STUDIES IN LOGIC, GRAMMAR AND RHETORIC 25 (38) 2011

Application of Hardy-Weinberg Law in biomedical research

Dorota Jankowska[1], **Robert Milewski**[1], **Urszula Górska**[1], **Anna Justyna Milewska**[1]

[1] Department of Statistics and Medical Informatics, Medical University of Bialystok

Abstract. The aim of this paper is to discuss the Hardy-Weinberg Law which is fundamental for population genetics. It discusses the intuitive expectations connected with the distribution of allele frequencies in a gene pool by using mathematical equations and defines the genetic equilibrium. The conclusions which are consistent with the Hardy-Weinberg Law and relate to biomedical applications seem necessary for the evaluation of data quality. Moreover, ways in which evolutionary forces break genetic equilibrium will be extensively discussed and will be presented using mathematical models of the dynamics of gene pool.

Introduction

Studies in biomedical research often require reference to the genetic characteristics of examined individuals. The genome of each organism provides the basic knowledge of its morphological, physiological, biochemical traits. Instructions contained on the carrier of genetic information like DNA and saved using the genetic code not only allow to specify the characteristics of organisms such as eye color, body build, blood group. They also determine the inheritance properties or allow the detection of susceptibility to certain diseases. Understanding the genetic structure plays a pivotal role in the prevention and treatment of genetic diseases. The current level of knowledge of genetics allows to accurately identify the causes of many diseases of genetic origin. What is more, there are already developed ways to treat genetic disorders by the so-called gene therapy. Genetic screening is also conducted to identify persons, who are carriers of genetic diseases. This research allows to determine the risk of inheritance of disease by offspring or to predict how a disease, which develops in old age, will proceed – and thus, take appropriate remedial steps. Moreover, it is interesting to analyze the genetic structure of some species of organisms owing to the ongoing changes in time. Determining which factors have an influence on the genetic composition of

ISBN 978–83–7431–296–7 ISSN 0860-150X

populations, it seems necessary to understand the course of evolution. It is worth to look at phenomena of dynamics shaping the genetic characteristics such as natural selection, mutation, migration or genetic drift.

Introductory concepts

A population is a set of organisms that belong to the same species, live simultaneously in a particular environment, influence each other and mate giving fertile offspring. These individuals have a certain set of alleles, which are forms of occurring genes. It is called the gene pool. The characteristics of the population are mainly determined by its genetic composition, and thus the genotypes or alleles frequencies.

If the population consists of N individuals, and there occur n different alleles of a particular gene, then the total number of possible genotypes will be:

$$G = \frac{n!}{2(n-2)} + n = \frac{n(n+1)}{2} \tag{1}$$

To simplify the consideration, let us suppose that there is a population in which there are only two different alleles of a gene – dominant A and recessive a. In accordance with the formula (1) it is possible to obtain 3 genotypes: AA, Aa and aa. Their cardinality in the population is denoted respectively N_{AA}, N_{Aa}, N_{aa}. Then the frequency of each of them will be calculated as a share of a given genotype in N-person population and will be marked as:

$$P_{AA} = \frac{N_{AA}}{N} \quad P_{Aa} = \frac{N_{Aa}}{N} \quad P_{aa} = \frac{N_{aa}}{N} \tag{2}$$

In addition, there is of course:

$$P_{AA} + P_{Aa} + P_{aa} = 1 \tag{3}$$

It is possible to calculate the particular allele frequencies in the gene pool too. However, it should be kept in mind that the population of N individuals, due to its diploidy, brings $2N$ alleles while each homozygote provides twice more alleles considered type than heterozygote. So, if we denote by p frequency of allele A and by q frequency of allele a we obtain the following relationships:

$$\begin{aligned} p &= \frac{2N_{AA} + N_{Aa}}{2N} = P_{AA} + \tfrac{1}{2}P_{Aa} \\ q &= \frac{2N_{aa} + N_{Aa}}{2N} = P_{aa} + \tfrac{1}{2}P_{Aa} \end{aligned} \tag{4}$$

and

$$p + q = 1 \tag{5}$$

Analogous calculations can lead to a greater number of alleles in the population. The field of genetics, which deals with the quantitative study of individual allele and whole genotypes frequencies in population is named population genetics. It analyzes the factors affecting the maintenance of these rates and any changes in time that lead to the evolution of organisms. So it describes the essence of the evolutionary mechanisms. Moreover, it attempts to explain such phenomena as the existence of genotypes associated with genetic diseases, their spread and any changes in these diseases.

Hardy-Weinberg Law (HWL)

A fundamental law, which is considered as the basis of population genetics, is the theorem formulated independently by the eminent mathematician, professor at the University of Cambridge, Godfrey Harold Hardy and German doctor Wilhelm Weinberg in 1908 [1]. It provides a theoretical support for the description of evolutionary phenomena. This law will be presented for a population of N individuals, where there are only two different alleles of a gene, A and a. Their proportions will be denoted as above, respectively p, q. Let us suppose that:

a) population is infinitely large,
b) organisms are diploid,
c) organisms reproduce sexually,
d) in population mating is random (so the population is panmictic) – individuals do not have preference during mating; each male has equal chances of crossing with each female and vice versa,
e) generations do not overlap – therefore, when the offspring matures into reproduction, the parental generation no longer has the ability to reproduce, and thus there is no intergenerational pairs,
f) organisms do not migrate – population is isolated, there is no exchange of individuals between populations of the same species, so there is no flow of genes from one population to another,
g) in the population is no mutation – a swap genetic material is not observed, which can cause the transformation of existing alleles into a completely new form; in addition, there is no transformation of allele A to a and conversely,

h) selection does not affect the tested locus – none of the combination of alleles has more than any other chance of surviving until give offspring, so there is no difference in the adaptation between genotypes.

