THE HARDY-WEINBERG LAW AND THE DISTRIBUTION OF POPULATIONS ON TERRA[1]

Eugenie POSDĂRĂSCU[*]

Abstract. *In the present paper, a genetic analysis of the distribution and compatibility of populations on Terra in Europe and in the Carpathian-Balkan-Danubian area is made, following human migrations, using Y-chromosome haplogroups (Y-DNA) and mitochondrial DNA haplogroups. For this we use the Hardy-Weinberg law, also known as the principle of genetic equilibrium. She states that the frequencies of alleles and genotypes in a population will remain constant from generation to generation in the absence of other evolutionary influences. These influences include partner selection, mutation, selection, genetic deviation, gene flow, and meiotic mechanism. This law eliminates the confusion between domination and selection. Today, tests for the frequencies of the Hardy-Weinberg genotype are mainly used to test stratification of the population and other non-random mating patterns.*

Keywords: *genetic analysis, heredity, Hardy-Weinberg law, haplogrup, Y-DNA, mtDNA, human migration, Europe, National Geographic, Carpathian-Balkan-Danube area.*

Ancient origins

Genetics studies the molecular and functional structure of genes, the behavior of genes in the context of a cell or organism, the way of transmitting the characters from the parents to the offspring, the distribution of genes and the variation and change of populations[2].

Genetics is the science of heredity, variability and reproduction of living beings.

The oldest information about genetics we have been still from ancient times during the Mesolithic period (about 15.000 – 5.000 ICs) when man, a plant cultivator and animal breeder, noticed their combinations and realized the importance of results. It seems that 13.000 years ago, the Sumerians managed to pollinate the plants artificially and create animal breeds through crushing. Since antiquity, character eradication has been

[*] Faculty of Exact and Engineering Sciences, "Hyperion" University, Bucharest.

[1] https://ro.wikipedia.org/wiki/Genetic%C4%83#ADN-ul_.C8.99i_cromozomii

observed in animals as well as in humans – children inherit certain features from parents (Observations made on horses are found on a stone tablet from 6.000 years ago).

Over time, the evolution of human society, corroborated with the need for comfort and well-being, has led to the accumulation of knowledge of heredity and its mode of exploitation for the benefit of both man's material and spiritual.

Genetics as a science has found its origins only in the middle of the nineteenth century due to the works and experiments of the Austrian monk Gregor Mendel that led to the establishment of the laws that bear his name. One of his theories argued that individuals inherit a mixture of traits from their parents, and that these traits occur as a result of distinct gene combinations and not a continuous mix.

Mendel's theories have not been taken into account for a good period of time, recognizing the importance of his discoveries being made only at the beginning of the twentieth century when Hugo de Vries, Carl Erich Correns and Erich von Tschermak got independently to the same conclusions as Mendel, rediscovering the laws of heredity. In 1903 the theory emerged that chromosomes are the bearers of heredity, and in 1909 that genes are located on the chromosomes.

The **genetic** term was introduced for the first time in 1906 by William Bateson, and in 1927 the term genetic mutation appeared used to describe the changes in the genetic material. In 1953, the helical structure was deciphered, double-stranded DNA structure (a substance bearing the hereditary material, present in all living organisms, mostly at the cell nucleus), and in 1957 the mechanism of its replication was described. The first half of the twentieth century was dedicated to the biochemistry and genetics experiments that confirmed that DNA is the genetic material.

All the cells of a body contain practically the same DNA. Genetic information is stored in the DNA as a code consisting of 4 macro-molecules: adenine, guanine, cytosine, and thymine. The order of these substances determines the essential genetic information for the development and maintenance of an organism.

The gene is the functional unit of DNA located in the chromosome that contains the information necessary for the processing of certain proteins and is responsible for the transmission of parent-child traits (phenotype).

Chromosomes are core structures of living cells that contain genes and are designed to preserve and correctly transmit the hereditary information from parental cells to descendants in the cell division process.

A living cell contains a well-established number of chromosomes in pairs. In the case of man, a normal cell contains 46 chromosomes, 23 inherited from the father and 23 from the mother.

Genetic trace is a particular configuration of DNA sequences specific to an individual. Two genetic traces from individuals have a probability of 10^{-7} being identical.

In the cell division process, DNA replication occurs by replicating its sequence. If DNA replication errors occur in the process, then we can say that genetic mutations occur. The causes of these mutations are numerous: from UV radiation to the chemicals we come into contact with, the food we eat or the quality of the environment we live in.

If the gene changes its structure, it will give rise to another alternative form called the allele, thus creating new protein structures and new characters. Any variation in the sequence of the DNA molecule can be considered as an allele. Allele is a discreet version of the same gene (red or yellow roses).

Heredity is a fundamental property of all living creatures; it is responsible for fidelity transmission of morphological (anatomical), physiological and behavioral (hereditary characters) from parents to descendants.

Variability is property belonging to the diversity of the living world, whereby each individual is unique; through which the offspring differ from parents and brothers due to qualitative or quantitative phenotypic differences.

Genotype is the genetic constitution of a cell, organism or individual, genetic material.

The phenotype represents the biochemical, morphological, physiological and behavioral, normal or pathological characteristics of an individual determined by its genetic constitution (genotype) and modulated by environmental factors.

Through his studies, Mendel discovered the difference between the genetic structure of organisms – the genotype and the appearance of organisms – the phenotype (the result of the interaction between the genotype and the environment).

Gametes are male and female, sexually differentiated. Gamete is a mature germ cell that, by fecundation another of the opposite sex, generates a zygote.

Each character of an organism is controlled by a pair of genes, called alleles. Each allele is transmitted from one generation to another as a

discrete unit. The dominant alleles (A) are highlighted in the uppercase and the lower case letters are hidden, recessive (a).

Diploid organisms with two copies of the same allele of a given gene are called homozygotes, whereas organisms with two different alleles of a given gene are called heterozygotes.

The allele set for a given organism is called genotype, while the observable features of the body are known to be its phenotype. When organisms have a heterozygous gene, often one allele is named a dominant allele since it determines the phenotype of the organism, while other alleles are called recessive alleles, since they are recessive in character and are not observed. Some alleles do not have complete dominance and then they have semi-dominance by the appearance of an intermediate phenotype, or they have a codominant by flushing both alleles at the same time.

Diploid cells are cells with a double set of similar chromosomes.

Haploid cells are cells with a single set of chromosomes.

The Heredity laws of Mendel

Purity law of gametes

Gametes are always genetically pure no matter whether they come from heterozygous (Aa) or homozygous (AA or aa) individuals because they contain only one of the hereditary factors (dominant or recessive).

Independent Segregation law of pairs of characters

During cell division, the allele pair is segregated and each gamma receives one gene from the parent pair.

The Hardy – Weinberg Law (Genetic Balance Principle)

Population genetics is the study of the distribution and variation of allele frequencies in a population and is subject to the four important evolutionary processes: natural selection, genetic drift, genetic mutation and recombination. It also takes into account factors such as genetic migration, environmental change, population categorization and structure. The discipline examines phenomena such as adaptation and specialization. It is traditionally a highly mathematical discipline, but the genetics of modern populations incorporate theoretical, laboratory and field research. Since the 1980s, computational approaches have also been introduced.

