The chicken is the first non-mammalian amniote whose genome was sequenced. The sequencing was carried out by the International Chicken Genome Sequencing Consortium and it was aimed at deciphering the genetic code of Red Jungle Fowl (Gallus gallus) and to compare it with known mammalian genomes, including the human genome (18). An individual, inbred female was chosen for sequencing, to reduce the heterozygosity and cover both chicken sex chromosomes (W and Z). The results of this project have created new possibilities for the analysis of the chicken genome. Conducted analyses showed that the total length of the sequence of the chicken genome is more than a billion nucleotides that encode about 15,500 genes (number of genes verified by Ensembl) (28). It was found that the chicken genome has a relatively low content of repetitive sequences (11%) compared to mammals, which typically contain 40-50% of repetitive sequences in their genome (8). The low level of repetitive sequences results from the decreased overall size of the Avian genome in comparison to mammalian genomes, e.g. size of the chicken genome is ~1.05 Gbp, whereas size of the human genome is ~2.8 Gbp. This characteristic of the chicken genome allowed for a relatively easy assembly of the entire DNA sequence and reduced the risk of errors (35). Comparative analysis of the DNA sequences of three chicken breeds (a broiler, a layer and a fancy breed – Silkie) revealed the presence of approx. 2.8 million single nucleotide polymorphisms (SNPs) (36).

The comparison between the three genomes: human (mammal), chicken (bird), and fugu (fish) showed that about one third of all genes are common to all three species. Those genes have been considered a common characteristic of vertebrates in general (8). During the evolution of the Aves, a number of gene families were lost from the genome, responsible for traits such as: vomeronasal receptors, casein milk proteins, proteins present in saliva and enamel proteins. On the other hand, new genes were gained, responsible for: scales, claws and feathers and genes encoding proteins specific to the eggshell (e.g., ovocleidin 116) (18). As a consequence, the chicken has become a model bird organism, with the genome sequence available, as well as ease at obtaining biological material resulting from a high reproduction rate and early sexual maturation of the individuals (9). Sequencing allows for deciphering an enormous amount of genetic information; as such it is the initial step in many experiments. Research on understanding the genetic bases of performance traits is one of

---

1) The review was supported by grant UMO-2014/13/B/NZ9/02123 funded by National Science Centre in Cracow (Poland).
the most dynamically developing branches of modern biology, medicine, veterinary and animal husbandry. New molecular and analytical technologies developed for medical use can be further adopted for model and livestock animals. Improvement in sequencing methods depends to a large extent on the competitiveness between leading producers of innovative sequencers, such as Illumina or Roche (27). Currently, the most innovative method of sequencing is the Next Generation Sequencing (NGS), which has already revolutionized molecular biology techniques. NGS is commonly used to sequence genome, but also to analyze the transcriptome, interactions between DNA, RNA and proteins and to carry out metagenomics and epigenetic analyses.

In this review the current state of NGS knowledge in poultry genetic research was analyzed, especially regarding chickens. We also present the potential practical applications of NGS in relation to health issues and production traits in chickens.

**Development of sequencing techniques**

Depending on the complexity and sophistication, the sequencing methods are classified as first or next (second, third) generation sequencing. The Sanger method, which is based on the detection of labeled, partially digested fragments of the two-dimensional fractionation (17), was classified as the first generation sequencing and has initiated the development of the next generation. Despite the improvement of the first generation sequencing techniques, they have become insufficient (26).

**Next generation sequencing.** Second generation sequencing is based on sequencing of amplified DNA and it includes Solexa’s sequencing by synthesis method (17). An exemplary strategy of the second generation sequencing is pyrosequencing, which allows for reading sequences containing up to one million base pairs. In this reaction new nucleotides are incorporated into a newly synthesized DNA strand, generating a chemiluminescence flash (1). Sequencing is carried out in three steps: 1) isolation of DNA and creation of libraries; 2) amplification of template DNA; and 3) parallel sequencing. Libraries are constructed by DNA fragmentation and ligation of the adapter with a known function. Such a library is replicated and ready for sequencing (17).

Third-generation sequencing allows for sequencing of a single DNA molecule without preamplification and includes such methods as tSMS (True Single Molecule Sequencing) and SMRT (Single Molecule Real-Time) (30). Manufacturers around the world are competing in the development of the most faithful sequencing method with the ultimate goal of the largest automation and miniaturization of the process (7). Nanoengineering brings new expectations to the development of molecular techniques. Nanopore sequencing differs from the above-mentioned strategies by not reading the nucleotide sequences derived from the synthesis, but from subsequently cleaved single nucleotide matrix (14).