Then the frequencies of individual alleles and genotypes in subsequent generations do not change [2–3]. In addition, proportions of particular genotypes will be respectively:

$$P_{AA} = p^2 \quad P_{Aa} = 2pq \quad P_{aa} = q^2$$

and they correspond to the distribution of binomial square:

$$(p+q)^2 = p^2 + 2pq + q^2 = 1 \tag{6}$$

Thus, the Hardy-Weinberg law by mathematical modeling describes, how the proportions of alleles and genotypes in the population should be arranged. It defines what we now call the genetic equilibrium [4]. It shows that the processes of reproduction do not cause changes in the frequencies of alleles in the population. If a) – h) assumptions are met, the proportions of alleles in subsequent generations remain constant and the population is not evolving. What is more, any deviations from the Hardy-Weinberg model are removed within one generation. A single random mating, in which there is no migration, mutation and selection pressure, is enough for the population to come back to equilibrium [5].

To prove the thesis posed in Hardy-Weinberg Law is worth recalling the inheritance properties of two different alleles of one locus formulated by Mendel. They will be illustrated on the so-called Punnett square [Tab. 1], which presents the possible combinations of gametes during procreation. This diagram is a method of predicting the frequency of individual genotypes in the offspring generation [6].

Tab. 1. Punnett square

		Male Genotype	
		A	a
Female Genotype	A	AA 1/4	Aa 1/4
	a	Aa 1/4	aa 1/4

All possible situations during reproduction in N-individuals population with two alleles of tested gene present [Tab. 2]. It shows how often members of a certain genotype will be crossed and gives the probability of formation of a particular genotype in the offspring.

Tab. 2. Frequencies and probabilities of offspring genotypes in bi-allelic population

Genotype		Frequency	The probability of genotypes in offspring		
Female	Male		AA	Aa	aa
AA	AA	P_{AA}^2	1	0	0
AA	Aa	$P_{AA}P_{Aa}$	1/2	1/2	0
AA	aa	$P_{AA}P_{aa}$	0	1	0
Aa	AA	$P_{Aa}P_{AA}$	1/2	1/2	0
Aa	Aa	P_{Aa}^2	1/4	1/2	1/4
Aa	aa	$P_{Aa}P_{aa}$	0	1/2	1/2
aa	AA	$P_{aa}P_{AA}$	0	1	0
aa	Aa	$P_{aa}P_{Aa}$	0	1/2	1/2
aa	aa	P_{aa}^2	0	0	1

The proportions of genotypes, which may occur in the offspring, are calculated as follows [7–8]:

$$P'_{AA} = 1 \cdot P_{AA}^2 + \tfrac{1}{2} \cdot P_{AA}P_{Aa} + \tfrac{1}{2} \cdot P_{Aa}P_{AA} + \tfrac{1}{4} \cdot P_{Aa}^2 = \left(P_{AA} + \tfrac{1}{2}P_{Aa}\right)^2 = p^2$$

$$P'_{Aa} = \tfrac{1}{2} \cdot P_{AA}P_{Aa} + 1 \cdot P_{AA}P_{aa} + \tfrac{1}{2} \cdot P_{Aa}P_{AA} + \tfrac{1}{2} \cdot P_{Aa}^2 + \tfrac{1}{2} \cdot P_{Aa}P_{aa} + \\ + 1 \cdot P_{aa}P_{AA} + \tfrac{1}{2} \cdot P_{aa}P_{Aa} = 2\left(P_{aa} + \tfrac{1}{2}P_{Aa}\right)\left(P_{AA} + \tfrac{1}{2}P_{Aa}\right) = 2pq$$

$$P'_{aa} = \tfrac{1}{4} \cdot P_{Aa}^2 + \tfrac{1}{2} \cdot P_{Aa}P_{aa} + \tfrac{1}{2} \cdot P_{aa}P_{Aa} + 1 \cdot P_{aa}^2 = \left(P_{aa} + \tfrac{1}{2}P_{Aa}\right)^2 = q^2$$

The frequencies of particular genotypes in the offspring generation can also be calculated directly from the available proportion of alleles in the parental population [7]. Due to the assumption of random mating, frequency with which male gametes containing allele A (with frequency p in a considered population) fuse in the process of reproduction with female gametes containing allele A, and create offspring with genotype AA is $P'_{AA} = p \times p = p^2$. Similarly, according to the probability theory, the frequency of homozygote aa can be determined as $P'_{aa} = q \times q = q^2$. Whereas setting a proportion of heterozygote Aa in the offspring requires taking into account the fact that they arise from the merger of gametes carrying different alleles, as it is presented in [Fig. 1]. The frequency of offspring with genotype Aa is $P'_{Aa} = 2pq$.

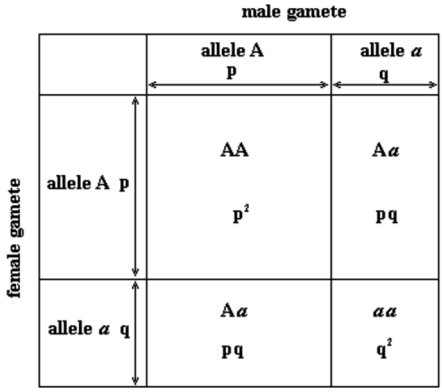

Fig. 1. Scheme of gametes fuse during the crossing of individuals in a population with two alleles of the tested gene

Additionally, it is worth noting that frequencies of genotypes correspond to area of rectangles, and the length of the sides of the figure correspond to frequencies different alleles.