The Hardy-Weinberg Law is also known as the principle of genetic balance. It states that the frequencies of alleles and genotypes in a population will remain constant from generation to generation in the

absence of other evolutionary influences. These influences include partner selection, mutation, selection, genetic deviation, gene flow, and meiotic mechanism.

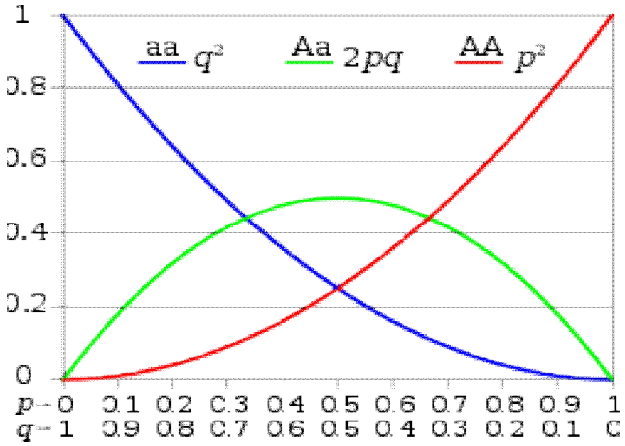


Figure 1. The model of the Hardy-Weinberg law in the presence of two alleles *A* and *a*: on the horizontal axis are the allele frequencies *p* and *q*, and on the vertical axis the frequencies of the three types of genotype.

In the simplest case of a single locus with two alleles denoted *A* and a with frequencies $f(A) = p$ and $f(a) = q$, respectively, the frequencies of the expected genotype under random pairing are $f(AA) = p^2$ for homozygous *AA*, $f(aa) = q^2$ for homozygotes *aa* and $f(Aa) = 2pq$ for heterozygotes. In the absence of selection, mutation, genetic drift or other forces, the frequencies of the *p* and *q* alleles are constant between generations, so that equilibrium is reached.

The principle is named after G. H. Hardy and Wilhelm Weinberg, who demonstrated it for the first time in Mathematics. Hardy's theory focuses on the fact that often a dominant allele tends to grow. This law eliminates the confusion between domination and selection. Today, tests for the frequencies of the Hardy-Weinberg genotype are mainly used to test stratification of the population and other non-random mating patterns.

The number and frequency of the locus are variable from one population to another and, over time, within the same population, depending on the evolutionary history of the population. The number and frequency of alleles define the genetic variability of the population. Alleles occur due to spontaneous mutations, defining themselves as inherent replication errors of DNA molecules over generations.

We suppose we have a population of hermaphrodite diploids, where each organism produces male and female gametes at the same frequency

and has two alleles at each gene locus. The organisms are reproduced by a random association of gametes ("the gene pool" population model). A locus in this population has two alleles, A and a, which occur at the initial frequencies $f_0(A) = p$ and respectively $f_0(a) = q$. The allelic frequencies of each generation are obtained by pooling the alleles of each genotype of the same generation according to the expected contribution of homozygous and heterozygous genotypes that are 1 and respectively 1/2:

$$f_t(A) = f_t(AA) + \frac{1}{2} f_t(Aa) \quad (1)$$

$$f_t(a) = f_t(aa) + \frac{1}{2} f_t(Aa) \quad (2)$$

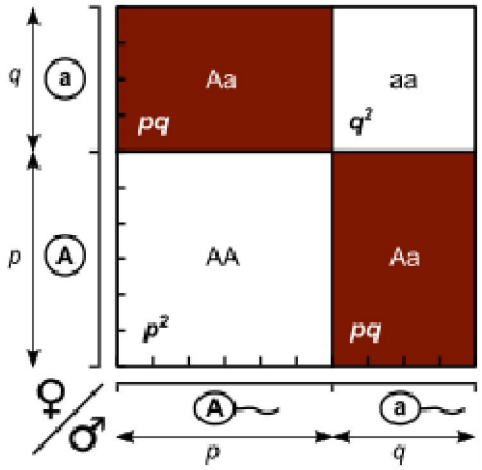


Figure 2. The length p, q corresponds to the allele frequencies (here $p = 0.6$, $q = 0.4$). Then the rectangle area represents the genotype frequencies (thus AA: Aa: aa = 0.36: 0.48: 0.16).

The ways in which it can form the genotypes for the next generation can be shown in a Punnett square, where the proportion is equal to the product of each genotype frequencies of alleles row and column of the current generation.

Punnett

		Female	
		***A*(*p*)**	***a*(*q*)**
Male	***A*(*p*)**	***AA*(p^2)**	***Aa*(*pq*)**
	***a*(*q*)**	***Aa*(*qp*)**	***aa*(q^2)**

Square

The amount of records is $p^2 + 2pq + q^2 = 1$, because the sum of the genotype frequencies is 1. The same relationships are obtained from:

$$f_0(A) + f_0(a) = p + q = 1$$
$$(p + q)^2 = p^2 + 2pq + q^2 = 1.$$

By adding elements of Punnett square or binomial expansion, we get the expected genotype ratios among offspring after a single generation:

$$f_1(AA) = p^2 = [f_0(A)]^2$$
$$f_1(AA) = pq + qp = 2pq = 2f_0(A)f_0(a)$$
$$f_1(aa) = q^2 = [f_0(a)]^2.$$

These frequencies define the balance of Hardy-Weinberg. It should be noted that the genotype frequencies after the first generation should not be equal to the genotype frequencies from the initial generation:

$$f_1(AA) \neq f_0(AA).$$

However, the frequencies of the genotype for all future times will equal Hardy-Weinberg frequencies. This is because the frequencies of the next generation genotype depend only on the frequencies of the alleles of the current generation which, as calculated by equations (1) and (2). From the original generation to $t \geq 1$:

$$f_1(A) = f_1(AA) + \frac{1}{2}f_1(Aa) = p^2 + pq = p(p + q) = p = f_0(A)$$
$$f_1(Aa) = f_1(aa) + \tfrac{1}{2}f_1(Aa) = q^2 + pq = q(q + p) = q = f_0(a).$$

For the general case of dioid diploids (organisms are either men or women) reproduced by random association of individuals, it is necessary to calculate for each sex the genotype frequencies of the nine possible connections between each parental genotype (*AA*, *Aa* and *aa*), weighted by the expected genotype contributions for each such mating. Equivalently, the six unique diploid-diploid combinations are considered: **(*AA, AA*); (*AA, Aa*); (*AA, aa*); (*Aa, Aa*); (*Aa, aa*); (*aa, aa*).**

