INTRODUCTION

In 1869, Friedrich Miescher discovered DNA, first named "nuclein". In 1918, Phoebus Levene identified the four bases, adenine (A), cytosine (C), guanine (G), and thymine (T), the building blocks of DNA (Levene, 1919). 30 years later, Erwin Chargaff found that the DNA base composition varied between species but determined that within a species the bases in DNA are always present in fixed ratios: the same number of As as Ts and the same number of Cs as Gs (Chargaff, 1950). However, the great career of DNA started in 1944 with the Avery, MacLeod and McCarty experiment (Avery et al., 1944), proving that not proteins or RNA, as previously suspected, but DNA is the substance responsible for bacterial transformation. The process of bacterial transformation was first described in 1928 by Griffith, who found that an extract from the virulent *Pneumococcus pneumoniae* strain could induce the conversion of a nonvirulent strain into a virulent one (Griffith, 1928). The next milestone in DNA studies was the model of DNA structure and replication proposed by Watson and Crick (1953). A few years later, Crick delivered a lecture during a symposium on the Biological Replication of Macromolecules held at University College London. The lecture, published in 1958 under the title "On the protein synthesis" (Crick, 1958), is considered as a very important article, formulating the so-called central dogma of molecular biology: DNA makes RNA, RNA makes protein. Since then, genes have not been considered as somewhat mythical "units of heredity" as it was common after Gregor Mendel's discoveries, but as fragments of the DNA chain on which information on the RNA and protein structure is written down. The main effort was then directed towards deciphering the mode of gene functioning.

In this respect, it is impossible not to mention the work of François Jacob and Jacques Monod, both working at the Pasteur Institute in Paris. They were studying genes involved in metabolism of lactose in the bacterium *Escherichia coli*. They found that these genes, *lacZ, lacY and lacA*, coding respectively for β-galactosidase, permease and transacetylase, are located close to each other on the *E. coli* chromosome and are transcribed on a single messenger RNA (mRNA) molecule. Transcription starts at the element called promoter, recognized by the RNA polymerase. Transcription depends on two other genes, the operator and the regulator. The operator is located close to the first structural gene (*lacZ*), while the regulator is a gene coding for a repressor, a protein that binds the operator and prevents the onset of transcription in the absence of lactose. The promoter, operator and structural genes form the so-called operon. Transcription of the lactose operon occurs only in the absence of glucose, the preferable carbon source, and in the presence of lactose being an inhibitor of the repressor. François Jacob and Jacques Monod together with André Lwoff were jointly awarded the Nobel Prize in Physiology in 1965 "for their discoveries concerning genetic control of enzyme and virus synthesis".

The following years brought further information on the structure and function of genomes. It was found that in eukaryotes, the majority of genes are not continuous. They are composed of "exons" and "introns". They are transcribed into pre-mRNA, which undergoes the process called "splicing", carried out by small nuclear ribonucleoproteins (snRNPs) which bind to both the 5′ and 3′ ends of the intron and cause the intron to form a loop. The intron is then removed from the sequence and the two remaining exons are linked together.

At the beginning of the seventies, a new branch of genetics named genetic engineering developed. It became possible to introduce precise changes in genetic material and to transfer genes from one organism to another. The classic example of a useful genetic manipulation is the construction of a bacterial strain containing the human gene coding for insulin. Insulin is a drug irreplaceable in diabetes treatment. Previously, it was acquired from the pancreases of pigs and cows, now it is easily available by extraction from bacteria growing in huge fermenters. The emergence of genetic engineering aroused some concerns connected with the safety of genetic engineering experiments. It was theoretically possible to use these

techniques for construction of dangerous microorganisms for military use. In 1975, Paul Berg, one of the creators of genetic engineering, organized a conference devoted to all possible applications of the new techniques at the Asilomar Conference Center (California, USA). Around 150 geneticists from all over the world discussed the dangers of genetic engineering. I attended this conference at the invitation of David Baltimore, the scientist from the Cancer Center of the Massachusetts Institute of Technology, who together with Renato Dulbecco and Howard Martin Temin was awarded the Nobel Prize "for their discoveries concerning the interaction between tumour viruses and the genetic material of the cell." The second Polish participant of the Asilomar Conference was professor Wacław Gajewski, head of the Department of Genetics at the University of Warsaw.

The final conclusion reached at the Asilomar Conference was to go ahead with genetic engineering, taking care to run experiments under laboratory conditions adequate to the potentially most dangerous element of this experiment. For example, construction of an *Eschericia coli* strain producing insulin can be performed in practically every biological lab, as neither this bacterial species nor insulin are dangerous to humans. On the other hand, experiments on genetically engineered strains used for the development of cures for infectious diseases should be performed only in specially equipped laboratories, giving confidence that bacteria will not get out. James Watson, the Nobel laureate for his work on DNA structure was a strong advocate of the genetic engineering expansion. At the time of the Asilomar Conference, he was an adviser of the US government on bacterial weapons. He claimed that he knew that the naturally occurring non-engineered bacteria strains stored in the US military warehouses could already kill every inhabitant of our planet in eight different ways and compared to that the products of genetic manipulations are the laughing stocks.