It should be noted that under the assumptions of the Hardy-Weinberg Law, allele frequencies in the population remain constant from generation to generation. If the p', q' denote the frequencies of allele A and a in the offspring, they can be calculated using formulas (3), (4), (5) as follows:

$$p' = P'_{AA} + \tfrac{1}{2}P'_{Aa} = p^2 + \tfrac{1}{2} \cdot 2pq = p^2 + p(1-p) = p$$

$$q' = P'_{aa} + \tfrac{1}{2}P'_{Aa} = q^2 + \tfrac{1}{2} \cdot 2pq = q^2 + q(1-q) = q$$

This allows to conclude that the frequency of genotypes also remain unchanged over time. They depend only on frequencies of alleles in the parental population which, as has been shown, are the same in subsequent generations.

Testing whether the population is in equilibrium

The relationships between frequencies of genotypes in a population, where a locus with two alleles is tested, can be represented graphically. An excellent representation is the De Finetti diagram [7].

Any point inside the triangle illustrates a combination of the proportions of genotypes occurring in the population. For example, a point shown in [Fig. 2] represents a population, in which genotype frequencies are as

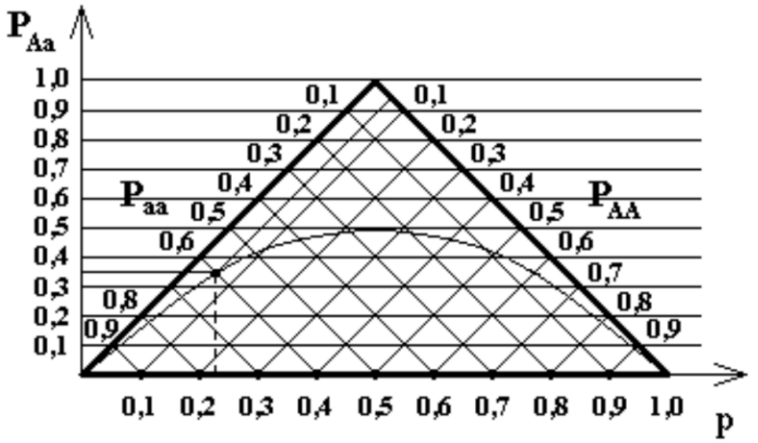

Fig. 2. De Finetti diagram with an exemplary point represents a population, which is characterized by the following distribution of the three genotype frequencies: $P_{AA} = 0.05$, $P_{aa} = 0.6$, $P_{Aa} = 0.35$

follows: $P_{AA} = 0.05$, $P_{aa} = 0.6$, $P_{Aa} = 0.35$. These values can be read by moving along the grid lines inside the triangle, which are routed perpendicular to the appropriate axis. The horizontal axis allows to determine the frequency of alleles occurring in the population. In the above example, you can read $p = 0.225$, which agrees with expectation for the population in the Hardy-Weinberg equilibrium:

$$p = P_{AA} + \tfrac{1}{2}P_{Aa} = 0.05 + 0.175 = 0.225$$

For each point inside the triangle, proportions of genotypes in the population, which it depicts, sum up to 1. Moreover the curve located in [Fig. 2] is the set of all those points that represent the populations in the genetic equilibrium state. Therefore, the De Finetti diagram makes it possible to determine whether the population meets Hardy-Weinberg equation and allows to declare, how much should the proportions of genotypes be, at specified frequencies of a particular alleles, for the population to be in a genetic equilibrium.

The first step during conducting clinical trials is usually to verify the distribution of allele and genotype frequencies in the population in accordance with those suggested by the Hardy-Weinberg Law. For this purpose the χ^2 test may be used. The value of χ^2_α statistics can be computed from the formula:

$$\chi^2_\alpha = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

where:

O – observed frequency of the i-th genotype,

E – frequency of the i-th genotype resulting from Hardy-Weinberg model.

The number of degrees of freedom is equal to the number of possible genotypes of the tested locus minus the number of parameters estimated on the basis of data plus one. This test is sometimes unreliable, especially when the checked gene has a numerous alleles or if the alleles are rare. Then it is proposed to perform an exact test, for example, exact test analogous to Fisher's exact test or the Monte Carlo method [9].

The Hardy-Weinberg Law formulates conditions to assume that allele frequencies in the population remain constant in time. Therefore, it can be used to predict the proportions of genotypes based on allele frequencies observed in the studied population. Conversely, HWL can also assess the frequency of a particular allele if the proportion of one of the genotypes is known. This ability may prove to be helpful, for example, in a conscious animal husbandry [13] to oversee the genetic composition of the herd or during making estimates associated with genetic diseases. If the proportion of organisms suffering from particular disease (homozygous recessive) is known then it is possible to approximate the frequency of allele responsible for the defect and to determine the amount of asymptomatic carriers in the population. For example, it has been stated that haemophilia A and B types, which is the bleeding disorder, affects more or less 1/12000 of the population. It allows to assess the number of people who carry the mutated allele but are free from symptoms of disease at 0.0181. Stating that in the gene pool are violations of alleles and genotypes frequencies expected by the Hardy-Weinberg model suggest the need for formulate a hypothesis on the causes of lack of genetic equilibrium. What is more, it demonstrates the validity of further studies, which seek to identify processes causing any deviations.

