An alternative pathway used by some microbes for the catabolism of glucose.

Epigenesis Is the development of plant or animal from an egg or spore through a series of processes in which unorganized cell masses differentiate into organs and organ systems.

Epigenetics (i) The study of heritable changes in gene expression that occur without a change in DNA sequence. Epigenetic phenomena such as imprinting and paramutation violate Mendelian principles of heredity. (ii) Is a new field in Genetics that deals with all mechanisms affecting gene expression not involving changes in the DNA sequence. Several

layers of information have been unraveled recently involving regulatory RNAs, methylation, acetylation, chromatin remodeling machines (CRM), etc. These layers lie outside DNA but are crucial to determine which genes are to be expressed. Altogether, these layers of information form the science of Epigenetics.

Epigenetic modifications Refer to any alteration in the genetic material leading to changes in gene expression patterns without affecting DNA sequences. In other terms, epigenetic phenomenon concerns any gene regulatory activity not involving changes to the DNA code and which can be heritable i.e. persists through one or more generations.

Epistasis (1) Is the phenomenon where the effects of one gene are modified by one or several other genes, which are sometimes called modifier genes. The gene whose phenotype is expressed is called epistatic, while the phenotype altered or suppressed is called hypostatic. Epistasis can be contrasted with dominance, which is an interaction between alleles at the same gene locus. Epistasis is often studied in relation to Quantitative Trait Loci (QTL) and polygenic inheritance. (2) double mutant where one mutation masks the phenotype of another mutation. Note that epistasis is not the same thing as dominance. With epistasis a mutation in one gene masks the expression of a different gene. With dominance, one allele of a gene masks the expression of another allele of the same gene.

EST Expressed sequence tag, obtained by performing a single raw sequence read from a random complementary DNA clone.

Esterification Refers to a reaction between an acid and an alcohol that eliminates a molecule of water and leads to the formation of ester. The latter can be represented by the formula RCOOR.

Eugenics The idea of improving the quality of human species by selective breeding. Encouraging breeding of those with supposedly good genes is positive eugenics, whereas discouraging those with genes for undesirable traits is negative eugenics.

Eukaryote An organism whose cells have a complex internal structure, including a nucleus. Animals, plants and fungi are all eukaryotes.

Ewens-Watterson Neutrality Test Also called E-W homozygosity statistics. Described by Ewens (1972) and Watterson (1978). A widely used test in population genetics to estimate the selection acting on a locus. It compares the sum of observed homozygosity for each allele of a given locus ($\boldsymbol{F_o}$) with the expected $\boldsymbol{F_e}$ value based on the number of alleles in the locus of interest, neutrality expectations and random mating assumption. A test of comparison yields an $\boldsymbol{F}$ value. Values close to zero mean that the locus is evolving under neutrality (genetic drift only) and there is no selection. Values of $\boldsymbol{F}$ significantly different from zero suggest selection. When $\boldsymbol{F_o > F_e}$, the locus is undergoing purifying selection, and when $\boldsymbol{F_e > F_o}$, the locus is under balancing selection.

Exocytosis Is a process by which membrane-bound vesicles move to the cell surface and fuse with the plasma membrane leading to three consequences: (1) restoration of the normal amount of plasma membrane lost during the endocytosis process; (2) dissolved molecules in the fluid contents of these vesicles are discharged into the extracellular fluid (this is termed secretion); (3) The integral membrane proteins that are exposed at the interior surface of the vesicles will be displayed at the cell surface.

Exonuclease Is a nuclease that sequentially removes nucleotides from one end of nucleic acid molecule. Hydrolysis of the backbone may occur on either side of the phosphodiester bond (the 3'-side termed ***a*** or the 5'-side termed ***b***, see chapter II).

Exosome A multiprotein complex involved in the degradation and maturation of RNA molecules. It contributes to a checkpoint that monitors proper 3'-end formation of mRNA and allows mRNA release from its transcription site.

Expressed sequence tag (EST) A short strand of DNA that is a part of a cDNA molecule and can act as identifier of a gene. Used in locating and mapping genes.

Extinction[3] In biology and ecology, is the end of an organism or of a group of organisms (taxon), normally a species. The moment of extinction is generally considered to be the death of the last individual of the species, although the capacity to breed and recover may have been lost before this point. Because a species' potential range may be very large, determining this moment is difficult, and is usually done retrospectively. This difficulty leads to phenomena such as Lazarus taxa, where a species presumed extinct abruptly "re-appears" (typically in the fossil record) after a period of apparent absence.

Extra-chromosomal inheritance Non-Mendelian inheritance due to extra-nuclear DNA (mitochondrial DNA in animals). The transmission of the trait only occurs from mothers.

F

FASTA A heuristic homology-searching program designed to compare a query sequence to database.

Ferredoxin An iron-sulfur protein containing iron and acid-labile sulfur in equimolar amounts. It was first discovered in an anaerobic, nitrogen fixing bacterium (*Clostridium pasteurianum*). It is also highly active in promoting photoreduction of $NADP^+$ in chloroplasts (photosynthetic electron transport).

F Factor (Fertility Factor) Transmissible plasmid (episome) in bacteria (such as *E. coli*) that acts as a sex factor. It is a circular DNA about 94 kb long. Conjugation and chromosomal gene transfer occur from F^+ or HFR (male) bacterium to F^- (female) bacterium.

F' (F-prime) factor Normally, the F factor contains genes related to conjugation/mating. The F' factor contains an additional portion of the bacterial genome.

Fingerprinting Is the identification of multiple specific alleles on a person's DNA to produce a unique identifier for that person.

Finished clone A large-insert clone that is entirely represented by finished sequence.

Finished sequence Complete sequence of a clone or genome, with an accuracy of at least 99.99% and no gaps.

Fitness Lifetime reproductive success of an individual (i.e., the total number of offspring who themselves survive to reproduce). It can be seen as the extent to which an individual

[3] Through evolution, new species arise through the process of speciation—where new varieties of organisms arise and thrive when they are able to find and exploit an ecological niche—and species become extinct when they are no longer able to survive in changing conditions or against superior competition. The relationship between animals and their ecological niches has been firmly established. A typical species becomes extinct within 10 million years of its first appearance, although some species, called living fossils, survive virtually unchanged for hundreds of millions of years. Most extinctions have occurred naturally, prior to Homo sapiens walking on Earth: it is estimated that 99.9% of all species that have ever existed are now extinct. Mass extinctions are relatively rare events; however, isolated extinctions are quite common. Only recently have extinctions been recorded and scientists have become alarmed at the high rates of recent extinctions. Most species that become extinct are never scientifically documented. Some scientists estimate that up to half of presently existing species may become extinct by 2100. It is difficult to estimate the trajectory that biodiversity might have taken without human impact but scientists at the University of Bristol estimate that biodiversity might increase exponentially without human influence.

successfully passes on its genes to the next generation. It has two components: survival (viability) and reproductive success (fecundity). Variation in fitness is the major driving force in biological evolution (see also genetic fitness).

Flow cytometry Analysis of biological material by detection of the light- absorbing or fluorescing properties of cells or subcellular fractions (i.e., chromosomes) passing in a narrow stream through a laser beam. An absorbance or fluorescence profile of the sample is produced. Automated sorting devices, used to fractionate samples, sort successive droplets of the analyzed stream into different fractions depending on the fluorescence emitted by each droplet.

Forensics The use of DNA for identification. Some examples of DNA use are to establish paternity in child support cases; establish the presence of a suspect at a crime scene, and identify accident victims.

Full shotgun coverage The coverage in random raw sequence needed from a large-insert clone to ensure that it is ready for finishing; this varies among centers but is typically 8–10-fold. Clones with full shotgun coverage can usually be assembled with only a handful of gaps per 100 kb.

Functional genomics A new field in biology that aims at obtaining an overall picture of genome functions, including the expression profiles at the mRNA level (transcriptome) and the protein level (proteomics).

G

Glucose-6-phosphate dehydrogenase deficiency Is an X-linked recessive hereditary disease characterised by abnormally low levels of glucose-6-phosphate dehydrogenase (abbreviated G6PD or G6PDH), a metabolic enzyme involved in the pentose phosphate pathway, especially important in red blood cell metabolism. G6PD deficiency is the most common human enzyme defect. Individuals with the disease may exhibit nonimmune hemolytic anemia in response to a number of causes, most commonly infection or exposure to certain medications or chemicals. G6PD deficiency is closely linked to favism, a disorder characterized by a hemolytic reaction to consumption of broad beans, with a name derived from the Italian name of the broad bean (fava). The name favism is sometimes used to refer to the enzyme deficiency as a whole, although this is misleading as not all people with G6PD deficiency will manifest a physically observable reaction to consumption of broad beans.

Gel electrophoresis Electrophoresis performed in an agarose or polyacrylamide gel so that molecules of similar electrical charge can be separated on the basis of size.

Gel retardation analysis A technique that identifies protein-binding sites on DNA molecules by virtue of the effect that a bound protein has on the mobility of the DNA fragments during gel electrophoresis.

Gel stretching A technique for preparing restricted DNA molecules for optical mapping by e.g. FISH.

Genetic code (1) The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. The DNA sequence of a gene can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence. (2) It is the code in which information for the synthesis of proteins is contained in the nucleotide sequence of a DNA molecule (or an RNA molecule in certain viruses). During gene expression an mRNA is produced by transcription, then the

mRNA is translated into polypeptide, each amino acid of the polypeptide being encoded by a sequence of three nucleotides (one codon).

Genetic map A genome map in which polymorphic loci are positioned relative to one another on the basis of the frequency with which they recombine during meiosis. The unit of distance is centimorgans (cM), denoting a 1% chance of recombination.

Gene The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e. a protein or RNA molecule). In addition to the coding regions (exons), a gene may have non-coding intervening sequences (introns) and transcription-control regions.

Gene chip technology Development of cDNA microarrays from a large number of genes. Used to monitor and measure changes in gene expression for each gene represented on the chip e.g. in a cell culture grown under different conditions or a wild type and a mutant grown under the same conditions.

Gene conversion Partial sequence transfer from one allele to another (interallelic recombination) converting one gene or allele to another one.

Gene expression (1) The series of events by which the biological information encoded in the nucleotide sequence of a gene is released and made available to the cell. The biological information is usually read by proteins that gain access and attach to the DNA at the appropriate positions and initiate a series of biochemical reactions referred to as gene expression. (2) The process by which gene coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Gene flow The movement of genes within a population or between two populations following genetic admixture. Gene flow creates new combinations of genes or alleles in individuals that can be tested against the environment. This way it is one of the sources of variation in the process of natural selection.