An enormous number of new information on genome structure was obtained due to the progress in the DNA sequencing techniques. 30 years ago, sequencing of the full human genome seemed to be a mission impossible, mainly because of the cost of this project, higher than the yearly budget of the Max Planck Institute. However, in spite of many objections, the Human Genome Project was launched in 1990. The project was run by the International Human Genome Sequencing Consortium (IHGSC) composed of scientists from 18 countries. One of the founding principles of the

consortium was that all human genome sequence information would be freely and publicly available within 24 hours of its assembly. In the meantime, a private biotechnology company called *Celera Genomics* also entered the race to sequence the human genome. Led by Dr. Craig Venter, *Celera* proclaimed that it would sequence the entire human genome within three years. In February 2001, both groups in two separate articles (Lander et al., 2001; Venter et al., 2001) simultaneously published drafts of the human genome sequence. Due to technical advances in DNA sequencing methods and a productive level of synergy between the two groups, they tied at the finish line, and both projects were completed ahead of schedule. The IHGSC pursued the second phase of the project: the finishing phase (IHGSC, 2004). During this phase, the researchers filled in gaps and resolved DNA sequences in ambiguous areas that were not solved during the first phase. The finishing phase yielded 99% of the human genome in the final form. The final form of the human genome contained 2.85 billion nucleotides, with a predicted error rate of 1 event per 100,000 bases sequenced. Furthermore, the IHGSC reduced the number of gaps by 400-fold; only 341 gaps out of 147,821 gaps remained. The remaining gaps were associated with technically challenging chromosomal regions. Although the earlier draft publications had predicted as many as 40,000 protein-encoding genes, the finishing phase reduced this estimate to between 20,000 and 25,000.

Today, the overwhelming success of the project is readily apparent. The completion of this project ushered in a new era in medicine. It is now possible to learn which of our illnesses have a genetic background and to identify genes responsible for them. The Human Genome Project also brought significant advances in technologies used to sequence DNA.

One particularly striking finding of the Human Genome Project research is that the human nucleotide sequence is nearly identical (99.9%) between any two individuals. However, even a single nucleotide change in a single gene can be responsible for a human disease. Because of this, our knowledge of the human genome sequence has also contributed immensely to our understanding of the molecular mechanisms underlying a multitude of human diseases. Furthermore, a merging of cytogenetic approaches with the human genome sequence will continue to propel our understanding of human diseases to an entirely new level. Thus, although it was met with skepticism at its inception, the Human Genome Project

will certainly be heralded as one of the most important scientific endeavors of our time.

Quite recently, a new, powerful method of gene engineering was developed. The method, named CRISPR-Cas (Fig. 1), is an efficient and reliable way to make precise, targeted changes to the genome of living cells. CRISPR (clustered regularly interspaced short palindromic repeats) and CRISPR-associated (Cas) gene coding for nuclease are essential in adaptive immunity of bacteria and archaea, enabling the organisms to respond to

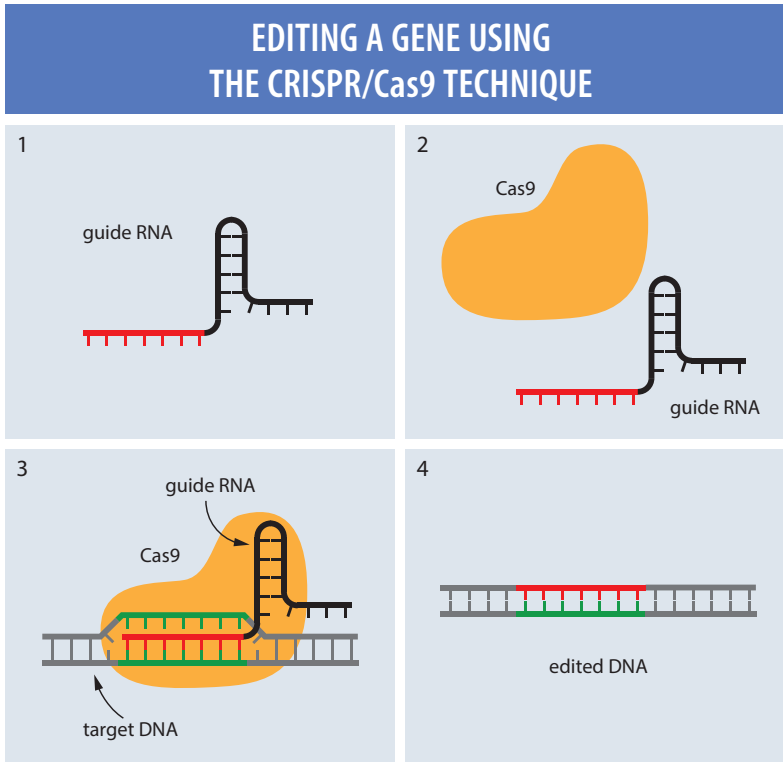

Fig. 1. An overview of the CRISPR-Cas9 method (CRISPR gene editing method). 1) A guide RNA with a fragment matching the target DNA sequence (red) is designed and created. 2) The guide RNA is added to the target cell along with the Cas9 protein. 3) Guide RNA pairs with the matching fragment of target host DNA (green), which gets cut by the Cas9 protein, creating a double-strand break at a precise target location. 4) Desired DNA sequence modifications can be introduced at the precise location within the genome.