Models of the dynamics of gene pool

Hardy-Weinberg Law is limited by several assumptions. In fact, very rarely is the situation that in the study population all the criteria are met [3]. Moreover, the evolution of the tested organisms is constantly observed. It would not be possible, if the frequency of genotypes in populations remain static from generation to generation. So, it is worth to consider the impact

of deviations from some assumptions of Hardy-Weinberg law on the gene pool. It is very interesting as evolutionary mechanisms, such as non-random mating, migration, natural selection, mutation and genetic drift, break genetic equilibrium.

Non-random mating

If individuals have a certain preference during choosing partners for reproduction, we have to deal with non-random mating. There are many causes of selective crossing. For example, an individual may favor those similar to himself/herself, and thus of the same genotype. The number of homozygote will increase in the offspring [10]. This is due, among others, to the fact that individuals with AA genotype crossing with genotypically identical homozygote AA procreate only homozygote AA. Considering this situation, let us suppose that:

$$\begin{aligned} P'_{AA} &= p^2 + M_1 \\ P'_{aa} &= q^2 + M_2 \end{aligned} \tag{7}$$

where M_1 and M_2 correspond to the values of deviations from genotype frequencies expected from the Hardy-Weinberg equilibrium.

Bearing in mind that the frequency of genotypes in the population sum up to one, it is:

$$P'_{Aa} = 1 - P'_{AA} - P'_{aa} = 2pq - M_1 - M_2 \tag{8}$$

It will affect the frequencies of alleles in a descendant generation, as follows:

$$\begin{aligned} p' &= P'_{AA} + \tfrac{1}{2}P'_{Aa} = p + \tfrac{1}{2}M_1 - \tfrac{1}{2}M_2 \\ q' &= P'_{aa} + \tfrac{1}{2}P'_{Aa} = q - \tfrac{1}{2}M_1 + \tfrac{1}{2}M_2 \end{aligned} \tag{9}$$

Therefore, these frequencies will depend on M_1 and M_2 parameters. If $M_1 > M_2$, so individuals with genotype AA often mate with genotypically similar to one another than it is among individuals with genotype aa, then allele A proportion will continue to increase ($p' > p$) and allele a proportion will decrease ($q' < q$) over generation. Otherwise, when $M_1 < M_2$, the situation is reversed, i.e. $p' < p$ and $q' > q$. What's more, if $M_1 = M_2$, then $p' = p$ and $q' = q$. This case is extremely interesting. It shows that, despite the change in the proportions of individual genotypes in the genetic composition of the population, allele frequencies remain constant.

Analogous reasoning can be applied for the population, in which there are clear preferences for individuals of the opposite genotype. Then the number of heterozygotes will increase. This is caused by the fact that individuals with genotype AA crossing with those of genotype aa can only beget heterozygote Aa.

It is possible to find examples of non-random mating both in the world of humans and animals. An important factor influencing the choice of partner is the appearance, for instance, tall individuals generally choose tall partners. Whereas, in the case of livestock husbandry, inbreeding is often practiced. It is the reproduction from the mating of two genetically related parents which consolidates the desired trait to cumulate particularly valuable genes. In nature, this phenomenon is not preferred. It often contributes to the disclose of unfavorable recessive alleles due to increased homozygosity in the population. For example, inbreeding takes place in the population of cheetahs. It is the result of isolation of areas where these predators live due to human activities. It results in this species' danger of genetic disease.

Migrations

In the wild populations are rarely isolated. So there is often movement of individuals from one population to another. Consequently, gene flow occurs between the genetic pools. Migration is responsible for genetic material transfer, and thus results in marked changes in the frequency of genotypes and alleles.

Mathematical analysis of the genetic structure of a population, in which migration occurs, can provide interesting conclusions. For this purpose let us suppose that in the examined sample the frequencies of alleles A and a were respectively p_1 and q_1. To this population were attached individuals from another population, in which tested locus have also allele A and a with proportions p_0, q_0. Let immigrants be the n-th part of the population formulated in this way. The frequencies of alleles after a single mating will be developed as follows:

$$\begin{aligned} p' &= (1-n)p_1 + np_0 = p_1 - n(p_1 - p_0) \\ q' &= (1-n)q_1 + nq_0 = q_1 - n(q_1 - q_0) \end{aligned} \tag{10}$$

Migration causes a change of alleles proportions in the studied population due to the fact that it is very unlikely that allele frequencies are the same in indigenous populations and in outer. However, if that migration is a single phenomenon then one mating with assumptions of the Hardy-Weinberg Law

is enough to return to a state of genetic equilibrium. But when the migration takes place continuously in t generations and in the ratio of n, the frequency of allele A can be calculated by a mathematical induction in the following way:

$$\begin{aligned} p' - p_0 &= (1-n)(p_1 - p_0) \\ p'' - p_0 &= (1-n)^2(p_1 - p_0) \\ p^{(t)} &= (1-n)^t(p_1 - p_0) + p_0 \end{aligned} \tag{11}$$

This formula allows, among others, to determine how quick is the movement of genes to the target population. For example, Afro-American citizens of the United States are descendants of the population coming from Africa as slave labor from XVII century and migrants flowing from a Caucasian population. R_0 allele frequency of the Rh gene in this population is $p^{(t)} = 0.446$. The proportion of this allele in African ancestors population was $p_1 = 0.630$ while its share in the Caucasian population is $p_0 = 0.028$. The migration started about 350 years ago which means about 14 generations. The transformation of the formula (11) allows to get:

$$n = 1 - \sqrt{\frac{p^{(t)} - p_0}{p_1 - p_0}}$$

It makes the possibility to calculate that gene flow occurs at a rate of $n = 2.6\%$ per generation.