General recombination Also called homologous recombination because the enzymatic machinery that mediates the exchange of strands can use any pair of DNA molecules as substrates. General recombination occurs most frequently during meiosis in diploid organisms generating gametes, which are genetically different (different allelic combinations).

Genocopy The same phenotype due to different genetic causes.

Genome The complete DNA sequence (or the whole set of genes) of an organism; it is also defined as the total genetic material in a set of haploid chromosomes in a given organism.

Genome project A project that aims at sequencing complete DNA that forms the genome of an organism.

Genotype The genetic constitution of an organism with respect to a trait. For a single trait on an autosome, an individual can be homozygous for the dominant trait, heterozygous, or homozygous for the recessive trait. Yellow seeds are dominant, but yellow seeded plants could have a genotype of either ***YY*** or ***Yy***.

Gyrase One of the bacterial DNA topoisomerases that functions during DNA replication to reduce molecular tension caused by supercoiling (supertwisting). DNA gyrase produces, then seals, double-stranded breaks. Gyrase introduces negative supercoils.

H

Halorhodopsin An archaeal rhodopsin that uses light energy to pump chloride through biological membranes.

Haploinsufficiency Situation where one normal copy of a gene alone is not sufficient to maintain normal function. It is observed as a dominant mutation on one allele (or deletion of it) resulting in total loss-of-function in a diploid cell because of the insufficient amount of the wild-type protein encoded by the normal allele on the other haplotype

Haplotype A particular combination of alleles (alternative forms of genes) or sequence variations that are closely linked -that is, are likely to be inherited together- on the same chromosome.

Hardy-Weinberg Law In an infinitely large population, gene and genotype frequencies remain stable as long as there is no selection, mutation, or migration. In a large population, the genotype frequencies will remain constant over time. For a biallelic locus where the allele frequencies are p and q, phenotype and genotype frequencies can be determined by Hardy-Weinberg equation: $(p + q)^2 = p^2 + 2pq + q^2 = 1$. The equation also applies for multiple-alleles genes.

Heat Shock Response Heat shock response is ubiquitous and highly conserved defense mechanism for protection of cells from harmful conditions such as heat shock, UV irradiation, toxic chemicals, infection, transformation and appearance of mutant and misfolded proteins.

Helicase An enzyme that unwinds the double DNA helix near the replication fork before DNA polymerase acts on it. Replication fork moves from 3' to 5' of the leading strand. Unwinding is also necessary for DNA repair. Mutations in the helicase genes on chromosome 2q and 19q are one group of causes of the DNA repair defect xeroderma pigmentosum (an autosomal recessive disease).

Hemizygous As in any X-linked trait in males, absence of a homologous counterpart for an allele. It may also result from deletion.

Hermaphroditism Having both male and female sexual organs in one individual. Most invertebrates and plants are hermaphrodites. Union of the gametes of the same individual (self-fertilization) is the most extreme example of inbreeding.

Heterochronic genes Genes that control the temporal dimension of development and can be thought of as the temporal analogs of the homeotic genes, which regulate spatial dimensions (e.g. anterior-posterior and dorsal-ventral axes) during development of metazoans.

Heterochrony (1) A genetic shift in timing of the development of a tissue or anatomical part, or in the onset of a physiological process, relative to an ancestor. (2) Is an evolutionary term that describes the comparatively common phylogenetic variation between species in the relative timing of developmental events. Heterochronic variation has also been induced by mutation to identify genes that regulate the timing of developmental events. (3) Heterochrony (from the Greek hetero meaning "other" and chronos meaning "time") describes a change in the timing of ontogenetic events between two taxa. These can be the result of relatively small genetic changes that may not even be alterations in DNA sequence, but in the timing of particular genes being expressed during development.

Heterosis or hybrid vigor[4] Increase in such characteristics as size, growth rate, fertility, and yield of a hybrid organism over those of its parents. Plant and animal breeders exploit heterosis by mating two different purebred lines that have desirable traits. The first-generation offspring generally show, in greater measure, the desired characteristics of both parents. Since this vigour may decrease if the hybrids are actually mated together, the parental lines must be maintained and crossed for each new crop or group desired. Increased vigor or other superior qualities arising from the crossbreeding of genetically different plants or animals.

Heterotroph An organism that depends on an external source of organic compounds for survival.

Hfr (high frequency of recombination) A male (donor) bacterial cell that has the F factor integrated into its chromosome is an Hfr cell. Crosses between Hfr cells and F^- females produce far more recombinant progeny than do crosses between F^+ males and F^- females.

Histones Five kinds of proteins forming the major components of chromatin in most eukaryotic DNAs.

Homologous chromosomes The pair of chromosomes in a diploid individual that have the same overall genetic content. One member of each homologous pair of chromosomes in inherited from each parent.

Homologous recombination *see General recombination.*

Hot-spot A region of DNA, which is particularly prone to transposition and mutation.

House-keeping genes Genes which are constitutively expressed in most cells because they provide basic functions.

Htf island Hpa Tiny Fragment island which are unmethylated CpG-rich regions in the genome. Eighty percent of these occur at or near genes, particularly housekeeping genes.

Hyperchromic Shift Refers to the change in the UV absorbance that occurs when DNA changes from the double-stranded to the single stranded condition or vice versa. Also referred to as thermal melting profile.

Hybrid vigor (heterosis) Unusual growth, strength, and health of heterozygous offspring from two less vigorous homozygous parents.

Hypothesis An unproven but testable scientific proposition. A theory is a statement with some confirmation.

Hypoxia[5] or hypoxiation Is a pathological condition in which the body as a whole (generalized hypoxia) or a region of the body (tissue hypoxia) is deprived of adequate oxygen supply.

[4] Two leading hypotheses explain the genetic basis for fitness advantage in heterosis. The overdominance hypothesis implies that the combination of divergent alleles at a particular locus will result in a higher fitness in the heterozygote than in the homozygote. Take the example of parasite resistance controlled by gene A, with two alleles A and a. The heterozygous individual will then be able to express a broader array of parasite resistance alleles and thus resist a broader array of parasites. The homozygous individual, on the other hand, will only express one allele of gene A (either A or a) and therefore will not resist as many parasites as the heterozygote. The second hypothesis involves avoidance of deleterious recessive genes (also called the general dominance hypothesis), such that heterozygous individuals will express less deleterious recessive alleles than its homozygous counterpart.

[5] A broad term meaning diminished availability of oxygen to the body tissues. Its causes are many and varied. There may be a deficiency of oxygen in the atmosphere, as in altitude sickness, or a pulmonary disorder that interferes with adequate ventilation of the lungs. Anemia or circulatory deficiencies can lead to inadequate transport and delivery of oxygen to the tissues. Finally, edema or other abnormal conditions of the tissues

Variations in arterial oxygen concentrations can be part of the normal physiology, for example, during strenuous physical exercise. A mismatch between oxygen supply and its demand at the cellular level may result in a hypoxic condition. Hypoxia in which there is complete deprivation of oxygen supply is referred to as anoxia.

I

Icosahedral symmetry Is the symmetry exhibited by an icosahedron and by the capsid of certain types of virus: 5-fold rotational symmetry through each of the 12 apexes, 3-fold rotational symmetry about an axis through the center of each of the 20 triangular faces, and 2-fold rotational symmetry about an axis through the center of each of the 30 edges.

Icosahedron A solid figure contained by 20 plane faces, all the faces being equilateral triangles of the same size. Note that each of the 20 triangular faces is threefold symmetric, so that there are a total of 60 (3x20) identical and equivalent positions in an icosahedron.

Illegitimate recombination Is a random recombination that occurs between two non-homologous DNA sequences (or showing very little homology). In Escherichia coli illegitimate recombination is *recA*-independent and may be mediated by gyrase (a type II topoisomerase). Indeed, strands cleaved by gyrase action at different sites may undergo crossing over by swapping of gyrase subunits covalently bound to the cut ends.

Imprinting Genomic imprinting is an epigenetic phenomenon that results in the differential expression of paternally and maternally inherited alleles of a gene whereby, either the paternal or the maternal gene are expressed. It is also defined as gene expression dependent of the parent of origin. During gametogenesis some genes, called imprinted genes, may acquire altered expression (via epigenetic modifications), thus becoming either activated of inactivated. Imprinted genes have a functional haploid state that makes them particularly vulnerable to being either inactivated of over expressed. Therefore, any mutation at such loci will have deleterious effect on the corresponding gene expression and as a result on the individual.

Inclusion bodies Discrete structures of various types present within cells. In virology the term generally refers to structures present in some virus-infected cells consisting of virions and/or viral components and/or cellular material. They may be characteristic in form and location for a given virus. In bacteriology the term refers to structures within a bacterial cell or to bacterial cells within eukaryotic host cell.

Inflammation Occurs in man and animals, usually localized response to cellular injury or to the presence of an allergen or certain types of microorganisms. It is characterized by dilatation and increased permeability of terminal arterioles and capillaries, leukocyte infiltration, redness, heat, and pain. It serves as a mechanism initiating the elimination of the noxious agents and of damaged tissue.

Informational proteins Proteins involved in DNA replication, transcription, and translation.

Initial sequence contigs Contigs produced by merging overlapping sequence reads obtained from a single clone, in a process called sequence assembly.

---

themselves may impair the exchange of oxygen and carbon dioxide between the capillaries and the tissues. The effect of hypoxia is to reduce the functional activity of tissues. The initial response may be one of temporarily increased activity. Terminally the tissue may be irreparably damaged.

Initiation complex A multi-protein complex that forms at the site of transcription initiation and is composed of RNA polymerase II, ubiquitous or general transcription or initiation factors and gene-specific enhancers/silencers.

Inner Cell Mass The cluster of cells located inside the blastocyst. These cells give rise to the embryonic disk of the later embryo and, ultimately, the fetus. These cells, also called embryonic stem cells, are sought because of their plasticity (totipotency).

Insertion sequence (IS) Mobile genetic elements discovered in prokaryotes. They possess inverted repeats at both extremities and comprise stop codons in their three reading frames. Their insertion leads to mutations and destabilizes the genome.

Integrase An enzyme that catalyzes a site-specific recombination (integration or excision) involving a prophage and a bacterial chromosome.

