

Elementary Genetic Terminology

Genome

The entirety of an organism's hereditary information. It is encoded in DNA (DeoxyRibonucleic Acid), a polymer consisting of four types of nucleotides: adenine (A), guanine (G), thymine (T), and cytosine (C). The nucleotides are organized in two polynucleotide chains that form a double helix with A-T and G-C base pairings. In eukaryotic cells (animal, plants, fungi), the cell nucleus contains several such DNA double strands, called **chromosomes**. In prokaryotic cells (bacteria), DNA typically forms a single ring (bacterial chromosome). Each chromosome can contain several Million base pairs (*megabases*). The length of entire genomes ranges from several thousand base pairs (RNA genomes of viruses, e.g. HIV around 10.000), to several 100.000 to millions in bacteria and several hundred billions in prokaryotes (human 3.2 billion: about 2 m total length, *Drosophila* 130 million).

Genes, Loci, and Alleles

Functional units of the DNA are called genes. Most important are protein coding genes. Proteins are polypeptide chains that are composed of (typically several hundred) amino acids. There are 20 different amino acids. A gene then corresponds to a DNA region that are translated (via RNA) into a sequence of amino acids. According to the *genetic code*, three consecutive nucleotides (= one *codon*) code for one amino acid. Due to the redundancies of the genetic code, several codons may correspond to one amino acid. There are also start and stop codons to define the begin and the end of the gene. The position of the gene on the DNA is called its *locus*. Variants in the specific sequence at a given gene locus are called different *alleles* of the gene. Allelic differences may or may not have consequences on the level of the amino acid sequence or the gene function. The total number of genes in higher organisms is typically several 10.000. In humans this number is with an estimated 20.000 not particularly high.

In population genetics, the terms *locus* and *allele* are also used in a slightly different way. Here, a locus refers to any stretch of DNA (protein coding or not) that can be treated as a recombinational unit. I.e. it is inherited as a whole from parents to offspring. Alleles correspond to different sequence variants or equivalence classes of sequences at the locus. At the most basic level, a molecular locus is a single nucleotide site with four alleles corresponding to the bases A, T, C, and G. In models, it is often sufficient to consider just two allelic classes, corresponding to an ancestral *wildtype* and a mutant.

Haploids and Diploids

Most simple life forms (bacteria, mosses, algae, fungi) have a single copy of each chromosome, they are haploid. All higher plants and animals are diploid, i.e. most of their chromosomes (the so-called *autosomes* = non-sex-chromosomes) are represented twice in each adult cell (many plants and some animals have even a higher ploidy level, e.g. *tetraploid* with a fourfold set). They thus carry two (or more) copies of each gene or any

genetic locus of interest. An exception are the sex chromosomes, where the heterogametic sex (XY-males in mammals, or ZW-females in birds) has a pair of two different chromosomes. In contrast to the somatic cells, the egg and sperm cells (the *gametes*) that are used for sexual reproduction contain only a single chromosome set, i.e. they are haploid.

Homozygote and Heterozygote

A zygote is a cell that results for the fusion of an egg and sperm cell and develops into a whole organism. A zygote thus possesses the full (diploid) chromosome set of the cells in the adult organism. A zygote that carries the same allele twice at a genetic locus of interest is called homozygote, if it carries two different alleles, it is heterozygote at this locus.

Genotype and Phenotype

The phenotype of an organism refers to its observable characteristics or traits. The phenotype includes very different aspects, such as morphological traits, biochemical processes or behavioral patterns. Depending on the study, only a (small) subset of all these characteristics is considered (e.g. the state of a single trait). The phenotype of an individual is influenced by heritable and non-heritable factors. Non-heritable factors are collectively referred to as the *environment*. The heritable factors due to the genetic composition of the individual are called the corresponding (underlying) genotype. Phenotypes may also be influenced by other heritable components, such as the genotypes of other individuals in the population.