Natural selection

An extremely important factor among the mechanisms responsible for evolution is natural selection. It takes place when some organisms have a greater ability to survive and reproduce, depending on their genetic traits. This leads to a gradual increase proportions of such individuals in the population, due to better adaptation to the environmental conditions where they live. The analysis of the selection process allows us to study the directions of its activity and the strength of the phenomenon.

The measure of the adaptation to life conditions of organisms characterized by certain combination of alleles in the tested locus is the fitness coefficient W. It is estimated by the extent of considered allele transfer from generation to generation as compared to other alleles [10]. The absolute fitness is defined as the product of the probability of survival of a particular genotype to reproductive age (F) and the number of offspring released per one parent (L). In the analysis of the phenomenon of selection it is more convenient to use the relative coefficient, the so called relative

fitness, referring to the index calculated for the best suited genotype in the population:

$$W = \frac{W^*}{W^*_{max}} \tag{12}$$

To express the strength of natural selection the selection coefficient is also used, which is the complement of fitness: $s = 1 - W$.

The preference of the environmental conditions of certain genotypes will inevitably affect alleles frequency in the population. To better understand the mechanism of selection, it is worth considering the general model of allele frequencies changes which do not depend on which of the genotypes is favored. Let us denote W_{AA}, W_{Aa}, W_{aa} fitness coefficients corresponding to individual genotypes. Then the frequencies of genotypes will change as follows:

$$\begin{aligned} P'_{AA} &= p^2 W_{AA} \\ P'_{Aa} &= 2pqW_{Aa} \\ P'_{aa} &= q^2 W_{aa} \end{aligned} \tag{13}$$

Moreover, average fitness of the whole offspring population is given by the formula:

$$\overline{W} = p^2 W_{AA} + 2pqW_{Aa} + q^2 W_{aa} \tag{14}$$

Now the total number of alleles in the population will be equal to $2N$. Therefore using the formula (3) it is possible to calculate the frequency of each allele in the gene pool:

$$\begin{aligned} p' &= \frac{p(pW_{AA} + qW_{Aa})}{\overline{W}} \\ q' &= \frac{q(pW_{AA} + qW_{Aa})}{\overline{W}} \end{aligned} \tag{15}$$

However, it seems interesting to determine the allele frequency growth rate that occurred between tested and offspring generation:

$$\begin{aligned} \Delta p = p' - p &= pq\frac{p(W_{AA} - W_{Aa}) + q(W_{Aa} - W_{aa})}{\overline{W}} \\ \Delta q = q' - q &= pq\frac{q(W_{aa} - W_{Aa}) + p(W_{Aa} - W_{AA})}{\overline{W}} \end{aligned} \tag{16}$$

Models presented above show that all changes depend on differences in fitness between individual genotypes. They are also proportional to the variance of allele frequencies, determined by the binomial distribution

$\sigma^2 = pq/2N$. Generally, one can say that the selection is stronger, the greater is the genetic variation in the studied population. Without such variation, selection is impossible [2].

Even more interesting conclusions may be provided by the analysis of the selection against particular genotypes occurring in the population. For example, it is interesting to determine, what effect will cause the weakest fitness of the recessive allele, whose presence is revealed only in the homozygote recessive. This situation is called *selection against homozygote recessive*. This phenomenon commonly occurs in nature. Many diseases or genetic defects are conditioned by the recessive allele. This allele results from a mutation and it causes lack of encoding of a functional protein. The deficiency is compensated by the second occurring allele, which thus becomes dominant. Mutated allele is removed from the population, as a less fitted to the prevailing conditions.

In the case of selection against homozygote recessive, fitness coefficients take values $W_{AA} = 1$, $W_{Aa} = 1$, $W_{aa} = 1 - s_{aa}$. According to the formula (16) the frequency of a recessive allele growth rate will be of negative value amounting to:

$$\Delta q = -\frac{pq^2 s_{aa}}{1 - s_{aa}q^2} \tag{17}$$

Then the proportion of recessive allele in the gene pool will systematically decrease and the strength of removing it from the population will depend, among others, on the initial frequency.

A classic example of selection against homozygote recessive was observed in England at the turn of centuries XIX and XX in reference to the population of the peppered moth (Biston Betularia). This variety of moth initially had mainly light colouration, determined by the recessive allele. It allowed them to effectively camouflage against predators in lichens growing on trees which were inhabited by these organisms. However, with time, as a result of widespread environmental pollution lichens were significantly reduced. Then, more suitable became a variety of dark-coloured because of their ability to hide on the darkened trees. So these individuals began to occur more often. This phenomenon is called the "industrial melanism" [2].

It is also worth analyzing in more detail a case of *selection causing complete elimination of a recessive allele* from gene pool, in a the situation when one of the alleles is lethal. Values of fitness coefficients will be $W_{AA} = 1$, $W_{Aa} = 1$, $W_{aa} = 0$. Therefore, in accordance with formula (15), recessive allele frequency in descendant generation is equal to:

$$q' = \frac{q}{1 + q} \tag{18}$$

Using the principle of mathematical induction it can be proved that, after n generations, the recessive allele frequency will take the form:

$$q^{(n)} = \frac{q}{1+nq} \tag{19}$$

An example illustrates how the changes of recessive allele frequency over generations in the situation when it works selection causing complete elimination of homozygote recessive may look, is given on [Fig. 3].