A single Punnett square is constructed for each one, so that it calculates its contribution to genotypes of the next generation. These contributions are weighted according to the probability of each diploid-diploid combination, which follows a multinomial distribution with $k = 3$. For example, the probability of mating pairing (*AA, aa*) is $2f_t(AA)f_t(aa)$) as a result only genotype *Aa*: [0,1,0]. Generally, the resulting genotype frequencies are calculated as:

$$[f_{t+1}(AA),\quad f_{t+1}(Aa),\quad f_{t+1}(aa)] =$$

$$= f_t(AA)f_t(AA)[1 \quad 0 \quad 0] + 2\, f_t(AA)f_t(Aa)\left[\tfrac{1}{2} \quad \tfrac{1}{2} \quad 0\right] +$$

$$+ 2\, f_t(AA)f_t(aa)[0 \quad 1 \quad 0] + f_t(Aa)f_t(Aa)\left[\tfrac{1}{4} \quad \tfrac{1}{2} \quad \tfrac{1}{4}\right] +$$

$$+ 2\, f_t(Aa)f_t(aa)\left[0 \quad \tfrac{1}{2} \quad \tfrac{1}{2}\right] + f_t(aa)f_t(aa)[0 \quad 0 \quad 1] =$$

$$[(f_t(AA) + \frac{1}{2}f_t(Aa))^2 \quad 2(f_t(AA) + \frac{1}{2}f_t(Aa)) \quad (f_t(AA) + \frac{1}{2}f_t(Aa))^2]$$

$$= [[f_t(A)]^2 \quad 2f_t(A)f_t(a) \quad [f_t(a)]^2].$$

As before, it can be shown that the frequencies of the alleles at time $t + 1$ are equal to those at time t and therefore are constant over time. Similarly, genotype frequencies depend only on allele frequencies and thus, after time $t = 1$, are also constant over time.

If in mono or dioic organisms, allele or genotype proportions are initially unequal to both genders, it can be demonstrated that after a random breeding generation, constant proportions are obtained.

Deviations from the law of genetic balance

The Hardy-Weinberg equilibrium is based on the following assumptions:

- Bodies (creatures) are diploid;
- Only sexual reproduction takes place;
- Generations do not overlap;
- Pairing is random;
- The size of the population is infinitely large;
- The allele frequencies are equal to the sexes;
- There is no migration, mutation or selection.

Violations of the Hardy-Weinberg hypotheses may cause deviations from expected results. How this affects the population depends on the hypotheses that are violated.

According to the HW principle, the population will have, after a single generation of random breeding within the population, certain given genotypic frequencies (called Hardy-Weinberg proportions). When the random mating hypothesis is violated, the population will not have the

proportions given by the Hardy-Weinberg law. A common cause of non-random mating is crossing, which results in an increase in homozygosity for all genes.

If a population violates one of the following four hypotheses, the population may continue to have Hardy-Weinberg proportions in each generation, but the allele frequencies will change over time.

- Selection generally causes alley frequencies to change, often quite fast. While directional selection eventually leads to the loss of all but the favored alleles, some selection forms, such as balanced selection, lead to balance without losing alleles.
- The mutation will have a very subtle effect on the allele frequencies. The mutation rates are in the order of 10^{-4} to 10^{-8}, and the change in the frequency of the alleles will, at most, be the same. Recurrent mutations will maintain alleles in the population, even if there is a strong selection against them.
- Migration links genetically two or more populations together. Generally, allele frequencies will become more homogeneous among populations. Some inherent migration patterns include nonrandom mating. For these models, the Hardy-Weinberg proportions will not normally be valid.
- The size of the small population can cause a random change of allele frequencies. This is due to a sampling effect and is called a genetic drift. The effects of sampling are most important when the allele is present in a small number of children.

Generalization

Generally, if we take into account alleles A_1, ..., A_n with the frequencies p_1 at p_n, we will write:

$$(p_1 + p_2 + \cdots + p_n)^2.$$

We will have frequencies for all homozygotes:

$$f(A_iA_i) = p_i^2$$

and for all the heterozygotes:

$$f(A_iA_{ij}) = 2p_ip_j.$$

The Hardy-Weinberg principle can also be generalized to polyploid systems, i.e. to organisms that have more than two copies of each chromosome. Considering only two alleles, we will have:

$$(p + q)^c.$$

For a tetraploid $c = 4$.

Genotype	Frequency
AAAA	p^4
AAAa	$4p^3q$
AAaa	$6p^2q^2$
Aaaa	$4pq^3$
aaaa	q^4

For distinct *c*-ploidy alleles, the frequency of genotypes according to the HW principle of genetic balance will be given by the expression:

$$(p_1 + p_2 + \cdots + p_n)^c.$$

In a population of infinite magnitude where the reproduction is of the panmyxic type (random crossing between individuals) and where there are no new migrations or mutations, the allelic frequencies, and therefore those of the corresponding genotypes, remain constant from one generation to the next (Hardy-Weinberg's law).

Although natural populations do not fulfill all the conditions imposed by the Hardy-Weinberg law, this law is used to calculate, with an acceptable error, allelic frequencies from phenotypes observed in the population.

Distribution of population on Terra

A haplotype is a group of genes in an organism are inherited together from one parent and a haploid (haploid from Greek: *πλούς, haploûs* "single, simple" and English: group) is a group of similar haplotypes which have a common ancestor with a single-nucleotide polymorphism mutation.

More specifically, a haplo-group is a combination of alleles at different regions of chromosomes that are closely linked and tend to be inherited together. A haplo-group consists of similar haplotypes. Haplogroup refers to a single line of descent, usually dating back thousands of years. As such, belonging to a haploid, by any person, is based on a relatively small proportion of the genetic material possessed by that individual.

Each haplogroup comes from and remains part of a single haploid (or paragrup). As such, any group of associated haplogruppes can be accurately modeled as a nested hierarchy, where each set (haplogroup) is

also a subset of a larger set (as opposed to biparental patterns such as human family trees).

Haplogroups are normally identified by an initial letter of the alphabet, and refinements consist of additional combinations of numbers and letters, such as (for example) $A \rightarrow A1 \rightarrow A1a$.

In human genetics, haplogroups most commonly studied are *Y*-chromosome haplogroups (*Y*-DNA) and mitochondrial DNA haplogroups (both can be used to define genetic populations).

Y-DNA is transmitted only along the patrilineal line from father to son while DNA mtDNA is transferred on the matrilineal line from the mother to the descendants of both sexes. *Y*-DNA and mtDNA DNA are not recombined, they change only by random mutation in each generation, without mixing between the genetic material of the parents.

It is usually assumed that there are few natural selections for or against a certain haplotypic mutation that has survived to this day. The prevalence of a particular marker in a population fluctuates further until it reaches either 100% or falls entirely out of the population. In a large population with efficient mixing, the rate of genetic diversion for the common alleles is very low. However, in a very small cross-section, the proportions may change much faster. Geographic variations and marked concentrations of haplotypes and groups of particular haplotypes thus mark the distinctive effects of repeated population blockages or founding events followed by population segregation and growth.

