18.4E: Noncoding DNA

Noncoding DNA are sequences of DNA that do not encode protein sequences but can be transcribed to produce important regulatory molecules.

Learning Objectives

- Summarize the importance of noncoding DNA

Key Points

- In the human genome, over 98% of DNA is classified as noncoding DNA and can be transcribed to regulatory noncoding RNAs (i.e. tRNAs, rRNAs), origins of DNA replication, centromeres, telomeres and scaffold attachment regions (SARs).
- Noncoding regions are most commonly referred to as ‘junk DNA’, however, this term is misleading as noncoding DNA does have functional importance.
- The proportion of coding and noncoding DNA within organisms varies and the amount of noncoding DNA typically correlates with organism complexity, though there are many notable exceptions.

Key Terms

- **intergenic**: describing the noncoding sections of nucleic acid between genes
- **noncoding**: DNA which does not code for protein
- **intron**: a portion of a split gene that is included in pre-RNA transcripts but is removed during RNA processing and rapidly degraded
Noncoding DNA

In genomics and related disciplines, noncoding DNA sequences are components of an organism’s DNA that do not encode protein sequences. Some noncoding DNA is transcribed into functional noncoding RNA molecules (e.g. transfer RNA, ribosomal RNA, and regulatory RNAs), while others are not transcribed or give rise to RNA transcripts of unknown function. The amount of noncoding DNA varies greatly among species. For example, over 98% of the human genome is noncoding DNA, while only about 2% of a typical bacterial genome is noncoding DNA.

Initially, a large proportion of noncoding DNA had no known biological function and was therefore sometimes referred to as “junk DNA”, particularly in the lay press. However, many types of noncoding DNA sequences do have important biological functions, including the transcriptional and translational regulation of protein-coding sequences, origins of DNA replication, centromeres, telomeres, scaffold attachment regions (SARs), genes for functional RNAs, and many others. Other noncoding sequences have likely, but as-yet undetermined, functions. Some sequences may have no biological function for the organism, such as endogenous retroviruses.

Genomic Variation between Organisms

The amount of total genomic DNA varies widely between organisms, and the proportion of coding and noncoding DNA within these genomes varies greatly as well. More than 98% of the human genome does not encode protein sequences, including most sequences within introns and most intergenic DNA. While overall genome size, and by extension the amount of noncoding DNA, are correlated to organism complexity, there are many exceptions. For example, the genome of the unicellular Polychaos dubium (formerly known as Amoeba dubia) has been reported to contain more than 200 times the amount of DNA in humans. The pufferfish Takifugu rubripes genome is only about one eighth the size of the human genome, yet seems to have a comparable number of genes; approximately 90% of the Takifugu genome is noncoding DNA.

In 2013, a new “record” for most efficient genome was discovered. Utricularia gibba, a bladderwort plant, has only 3% noncoding DNA. The extensive variation in nuclear genome size among eukaryotic species is known as the C-value enigma or C-value paradox. Most of the genome size difference appears to lie in the noncoding DNA. About 80 percent of the nucleotide bases in the human genome may be transcribed, but transcription does not necessarily imply function.
Utricularia gibba flower: Utricularia gibba has 3% noncoding DNA, which is low for flowering plants. This 3% has given this plant the ‘most efficient’ genome.