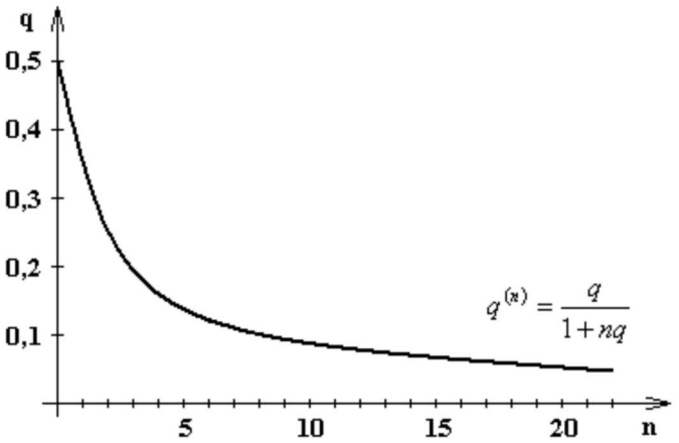

Fig. 3. The function depicting the decrease of the recessive allele frequency (in a population with initial frequency of $q = 0.5$), which presents the effect of actions the natural selection causing elimination of homozygote recessive

For example, if value of recessive allele frequency is equal to 0.5 in the tested population, [Fig. 3] shows how this frequency will change over successive generations. It allows to make the conclusion that the smaller frequency of the allele, the slower it is decreased [2].

These observations lead to the negation of the theses formulated by the theory of eugenics. Its idea was based on the necessity of improving the population by encouraging the reproduction of individuals with desirable heredity traits and discouraging the reproduction of organisms less genetically valuable. It disseminated the fear that the mating of individuals with certain genetic defects may lead to the degeneration of the population by the spread of these defects [10]. As a solution of this problem, eugenics proposed sterilization of people with genetic disorders. Model depicted in [Fig. 3] shows that the postulates of this theory are unfounded. Prohibition of reproduction of organisms with heritable disorders will not remove defective alleles from the population. It can only reduce their frequency. In addition, this decrease will

be small, because of the low initial frequency of the allele responsible for the disease. For example, cystic fibrosis affects about one in 2500 Caucasian individuals (i.e. $q^2 = 0.0004$). Hence, in this population the frequency of the recessive allele causing the disease is $q = 0.02$. According to formula (19), to double reduce the proportion of patients in that population, 21 generations are needed. Tenfold decrease would require 109 generations.

Quite interesting phenomenon occurring in nature is when the best adapted to life are heterozygous organisms. The situation of heterozygote advantage is known as overdominance. So *a selection against both homozygotes* takes place. It means that fitness coefficients are equal to: $W_{AA} = 1 - s_{AA}$, $W_{Aa} = 1$, $W_{aa} = 1 - s_{aa}$. Substituting these values into formula (16) it is possible to obtain:

$$\Delta q = \frac{pq(ps_{AA} - qs_{aa})}{\overline{W}} \tag{20}$$

This allows us to conclude, that the growth rate Δq will be positive, so the frequency of the recessive allele will increase from generation to generation, when $ps_{AA} - qs_{aa} > 0$. After simple transformations this formula gives the relationship:

$$q < \frac{s_{AA}}{s_{AA} + s_{aa}} \tag{21}$$

Otherwise, the proportion of recessive allele in population will decrease with time. However, when the recessive allele frequency will be:

$$q_r = \frac{s_{AA}}{s_{AA} + s_{aa}} \tag{22}$$

then the population reaches a genetic equilibrium. So proportions of alleles in population will remain constant in subsequent generations. It is worth emphasizing that this equilibrium does not depend on the initial allele frequencies, but only on their fitness [2]. Graphical representation of function of recessive allele growth rate depending on the proportion of this allele in the studied population is shown in [Fig. 4].

The pressure of selection against both homozygotes does not remove any of alleles from the population. Each deviation from genetic equilibrium q_r makes that population strives to return to a state of stability.

The only example of overdominance which was previously explored and explained in a widely recognized way is the phenomenon of persistence of the allele causing sickle-cell anaemia in the African population exposed to malaria [11]. Homozygotes recessive suffer from anemia and in 80% of cases die not surviving to the reproductive period (so their fitness is $W_{aa} = 0.20$).

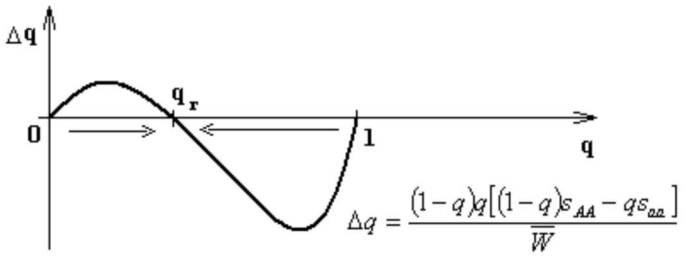

Fig. 4. The function of recessive allele growth rate depending on the frequency of this allele, which is a result of acting a selection against both homozygotes [2]

Although homozygotes dominant are devoid of the defective allele and they do not have immunity to malaria, which is high mortality disease. Therefore, heterozygotes are the best adapted to African life conditions. Admittedly, such individuals are sickle-cell anaemia allele carriers, but do not suffer from malaria.