**Paired-end sequencing.** In general, NGS technology, which allows for a parallel sequencing of vast numbers of DNA fragments, has replaced other methods of sequencing (17). During process of parallel sequencing, nucleotides are added to the matrix in sequential order, which represents sequence of the DNA chain, while the nucleotide sequence is recorded by computer. One of the methods which significantly improves quality of the reads is paired – end sequencing, which generates sequences from both 3’ end and 5’ end side of the template DNA. This way sequencing efficiency is nearly doubled (31). After generating sequence reads, bioinformatic analysis is performed. Preliminary analysis involves removal of the adaptors and the low quality sequence. This is followed by mapping of the sequence to the reference genome or de novo genome assembly (15). At the end an information at the whole-genome level is collected, such as detection of new genes, estimation of gene expression level or the identification of genetic variants. Beside sequencing of the entire genome, it is also possible to read selected fragments only, using methods such as: exome sequencing (still unavailable for poultry) and targeted sequencing (targeted NGS). The obtained results are more detailed and the process is more economical in comparison to sequencing of the entire genome (2).

**Application of next generation sequencing to chicken genome research**

**SNP discovery.** Resequencing of the chicken genome based on NGS made it possible to identify 57,636 SNP markers (16). Among newly discovered SNPs, 328 of the markers were mapped to microchromosomes that have not been analyzed earlier. All of these newly discovered SNPs were used to construct a DNA microarray (chip), used to detect polymorphisms. Identification of SNPs which are based on NGS has made it possible to improve coverage of the chicken genome (index validation was 94%) and the discovery of SNPs in regions where they had not been identified earlier (16).

Boscheiro et al. (5) conducted comparative studies of 11 commercial and 5 experimental chicken lines with the aim of detecting genetic differences using the NGS approach. Altogether 883’000 mutations were detected, including insertions and deletions, which are the most common polymorphisms in the genome, that exert significant impact on many production traits and genetic disorders. These mutations were unevenly distributed throughout the genome. The lowest density of single nucleotide mutations was detected on chromosomes: Gallus gallus Z chromosome (GGAZ), GGA16, GGA22 and GGA25, while the highest, on GGA6. Approximately 30% of polymorphisms were non-synonymous mutations, resulting in the change of encoded amino acids and reducing their biological activity (eg., proliferation, construction and operation of the Golgi apparatus, spermatogenesis, muscle contractions). About 2.5% of the identified point mutations were found in the genes responsible for production traits.
Almost 81% of the mutations detected by NGS overlapped with the sequencing results obtained previously with Whole Genome Shotgun method (6). NGS based analysis extended the number of known mutations with 168’000 newly identified. Deep sequencing technology allows for an accurate detection of insertions/deletions, particularly rare ones, which improves overall precision of genomic studies. The wide range of the detected point mutations in the genome made it possible to undertake further analysis of their impact on the phenotypic traits or genetic diseases in birds.

Male birds from fat and lean chicken lines were used as a resource population for the determination of significant epistatic effects related to carcass and growth traits using Illumina SNP chip. Several epistatic interactions related to testis weight in chickens were detected. miR-142-5p with ICF12 (its target gene) and with other genes located on GGA21 and GGAZ might be connected with development and testis growth. This type of research analyzing the epistatic interaction can play a very important role in the phenotypic variation of various traits in chickens (39).

Copy number variation. Copy number variation (CNV) is a repetition of the section of the genome and the number of this repeats is different among individuals in the population. Known chicken phenotypes related to CNV include pea-comb, dermal hyperpigmentation and dark brown plumage color or late-feathering on chromosome Z (34). CNV can lead to phenotypic variability, such as genetic resistance to infectious diseases. For instance, experimental lines of laying hens (6, and 7) demonstrated low susceptibility to Marek’s disease and a high coefficient of inbreeding. Both of these lines have been used on a wide scale for the identification of the molecular and genetic bases of immune resistance. These chickens carry B2 haplotype of major histocompatibility complex (MHC), which is the main receptor for Marek’s disease virus (MDV) (20). MDV causes paralysis of the limbs and a high mortality rate, whereby the identification of the genetic basis of the disease is highly important particularly from the point of view of the poultry industry (38). The next generation sequencing was used to screen both lines for polymorphism within CNVs and associate it with genetic resistance to MDV virus. As a result, 5’680 CNVs were localized in both genomes. Genomic regions with a high degree of CNV polymorphism and associated with resistance to MDV encode mitogen-activated protein kinase (MAPK) (38). These kinases are a group of proteins responsible for receiving and transmitting information from the cell surface, which play an important role in the regulation of gene expression (23). Conducted analysis revealed 67 CNVs overlapping with 62 genes related to the response to MDV. The rich source of line-specific CNVs is valuable for the study of diseases involving two lines of lying hens (38).