Intein An internal part of the protein that is removed from by a posttranslational process termed intein splicing.

Intein homing The conversion of a gene coding for a polypeptide that lacks an intein into one coding for an intein-plus protein, catalyzed by the spliced component of the intein.

Intein splicing In prokaryotes and eukaryotes some genes contain an in-frame open reading frame that codes for an internal protein, termed intein. The latter autosplices post-translationally and ligates the external sequences to yield a functional external protein, called extein, and an intein. In addition to the splicing domain most of inteins contain a separate endonucleolytic domain that enables them to maintain their presence by a homing mechanism called intein homing.

Interactome Is the complete protein interaction maps that can be generated by large scale approaches of protein-protein interactions.

Intercalating agents Are flat hydrophobic molecules composed of fused heterocyclic rings that insert between stacked base pairs of DNA or base-paired regions of ssDNA. These agents, such as ethidium bromide and actinomycin D, cause a distortion in DNA double helix by forcing the bases apart, thus DNA unwinds and appear as a ladder-like structure. Many intercalating agents have anti-microbial and anti-tumor activity, inhibiting transcription, DNA replication etc. and inducing frameshift mutations.

Internal ribosome entry site (IRES) Refers to the sequence that allows eukaryotic ribosomes to enter directly at an internal AUG codon rather than the usual scanning of the mRNA from the capped 5' end.

Interrupted mating It is a technique used to map the genes of the bacterial chromosome. Usually appropriate donor (*Hfr*) and recipient (F^-) cells with different growth requirements are used. Matting is allowed to proceed, then it is interrupted by the shearing of the joined cells in a blender at chosen intervals. This procedure separates the cells at various stages in the transfer of the *Hfr* chromosome. The cells are then plated on selective media and tested for their growth requirements.

Introns and exons Genes are transcribed as continuous sequences, but only some segments of the resulting messenger RNA molecules contain information that codes for the gene's protein product. These segments are called exons. The regions between exons are known as introns, and are spliced from the RNA before the product is made.

Intron homing The conversion of a gene that lacks an intron into one that contains an intron, catalyzed by a protein coded by the intron

Isometric growth Many animals undergo isometric growth as they mature from new hatchling to adult. This means that all the body parts grow at approximately the same rate, and the adult proportions are not significantly different from those of the juvenile. For example, see our pal *Batrachoseps*, one of the few salamanders that has a terrestrial (not a gilled, aquatic) larva.

K

Kaposi's sarcoma A rare multifocal neoplastic disease, which occurs in two forms: slow and indolent, limited mainly to the skin, and rapid and fulminant, involving the skin and gastrointestinal tract. The aggressive form occurs in children in tropical Africa and is a common feature in about 30% of AIDS patients. Incidence of Kaposi's sarcoma is also increased in immunosuppressed organ transplant patients.

Karyotype A photomicrograph of individual chromosomes arranged in a standard format showing the number, size, and shape of each chromosome type; used in low- resolution physical mapping to correlate gross chromosomal abnormalities with the characteristics of specific diseases.

Kilobase (kb) Unit of length for DNA fragments equal to 1000 nucleotides.

kinetochore Structure forming at centromere during Mitosis for binding microtubules.

Kingdom The major taxonomic group in the current classification of living organisms. The five Kingdoms are Monera, Protoctista, Fungi, Plants and Animals.

Kozak sequence In some viral mRNAs, the consensus sequence surrounding the initiating codon 5'ACCAUGG3' . It facilitates ribosomal binding and therefore, protein synthesis. The most consistent position is located three nucleotides before the initiation codon (ATG) and is almost always an adenine nucleotide

L

Lamprush chromosome (1) A large chromosome found especially in the immature eggs of amphibians, consisting of two long strands that form many brushlike loops along the main axis of the chromosome. So named because of the bristling appearance given them by many open loops of chromatin along the extended chromosome. (2) A greatly enlarged diplotene chromosome that has apparently filamentous granular loops extending from the chromomeres and is characteristic of some animal oocytes.

Lethal alleles Mutated genes that are capable of causing death.

Ligase An enzyme which is of vital importance in recombinant DNA technology. It joins nucleotides together by a phosphodiester bond between the 5'-P end of a polynucleotide chain and the 3'-OH end of another one.

Linkage disequilibrium (LD) The tendency for two 'alleles' to be present on the same chromosome (positive LD), or not to segregate together (negative LD). As a result, specific alleles at two different loci are found together more or less than expected by chance. The same situation may exist for more than two alleles. Its magnitude is expressed as the delta (Δ) value and corresponds to the difference between the expected and the observed haplotype frequency. It can have positive or negative values. LD is decreased by recombination. Thus, it decreases every generation of random mating unless some process opposing the approach to linkage equilibrium. Permanent LD may result from natural selection if some gametic combinations result in higher fitness than other combinations.

Light reaction Refers, collectively, to those photochemical events involved in the conversion of radiant energy into chemical energy.

Lithoautotroph An organism that is capable of utilizing rocky minerals as a source of food for energy production.

Long and short arms The regions either side of the centromere, a compact part of a chromosome, are known as arms. As the centromere is not in the center of the chromosome, one arm is longer than the other is.

Loss-of-heterozygosity (LOH) Refers to the disappearance of polymorphic marker alleles when constitutional DNA and tumor DNA from cancer patients are compared. The consequence is usually genomic deletion discarding the normal copies of tumor suppressor genes. Such deletion (or functional deletion through methylation) may uncover existing mutations in the homologue copy.

Lyon hypothesis The proposition by Mary F. Lyon that random inactivation of one X chromosome in the somatic cells of mammalian females is responsible for dosage compensation and mosaicism.

Lysosomes are membrane bound vesicles (0.05 to 0.5 micron) containing more than 40 hydrolytic enzymes that can digest most biological macromolecules. These organelles are the sites of intracellular digestion (phagocytic vacuole and turnover of cellular organelles) that are more numerous in cells performing phagocytosis. The limiting membrane keeps the digestive enzymes separate from the cytoplasm. The most common lysosomal enzymes are acid phosphatase, ribonuclease, deoxyribonuclease, proteases, sulfatases, and lipases. The enzymes function optimally at pH 5 and are mostly inactive at the pH of the cytosol (7.2). This taken with the limiting membrane protects the cell from digesting itself. Lysosomal enzymes are synthesized on the rough ER, transferred to the Golgi for modification and packaging. The cellular machinery attaches a directional signal to the enzymes (mannose-6-phosphate) that allows the ER and Golgi to sort these proteins and, via a receptor mediated process, segregate them to forming lysosomes.

M

Malpighian tubule system is a type of excretory and osmoregulatory system found in some Uniramia (Insects and Myriapoda), arachnids and tardigrades. The system consists of branching tubules extending from the alimentary canal that absorbs solutes, water, and wastes from the surrounding hemolymph. The wastes then are released from the organism in the form of solid nitrogenous compounds. The system is named after Marcello Malpighi, a seventeenth century anatomist.

Marker An identifiable physical location on a chromosome (e.g., restriction enzyme cutting site, gene) whose inheritance can be monitored. Markers can be expressed regions of DNA (genes) or some segment of DNA with no known coding function but whose pattern of inheritance can be determined. See RFLP, restriction fragment length polymorphism.

Mass Spectrometry (MS) A basic technology in proteomics for protein identification. MS consists of three essential parts, an ionization source that converts molecules into gas-phase ions. Once ions are created, individual mass-to-charge ratios are separated by a second device-mass analyzer and transferred to the third, an ion detector.

Maternal inheritance Diseases due to mutations in mtDNA are transmitted only by mothers because all mitochondria are inherited via the egg. Thus, all offspring of an affected female are at risk of inheriting the abnormality, whereas no offspring of an affected male are at risk.

Clinical manifestations are variable and may be due to variable mixtures of mutant and normal mitochondrial genomes (heteroplasmy) within cells and tissues.

Mating type Genetically determined characteristics of bacteria, ciliates, fungi and algae, determining their ability to conjugate and undergo sexual reproduction with other members of the species. In yeasts (*S. cerevisiae*) which have only two types, **a** and α, only cells of opposite types can conjugate.

MEDLINE A free one-line literature database of papers in biomedical sciences (http://www.ncbi.nlm.nih.gov/Entrez/medline.html).

Megabase (Mb) Unit of length for DNA fragments equal to 1 million nucleotides and roughly equal to 1 cM.

Melting temperature (T_m) The temperature at which the two strands of a double-stranded DNA molecule come apart. For short (<18 nucleotides) oligonucleotides T_m value (°C) is estimated by the formula: T_m = (number of A + T) x 2 + (number of G + C) x 4. See Chapter II for further details.

Merged sequence contigs Contigs produced by taking the initial sequence contigs contained in overlapping clones and merging those found to overlap. These are also referred to simply as 'sequence contigs' where no confusion will result.

Merozygote A bacterial cell which is part haploid part diploid. It may result from sexual conjugation, transduction, or transformation. A merozygote contains its own chromosome (termed *endogenote*) and the fragment of genetic material received from the donor cell (called *exogenote*). If the corresponding alleles on both *endogenote* and *exogenote* are identical the merozygote is termed *homogenote*. Otherwise, the merozygote is said to be *heterogenote*.

Messenger RNA (mRNA) Proteins are not synthesized directly from genomic DNA. Instead, an RNA template (a precursor mRNA) is constructed from the sequence of the gene. This RNA is then processed in various ways, including splicing. Spliced RNAs destined to become templates for protein synthesis are known as mRNAs.

Methanogen An organism that utilizes CO_2 and H_2 in order to produce methane.

Microaerophilic organisms Are prokaryotes that use oxygen but at concentration lower than that in the atmospheric air.

Microorganism (or microbe) Refers to every microscopic organism that belongs to the categories of algae, bacteria, fungi, protozoa, and viruses. Under this definition, an organism correspond to a biological entity that encompasses the genetic information that is needed for its own replication or reproduction. In this context, an organism may exist under natural conditions either as independent or autonomous entity or (this the case or viruses and intracellular obligate parasites) as a stable form faraway from the site of its replication reproduction. Some authors do not consider viruses as microorganisms and prefer to reserve the term "microorganism" to every biological entity that is formed of one or more cells.

Microsatellites Referred to as Simple Tandem Repeats (STR); they form a class of simple sequence length polymorphism (SSLP), whose repeats are much shorter than those of microsatellites. In fact, microsatellites clusters are usually of less than 150bp and the repeat unit is usually 4bp or less.

