Chapter 1

Living systems and basic genetic concepts

Music is in my DNA! Rihanna

1.1 The organism and genetics

We here present a number of facts about organisms and some background material about genetics that will be used in this course.

Life on Earth presumably finds its origins at the bottom of the ocean, in a phase transition that took place in hydrothermal vents (fissures in earths surface from which geothermally heated water is released). Regardless of its exact nature, life's origin led to the emergence of **self-replicating molecules**, which evolved over millions of years into the first reproducing organisms (about 3.9×10^9 years ago). Over billions of years, these organisms subsequently evolved into the organisms we encounter today. Earth is now populated by about 10 to 50 millions of plant and animal **species**, which are groups of distinct organisms that cannot interbreed. There are also billions or trillions of bacteria types and untold numbers of fungi, algae, and archae.

Every living organism, from bacteria to plants to animals, is made of **cells**, whose number is about 3×10^{13} in an adult human. Cells are the basic structural and functional biological units. While of different types (skin cells, muscle cells, neurons, blood cells, stem cells, and others), all cells have common characteristics, namely an outer envelope, a membrane that surrounds the cytoplasm, which in turn contains the many biomolecules making up the cell. Cells of higher organisms (such as plants and animals) also contain a **nucleus**, a centrally located spherical body that contains the molecule **deoxyribonucleic acid** or, for

short, **DNA**. This is the molecule that carries the **hereditary information** necessary for producing various types of cells. DNA regulates cellular function and interactions with other cells, and ensures the transmission of that information to the next generation of cells. Thus, the DNA molecule carries the hereditary information used in the growth, development, and reproduction of living organisms, and is consequently behind the synthesis of an individual.

Hereditary information is encoded as follows. DNA is a long and complex molecule for which the key ingredients are four smaller and canonical units called **nucleobases** (or bases for short): adenine (A), guanine (G), thymine (T), and cytosine (C). These (nucleo)bases are linked to each other in a sequence forming a single strand of DNA. Pairs of strands are arranged in a double helix, thus binding pairs of bases together, these pairs being always A-T and G-C. Therefore, although DNA is a double-stranded sequence of nucleobases (sometimes called sites), the information carried by DNA lies in the sequence along a single strand, say a sequence ATGCACGTATAG (the actual total DNA sequence in a human is about 3×10^9 nucleotides long). In specific regions of DNA, called the **coding regions**, three bases in a row, say GCA, code for an amino acid, which is a basic organic compounds and there are twenty different types of amino acids. Thus, DNA strands in the coding regions are **translated** into chains of amino acids. One or several folded chains of amino acids form a **protein**, which is the canonical biological molecule. The proteines (about 90 000 in a human body) are the biomolecules performing the functions within the cell, including metabolic reactions, DNA replication, responses to stimuli, response to pathogen, and transport of molecules from one location to another, possibly across cells.

Overall DNA is thus translated by the cell machinery into proteins, and proteins themselves govern the main biological functions of the organism. We can summarize this by the following arrow of causation, which represents the **central dogma of molecular biology**:

$$\text{DNA} \quad \longrightarrow \quad \text{proteins} \quad \longrightarrow \quad \text{organismic features.} \tag{1.1}$$

This says that DNA influences organismic features, but not the other way around (there are exceptions to this rule but they are beyond the scope of this course).