Structural variations. Structural variations in the genome can cover a wide range of polymorphisms, from SNPs to chromosomal abnormalities. Until now it was possible to detect the structural variations using comparative genomic hybridization (CGH) (11). However, the main drawback of this method is its low resolution, allowing for the detection of changes larger than 8’000 nucleotides (33). With the introduction of the latest sequencing technologies such as paired-end NGS, it has been possible to detect even small mutations. Detection of structural variations by the paired-end method enables identification of inversion and translocation (22). In studies investigating structural differences in the chicken genome, four different lines of broilers and laying hens were used. DNA was amplified using emulsion PCR and genomic libraries were subsequently sequenced using paired-end sequencing. DNA sequences obtained from each line was compared with the reference chicken genome of Red Jungle Fowl (Gallus gallus). The comparison showed hundreds of differences both between the analyzed lines and with regards to the reference genome (22).

Integration of viral DNA in chicken genome. Avian Endogenous Retrovirus-HP (EA V-HP) is a retrovirus, which integrated with the chicken genome before the domestication event and until today is an integral part of its genome. EA V-HP is able to mutate and give rise to the Avian Leukosis Virus (ALV), which confers epidemiological problems in the poultry industry (29). Retroviruses are able to integrate into the host genome, using the molecular mechanisms present in the host cell. In this case, the virus undergoes natural processes such as mutagenesis, which can lead to changes in gene expression or activity of the viral proteins. Under the influence of the defense mechanisms of the host cells, the viral genome may be in the residual separate form, but very often it rather constitutes an integral part of the host genome. Therefore, the whole genome sequencing enables revealing the sequence of the host genome and the viral genome (10, 19). Research shows that the average vertebrate genome consists of 4-10% of residual viral genomes, and over the time, viral and host genomes are gradually fused in one (32).

Comparative analysis using paired-end NGS was applied to several genomes: eight inbred Leghorn lines, Silkie chicken, Thai chicken, Red Junglefowl and two rural hens from the Horro and Jarso regions in Ethiopia. Two reference genomes were used: Red Junglefowl genome and viral EA V-HP genome. It was determined that the virus integrated into the chicken genome seventy-five times on average. The genomic integration rate is closely correlated with the size of chromosomes. The larger chromosome, the more integration sites it contains. The results confirm integration of the EA V-HP viral genome into contemporary chicken breeds. But, more interestingly, occurrence of the viral genome in Red Junglefowl DNA indicates its integration into avian genome even before the event of domestication (37).

Identification of genes encoding production traits. Next generation sequencing has also proved useful in finding genes responsible for production traits, for example egg shell quality. Damaged egg shells not only
cause large losses in the poultry industry, but can also pose health problems to the consumers. For analysis of the thickness and strength of the shell, uterus tissues were taken from 49-week-old Rhode Island hens at 18 h after previous oviposition. Hens of this line lay eggs with shells of different thicknesses and strengths. NGS was used for genome sequencing and further search for genes and mutations responsible for the level of shell calcification. A total of 14'234 genes were identified, 889 out of which were characterized as genes responsible for shell endurance, both determining its low and high resistance to mechanical damage. These genes are closely related to the degree of mineralization and calcium pathways in the chicken organism. This study also allowed for resequencing of the chicken genome followed by identification of 3'671'919 SNPs (DNA polymorphisms) and 508'035 point mutations, all located in protein coding regions. Obtained data can be used in SNP association studies with the egg shell quality traits (40).

The attention of scientists is focused on defining various potential genetic markers demonstrating amino acid changes responsible, for example, for meat color. Whole genome re-sequencing analysis using HiSeq paired-end read method was performed on two broiler chicken lines selected for muscle color (high and low muscle color defined based on the indicator to determine the muscle color of breast fillet). 2,884 SNPs were identified (1,307 for high muscle color and 1,577 for low muscle color). Analysis suggested that chicken meat color might be associated with chromosomal DNA stability and quantity of various collagen subtypes. This information might be helpful in further selection strategy (24).