Minisatellites Also termed Variable Number of Tandem Repeats (VNTRs), form a type of simple sequence length polymorphism (SSLP) in which the repeat unit is a few tens of nucleotides in length. Minisatellites form clusters up to 20kb in length with units up to 25bp.

Mismatch repair A DNA repair system that can correct mismatches resulting (non standard or non Watson-Crick base pairs) e.g. from the incorporation of the wrong nucleotides during DNA synthesis. These mismatches usually escape the proofreading activity of the DNA polymerase. These mismatches may result from recombination in the heteroduplex regions.

Mitomycin C An antibiotic purified from *Streptomyces Spp.* with general cytotoxicity. It contains a quinone group and an aziridine ring. It binds to DNA leading to inhibition of DNA replication followed by degradation of DNA. Despite this effect, synthesis of RNA and proteins may pursue for a while. It behaves as a bifunctional alkylating agent and forms covalent cross-links between complementary strands of dsDNA. Preferred sites for alkylation are the O-6 groups of guanine residues. Alkylation of B-DNA by mitomycin can cause conformational changes that may involve the transition to Z-DNA. Mitomycin is a potent inducer of lysogenic bacteriophages.

Modulon Is a regulon (or regulons) concerned with multiple pathways or functions, in which operons may be under individual controls as well as common, pleiotropic regulatory protein. For example: the CAP modulon contains all regulons/operons, such as the lac operon and ara regulon, that are regulated by CAP/cAMP, but each operon has other regulators as well.

Monera One of the five Kingdoms which contains all prokaryotes. It contains archaebacteria, eubacteria and cyanobacteria. The first life form emerged over 3,500 Mya. Eukaryotes are believed to have evolved from members of this Kingdom.

Morphogenesis is the origin of form (from the Greek morph meaning "shape" and genesis, meaning "origin"). It is the process by which embryos grow into their proper form. As an embryo grows from a zygote through its various forms, genes in each cell turn on and off, changing the identity and fate of each cell. If the timing of these changes is modified, even slightly, profound effects on the finished product can arise. In short, if ontogenies diverge, then adult morphologies may also diverge.

Motif A short conserved region in a protein sequence. Motifs frequently form recognition sequences or are highly conserved parts of domains. Motif is sometimes used in a broader sense for all localized homology regions, independent of their size.

Multiple alignment An alignment of three or more sequences, with gaps (spaces) inserted in the sequences such that residues with common structural positions are aligned in the same column of the multiple alignment.

Murein A peptidoglycan found in the cell wall of bacteria consisting of a repeated disaccharide (N-acetylglucoseamine and N-acetylmuramic acid) attached to chains of four or five amino acids.

Mutasome A protein complex that is constructed during the SOS response of Escherichia coli. This complex comprises several copies of the RecA protein and the $UmuD'_2C$ complex, the latter is a timer made up of two UmuD' proteins and one copy of UmuC.

Mutagen (Latin, literally origin of change) Is a physical or chemical agent that changes the genetic material, usually DNA, of an organism and thus increases the frequency of mutations above the natural background level. As many mutations cause cancer, mutagens are therefore also likely to be carcinogens. Not all mutations are caused by mutagens: so-called "spontaneous mutations" occur due to spontaneous hydrolysis, errors in DNA replication, repair and recombination.

Mutation Change in the DNA sequence of a gene to some new, heritable form. An alteration in a genome compared to some reference state. Mutations do not always have harmful effects.

N

Nalidixic acid One of many quinolone antibiotics. Nalidixic acid inhibits bacterial DNA replication and causes degradation of DNA in the cell. It acts as a specific inhibitor of gyrase, preventing the breakage-and-reunion activity of the enzyme.

Neurospora crassa Haploid, heterothallic, filamentous Ascomycete fungus (bread mold). It has two mating types (A and a) operating as sexual compatibility system, and 11 het loci operating as heterokaryon compatibility system in vegetative phase.

Nuclear Matrix Within the nucleus among the nucleosomes is a matrix composed of proteins, metabolites and ions. When all else is removed there remains a fibrillar nucleoskeleton that may provide a scaffold for the folded DNA. A fibrous lamina (80 to 300 nm thick) lies just under the nuclear envelope. This lamina dissociates just prior to mitosis due to the phosphorylation of the laminins. Post-mitotic laminin dephosphorylation results in reformation of the nuclear lamina.

Nucleolar organizer A region on a chromosome that is associated with formation of a new nucleolus following cell division. It contains the genes for several species of ribosomal RNA (rRNA), i.e., 18S, 5.8S, 5S and 28S in eukaryotes.

Nuclease Refers to any enzyme that cleaves the polynucleotide backbone of a nucleic acid. We distinguish two main types: endonuclease and exonuclease.

Nucleoid An aggregate of DNA molecules that occur in bacteria or in cytoplasmic organelles such as mitochondria or chloroplasts (up to 10 molecules may be found per mitochondrial or plastid nucleoid).

Nucleolus This spherical structure within the nucleus is composed of DNA, RNA and protein. The DNA present is that which codes for rRNA, known as nucleolar organizers. Proteins synthesized in the cytoplasm enter through the nuclear pores and become associated with the newly made rRNA in the nucleolus. Afterwards the ribosomal subunits migrate to the cytoplasm. The nucleolus disappears during cell division but reappears in the final stage of mitosis.

Nucleosomes Basic structural unit of eukaryotic chromosome forming "beads on a string."

O

Oncogene Refers to any gene that can induce or stimulate cell growth and division. A gene that can potentially lead to neoplastic transformation in the cell in which it occurs or into which it is introduced.

Oncogenic Refers to any agent, termed oncogen, that can induce the formation of cancer.

Open reading frame (ORF) A nucleotide sequence encoding a polypeptide starting with a start and ending with a stop codon.

Operational proteins Comprise metabolic enzymes, membrane receptors, transporters, some cell division proteins, etc. They are not linked to gene expression.

Operon A group of prokaryotic structural genes that are controlled by a common promoter and are transcribed as a single transcription unit. The structural genes in an operon are often involved in the same metabolic pathway. This organization allows the corresponding genes

to be expressed in a coordinated fashion, they are either turned on (by induction) or turned off (by repression) depending on the needs of the cells. For instance, the three structural genes of the *lac* operon are induced in presence of lactose (the substrate to be degraded by this catabolic pathway) and repressed when the substrate is no longer available. Regulation of gene expression in operons may occur at the level of transcription (promoter control), transcription termination (attenuator), or a combination of both.

Organotroph Refers to any organism that uses organic substrates to produce energy. A chemoorganotroph obtains the required energy via the metabolism of organic substrates. A photoorganotroph uses an organic substrate as an electron donor in phototrophic metabolism.

Orthologs Designate homologous genes in different species that derived from a common ancestor gene during the speciation event; orthologs are likely to perform the same function in both species.

P

Palindrome Corresponds to a region of a nucleic acid that contains two inverted repeats. The latter have the potential to form a tertiary structure called cruciform when the normal interstrand base pairing is replaced by intrastrand pairing.

Parabasalidea Single-celled eukaryotes that usually live as parasites or symbionts in animals. They lack mitochondria but have hydrogen-producing organelles called hydrogenosomes. *Trichonomos vaginals* is a member species.

Paralogs Are homologous genes homologous sequences that derived from a common ancestry by duplication. The duplication event may have occurred before or after the speciation event. Because of genetic variation and diversification paralogs are unlikely to perform the same function.

Paramutation In paramutation, two alleles of a gene interact so that one of the alleles is epigenetically silenced. The silenced state is then genetically transmissible for many generations.

Parkinson's disease A chronic progressive nervous disease chiefly of later life that is linked to decreased dopamine production in the substania migra and is marked by tremor and weakness of resting muscles and by a shuffling gait.

Parthenogenesis (virgin birth) Reproduction involving unfertilized eggs. The offspring of parthenogenetic parents are less diverse than those of sexual parents.

Penetrance The proportion of individuals with a given genotype (heterozygotes for a dominant gene) who express an expected trait, even if mildly. If a disease gene is not causing the disease in all its carriers, its penetrance is low (not to be mixed with variable expression).

Peroxisome Small microbody (0.5 to 1.2 microns) that is, like lysosome, single-membrane-bounded vesicle containing oxidative enzymes. Peroxisomes contain amino oxidases, hydroxyacid oxidase and catalase that use the molecular oxygen to generate peroxides, which are toxic. Catalase protects the cell from hydrogen peroxide H_2O_2 damage by decomposing it to H_2O and O_2. Enzymes involved in lipid metabolism are also found in peroxisomes. Peroxisomal enzymes are synthesized on the free cytosolic ribosomes with a signal sequence that directs them to peroxisomes. As enzymes are added the peroxisome grows and then splits into two smaller peroxisomes.

Phagocytosis A type of endocytosis. In phagocytosis ("cell eating"), the drop engulfed is relatively large (and the membrane-enclosed bubble is often called a phagosome or a vacuole). It results in the ingestion of particulate matter (e.g., bacteria) from the extracellular fluid (ECF) by certain specialized cells (e.g., neutrophils, macrophages, and the ameba). It occurs sporadically. In due course, phagosomes deliver their contents to lysosomes. The membranes of the two organelles fuse. Once inside the lysosome, the contents of the phagosome, e.g. ingested bacteria, are destroyed by the degradative enzymes of the lysosome.

Phenetics A system of classification based on the typing of as many variable characters as possible. These characters are scored numerically and analyzed by rigorous mathematical methods.

Phenocopy A condition which is due to environmental factors but resembles one which is genetic. A non-hereditary phenotypic modification caused by special environmental conditions that mimics a similar phenotype caused by a known gene mutation.

Phenotype The physical appearance of an organism with respect to a trait, i.e. yellow (Y) or green (y) seeds in garden peas. The dominant trait is normally represented with a capital letter, and the recessive trait with the same lower case letter.

Pheromone (formerly called mating-type factors, sex factors or gamones) Species- or mating-type-specific chemical produced by an animal to communicate with an effect on their behavior without being consciously perceived as smell. Probably the most ancient communication system in living organisms. It has been noted in bacteria (Streptomyces faecalis); protists such as ciliates (Euplotes raikovi); amoeba (Dictyostelium); algae (Fucus vesiculosus); and in fungi (for example Basidiomycetous).