The **genome** of an organism is the totality of its DNA[1], but not all of the DNA encodes biological information. DNA is usually decomposed into regions of interest, whose specific structure and functionality is studied by **genomics**. A **gene** is roughly the region of DNA that codes for an amino acid chain and thus for a single and simple protein, the cell's functional unit (about 20 000 genes in a human body, but some genes affect several proteines so the number of proteines is larger). The particular position of a gene in the DNA is its **locus** and a particular DNA sequence is an **allele**. In other words, an allele is thus a variant of a gene. The linear arrangement of genes interspaced with non-coding regions (sometimes called the "junk DNA"), along molecules of DNA, forms a **chromosome**. The number of chromosomes is characteristic to each species and the total number of chromosomes make

[1]Some viruses have a slightly different encoding of heredity information and use RNA (ribonucleic acid) molecules instead of DNA.

up about the whole DNA sequence of an organism. If an organism carries only a single set of chromosomes, then the organism is called **haploid**. Organisms like mosses and bacteria are haploid and tend to reproduce **asexually**, by which offspring arise from a single parent whose chromosomes they inherit, and thus their genes originate from that parent only.

Other organisms, like trees, humans and other animals carry in each cell two copies of each chromosome and are called **diploid** (human have 23 pairs of chromosomes: 22 pairs of autosomes and one pair of sex chromosomes, which determines the sex of an individual). Diploids tend to reproduce **sexually**, meaning offspring arise from the fusion of reproductive cells of two parents and thus inherit the genes of two distinct individuals. Under sexual reproduction, the pair of chromosomes within an individual separate at **meiosis** to form gametes, and each gamete is haploid (sperm or egg) carrying a single set of chromosomes. Fertilization between a sperm (gamete carrying genes from a Dad) and an egg (gamete carrying genes from a Mom) form the zygote, the initial cell out of which the organism develops. The distinctive consequence of sexual reproduction is that **recombination** during meiosis results in the formation of non-parental gametes by combining on a single gamete some of an individual's maternal alleles and paternal alleles. Hence, genes are inherited but not whole chromosomes, since alleles from two different chromosomes may be reshuffled at meiosis via crossing-over (exchange of genetic material).

The **phenotype** of an individual comprises any molecular, physiological, morphological, or behavioral trait that can be observed (possibly using a microscope for the cellular machinery). The **genotype** of an individual is the set of all of its genes and thus, consists of the DNA sequences affecting its phenotype. The concepts of genotype and phenotype allow us to abstract away from molecular details about DNA and instead look at the central dogma of molecular biology (eq. 1.1) in the following simplified way:

$$\text{genotype} \quad \longrightarrow \quad \text{phenotype(s)}. \tag{1.2}$$

This implies that the genotype (the set of an individual's genes) influences the phenotype, but not the other way around. The phenotype can of course also be influenced by the environment, and it is more accurate to write

$$\text{genotype} \times \text{environment} \quad \longrightarrow \quad \text{phenotype(s)}, \tag{1.3}$$

which means that the genotype interacts with the enviroment to produce an individual's phenotype, an important point to which we will return later in this course.

1.2 The two hallmarks of living systems

In addition to the distinctive arrow of causation going from genotype to phenotype, living systems are also characterized by two striking hallmarks.

• **Diversity**. Organisms come in a tremendous diversity of forms. From bacteria, to spiders, to elephants, the diversity of phenotypes observed in nature is staggering. Probably up to 10 billion species have existed on Earth since the origins of life. In addition, every individual goes through an impressive number of changes during its lifespan and thus itself expresses diversity in physiology, morphology, and behavior (e.g., we all go through the stages of toddlerhood, adolescence, adulthood, etc).

• **Functional organization (agency)**. From molecular, cellular, and physiological structures within individuals to behavioral interactions between them, organisms display purposefulness in form and goal-directedness in behavior. In other words, living systems display agency and functionality (e.g., the molecule haemoglobin is built to capture oxygen, eyes are built to see, secondary sexual traits to attract mates, etc.). This functionality is so unequivocal that humanity has attributed purpose to animals and plants since the mists of time.

Are there simple and coherent explanations to these observations? The broad goal of this "Genes, Populations, and Evolution" course is to provide an introduction to the quantitative theory of evolution. That means a stylized understanding of the forces changing the genetic components of morphological, physiological, and behavioral phenotypes in a population over time, eventually leading to diversification and individual phenotypes having the appearance of functional organization.