Important conclusions may be supplied by a case study, when heterozygous individuals are the least adapted to the specific environment. Then *a selection against heterozygote* takes place. For example, such situation may occurs when two populations homozygous to different alleles of the tested gene merge. As a result of mating individuals from such created population will be heterozygous organisms. But their fitness is the lowest in the population [11]. For selection coefficients equal to s_{AA}, s_{Aa}, s_{aa}, where $s_{Aa} \geq \max\{s_{AA}, s_{aa}\}$, it is possible to perform a similar reasoning as above, to get recessive allele growth rate in subsequent generations as:

$$\Delta q = pq\frac{q(s_{Aa} - s_{aa}) + p(s_{AA} - s_{Aa})}{\overline{W}} \tag{23}$$

Genetic equilibrium will be obtained in the population when:

$$q_r = \frac{s_{Aa} - s_{AA}}{2s_{Aa} - s_{AA} - s_{aa}} \tag{24}$$

What is worth emphasising – this is not a stable equilibrium point. Each value of recessive allele a frequency below the value that gives the equilibrium, and hence $q < 0$, will lead to the elimination of this allele from the population due to the selection pressure. On the other hand, such proportions above the value in the equilibrium cause further move away from this state until the fixation of the allele in the population (due to $\Delta q > 0$), and thus result in the removal the alternate allele A from the population. Thus, if on alleles act only the selection against heterozygote, it results in the elimination of one of them.

Mutations

The main cause of any genetic variation is mutations. They can be defined as sudden changes in DNA sequence arising as a result of errors during the replication process or under the influence of physical and chemical factors [10]. Point mutations, such as silent or missense, that change only a single nucleobase, are often neutral. They do not cause dysfunctions of the genome and do not affect the phenotype. On the frequency of allele, which have resulted from mutations, does not interact the pressure of selection. But the majority of mutations, including nonsense mutations, deletions, insertions and reading frame shift can produce dangerous effects. It can result in the rise of alleles with slightly modified functions or entirely harmful and even lethal in homozygous individuals. These alleles are recessive when this error is compensated by the alternative allele, and in homozygous organisms they cause many genetic disorders. An example of this phenomenon is the Tay-Sachs disease occurring mainly among the Ashkenazi Jews population. It is usually the result of insertion of four nucleobase pairs in 11 exon HEXA gene (606869.0001). It makes the reduction of the activity or lack of synthesis of the beta-hexosaminidase A enzyme, what leads to the accumulation of fatty acid-gangliosides GM2 in brain nerve cells. This disease is lethal in homozygous recessive organisms and it causes the death of a 3–4 years old child.

In conclusion it should be emphasized that all types of mutations affect the allele frequencies of particular gene in the studied population. They contribute to creating new alleles, and thus changing the composition of the gene pool. However, for these considerations a situation may be important when in the bi-allelic population occurs a transformation from one allele to another and vice versa. Let allele A be converted by mutation in allele a with probability u, and let the reverse transformation take place with probability v. Then, if there is no pressure of natural selection then a recessive allele growth rate over successive generations is given by formula:

$$\Delta q = up - vq \tag{25}$$

The population will reach genetic equilibrium, which is constantly trying to achieve, when the proportion of allele a will be:

$$q_r = \frac{u}{u+v} \tag{26}$$

Mutant alleles, which have lost their functionality, are recessive. They are revealed in the homozygote recessive [10]. So, selection against homozygote recessive interacts on them. Taking into account both the pressure of muta-

tion and natural selection, changes of recessive allele frequency over generations are as follows:

$$\Delta q = u(1-q) - vq - \frac{(1-q)q^2 s_{aa}}{1-q^2 s_{aa}} \tag{27}$$

Due to the fact that the frequency of the recessive allele in the population is small, the component vq can be neglected. It has only a little impact on the growth rate. Assuming $\Delta q = 0$ and making simple transformations in formula (27) it can be seen that the tested forces will be in equilibrium at the frequency of allele a equal to:

$$q_r = \sqrt{\frac{u}{(1+u)s_{aa}}} \tag{28}$$

In conclusion, it should be noted that the mutation is a slowly progressing factor in evolution. It occurs between once per 10 thousand and once per 100 thousand on allele at generation. Only the mathematical analysis of mechanism of mutations model reveals that they play a really important role. Indeed, the rarer recessive allele a in the population, the higher is its growth rate over successive generations, which is perfectly shown in the formula (27). Then, there are more frequent transformations from allele A to a. Moreover, the influence of selection pressure is less important [11].

Genetic drift

Gene transfer from parents to offspring is random. Any changes in time of allele frequencies in the gene pool, which are not the result of the above described phenomena but are caused by the accidental fuse of gametes in the reproduction, are referred as genetic drift. In a finite population, where frequency of allele a is p and there is no pressure causing evolution, tested allele frequency will not be exactly replicated after a single mating, but will be equal approximately to p. In the next generation such situation will be repeated, but this time studied proportion of allele will endeavor to the new value. Such fluctuation of allele frequencies is unpredictable – the direction or the strength of these changes can not be determined. Sometimes it leads to the elimination or fixation of one of the alleles in the population regardless of its fitness. Thus it contributes to the homogeneity or loss of variability in the population. The speed of this phenomenon depends on the size of the population. This fact is exactly presented on [Fig. 5], where fluctuations were observed in small population of five individuals [Fig. 5a] and large, 100-members populations [Fig. 5b]. In each population, the initial proportion of the tested allele is 0.4.

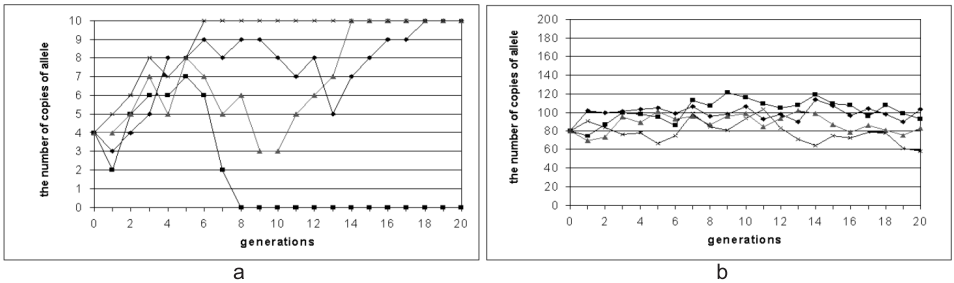

Fig. 5. Computer simulations of changes in the frequency of allele in four bi-allelic populations as a result of genetic drift.