Phosphodiesterase Is the name given to nucleases because the hydrolysis reaction is carried out by the enzyme to cleave the phosphodiester bonds of the polynucleotide backbone.

Phosphorylation Is the process of phosphorylating a chemical compound either by reaction with inorganic phosphate or by transfer of phosphate from another organic phosphate. Proteins are either activated or inactivated by phosphorylation.

Photolithoautotroph Organism that get energy from light, its electrons from an inorganic chemical and its carbon source is from carbon dioxide.

Photolithotroph Lithotroph that uses inorganic substrates such as sulfur, sulfide, H_2, etc. as electron donors in photosynthesis.

Phototroph An organism that uses light as a primary source of energy for growth. Note that some phototrophs are facultative chemotrophs.

Physical map A map of the locations of identifiable landmarks on DNA (e.g., restriction enzyme cutting sites, genes), regardless of inheritance. Distance is measured in base pairs. For the human genome, the lowest- resolution physical map is the banding patterns on the 24 different chromosomes; the highest- resolution map would be the complete nucleotide sequence of the chromosomes.

Pinocytosis A type of endocytosis. In pinocytosis ("cell drinking"), the drop engulfed is relatively small. It occurs continuously in almost all cells. A cell sipping away at the extracellular fluid (ECF) by pinocytosis acquires a representative sample of the molecules and ions dissolved in the ECF. But endocytosis also provides a much more elegant method for cells to pick up critical components of the ECF that may be in scant supply.

Plasmids Are naturally occurring, self-replicating, circular and extrachromosomal DNA molecules that are found mainly in bacteria. They carry genes providing some metabolic advantages to the host cell such as antibiotic resistance, enzymatic activity etc...

Pleiotropic A single gene determines more than one phenotype for an organism. More than one effect of a gene on the phenotype. The effects may occur simultaneously or sequentially. An example is the DNA repair enzymes which have several other functions (transcription, cell cycle regulation, regulation of gene rearrangements).

Polycistronic mRNA An mRNA molecule that contains the transcript of two or more genes. A polycistronic mRNA contains several distinct regions: a 5' leader, two or more coding regions, each of which begins with a translation initiation codon, and a 3' trailer.

Polygenic Traits controlled by two or more genetic loci. They are usually influenced by environment as well (multifactorial).

Polymerase chain reaction (PCR) A method for amplifying a DNA base sequence using a heat-stable polymerase and two 20-base primers, one complementary to the (+)-strand at one end of the sequence to be amplified and the other complementary to the (−)-strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample.

Polymorphism A region of the genome that varies between individual members of a population. To be called a polymorphism, a variant should be present in a significant number of people in the population. It is defined as a Mendelian trait that exists in the population in at least two phenotypes, neither of which occurs at a frequency of less than 1%. Polymorphism at a genetic locus is due to either balanced polymorphism (heterozygous advantage, frequency-dependent selection) or unequilibrium states (temporary polymorphism) as occurs during frequency-dependent selection and genetic drift (alleles becoming fixed or extinct).

Polyprotein A translation product consisting of a series of linked proteins which are processed by proteolytic cleavage to release the mature proteins.

Polyteny When replication of DNA proceeds repeatedly without forming new nuclei in telophase, polyteny is generated if the replicated chromosomes remain in precise alignment forming giant chromosomes (polytene chromosomes in a salivary gland cell of *Drosophila melanogaster*).

Position effect A difference in phenotype that is dependent on the position of a gene or a group of genes, often caused by heterochromatin nearby. Thus, the change in a gene's location may cause a change in its expression (a problem that has to be overcome in gene therapy).

Post-transcriptional gene silencing (or PTGS) Is a gene silencing process that operates in plants. It refers to the inhibition of gene expression in a sequence-specific manner by double-stranded RNAs (dsRNAs). The latter are believed to trigger degradation of the corresponding mRNAs. PTGS serves as an immune system to protect plant cells by degrading the nucleic acid of RNA viruses. It is related to the RNA interference (RNAi) in vertebrates and invertebrates and to quelling in fungi.

Post-translational modifications Cleavage of amino terminal peptide, hydroxylation and oxidation of amino acids in the polypeptide chain for cross-linking, covalent modifications by acetylation, phosphorylation and glycosylation.

Primary and secondary metabolism Primary or basic metabolism refers to all biochemical processes for the normal anabolic and catabolic pathways, which result in assimilation, respiration, transport, and differentiation. Secondary metabolism generates diverse and seemingly less essential or non-essential byproducts called secondary products that are sources of fine chemicals, such as drugs, dyes, flavors, and fragrances. While primary metabolism consists of biochemical pathways that are in general common to all cells of a multicellular organism, secondary metabolisms consist of a large number of diverse processes that are specific to certain cell types.

Prion The causative agent of many diseases that affect mammals e.g. fatal insomnia, kuru and Creutzfeldt-Jakob diseases in humans, scrapie in sheep, etc. This agent is made up purely of protein (it does not seem to have any nucleic acid).

Processivity Refers to the length of the polynucleotide that is synthesized by a DNA polymerase before it dissociates from the template. It is the ability of a DNA polymerase to remain associated with a DNA template for a relatively long time.

Prokaryote A single-celled organism with a simple internal structure and no nucleus. Bacteria and archaebacteria are prokaryotes.

Promoter A DNA regulatory sequence of a gene required for the expression of the downstream open reading frame (the coding part of the gene). Also, a nucleotide sequence that is recognized and bound by a DNA-dependent RNA polymerase (directly or indirectly) during the initiation of transcription. Transcription is initiated at the start point or site located within the promoter sequence. The first nucleotide to be transcribed is designated +1. Nucleotides downstream of +1 position are numbered +2, +3 etc and those upstream of this position are numbered –1, -2, -3 etc.

Proofreading In DNA synthesis, the ability of DNA polymerase to recognize mismatched bases. DNA polymerase corrects mistakes with its exonuclease activity. RNA editing is also possible at the mRNA level in some simple organisms.

Proteasome A multisubunit protein structure that is involved in the degradation of other proteins.

Proteobacteria[6] A group of Gram-negative bacteria, often called purple bacteria that include several phylogenetic subdivisions: the alpha (α), beta (β), gamma (γ), delta (δ) and epsilon (ε) proteobacteria. The α-proteobacteria are believed to be the endosymbiont progenitors of mitochondria.

Proteomics The proteome is the whole set of proteins at work in a given cell under some conditions. A new discipline that belongs to functional genomics and aims at the studying of biological aspects of all proteins at once in a systematic manner.

[6] The Proteobacteria are a major group (phylum) of bacteria. They include a wide variety of pathogens, such as *Escherichia*, *Salmonella*, *Vibrio*, *Helicobacter*, and many other notable genera. Others are free-living, and include many of the bacteria responsible for nitrogen fixation. In 1987, Carl Woese established this grouping, calling it informally the "purple bacteria and their relatives". Because of the great diversity of forms found in this group, the Proteobacteria are named after Proteus, a Greek god of the sea, capable of assuming many different shapes, and it is therefore not named after the genus Proteus.

Protein synthesis Is a complex process that links the nucleotide sequence in a DNA molecule to the amino acid sequence in the polypeptide. This requires the synthesis of a mRNA molecule which then determine the order of the amino acid in the polypeptide chain. Amino acids cannot directly recognize their specific codons in the mRNA, and each need to be attached to a specific tRNA molecule that contain the anticodon (a three nucleotide sequence that is complementary to the codon sequence). The charging of a tRNA with its specific amino acid occurs via a two-step reaction. The first step, called activation, sees the amino acid react with ATP in the presence of a specific aminoacyl-tRNA synthetase to form an enzyme-bound aminoacyl-AMP complex. In the second step, called transfer, the aminoacyl group is transferred to the 2' or 3' position of the terminal adenosine residue in the appropriate tRNA. Interaction between charged tRNAs and the corresponding codons in the mRNA. as well as peptide bond formation, are mediated by a high molecular weight nucleoprotein complex called ribosome. The latter has two binding sites for charged tRNA: the A site (entry or acceptor site) and the P site (peptidyl or donor site). In short, ribosome binds to the mRNA and moves along in such a way that each codon is exposed to the A site and amino acids are linked together in the order dictated by the nucleotide sequence to produce the polypeptide.

Protoctista The modern name for the Kingdom Protista. Probably the first double-chromosomed beings. Well-known members are: seaweeds (algae), amoebas, ciliates, mastigotes, water molds, slime molds and slime nets.

Proto-oncogene A normal cellular gene that, when mutated or inappropriately expressed, can cause a cell to become cancerous. They normally participate in the control of the cell cycle.

Pseudogene A region of DNA that shows extensive similarity to a known gene, but which cannot itself function, either because it has lost the signal required for transcription (the promoter sequence) or because it carries mutations that prevent it from being translated into protein. Pseudogenes are inactivated but stable components of the genome derived by mutation of an ancestral active gene.

Proteome The complete set of proteins encoded by the genome. The complete set of proteins at work in a given cell grown under specific conditions.

Pseudoalleles Genes that behave like alleles but can be separated by crossing over. The eye color genes on the X chromosome of Drosophila are for example closely adjacent but separable loci and not alleles of a single gene.

Pseudogene A sequence of DNA similar to a gene but nonfunctional; probably the remnant of a once-functional gene that accumulated mutations.

Pseudomurein A substance similar to murein but contains N-acetyltalosaminuronic acid instead of N-acetylmuramic acid, characteristically found in Archaea. It lacks the D amino acids found in bacterial cell walls.

Public sequence databases The three coordinated international sequence databases: GenBank, the EMBL data library and DDBJ.

Purple membrane (and red membrane) Corresponds to functionally and structurally differentiated, purple-pigmented regions of the cytoplasmic membrane that develops under microaerobic or anaerobic conditions in the light. The "purple membrane" may occupy up to 50% of the total membrane surface. The remaining area, which contains red carotenoid pigments, is usually referred to as "red membrane".

Q

Quaternary Structure Complexes of 2 or more polypeptide chains held together by noncovalent forces but in precise ratios and with a precise 3-D configuration.

Quelling Is an RNA silencing process that operates in fungi. It is related to post-transcriptional gene silencing in plants and to RNA interference in vertebrates and invertebrates.

Quorum sensing[7] (1) Is a type of decision-making process used by decentralized groups to coordinate behavior. Many species of bacteria use quorum sensing to coordinate their gene expression according to the local density of their population. In similar fashion, some social insects use quorum sensing to make collective decisions about where to nest. In addition to its function in biological systems, quorum sensing has several useful applications for computing and robotics. (2) The phenomenon whereby the accumulation of signalling molecules enable a single cell to sense the number of bacteria (cell density).

