LEARNING OBJECTIVES

- Distinguish homologs, orthologs and paralogs

Homology forms the basis of organization for comparative biology. A homologous trait is often called a homolog (also spelled homologue). In genetics, the term “homolog” is used both to refer to a homologous protein and to the gene (DNA sequence) encoding it. As with anatomical structures, homology between protein or DNA sequences is defined in terms of shared ancestry. Two segments of DNA can have shared ancestry because of either a speciation event (orthologs) or a duplication event (paralogs). Homology among proteins or DNA is often incorrectly concluded on the basis of sequence similarity. The terms “percent homology” and “sequence similarity” are often used interchangeably. As with anatomical structures, high sequence similarity might occur because of convergent evolution, or, as with shorter sequences, because of chance. Such sequences are similar, but not homologous. Sequence regions that are homologous are also called conserved. This is not to be confused with conservation in amino acid sequences in which the amino acid at a specific position has been substituted with a different one with functionally equivalent physicochemical properties. One can, however, refer to partial homology where a fraction of the sequences compared (are presumed to) share descent, while the rest does not. For example, partial homology may result from a gene fusion event.

**Figure: Example of Homologous DNA:** This is the sequence alignment of a homologous protein from two different species. The “**” represent a conserved amino acid in the two proteins.

Homologous sequences are orthologous if they were separated by a speciation event: when a species diverges into two
separate species, the copies of a single gene in the two resulting species are said to be orthologous. Orthologs, or orthologous genes, are genes in different species that originated by vertical descent from a single gene of the last common ancestor. For instance, the plant Flu regulatory protein is present both in *Arabidopsis* (multicellular higher plant) and *Chlamydomonas* (single cell green algae). The *Chlamydomonas* version is more complex: it crosses the membrane twice rather than once, contains additional domains, and undergoes alternative splicing. However, it can fully substitute the much simpler *Arabidopsis* protein, if transferred from algae to plant genome by means of gene engineering.

Significant sequence similarity and shared functional domains indicate that these two genes are orthologous genes, inherited from the shared ancestor. Orthologous sequences provide useful information in taxonomic classification and phylogenetic studies of organisms. The pattern of genetic divergence can be used to trace the relatedness of organisms. Two organisms that are very closely related are likely to display very similar DNA sequences between two orthologs. Conversely, an organism that is further removed evolutionarily from another organism is likely to display a greater divergence in the sequence of the orthologs being studied.

Homologous sequences are paralogous if they were separated by a gene duplication event: if a gene in an organism is duplicated to occupy two different positions in the same genome, then the two copies are paralogous. Paralogous genes often belong to the same species, but this is not necessary. For example, the hemoglobin gene of humans and the myoglobin gene of chimpanzees are paralogs. Paralogs can be split into in-paralogs (paralogous pairs that arose after a speciation event) and out-paralogs (paralogous pairs that arose before a speciation event). Between species out-paralogs are pairs of paralogs that exist between two organisms due to duplication before speciation. Within species out-paralogs are pairs of paralogs that exist in the same organism, but whose duplication event happened after speciation. Paralogs typically have the same or similar function, but sometimes do not. Due to lack of the original selective pressure upon one copy of the duplicated gene, this copy is free to mutate and acquire new functions. Paralogous sequences provide useful insight into the way genomes evolve. The genes encoding myoglobin and hemoglobin are considered to be ancient paralogs. Similarly, the four known classes of hemoglobins (hemoglobin A, hemoglobin A2, hemoglobin B, and hemoglobin F) are paralogs of each other. While each of these proteins serves the same basic function of oxygen transport, they have already diverged slightly in function: fetal hemoglobin (hemoglobin F) has a higher affinity for oxygen than adult hemoglobin. However, function is not always conserved. Human angiogenin diverged from ribonuclease, for example, and while the two paralogs remain similar in tertiary structure, their functions within the cell are now quite different.

---

**Key Points**

- A homologous gene (or homolog) is a gene inherited in two species by a common ancestor. While homologous genes can be similar in sequence, similar sequences are not necessarily homologous.
- Orthologs are homologous genes where a gene diverges after a speciation event, but the gene and its main function are conserved.
- If a gene is duplicated in a species, the resulting duplicated genes are paralogs of each other, even though over time they might become different in sequence composition and function.

---

**Key Terms**

- **conserved**: In biology, conserved sequences are similar or identical sequences that occur within nucleic acid
sequences (such as RNA and DNA sequences), protein sequences, protein structures.

• **selective pressure**: Any cause that reduces reproductive success in a proportion of a population, potentially exerts evolutionary pressure or selection pressure.