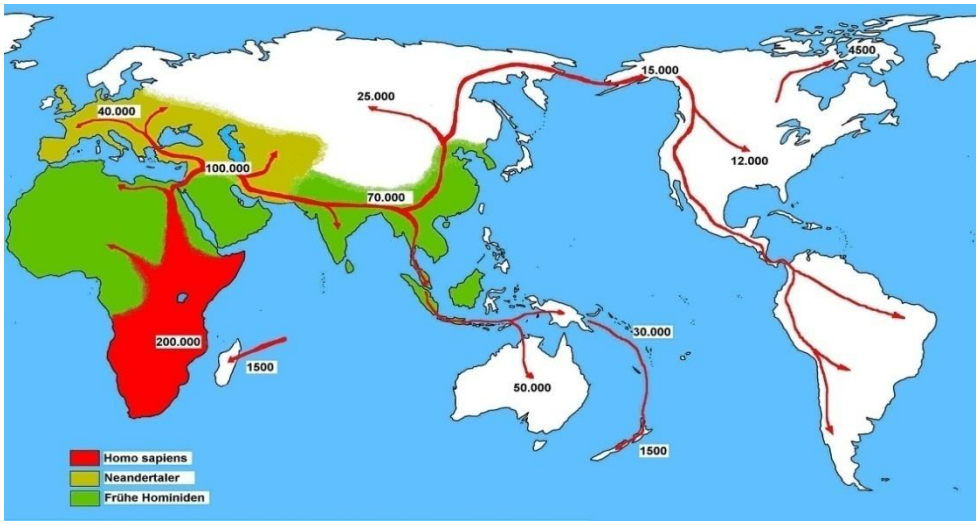


Figure 3. Expansion of Homo Sapiens.

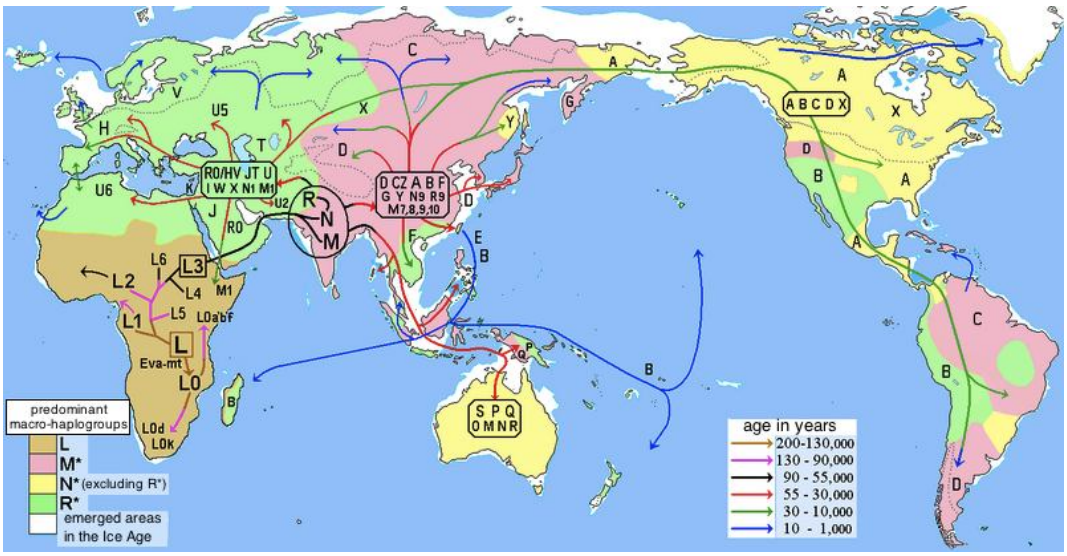


Figure 4. Human migration and mitochondrial mtDNA haplogrups.

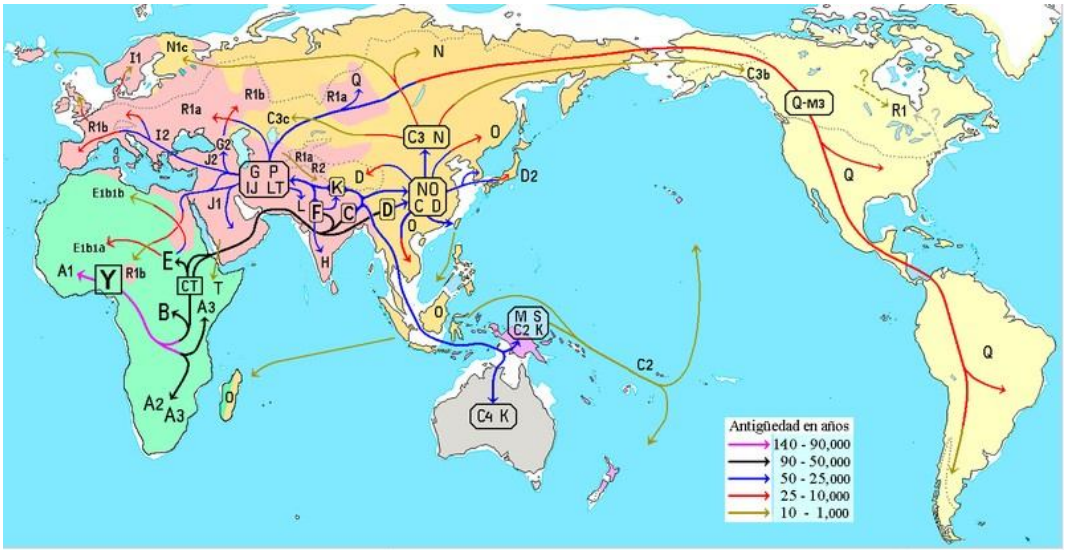


Figure 5. Human migration and Y-DNA haplogroups.

Distribution of population on Europe

Figure 6. Old Europe.

If we take into account the known data on the occurrence of Homo Sapiens in Europe, about 40 to 45.000 years ago, and their arrival on the Africa-Levant-Anatolia-Balkans route from where spreading to Europe continued, we could say that the haplogrup I descends from these primes the first division of the IJ trunk, taking place 38.000 years ago in the Balkans or Anatolia or both.

Considering that at that time we were at the peak of glaciation, the Balkan Peninsula has functioned as a shelter and hearth for a long time for these populations, enough to show the mutations that led to the emergence of the haplogroup I followed by its divisions.

The fact that Homo Sapiens has been found, dating from 40-45.000 years ago in Kents Cavern, England, or 43-45.000 years ago in Grotta del Cavallo, Italy, it shows that Homo Sapiens migration continued to the west, the more likely on a small scale due to extreme climatic conditions, the vast majority of populations remaining sheltered in the Danubian-Balkan area.