R

Raw sequence Individual unassembled sequence reads, produced by sequencing of clones containing DNA inserts.

Recessive trait The opposite of dominant. A trait that is preferentially masked.

Recombinase A group of enzymes that catalyze the joining of two DNA molecules after recognizing the recombination sites.

Recombinant DNA technologies Procedures used to join together DNA segments in a cell-free system (an environment outside a cell or organism). Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular chromosome.

Recombination Genetic recombination is a process in which one or more DNA molecules are rearranged to generate new allelic combinations. It involves an exchange of genetic material between two homologous DNA molecules (nonsister chromatids), the incorporation of one molecule into another, or the inversion of the orientation of a segment within a DNA molecule (resulting from the recombination between two inverted repeated sequences). We distinguish different types of recombination: general recombination, site-specific recombination, transposition and illegitimate recombination.

Recombination frequency (RF) Refers to the number of recombination phenotypes (as in a test-cross ***AaBb*** x ***aabb***, where **Ab** and **aB** represents the recombinant class when genes ***A*** and ***B*** are in coupling phase) to the total progeny, expressed as percentage. RF gives an indication of the relative position of markers (genes or mutations) on the chromosome. A RF of 50% implies that the markers are independent (either located on different chromosomes or present on the same chromosome but separated by a distance higher than 50 centi-Morgan (cM). Only markers located within the 50cM distance limit can be mapped via appropriate test-crosses. The lower the RF the smaller the distance between markers.

[7] Some of the best-known examples of quorum sensing come from studies of bacteria. Bacteria use quorum sensing to coordinate certain behaviors based on the local density of the bacterial population. Quorum sensing can occur within a single bacterial species as well as between diverse species, and can regulate a host of different processes, in essence, serving as a simple communication network. A variety of different molecules can be used as signals. Common classes of signaling molecules are oligopeptides in Gram-positive bacteria, N-Acyl Homoserine Lactones (AHL) in Gram-negative bacteria, and a family of autoinducers known as autoinducer-2 (AI-2) in both Gram-negative and Gram-positive bacteria.

Regulatory protein A protein that is capable of recognizing a specific sequence of within the DNA and binding to that sequence with high affinity thereby altering gene expression.

Regulon (1) A system in which two or more structural genes (usually non-contiguous) and/or operons, each with its own promoter, are subject to coordinated regulation by a common regulator molecule. These genes and operons share common regulatory sequences which can be recognized by the same regulatory molecule e.g. *arg* regulon and *SOS* system. (2) In cell biology and genetics, a regulon is a collection of genes or operons under regulation by the same regulatory protein. This term is generally used for prokaryotic systems, for example quorum sensing in bacteria. It is a group of operons/genes spread around the chromosome but controlled by a common factor or stimulus. Multiple regulons can form a modulon.

Reoviridae A family of viruses in which the virions non-enveloped (about 60-80nm diameter) and contains a segmented genome of 10-12 different linear dsRNA molecules. The capsid contains a total of 6-10 polypeptide species and usually comprises two concentric layers: an inner, apparently icosahedral core surrounded by a more or less labile, icosahedral or amorphous outer layer. The viruses replicate in viroplasms in the host cell cytoplasm, sometimes forming crystalline arrays.

Replichore A term used to designate half chromosome (as in *E. coli*) comprised between the origin and terminus of replication.

Replicon Refers to any DNA sequence that possesses an origin of replication and that has the ability to be replicated in a suitable cell. A unit of genetic material which behaves autonomously during replication of DNA. In bacteria, a whole chromosome is a replicon. In eukaryotes, chromosomes are divided into hundreds of replicons. Each replicon contains a segment beginning with a binding site for RNA polymerase.

Repliconation The process in which a circular double-stranded DNA molecule, capable of replication, is formed within a cell from DNA that has been taken up e.g. conjugation.

Restriction enzyme (endonuclease) A protein that recognizes specific, short nucleotide sequences and cuts DNA at those sites. Bacteria contain over 400 such enzymes that recognize and cut over 100 different DNA sequences. See restriction enzyme cutting site.

Restriction enzyme cutting site A specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA. Some sites occur frequently in DNA (e.g., every several hundred base pairs), others much less frequently (rare- cutter; e.g., every 10,000 base pairs).

Restriction mapping A physical mapping of the position of restriction sites in a DNA molecule by analyzing the sizes of restriction fragments using gel electrophoresis.

Retotransposons They refers to mobile elements that transpose or move via an RNA intermediate molecule; the DNA element is transcribed into RNA, and then reverse-transcribed into DNA, which is then inserted at a new site in the genome.

Riboswitch Protein repressors and corepressors are not the only way in which bacteria control gene transcription. It turns out that the regulation of the level of certain metabolites can also be controlled by riboswitches. A riboswitch is section of the 5'-untranslated region (5'-UTR) in a molecule of messenger RNA (mRNA) which has a specific binding site for the metabolite (or a close relative). Some of the metabolites that bind to riboswitches: the purines adenine and guanine, the amino acids glycine and lysine, and flavin mononucleotide (the prosthetic group of NADH dehydrogenase). In each case, the riboswitch regulates transcription of genes involved in the metabolism of that molecule. The metabolite binds to

the growing mRNA and induces an allosteric change that for some genes causes further synthesis of the mRNA to terminate before forming a functional product and for other genes, enhances completion of synthesis of the mRNA. In both cases, one result is to control the level of that metabolite (a kind of feedback inhibition). Some riboswitches control mRNA translation rather than its transcription. It has been suggested that these regulatory mechanisms, which do not involve any protein, are a relict from an "RNA world".

RFLP Restriction Fragment Length Polymorphism: genetic polymorphism in the restriction sites resulting in a variation in the length of DNA fragments generated by the corresponding restriction enzymes. RFLP are markers usually used in mapping. RFLP are biallelic.

RFLP and disease diagnosis Restriction-Fragment-Length Polymorphism is a powerful technique in molecular biology that is particularly useful in diagnosing dominant and sex-linked genetic diseases. By treating the DNA from somatic cells of different individuals with restriction enzymes and then subjecting the resultant fragments to electrophoresis on a gel strip, one can sort the DNA fragments according to size. If radioactive probes are then applied to these gels, one can recognize specific DNA fragments. Because of the extensive DNA polymorphism in different individuals, the patterns so obtained will generally be different. By performing this procedure on various members of a family in which, for example, a dominant genetic disease occurs, it may be possible to identify a particular fragment or set of fragments that appear only in those family members who display the disease. This approach is being applied to the identification of specific gene defects in such conditions as Huntington's chorea.

RNA interference (RNAi) Is a gene silencing process that inhibits the activity of a gene by the introduction of double-stranded RNAs (dsRNAs) of sequence specific to the target gene. The specificity and potency of RNAi make it an ideal approach to study the function of genes beginning only with the genomic sequence. Recently, the RNAi strategy has been used at a genomic scale to analyze the function of nearly all the predicted genes of some chromosomes of the nematode *C. elegans*.

RNA polymerase IV Is an enzyme which synthesizes small interfering RNA (siRNA) in plants. Polymerase IV is specific to plants genomes and is required for the synthesis of over 90% of 24-nt heterochromatic siRNA. RNA polymerase silences the transposons and repetitive DNA in the siRNA pathway. The siRNA plays a major role in defending the genome against the invading viruses and transposable elements by RNA directed DNA methylation. Polymerase IV and ROS1 demethylase unlocks and recondenses the 5S rDNA chromatin, which is present in seed and used for the development of adult features in plants. Polymerase IV is involved in setting the methylation patterns in the 5S genes during plant maturation. In arabidopsis polymerase IV works with binding protein DCL3 and a RNA polymerase II RDR2 in a silencing pathway which Polymerase IV would produce RNA, which is changed to dsRNA by RDR2 then converted to siRNA by DCL3.

RNA Polymerase V[8] Is a nuclear RNA silencing enzyme recently shown to generate noncoding transcripts at loci silenced by 24nt siRNAs.

[8] (1) Arabidopsis Pol V, AGO4, DMS3 and the putative chromatin remodeller, DRD1 function in the silencing of siRNA-homologous loci at one or more steps downstream of siRNA biogenesis. Recently, we showed that DRD1 facilitates Pol V transcription of noncoding RNAs at target loci, revealing a functional relationship between these two activities. However, the functional relationships, if any between AGO4, DMS3 and Pol V transcription are unclear. (2) Plants encode subunits for a fourth RNA polymerase (Pol IV) in addition to the well-known DNA-dependent RNA polymerases I, II, and III. By mutation of the two largest subunits (NRPD1a and NRPD2), we show that Pol IV silences certain transposons and repetitive DNA in a short interfering RNA pathway involving

RNases in *E. coli* RNase I is an endoribonuclease, which can degrade most RNA molecules to oligonucleotides with 3'-phosphate and 5'-hydroxyl termini. RNase II is an exoribonuclease, which removes nucleotides (nucleoside 5'-PO_4) from the 3' end of an RNA molecule. RNase III is an endoribonucleotide, which cleaves double-stranded regions in RNA molecules i.e. the target site resides in stem-and-loop or hairpin structures in RNA.

Rolling-circle replication A mode of replication used by circular DNAs, thus generating molecules that look like lariats. It was traditionally associated with certain bacterial plasmids and viruses, but is increasingly recognized as having a wider distribution.

S

Saccharomyces cerevisiae Unicellular Ascomycete yeast known as the baker's or brewer's yeast. Widely used as a simple eukaryotic model, particularly in recombinant DNA and cell cycle studies as well as in mating type and heterokaryon compatibility studies. It has most advantages of a prokaryotic system but is a true eukaryote. It is considered as the *E .coli* of the eukaryotes. *S. cerevisiae* can reproduce both asexually and sexually, and can be cultured in either the haploid or the diploid state. One major advantage of yeast is the ease with which specific gene disruptions, gene replacements, and gene retrievals can be accomplished. Its complete genome was sequenced in 1997 and contains 12,057,500 bp, 6,000 genes in 16 chromosomes. It is used in the creation of YACs.

Scaffold The result of connecting contigs by linking information from paired-end reads from plasmids, paired-end reads from BACs, known messenger RNAs or other sources. The contigs in a scaffold are ordered and oriented with respect to one another.