and eliminate invading genetic material. Palindromic repeats were initially discovered in the 1980s in *E. coli* (Ishino et al., 1987), but their function was not confirmed until 2007 by Barrangou and colleagues, who demonstrated that *S. thermophilus* can acquire resistance against a bacteriophage by integrating a genome fragment of an infecting virus into its CRISPR locus (Barrangou et al., 2007). Three types of CRISPR mechanisms have been identified, of which type II is most studied. In this case, invading DNA from viruses or plasmids is cut into small fragments and incorporated into a CRISPR locus amidst a series of short repeats – around 20 base pairs (bp). The loci are transcribed and transcripts are then processed to generate small RNAs (crRNA – CRISPR RNA), which are used to guide endonucleases to the complementary sequences within invading DNA.

Using the CRISPR-Cas technique, one can precisely destroy the target gene or change the mutated gene into the wild type one. It is expected that the technique will be used to "repair" genes responsible for hereditary diseases; however, there are many scientists claiming that the technique is yet not one hundred percent safe and could lead to unwanted side effects. Despite that, a group of Chinese scientists undertook the first trials and injected a person with cells that contain edited genes. They believed that gene editing could improve the ability of immune cells to attack cancer. A team led by oncologist Lu You at Sichuan University in Chengdu delivered the modified cells into a patient with aggressive lung cancer at the West China Hospital (Liang et al., 2015). The researchers removed immune cells from the recipient's blood and then disrupted the gene coding for the protein PD-1, which normally puts the brakes on a cell's immune response: cancers take advantage of that function to proliferate. The edited cells were cultured to increase their number and injected back into the patient. The hope was that, without PD-1, the edited cells would attack and defeat the cancer. There are plans to use the same technique, in both China and the US, for treatment of patients with other types of cancer (Cyranoski, 2016).

Except for applications of the new techniques of DNA analysis in medicine, they have been a very important tool for studying evolution of living organisms. The discovery that DNA survives in ancient plant or animal remains and can be amplified by the polymerase chain reaction and sequenced has greatly affected studies on evolution. The first published

articles suggested that it would become possible to go back even hundreds of millions of years by studying fossil DNA. However, the reports (Cano et al., 1993) on DNA sequences obtained from insects embedded in amber appeared to be not true. Such insects are dated to more than one hundred million years, while it is known that DNA is unlikely to survive for more than about 1,000,000 years. However, even by studying such a geologically short period, one can obtain answers to important questions concerning evolution, systematics, paleoecology, the origin of diseases, and evolutionary processes at the population level.

Ancient DNA

Every DNA sample obtained from excavated human, plant or animal remains is called ancient DNA (aDNA). Studies on aDNA attracted attention not only of molecular biologists, but also of the general public. This was a result of several spectacular discoveries, especially those concerning human history. Ancient DNA studies revealed complexity of the history of modern humans. They also proved that our forebears interbred during the Middle Paleolithic and early Upper Paleolithic with Neanderthals – each contemporary non-African has in their genome 1–4 % admixture of the Neanderthal DNA. The interbreeding happened in several independent events that included not only Neanderthals but also Denisovans as well as several unidentified hominins. aDNA studies revealed, among others, Scandinavian origins of the Rurik Dynasty as well as hint to such origins for the Piast Dynasty. Through analyses of aDNA, it was possible to trace the paths of human migrations and the relationships between populations now inhabiting the Earth.

Almost equally interesting are the results of studies on history of animal and plant species, including those which became extinct either thousand years ago or in quite recent times. There are numerous animal species we know only from their remnants preserved in permafrost or deep in caves; some iconic examples being the mammoth, cave bear, wooly rhinoceros or the Mauritian bird dodo. By studying their DNA, we can establish their relationships with their still living cousins. There are even verge of science-fiction projects to revive some of these species in a process named

de-extinction; for example by introducing DNA isolated from the mammoth remains into the egg cells of an elephant.

The progress in the field of genetic engineering made it possible to transfer genes from one individual to another, using as a donor either DNA isolated from living species or from the remnants of the extinct ones. It is now a common practice to improve the plant species by equipping them with a bacterial gene and thus making them resistant to insect pests.

In the following chapters, methods of the aDNA analysis are presented as well as their use in studies of evolution of humans and the plant and animal species.