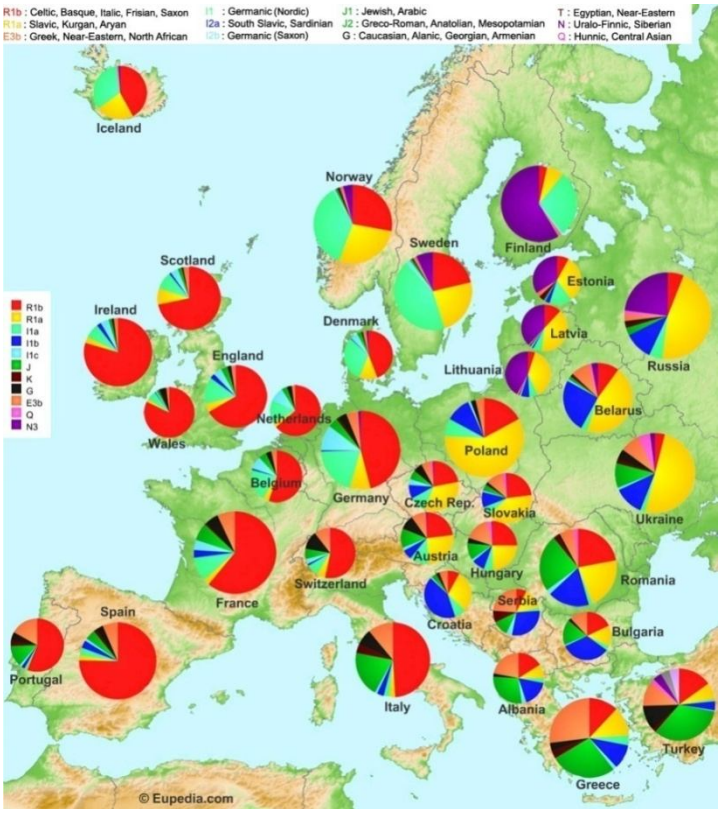
It is unclear what caused the Neanderthal population to disappear if Homo Sapiens was the cause or just took advantage of its extinction caused by the extreme weather conditions of 39-41.000 (the peak of glaciation). They have coexisted and even crossed the Neanderthals for at least 5000 years.

As soon as the glaciation ended, the expansion of the haplogroup I and its divisions into Western, Northern, Central and Mediterranean Europe began, starting from about 13.000 years ago. From this moment we can speak of a Danubian population.

The only populations that could compete with them would have been "wintering" in the Iberian Peninsula or the dwarf shelters in southern England, but the fact that we cannot find haplogroups Y that originate in those areas, but only a few "widowed" mitochondrial haplogroups "Shows that, in their expansion, the Danubians had no problem eliminating them, taking their women and continuing to colonize Europe".

Considering that we are not talking about densely populated populations, we can still say that the Danubians have been authoritatively controlling a huge territory for many thousands of years, even if they have not developed a unitary culture but several related cultures. Throughout this period they have assimilated, in successive stages, the populations that brought agriculture to Europe and the cattle breeders who arrived from Anatolia. At the same time, they have liquidated the populations they considered a threat, and the evidence is again "widowed" mitochondrial haplogroups to which Y haplogroups cannot be associated.

The R1a populations, after pushing northward by R1b, will return to areas in the northern Black Sea after their departure, but only around 3300 BC will come back into contact with "Old Europe". Their entry into Danubian territories will take place after the decline of the Cucuteni-Tripolie civilization, most likely due to the drought, which has been establishing for a long time, starting with about 3200 BC (the Subboreal climatic phase), which determines their transition from agriculture to shepherd. Around this period it began to find mixed tombs in the west of Dnieper, belonging to both funerary rituals, and it is estimated that the Thracian ethnogenesis will begin and will end around 1500 BC.



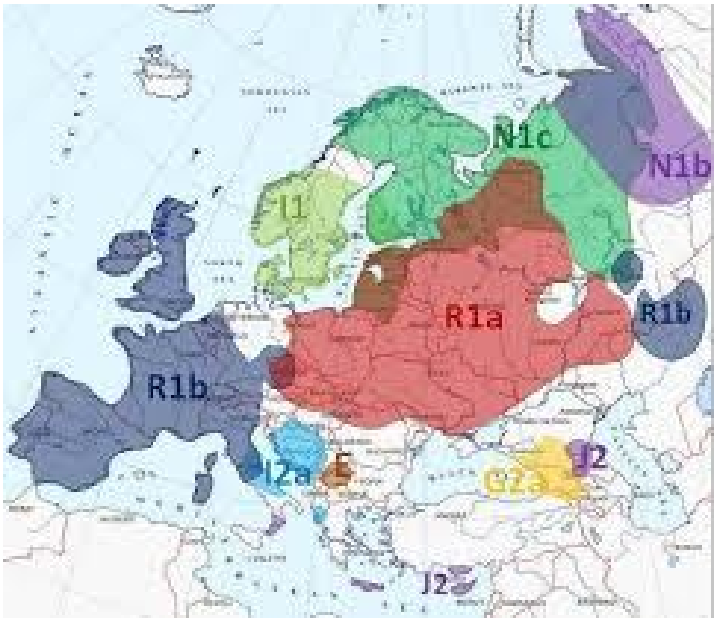


Figure 7. Y-DNA haplogrups in Europe.

In order to give us an idea of how close are the genetic compositions of the current European populations, please refer to the table of haplogroup frequency compositions on:

http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml.

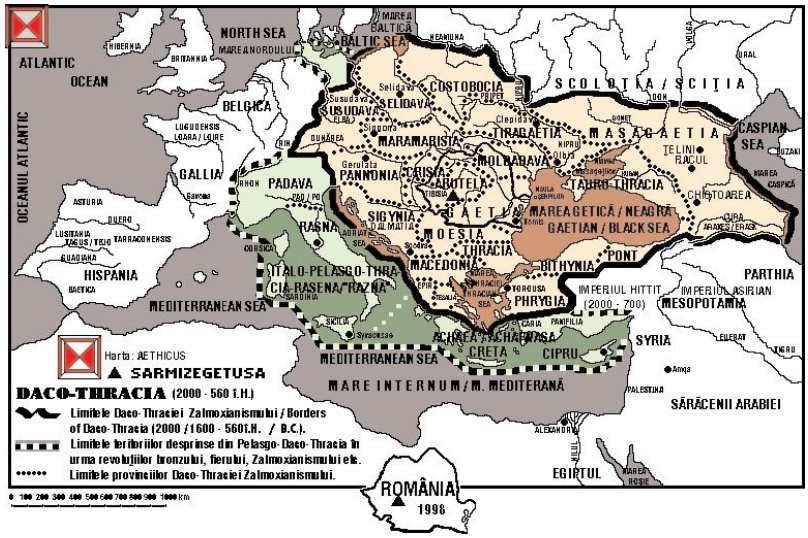


Figure 8. Ethnogenesis of Thracians.

Population distribution in the Carpathian-Balkan-Danube area

An analysis done on the data provided by eupedia.com shows that the population of Romania and its neighbors, after so many invasions and destructions, includes:

- On the paternal line, approximately 30% of the descendants of the native population 25.000 years ago, to whom more than 60% are added, descendants of some farmers and livestock breeders, arrived and assimilated 6.000-8.000 years ago, the rest represent more recent contributions, minor in importance.
- On the maternal line, about 50% descendants of populations with at least 25.000 years of age on this territory, plus 15% descendants of hunter-gatherers aged 13-17.000 years ago and about 20% descendants of farmers and breeders of cattle arriving from 6.000 to 8.000 years ago, with the remainder being insignificant.