Satellite DNA Repetitive DNA that forms a satellite band in a density gradient. In the satellite DNA is made up of fragments containing long series of tandem repeats. Different types of satellite DNAs have been described, each with a different repeat unit; these units are anything ranging from <5 to >200bp.

Second messenger Is an intermediate in a certain type of signal transduction pathway. Second messengers are considered as less specific signaling compounds that transduce the signal from a cell surface receptor in several directions, so that a variety of cellular activities is affected.

Selection Coefficient (s) s = 1 - W where W is relative fitness. This coefficient represents the relative penalty incurred by selection to those genotypes that are less fit than others. When the genotype is the one most strongly favored by selection its s value is 0.

Sequence-contig scaffolds Scaffolds produced by connecting sequence contigs on the basis of linking information.

Sequence motif A short conserved amino acid (or nucleotide) sequence pattern that represents a specific functional site of a protein (or nucleic acid) molecule.

Sex-linked A gene coded on a sex chromosome, such as the X-chromosome linked genes of man.

Shine-Dalgarno (S-D) sequence An eight nucleotide consensus sequence 5' UAAGGAGG 3' found in bacterial mRNAs five to ten bases before the translation initiation codon (AUG). It is thought to be involved in initiation of translation by helping the mRNA bind to the

---

RNA-dependent RNA polymerase 2 and Dicer-like 3. The existence of this distinct silencing polymerase may explain the paradoxical involvement of an RNA silencing pathway in maintenance of transcriptional silencing. Herr *et al.* RNA polymerase IV directs silencing of endogenous DNA. 2005. *Science* **308**:118-20.

ribosome (16S rRNA), thus it can be called the Ribosomal Binding Site (RBS). In eukaryotic DNA, there is no such sequence. The 5' cap present on all eukaryotic mRNAs seems to be the first signal to start protein synthesis.

Shotgun sequencing Breaking total genomic DNA (or simply a part of the genome) into a large number of small pieces, sequencing the pieces and assembling all the fragment to generate the final continuous sequence of the DNA molecule.

Signal transduction A complex multistep pathway by which extracellular signals are transduced from plasma membrane receptors to the transcription machinery in the nucleus and the translation machinery in the cytoplasm, subsequently to regulate cell proliferation and differentiation. The components are growth factors, growth factor receptors, membrane and cytoplasmic tyrosine kinases, GTP-binding (G) proteins, nuclear binding proteins and transcription factors.

Signal transduction pathways Pathways that respond to signals such as nonsteroid hormones, which bind to specific receptors present on the cell surface. Once activated, these pathways mobilize various second messengers e.g. cyclic nucleotides, Ca^{2+} ions, and other substances leading to the activation or inhibition of enzymes or enzymatic pathways in a very specific fashion.

Slime moulds A category of eukaryotic organisms, which typically have some fungus-like attributes and some animal-like attributes. As commonly used, the term slime moulds refers specifically to the so-called true or acellular slime moulds and the cellular slime molds.

Small nuclear ribonucleoproteins or snRNPs Are particles that occur in the nucleus of eukaryotic cells. They consist of one snRNA plus one or more proteins. SnRNPs are involved in a variety of functions such as pre-mRNA splicing, histone mRNA 3' end processing, ribosomal RNA processing, telomere replication and tRNA maturation.

SNP Single nucleotide polymorphism, or a single nucleotide position in the genome sequence for which two or more alternative alleles are present at appreciable frequency (traditionally, at least 1%) in the human population. Most genetic variation between individual humans is believed to be due to SNPs.

SOS response A series of biochemical changes that occur in *E. coli* in response to damage to the genome and other stimuli. The SOS response is induced when the genome of the bacterium suffers extensive damage as a result of exposure to e.g. UV radiation or chemical mutagens. The SOS response enables the cell to repair and replicate its damaged DNA.

SSLPs or Simple Sequence Length Polymorphisms are arrays of repeat sequences that display length variations i.e. different alleles (different length variants) containing different numbers of repeat units.

SSR Simple Sequence Repeat, a sequence consisting largely of a tandem repeat of a specific *k*-mer [such as $(CA)_{15}$]. Many SSRs are polymorphic and have been widely used in genetic mapping.

Stationary phase In a batch culture (or closed culture) refers to the phase where the growth rate declines and eventually reaches zero.

Stimulon Is a collection of genes (which may be in operons and regulons) under regulation by the same stimulus. This term is generally used for prokaryotic systems, for example quorum sensing in bacteria.

Stringent control In bacteria, a control mechanism in which cells deprived of an essential amino acid show a rapid decrease in the synthesis of rRNA and tRNA and hence in protein

synthesis. The cells may also exhibit increased rates of protein degradation, inhibition of DNA replication and induction of heat shock proteins. The trigger for the stringent response is the presence of uncharged tRNA in the A site of a ribosome. The stringent control is dependent on the product of the *relA* gene, also called stringent factor that is necessary for the synthesis of guanosine 5' -diphosphate 3' -diphosphate (ppGpp or guanosine tetraphosphate). Mutants *relA*$^-$ do not exhibit the stringent response when starved for amino acids and are said to be relaxed. Note that ppGpp apparently acts as an alarmone that inhibits e.g. transcription; yet the precise mode of action remains unknown.

Structural gene Is a gene that codes for an RNA molecule or a protein (structural or enzymatic) other than a regulator.

STS Sequence Tagged Site, corresponding to a short (typically less than 500 bp) DNA sequence that has a single occurrence in the human genome, whose location and base sequence are known, and for which a polymerase chain reaction assay has been developed. STSs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as landmarks on the developing physical map of the human genome. Expressed sequence tags (ESTs) are STSs derived from cDNAs.

Subgenomic mRNA Is an mRNA of an RNA virus that is produced in an infected eukaryotic cell and is shorter that the genomic RNA. In eukaryotic cells, translation cannot be initiated from an internal site in a mRNA, so that the formation of a subgenomic mRNA allows an internally-located initiation site to become terminally. This permits independent expression of genes within the genome. Subgenomic mRNA may be monocistronic (or monogenic when the transcript contains a single gene) or polycistronic (or polygenic when the transcript entails two or more genes).

Supercoiling A conformational state in which a double helix is overwound or underwound so that superhelical coiling occurs. DNA supercoiling is analogous to twisting or untwisting a multistrand rope so that it is torsionally stressed. Negative supercoiling introduces a torsional stress that favors unwinding of the right handed double helix, whereas positive supercoiling overwinds such a helix. Both forms of supercoils tend to compact DNA, thus affecting its sedimentation speed during ultracentrifugation and its migration on an electrophoretic gel. Both sedimentation and migration of supercoiled DNA will be faster than the relaxed form.

Surface exclusion (also called entry exclusion) the phenomenon in which the presence of a conjugative plasmid is a cell reduces the ability of that cell to receive an identical or related plasmid, by conjugation, from another cell. In *E. coli F* plasmid conjugal system, surface exclusion involves the products of genes *traS* (a cytoplasmic membrane protein) and *traT* (an outer membrane protein). The *traT* product appears to inhibit stable contact between donor cells. The effect of surface exclusion may be lost if e.g. donor cells progress to the stationary phase of growth or are starved of amino acids. Such cells, which can act as recipients, are called F^- phenocopies.

Svedberg Unit (S) Is the unit in which the sedimentation coefficient of a molecule or particle is commonly quoted. The sedimentation constant ***s*** is defined by the following formula: $s=dx/dt \times 1/\omega^2 r$, in which ***dx/dt*** is the measured sedimentation rate (in the ultracentrifuge), ***w*** is the angular velocity (radians sec^{-1}), and ***r*** is the distance in cm between a point within the sample and the axis of rotation. s is expressed in sec-1. The Svedberg unit is equal to 10^{-13} sec. Therefore, if the bacterial small ribosomal subunit has a sedimentation coefficient of 30S, this means that s is equal to 30×10^{-13} sec. This is obtained if the viscosity of the liquid medium were equal to that of water at 20ºC ($S^{20}{}_w$).

Symbiosis Is a term used to describe any stable condition in which two different organisms, called symbionts, live in more of less close physical association. The nature of relationship between symbionts ranges from mutually beneficial (mutualism) to antagonistic (parasitism).

Synapomorphy An apomorphy (derived or specialized character) shared by two or more groups, which originated in their last common ancestor.

Synteny Refers to the partial or complete conservation of gene order when two genomes are compared. Two genomes are considered syntenic when a least some of the genes are located at similar map positions.

T

Tandem mass spectrometry (MS/MS) In MS/MS an ionized peptide of interest is selected by the first MS and fragmented by collision with inert gas and the resulting fragments are then analyzed by the second MS. It generates information about the amino acid sequence.

Tetrad The duplicated homologous chromosomes held together by chiasmata prior to anaphase of meiosis I. Each tetrad contains four chromatids. The term tetrad refers also to the four products of meiosis kept together like the four spores in the ascus of budding yeast.

Thermal Melting Profile of DNA (TMP) Refers to a plot of UV-absorption of DNA at 260nm against temperature for a given sample of dsDNA. Note that melting and production of ssDNA is concomitant with an increase as much as 40% in the UV absorption. This increase is also termed hyperchromic shift and depends on the higher potential of π electron transition of the aromatic bases in ssDNA compared to dsDNA; the midpoint of the TMP corresponds to the T_m.

Topoisomers Are topologically different forms of the same DNA molecule. For instance, a circular dsDNA may occur in distinct forms having different topological properties such as relaxed form, negatively and positively supercoiled forms. Interconversion between topoisomers (having different linking number) involves breakage and rejoining steps; these reactions are catalyzed by enzymes called topoisomerases.

Topoisomerase Refers to any enzyme that is able to convert one topoisomer into another, thus altering its supercoiling degree. This changes the linking number of the dscccDNA molecule. We distinguish two main types of topoisomerases. Type I, termed untwisting, relaxing or nick closing enzymes, introduce a single nick in one strand of a DNA duplex, then the unbroken strand is passed through the gap before resealing; this reaction leads to a change in the linking number in steps of one. Type II break both strands (ds cut), then a double-stranded segment from elsewhere in the molecule is passed through the break and free ends are sealed without strand rotation or twisting; this changes the linking number in steps of two.