Another trait of interest from the production point of view is disease resistance. Selective breeding for improved immune response including natural antibodies (NAb) may increase chicken survival by improving overall disease resistance. NAb are antigen-binding antibodies present in individuals without prior exposure to the antigen and they play an important role in innate and adaptive immunity. Genome-wide association studies were conducted to identify genes underlying the genetic variation of keyhole limpet hemocyanin (KLH)-binding NAb. Identification of genomic regions related with to KLH binding NAb level showed a region on GGA4 associated with the trait of interest. Candidate genes located in this genomic region are responsible for organism development, metabolism, cancer development and viral infections (4).

**Next-generation sequencing studies in other poultry species**

Poultry farming is a rapidly developing sector of the economy around the world. Further development of poultry at a high level requires facing the new challenges posed by consumers and producers of poultry companies. Thanks to DNA sequencing, it is possible to analyze genetic variability of production traits and phenotypic characteristics of animals. Analysis at the molecular level will facilitate selection of the animals resistant to pathogens or carrying favorable production traits (3).

Japanese quail is an excellent model in the biological studies mainly due to its rapid reproduction and maturation. This species is farmed around the world as an alternative to chicken source of poultry meat and eggs. The main producers of quail arc: China, Japan, France and Brazil and in these countries research on quail is of great interest. The main limitation that does not allow the use of quail in biomedical research is the lack of detailed information on the genomic sequence of these birds. To address this problem, studies have begun which sequence the genome of quail using the next generation DNA sequencing. The obtained sequence of Japanese quail was compared to the reference chicken genome. On the basis of the draft of the genome sequence comparison, 100 microsatellite markers were developed, which may in the future be used to evaluate the genetic variability of 11 lines of Japanese quail. Based on genomic sequences quail orthologs of spermatogenesis markers were identified (21).

Next-generation sequencing technology was also used in the turkey genome sequencing project, which was launched in 2008. At that time obtained DNA sequence covered 89% of the entire turkey genome. Scientists have made a great effort to refine and enlarge the already existing sequence. For this purpose a study was performed which generated turkey transcriptome sequences using RNA-seq. The current version of the DNA sequence was extended by the transcriptome sequences, and it finally covers 95% of the entire genome. Turkey genome was the first genome of a vertebrate sequenced using a combination of two NGS platforms (Roche 454 and Illumina GAII) (12). The DNA sequence was used in the research project focused on specific transcripts associated with response to aflatoxin AFB1. Consumption of feed contaminated by mycotoxins (toxins produced by fungi) adversely affects poultry health and production efficiency. Aflatoxin AFB1 is one of the most dangerous mycotoxins, and its consumption at a high dose causes death or in a smaller concentration, liver damage. The experiment aimed at examining the response of the liver transcriptome to aflatoxin AFB1 poisoning. Moreover, the influence of probiotic *Lactobacillus* bacteria which could inhibit the absorption of toxins in the intestines was studied as well. Several genes: *LPL*, *MAT2A* and *MDM2* were identified, potentially related to the aflatoxin poisoning that might be used as biomarkers for AFB1. Administration of the probiotic resulted in decreased expression of certain genes. The knowledge gained might be useful for seeking protective alleles, supporting the mitigation of the effects of mycotoxin poisoning (25).

Another poultry species analyzed with NGS technologies is guinea fowl. Deciphering the genome of guinea fowls was based on sequencing of transcriptomes from different tissues (pancreas, liver, hypothalamus and bone marrow). Results of RNA-seq showed that 60%
of long segments of guinea fowl DNA (contigs) are not homologous with the chicken DNA sequences available in the database. Information derived from the analysis of the transcriptome will further support progress in research on improving the efficiency of the production of guinea fowl and other birds (13).

The commonness of genome research of poultry provides a great deal of sequence datasets which are still not enough bioinformatically processed. Millions of sequences have not been annotated to specific genes and/or other characteristics. In the next few years, analysis of the chicken genome will be an important direction of research. Next generation sequencing is widely used in poultry breeding. Recently expanding directions of research is the analysis of the composition of the chicken intestinal microflora and its impact on the health of these birds. The sequence of bacterial genomes obtained through NGS has allowed for the identification of many new species of microorganisms, which might be important for animal and human health.

References


Corresponding author: Ms. Eng. Aleksandra Dunislawka, ul. Mazowiecka 28, 85-084 Bydgoszcz, Poland; e-mail: aleksandra.dunislawka@uta.edu.pl