The table below shows the genetic composition of Romanians, Bulgarians and Hungarians, as well as the Romanians' common composition with the Bulgarians:

Haplogroup	RO%	BG%	Common	H%	Observations
I1	4,5	4	4	8,5	Gothic Tribes, II-IV AD
I2/I2a	26	20	20	16	Descendants of the Danube 25ka
I2b	2,5	2	2	2	Descendants of the Danube 25ka
R1a	17,5	17	17	29,5	Pre-Indo-Europeans (Thracians, Germans, Iranians, Slavs), 3300-1500BC
R1b	12	11	11	18,5	Pre Celts, 4200-2500 and Celts, 2500 – 1200BC
G	5	5	5	3,5	The tribes that bring agriculture, 7000-6500, 6300-5800BC
J1	1,5	3	1,5	3	Cattle and goat breeders, 6000-5000BC
J2	13,5	11	11	6,5	Cattle and goat breeders, 6000-5000BC
E1b1b	15	23,5	15	8	North African hunter-gatherer slaves, 3000BC
T	0,5	1,5	0,5	0,5	Eastern-Anatolian cattle and goat breeders, 6000-5000BC
Q	0,5	0,5	0,5	0	Descendants of Huns, Mongols, Turks, the Soviet army
N	0,5	0,5	0,5	0,5	Descendants of Hungarians, Tartars, Turks and the Soviet army
Total	99	99	88		

The conclusion is that Bulgarians are rather our blood brothers (a little black E1b1b); as Uralian Bulgars have 0.5N and as Slavs have 17R1a as we do. A similarity / compatibility of 88% was found between the Romanians 'and the Bulgarians' "genetic mix".

The genetic composition in Hungarian is: I (all) – 26.5%, R1a – 29.5%, R1b – 18.5%, G – 3.5%, J1 + J2 – 9.5%, E1b1b – 8%, Q – 0%, N –0.5%, very similar composition as ours. The Hungarians have virtually no upsurge of honeysuckle and very small Ural (0.5N), just like us. The Finno-Ugric populations with which they are related have, for N: 61.5% – Finland, 34% – Estonia, 17% Bashkiria.

Comparing Romania with other countries in the Balkan region similarly follows the following regional compatibility: Macedonia – 88.5%, Bulgaria – 88%, Northern Greece, (Macedonia + Thracia) – 86%, Serbia – 82%, Gagauzia – 80%, Croatia – 78.5%, Moldavia – 78%, Montenegro – 77%, Bosnian Serbians – 76.5%. Other results that we are interested in: Hungary – 72%, Central Greece – 71%, Albania – 68.5%, Russia – 44.5%, Ukraine – 59.5%.

Serbia's Compatibility: Bosnian Serbs – 85%, Croatia – 83.5%, Montenegro – 83%, Romania – 82%, Bulgaria – 79%, Macedonia – 78.5%, Northern Greece – 75.5%.

Macedonia's Compatibility: Bulgaria – 90%, Romania – 88.5%, Northern Greece (Macedonia + Thracia) – 88%, Bosnian Serbs – 80%, Montenegro – 79%, Serbia – 78.5%, Central Greece and Albania – 76%.

Bulgaria's Compatibility: Macedonia – 90%, Northern Greece – 89%, Romania – 88%, Gagauzia – 79.5%, Montenegro and Serbia – 79%, Central Greece and Bosnian Serbians – 78.5%, Moldovia – 76.5%.

The genetic rapprochement of the Bulgarians and the Gagauz Romanians shows that the latter are not Turks but Masons from the Turks; some of them went to Islam and were brought to Islam in the 19th century in Bugeac by the Tsarist Empire.

Compatibility of Northern Greece (Macedonia + Thracia): Bulgaria – 89%, Macedonia – 88%, Romania – 86%, Central Greece – 80.5%, South Greece – 79%, Gagauzia – 77.5%, Albania – 77%, Aegean Islands – 76.5%.

Northern Greece (Macedonia + Thracia) forms a compact block with Macedonia, Bulgaria and Romania, over 86%, followed by Central and Southern Greece.

The fact that after this implant the composition of Northern Greece is closer to Bulgaria, Macedonia and Romania than to the rest of Greece leads us to the conclusion that "the Anatolian Greeks" were in fact descendants of the Thracians who formed the Anatolian kingdoms in antiquity. The idea that the Macedonians would be Latinized Greeks is not good.

The National Geographic References concerning the origins of the populations show us the genetic traces for Romania, Bulgaria and Greece:

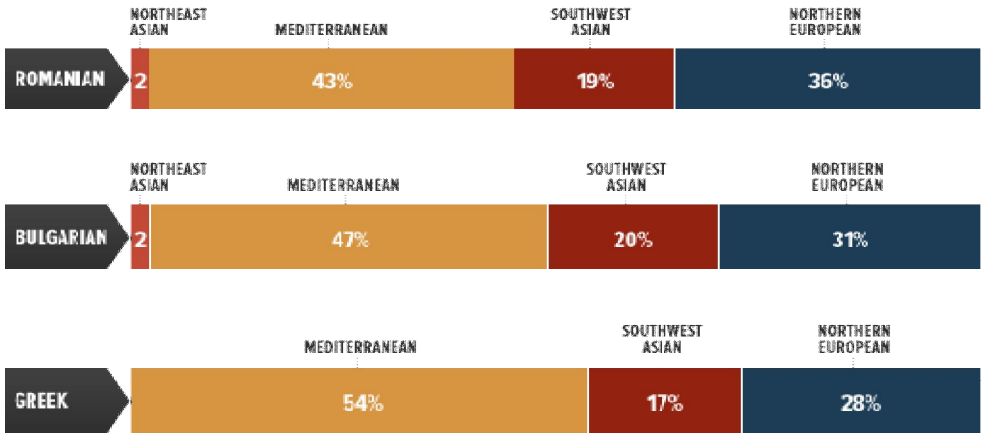


Figure 9. National Geographic References.

This analysis is based on samples collected from people from Romania, Bulgaria and Greece. The percentages of the Mediterranean and Southeast Asia reflect the strong influence of Middle Eastern farmers who arrived in the area more than 7.000 years ago. The North European component probably comes from the pre-agricultural population of Europe – the oldest colony that arrived more than 35.000 years ago during the Upper Paleolithic. The 2% Northeast Asia component shows that there has been a mix with groups in the East and is typical of Eastern European populations, such as Romanians, Bulgarians, Russians and Nordic Caucasians (https://genographic.nationalgeographic.com/reference-populations/).

Interdisciplinary research – Romanians between Europe and Asia

The study made by a group of Romanian researchers and published in an international scientific journal reveals that, from the point of view of mitochondrial DNA, "Transylvania is closer to the populations of Central Europe than the population of the other Romanian provinces, which is more close to the Balkans".

