Genome sizes (nucleotide pairs)

Mouse 3 x 10^9 nucleotide pairs

Human 2.9 x 10^9 nucleotide pairs

Onion 1.5 x 10^{10} nucleotide pairs

If you find it difficult to remember the exact numbers, we can say that humans have a genome size of approximately 3 billion nucleotide pairs.

The features of the genomes of eukaryotes and humans

1. The main difference between the genomes of eukaryotes and prokaryotes is the presence of redundant DNA in eukaryotes.

The genome of E. coli consists of 4.6 x 10^6 nucleotide pairs. Protein-coding genes account for approximately 88% of the genome. About 1% of the genome is occupied by genes encoding RNA molecules, and about 1% is made up of non-protein-coding repeats. Intergenic gaps occupy about 11% of the genome. Regions that play a regulatory role are found within intergenic gaps.

In humans, exons (regions that code directly for proteins) occupy approximately 1-1.5% of the genome. 24% of the genome is occupied by intronic regions of genes. Intergenic gaps occupy 75% of the genome.

In ciliate infusoria, intergenic gaps occupy about 95% of the genome.

2. Proteins encode approximately 1% of the genome. However, approximately 80% of the genome is transcribed to some extent. In other words, about 80% of the genome is involved in RNA synthesis. However, most transcripts (RNA) are not translated into proteins. These include: introns (which are excised during splicing), multiple ribosomal RNAs, transfer RNAs, and numerous small RNAs.

The majority of human non-coding RNAs are the ‘housekeeping’ RNAs of the cell. These RNAs are involved in protein synthesis. Such RNAs include ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), RNAs that are part of spliceosomes, and others.

The human genome contains approximately 500 genes that encode tRNA and around 200 that encode rRNA. The rRNA-encoding genes are located on the short arms of 5 acrocentric chromosomes (13, 14, 15, 21, 22).

3 Another regularity of the eukaryotic genome structure is the presence of nucleotide sequences with different frequencies of repeats.

The following variants are distinguished

1. Unique sequences. These are not repeated. They are mainly rep-

resented by genes.

2. Medium-frequency repeats. These repeats are repeated tens or even hundreds of times. These repeats form families. The repeats almost completely repeat each other. Genes encoding transfer and ribosomal RNAs belong to this group.

3. High-frequency repeats, their number can reach 1 million. They mainly include mobile genetic elements.

In humans, mobile genetic elements make up around 45% of the genome. Most of these are retrotransposons that use RNA in their life cycle.

Classes of moderately repetitive SINE, LINE, and LTR elements have been found in humans (and other mammals).

SINE (short interspersed nuclear elements). These are fragments of 100 to 300 nucleotide pairs interspersed with unique sequences of 1000 to 2000 nucleotide pairs. These fragments account for 10% of the total length of the human genome.

LINE (long interspersed nuclear elements). Here repeats of 5000 nucleotide pairs alternate with unique sequences of 35000 nucleotide pairs. They make up 21% of the genome’s total length.

LTR transposons are the least numerous, occupying only 8% of the genome’s total length.

The importance of mobile genetic elements

1. They can lead to mutations if they are incorporated into genes.

2 Mobile elements can alter gene expression activity. They contain elements similar to enhancers and stimulate gene transcription. One situation that has been described is when a mobile element moves to a DNA site near a protoncogene and stimulates its expression. Protein overproduction occurs, causing the cell degenerate into a malignant cell.

3 Occurrence of chromosomal mutations. If there are mobile elements in homologous chromosomes, duplications, deletions and inversions may occur during crossing over.

4. The next feature of the genome. Currently, around 2000 DNA sites have been described that are present in different numbers of copies in the genome. The term ‘copy number variations’ (CNVs) has appeared. Each individual genome contains more than 100 such variations, each averaging 250,000 nucleotide pairs in size. Often, whole genes fall within these variant sections. As a result, a particular person’s gene may have one or three copies instead of two. This changes the amount of a given gene (or gene

dosage). In some cases, a gene may be completely absent without a significant change in phenotype.

Some fairly large portions of the genome are missing altogether in a number of healthy individuals. These ‘optional’ sections can range in length from several thousand to several million nucleotide pairs.

In some cases, changes in the number of copies of genome regions can lead to the pathology forming. For nervous system diseases alone, around 20 copy number variations (CNVs) have been identified that lead to various pathologies, including Alzheimer's or Parkinson's disease, autism, and schizophrenia. A special DECIPHER database (http://www.sanger.ac.uk/PostGenomics/decipher/) has been created to capture new information regarding copy number variations in health and disease.

Please note!!! The human and chimpanzee genomes differ by approximately 1.23%. About 0.25% of the nucleotide pairs that differ between humans and chimpanzees are normal polymorphisms. The true differences between chimpanzees and humans are less than 1%. If we consider only genes, approximately 6% of all functional genes are unique to either humans or chimpanzees.