Genetic drift is one of the main mechanisms of evolution, as important as mutations. It plays a pivotal role especially in promoting favorable and rare mutations, before they have stabilized [10]. The effect of drift is clearly visible in a small, isolated population. Such populations may arise, for example, as a result of a bottleneck effect. It takes place when the population drastically reduced their abundance, because of some disaster (drought, flood, earthquake, disease, etc.). The new gene pool is different from the former one because it consists of accidentally selected individuals who do not have all alleles, which were in the initial population, or have them with completely new proportions. Another way to create a population is when several individuals leave the parent population and then inhabit an entirely new area, so form a new population by mating. This phenomenon is called the founder effect. As an example it may be consider a situation of 15 British colonists, who in 1814 established a settlement on the islands of Tristan da Cucha in the Atlantic Ocean. After 150 years, it became clear that in the offspring population disproportionately frequent is recessive allele *retinitis pigmentosa* responsible for a progressive form of blindness. The reason for this phenomenon is the fact that one of the colonists was the carrier of the mutated allele. The difference between the bottleneck effect and the founder effect is that in the case of the first phenomenon one population is transformed into a completely new, and the second one causes the formation of two populations, which occur side by side.

Conclusions

The Hardy-Weinberg Law is still interesting to researchers. For many scientists, who study this principle, it provides new and important conclusions. For example, C. C. Li showed [12] that non-random mating is only

a sufficient condition, not a necessary one, to static frequencies of genotypes in a population across generations.

Knowledge of the Hardy-Weinberg Law and its conclusions is very important in medical science. It is the basis of population genetic research. Moreover, it is used during clinical trials to establish the quality of data by comparing the observed frequency of each genotype with those, which are expected under the Hardy-Weinberg model. When there is no genetic equilibrium, conventional statistical analysis can not be carried out and the data should be usually excluded from further analysis [13]. But, in practice, testing the compliance of genotypes frequencies distributions in the population is often neglected. Any deviations from genetic equilibrium are rarely admitted in the published reports, although they may be crucial for the course of the study [14–15]. They may indicate problems, errors or oddities in the analyzed data. Consequently, any conclusions drawn for the population, in which there is no Hardy-Weinberg equilibrium, can be challenged.

The findings, during trials that the frequencies of alleles and genotypes in the study population are incompatible with the Hardy-Weinberg model demand the necessity to look for reasons [16]. Factors, discussed in the paper, affecting allele frequencies, the conditions under which they may take place, their strength and direction of interaction allow to formulate the assumptions about the phenomenon which cause the disorder. Knowledge related to them is the basis for further research which may lead to the verification of many hypotheses.

REFERENCES

[1] Edwards A. W. F., G. H. Hardy (1908) and Hardy-Weinberg equilibrium, Genetics, 179(3), pp. 1143–1150, 2008.
[2] Krzanowska H., Łomnicki A., Rabiński J., et al., Zarys mechanizmów ewolucji, PWN, 2002.
[3] Korzeniewski B., Powstanie i ewolucja życia, EREM FOSZE, Rzeszów, 1996.
[4] Łomnicki A., Spotkanie teorii Darwina z genetyką, KOSMOS, T. 58, nr 3–4, pp. 315–317, 2009.
[5] Crow J. F., Eighty years ago: The beginnings of population genetics, Genetics, 119 (3), pp. 473–476, 1988.
[6] Salomon E. P., Berg L. R., Martin D. W., Biologia, MULTICO, 2007.
[7] Ziegler A., Konig I. R., A statistical approach to genetic epidemiology, WILEY-VCH, 2010.
[8] Evens W. J., Mathematical Population Genetics, Springer, 2004.
[9] Engels W. R., Exact test for Hardy-Weinberg proportions, Genetics, 183(4), pp. 1431–1441, 2009.
[10] Fletcher H., Hickey I., Winter P., Genetyka, PWN, Warszawa, 2010.

[11] Łomnicki A., Dobór naturalny. Dostosowanie i współczynnik doboru oraz inne miary dostosowania, KOSMOS, T. 58, nr 3–4, pp. 323–328, 2009.
[12] Li C. C., Pseudo-Random Mating Populations. In Celebrations of the 80^{th} Anniversary of the Hardy-Weinberg Law, Genetics, 119 (3), pp. 731–737, 1988.
[13] Short A. D., Kennedy L. J., Barnes A., et al., Hardy-Weinberg expectations in canine breeds: Implications for Genetic Studies, Journal of Heredity, 98 (5), pp. 445–451, 2007.
[14] Salantil G., Amountza G., Ntzani E. E., Ioannidis J. P. A., Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power, European Journal of Human Genetics, 13, pp. 840–848, 2005.
[15] Kotsis I., Gyorffy B., Nemeth E., Vasarhelyi B., Examination of Hardy-Weinberg equilibrium in papers of Kidney International: an underused tool, Kidney International, 65, pp. 1956–1958, 2004.
[16] Xu J., Turner A., Littre J., et al., Positive results in association studies are associates with departure from Hardy-Weinberg equilibrium: hint for genotyping error?, Human Genetics, 111, pp. 573–574, 2002.