"At a crossroads of important roads between Asia and Europe, Romania experienced episodes of continuous migration and invasion. The precise routes could have been modeled by the topology of the territory and had various effects on the genetic structure of mitochondrial DNA (mtDNA) in the Romanian historical provinces", according to a study published on March 7 by several researchers Romanians in BMC Genetics- https://bmcgenet.biomedcentral.com/articles/10.1186/s12863-017-0487-5

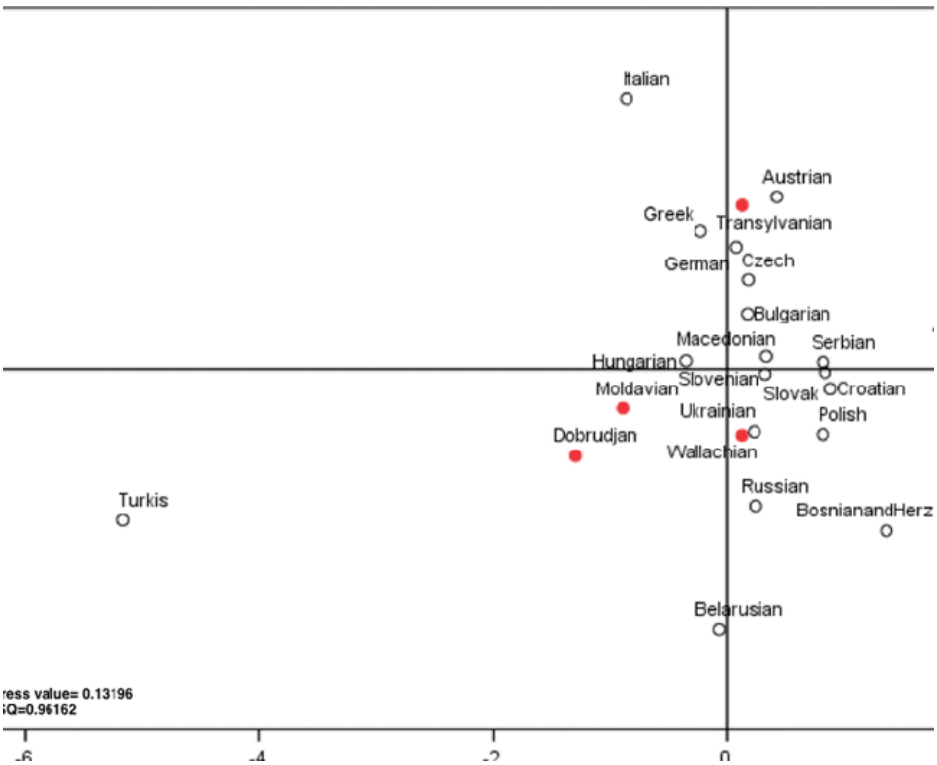


Figure 10. Analyzing the populations' proximity of the four Romanian provinces of 19 neighboring populations, from the point of view of mitochondrial DNA.

Researchers classified the vast majority of individuals from the four Romanian populations into nine Eurasian mitochondrial haplogroups (*H*, *U, K, T, J, HV, V, W,* and *X*). The Romanian populations were also exposed to Asian haplogruppes (*A, C, D, I, M* and *N*) and to an African haplogroup – L. The *A, C, D* and I Asian haplogroups were degraded in the Romanian sample with a global frequency of 2.24%, in line with the frequency in other European populations. In Transylvania, the *M* and *N* Asian haloes had a low frequency, unlike Walachia, Dobrudja and Moldavia. Frequency of haplogroup H (40.98%) in Romanians – the most common haplogroup in Europe – was in line with the frequency observed in most European populations. Halogroup H ranged from a relatively high frequency in Transylvania to a lower frequency in Dobrudja. Also, the U5 subhaplogue, the oldest in Europe, had an increased frequency in all four provinces.

In the article there is a brief history of the evolution of the population of Romania from the Paleolithic to the Union of 1918. "As a segment of

the Danube Basin, this territory was a crossroads of major roads between Asia and South-Eastern Europe, North, and one of the direct eastbound routes connecting the North Pontic steppe (...) The population circulation during the Neolithic and Bronze Age epitomized the genetic variation of the Romanian population today, especially during the Middle Neolithic. During the Iron Age, on the territory defined by the Danube Basin and the Carpathian Arch, the archaeological discoveries and historical sources revealed the presence of Indo-European populations close to the Thracians who arrived probably in the second millennium BC, in Transylvania, bearing the name the Dacians and the Getians in Walachia, Dobrudja and Bessarabia", according to the study. Another important step was the partial conquest of Dacia by the Romans (the 1st and 2nd centuries after Christ), with the exception of the present regions of Moldova, northeastern Transylvania and eastern Muntenia. A period of colonization followed by different groups in the Roman Empire (Italians, Illyrians, Thracians, Greeks, Celts, Germans and East or North Africans) that settled mainly in Transylvania.

"Beginning with the 6th century, the Slavs penetrated massively on the present territory of Romania and were eventually assimilated by the Dacian-Romanian, Proto-Romanian and Romanian populations until the end of the 12th century. Romania's current territory was divided into the Middle Age (around the 14th century AD) in three different political structures: Walachia, Moldavia and Transylvania. Dobrudja was, for most of its medieval history, under Ottoman domination. Over the next few centuries, the historical provinces of Romania have been continuously under Byzantine, Austro-Hungarian, Habsburg, Ottoman and Russian suzerainty. However, these empires did not have a major demographic impact, with the exception of Transylvania, which underwent massive colonization with the Szeklers, Saxons and Hungarians under the rule of the Kingdom of Hungary between the 10th and 13th centuries."

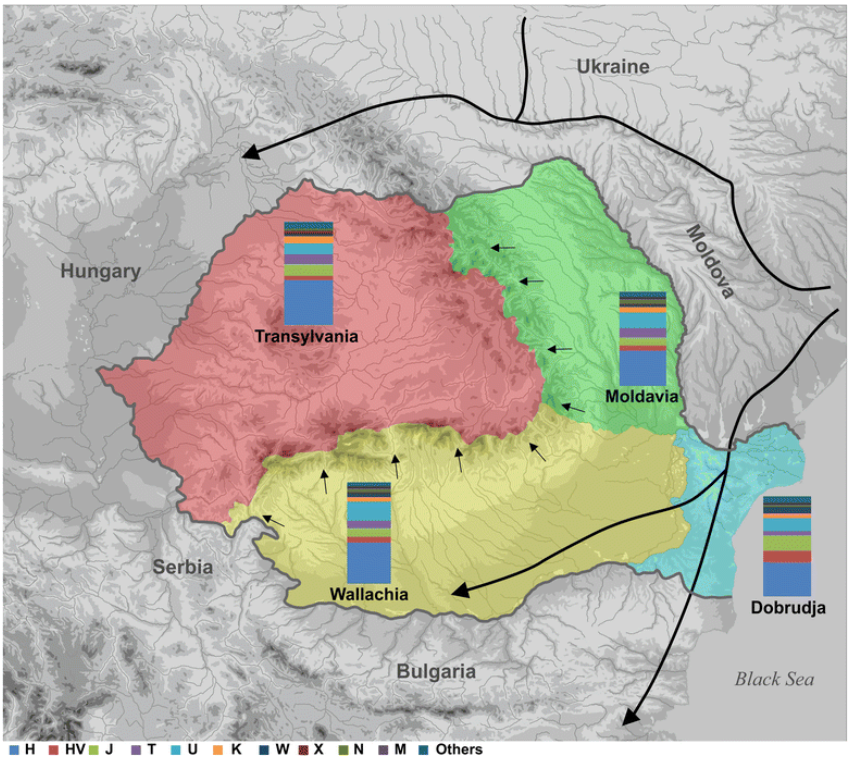


Figure 11. The combinations of haplogroups present in the Romanians in the four historical provinces. Blue is the haplogroup H, the most common haplogrup of European populations.