Transcription (1) The process of copying a gene into RNA. This is the first step in turning a gene into a protein, although not all transcripts lead to proteins. (2) Before the synthesis of a protein begins, the corresponding RNA molecule is produced by RNA transcription. One strand of the DNA double helix is used as a template by the RNA polymerase to synthesize a messenger RNA (mRNA). This mRNA migrates from the nucleus to the cytoplasm. During this step, mRNA goes through different types of maturation including one called splicing when the non-coding sequences are eliminated. The coding mRNA sequence can be described as a unit of three nucleotides called a codon.

Transcription factor A protein needed to initiate the transcription of a gene. Some transcription factors bind to specific sequences of DNA (promoters and enhancers); others bind to each other; many bind both to DNA as well as to other transcription factors.

Transcriptome The complete set of RNAs transcribed from a genome in a particular cell at a particular time. The trancriptome should be considered as a dynamic link between an organism's genome and its physical characteristics; it is modulated by both external and internal factors conveying thus the identity of each expressed gene and its level of expression for a defined population of cells.

Transduction Is a type of genetic transfer from donor cell to recipient cell mediated by viruses. When a phage replicates within the host cell, a few progeny phages receive (encapsidate) pieces of host DNA. These virions can adsorb to the host cell and inject their DNA in the normal way.

Transformation The process during which many bacteria can acquire new genes by taking up DNA molecules (e.g., a plasmid or linear DNA molecule) from their surroundings. For instance, the development of protocols to deliberately transform the gram-negative bacterium *E. coli* has made possible the cloning of many genes and the development of the biotechnology industry.

Translation (1) The process of using a messenger RNA sequence to build a protein. The messenger RNA serves as a template on which transfer RNA molecules, carrying amino acids, are lined up. The amino acids are then linked together to form a protein chain. (2) The ribosome binds to the mRNA at the start codon (AUG) that is recognized only by the initiator tRNA. The ribosome proceeds to the elongation phase of protein synthesis. During this stage, complexes, composed of an amino acid linked to tRNA, sequentially bind to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. The ribosome moves from codon to codon along the mRNA. Amino acids are added one by one, translated into polypeptidic sequences dictated by DNA and represented by mRNA. At the end, a release factor binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.

Transposable element (TE) Also termed jumping gene, jumping element or transposon, is a DNA segment that has the capacity to translocate or "jump" from one site to another site, called target site, within the same or between two different DNA molecules. This process called transposition does not require extensive sequence homology between the TE and the target site. Moreover, transposition may be either replicative (a copy of the TE remains in place while another copy appears at another position in the genome) or conservative (the original TE jumps to a new position in the genome). TE occurs in prokaryotes and eukaryotes. Most TE have at least one open reading frame that codes for the transposase, a protein required for transposition. Also, each TE is flanked by two direct repeats that seem necessary for the recognition of the TE by its transposase.

Transposition Is a process that uses recombination (rather than a type of recombination) to transfer a DNA segment, termed transposon or transposable element TE), from one site to a new position in the genome. Usually, transposition is a rare event but its actual frequency depends on the TE, the physiological state of the cell, the growth conditions, etc. The functions required for transposition are usually encoded by the TE itself. However, certain host functions may also be needed.

Triplet repeat In this situation, a triplet of nucleotides increases in number within a gene. A mutation especially occurring in central nervous system disorders is the increased number of

triplets repeats. Examples include myotonic dystrophy, Huntington's disease, Friedreich's ataxia and fragile X syndrome. Also in polycystic ovary syndrome, androgen receptor gene has increased number of CAG repeats. Expansion may be greater depending on the transmitting parent (eg, the mother in myotonic dystrophy, the father in Huntington's disease); thus, a parent-of-origin effect and genetic anticipation can be observed. Increased number of repeats of a triplet may trigger methylation of the gene that causes the disease.

Typhoid fever Not to confuse with typhus fever, is an acute infectious human disease caused by the gram-negative bacterium Salmonella typhi. Transmission occurs via contaminated food or water. The incubation period ranges from 7 to 15 days. Symptoms include malaise, headache, and fever that increases stepwise over several days. Inflammation of the lymphoid tissues in the ileum is characteristic and may lead to necrosis and bowel perforation or hemorrhage which may be fatal. The chemotherapy is based on chroramphenicol, ampicillin, amoxycillin, etc. Inactivated typhoid vaccines are available.

Two-component signal-transduction systems Comprise two interacting proteins: a histidine kinase sensor with an extracellular domain and an intracellular response regulator. Following binding of the adequate ligand to the input domain of the first component, auto-phosphorylation of the second domain ensues, leading to the activation by phosphorylation of the response regulator. The latter may be a transcription factor or an intermediate in a regulatory pathway. Induction of the two-component system results in adaptive response via changes in gene expression profile.

Typhus fever Also called louse-borne typhus is an acute infectious disease caused by *Rickettsia prowazekii*. Mortality rates in untreated cases may be 5 to 70%. Treatment is based on the use of tetracycline and chloramphenicol.

U

Ubiquitin A protein-regulating modification. Usually it is a small protein that becomes covalently linked to a protein targeted for degradation. However, and since its discovery 30 years ago, this molecule has been linked to many cellular processes including transcriptional regulation most likely via histone ubiquitination.

Underdominance Also called heterozygous disadvantage. This unusual selection process occurs when heterozygotes are less fit than either homozygote. This situation is likely to arise when two adjacent populations are isolated and become homozygous for different alleles, and then come into secondary contact at the borders of their ranges. This is the opposite of overdominance.

Uniparental disomy Inheritance of both homologues of a chromosome from one parent, with loss of the corresponding homologue from the other parent.

Universal tree of life The relationships of all living species based on the phylogeny of rRNA sequences. The universal tree topology shows the three domains (Eukarya, Bacteria and Archaea) of life to be separated but have a unique ancestor (monophyletic). Independent protein phylogenies show that the Archaea and Eukarya are sister groups, so the tree must be rooted in the Bacteria.

V

Variable expression A variation in phenotype between affected members of the same family (i.e. individuals carrying identical mutations). It occurs in many dominant conditions and may be associated with reduced penetrance

V-DNA An *in vitro* form of double-stranded covalently closed circular DNA (ds cccDNA) produced by annealing two complementary ss cccDNA molecules without strand breakage. The resulting V-DNA is characterized by a linking number of zero. Any regions of right-handed configuration must be compensated for by negative supercoiling and/or regions of left-handed conformation.

Vector A plasmid, phage, or cosmid into which foreign DNA may be inserted for cloning.

Virion The extracellular form of an infective virus. It refers to a complete virus particle that consists of RNA or DNA core with a protein coat and sometimes with external envelopes.

Viroid Any of several causative agents of plant disease that consist solely of a single-stranded RNA of low molecular weight arranged in a closed loop or a linear chain.

Viroplasm Also called virus factory or viroplasmic matrix, is a type of inclusion body that is the site of virus replication and/or assembly in a virus-infected cell. A viroplasm generally consists of accumulations of virions and/or virus components.

Virus A submicroscopic infective agent belonging to a large group. Viruses, which are considered either as extremely simple microorganisms or as extremely complex molecules, contain a protein coat surrounding an RNA or DNA genome. They lack any semipermeable membrane, are able to grow and multiply only in living cells and cause various important diseases in humans, animal and plants.

W

Weismann's hypothesis A theory proposed by August Weismann in 1880 stating that genetic information of a cell diminishes with each cell division. According to this theory, only the first cell (the zygote) is totipotent; then as the cell divides the progressive loss of genetic material would account for differentiation since each cell type must keep the very genes that are needed for its functions. Weismann hypothesis was turned down by Hans Spemann in 1902. Spemann divided a salamander embryo in two and showed early embryo cells retain all the genetic information necessary to create a new organism.

Wild-type allele The non-mutant form of a gene, encoding the normal genetic function. Generally, but not always a dominant allele.

Wobble hypothesis It accounts for the third base degeneracy in the genetic code i.e. it explains why the same tRNA recognizes many codons that differ only in the 3rd position of a codon. According to this hypothesis normal base-pairing occur between the 1^{st} and 2^{nd} bases of the codon and their complementary 3^{rd} and 2^{nd} bases of the anticodon, respectively. The 1^{st} position of the anticodon, also termed, wobble position, can undergo a nonstandard base-pairing with the 3^{rd} base of the codon. For instance, G in the wobble position can base-pair with either C or U in the 3^{rd} position of the codon, and similarly U can base-pairs with either A or G.

X

Xenopus An amphibian (frog) who shared a common ancestor with mammals about 350 million years ago. The oldest species in which all three regions of the MHC are linked. Its eggs are very large and have front-to-back orientation even before they are fertilized.

X-gal it is the 5-bromo-4-chloro-3-indolyl-β-D-galactoside. Hydrolysis of X-gal by β-galactosidase yields a blue-green substance that adheres to the bacterial cells, so that cells that express β-galactosidase form blue-green colonies on X-gal containing plates.

Y

YAC Yeast Artificial Chromosome: a yeast DNA molecule containing all the components of a normal chromosome, it is constructed from the telomeric, centromeric, and replication origin sequences needed for replication in yeast cells. It is used as a vector to clone and amplify large fragment of a genome (up to 400 kb).

Yeast artificial chromosomes (YAC) An artificial chromosome created from DNA, centromere and telomere of yeast chromosomes. Heavily used in cloning of very large genomic fragments.

Z

Z-DNA A left-handed helical form of dsDNA first observed in oligomers of alternating dC and dG in the presence of high salt concentrations or ethanol. In these oligomers the strands are connected by Watson-Crick base pairing, the sugar-phosphate backbones form an irregular zig-zag (hence the name Z-DNA that contains approximately 12 bp/turn). This form of DNA has a single very deep helical groove corresponding to the minor groove of B-DNA. Stretches of alternating dC and dG in a plasmid can undergo transition from right to left-handed helical form under physiological levels of ionic strength and superhelical density. Whether or not Z-DNA occurs *in vivo* remains controversial. However, the isolation of proteins that bind preferentially to Z-DNA rather than B-DNA argues in favor of the existence of Z-DNA form *in vivo*.

Zebrafish A model organism (*Brachydanio rerio*) to study vertebrate biology, physiology and human disease. Its high fecundity and short generation time make it useful for genetic studies as well. Another useful feature is that their fry are transparent. Hundreds of mutants resembling human diseases have been identified.

Zinc finger protein A DNA-binding domain of a protein that has a characteristic pattern of cysteine and histidine residues that complex with zinc ions. This motif occurs in several types of eukaryotic transcription factors.

*****