Final conclusions

In a population of infinite magnitude where reproduction is of the panmyxic type (random crossing between individuals) and where there are no new migrations or mutations, the allelic frequencies, and therefore those of the corresponding genotypes, remain constant from one generation to the next (Hardy-Weinberg's law). Alleles occur due to spontaneous mutations that are inherent replication errors of DNA molecules over generations. Natural populations do not exactly meet all the conditions imposed by the Hardy-Weinberg law, so the law is used to calculate, with an acceptable error, allelic frequencies from the phenotypes observed in the population. The number and frequency of alleles of a locus, which define the genetic variability of the respective population, are variable from one population to another and, over time, within the same population, depending on the evolutionary history of that population.

In human genetics, the most commonly studied haplogroups are Y-chromosomal haplogroups (Y-DNA) and mitochondrial DNA haplogroups. Y-DNA is transmitted only along the patrilineal line from father to son while DNA mtDNA is transferred on the matrilineal line from the mother to the descendants of both sexes. Y-DNA and mtDNA DNA are not recombined; they change only by random mutation in each generation, without mixing between the genetic materials of the parents. It is usually assumed that there are few natural selections for or against a certain haplotipic mutation that has survived to this day. The prevalence of a particular marker in a population fluctuates further until it reaches either 100% or falls entirely out of the population. In a large population with efficient mixing, the rate of genetic diversion for the common alleles is very low. However, in a very small cross-section, the proportions may change much faster. Geographic variations and marked concentrations of haplotypes and particular haplotype groups therefore mark the distinctive effects of repeated population blockages or founding events followed by population segregations and increases.

If we take into account the known data on the occurrence of Homo Sapiens in Europe, about 40 to 45.000 years ago, and their arrival on the Africa – Levant – Anatolia – Balkans route from where the spread to Europe, we could say that the haplogrup I descend from these primes the first division of the IJ trunk, taking place 38.000 years ago in the Balkans or Anatolia or both. Taking into account that at that time we were at the peak of glaciation, the Balkan Peninsula has functioned as a shelter and fireplace for a long time for these populations, enough to show the mutations that led to the emergence of the haplogroup I followed by its divisions. The fact that Homo Sapiens had been found, dating from 40-45.000 years ago in Kents Cavern, England, or 43-45.000 years ago in Grotta del Cavallo, Italy, shows that Homo Sapiens migration continued to the west more likely on a small scale due to extreme climatic conditions, the vast majority of populations remaining sheltered in the Danubian-Balkan area. The disappearance of the Neanderthal population is still unknown, the cause being Homo Sapiens or the extreme climatic conditions from 39 to 41.000 (peak of glaciation). It is certain that Homo Sapiens have coexisted and have even crossed the Neanderthal for at least 5000 years.

As the glaciation ended, about 13.000 years ago, the expansion of the haplogrup I and its divisions into Western, Northern, Central and

Mediterranean Europe began. From now on, we can speak of a Danubian population that has been authoritatively controlling a huge territory for many thousands of years. Throughout this period they have assimilated, in successive stages, the populations that brought agriculture to Europe and the cattle breeders who arrived from Anatolia. At the same time, they have liquidated the populations they considered a threat, and the evidence is again "widowed" mitochondrial haplogroups to which Y haplogroups cannot be associated.

Comparing the "genetic mix" of Romanians and Bulgarians, there is a similarity / compatibility of 88%, which shows us that Bulgarians are our blood brothers. The analysis of the data provided by eupedia.com shows that the population of Romania and its neighbors, after so many invasions and destructions, today includes descendants of the population over 6.000-8.000 years of about 90% on paternal lines and about 85% on maternal line. The study by a group of Romanian researchers reveals that, from the point of view of mitochondrial DNA, "Transylvania is closer to the populations of Central Europe than to the population of the other Romanian provinces, which is closer to the Balkans"[3].

REFERENCES

[1] https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle
[2] http://www.nature.com/scitable/knowledge/library/the-hardy-weinberg-principle-13235724
[3] http://www.nature.com/scitable/definition/hardy-weinberg-equilibrium-122
[4] https://www.khanacademy.org/science/biology/her/heredity-and-genetics/v/hardy-weinberg
[5] http://anthro.palomar.edu/synthetic/synth_2.htm
[6] https://www.thoughtco.com/what-is-the-hardy-weinberg-principle-1224766
[7] https://www.youtube.com/watch?v=oEBNom3K9cQ
[8] https://www.youtube.com/watch?v=L7rpgdOsq_s
[9] https://www.youtube.com/watch?v=oG7ob-MtO8c
[10] http://wallace.genetics.uga.edu/groups/evol3000/wiki/adcc5/HardyWeinberg_Principle.html
[11] http://www.biology-pages.info/H/Hardy_Weinberg.html
[12] https://www.ndsu.edu/pubweb/~mcclean/plsc431/popgen/popgen3.htm
[13] http://education.seattlepi.com/five-parts-hardyweinberg-equilibrium-3968.html
[14] http://www.sparknotes.com/biology/evolution/populationgenetics/section1.rhtml
[15] http://anatomie.romedic.ro/genetica
[16] http://www.eupedia.com/europe/maps_Y-DNA_haplogroups.shtml
[17] http://www.eupedia.com/europe/Haplogroup_R1a_Y-DNA.shtml

[3] http://adevarul.ro/locale/cluj-napoca/adn-ul-romanilor-transilvania-mai-apropiat-austriecilor-decat-muntenilor-moldovenilor-1_58c097625ab6550cb82627cf/index.html

[18] http://www.eupedia.com/europe/Haplogroup_R1b_Y-DNA.shtml
[19] https://genographic.nationalgeographic.com/reference-populations/
[20] http://adevarul.ro/locale/cluj-napoca/adn-ul-romanilor-transilvania-mai-apropiat-austriecilor-decat-muntenilor-moldovenilor-1_58c097625ab6550cb82627cf/index.html
[21] http://gazetadecluj.ro/cercetator-implicat-studiul-adn-ului-romanilor-transilvaneni-nu-sunt-diferiti-de-romanii-din-celelalte-provincii/
[22] https://thraxusares.wordpress.com/2015/02/05/agricultorii-din-trecutul-romanilor/
[23] http://politeia.org.ro/magazin-istoric/balcanica-1-ilirii-o-etnogeneza-balcanica/13474/
[24] http://politeia.org.ro/magazin-istoric/balcanica-3-tracii-o-etnogeneza-balcanica/13925/
[25] http://www.thefullwiki.org/Literary_sources_for_the_origin_of_the_Romanians