Developing an Online Portal for Determining the Genomic Signature of Archaic DNA that are Associated to Modern Human Genetic Diseases: A Meta-Analysis Study

Niyazi Senturk[1], Mahmut Cerkez Ergoren[2,3]

ABSTRACT

Objective: Mutations or introgression can cause and rise adaptive alleles of which some can be beneficial. Archaic humans lived more than 200,000 years ago in Europe and Western Asia. They were adapted to the environment and pathogens that prevailed in these locations. It can therefore be thought that modern humans obtained significant immune advantage from the archaic alleles.

Materials and Methods: First, data were collected by meta-analysis from previously identified genetic diseases caused by alleles that were introgressed from archaics. Second, the in silico model portal (http://www.archaics2phenotype.xxx.edu.tr) was designed to trace the history of the Neanderthal allele. The portal also shows the current distribution of the genotypes of the selected alleles within different populations and correlates with the individuals phenotype.

Results: Our developed model provides a better understanding for the origin of genetic diseases or traits that are associated with the Neanderthal genome.

Conclusion: The developed medicine model will help individuals and their populations to receive the best treatment. It also clarifies why there are differences in disease phenotypes in modern humans.

Keywords: Archaic DNA, single nucleotide polymorphism, toll-like receptor

Cite this article as: *Senturk N, Ergoren M.C. Developing an Online Portal for Determining the Genomic Signature of Archaic DNA That Are Associated to Modern Human Genetic Diseases: A Meta-Analysis Study. Eurasian J Med 2020; 52(2): 153-60.*

[1]Department of Biomedical Engineering, Near East University Faculty of Engineering, Nicosia, Cyprus
[2]Department of Medical Biology, Near East University School of Medicine, Nicosia, Cyprus
[3]DESAM Institute, Near East University, Nicosia, Cyprus

Received: January 9, 2019
Accepted: August 22, 2019
Available Online Date: June 2, 2020

Correspondence to: Mahmut Cerkez Ergoren
E-mail: mahmutcerkez.ergoren@neu.edu.tr

DOI 10.5152/eurasianjmed.2019.18424

Introduction

Archaic humans lived more than 200.000 years ago in Europe and Western Asia [1]. They were well adapted to the surrounding environment and pathogens [2]. Archaic humans are the subspecies of *Homo sapiens*, and include *Homo heidelbergensis*, *Homo rhodesiensis*, *Homo neanderthalensis*, and *Homo antecessor*. There are anatomical differences between archaic and modern humans. Modern humans have evolved from archaic humans and *Homo erectus*. When modern humans migrated from Africa, they were faced with some difficulties such as different climate, environmental challenges, and pathogens in the new region [1]. In the regions where they migrated from Africa, they hybridized with Neanderthals and Denisovans. Thus, some alleles passed from Neanderthals to modern humans.

Neanderthals evolved 250.000 years ago and were known as *H. neanderthalensis* [3]. Neanderthals were geographically spread from England to Siberia. They were powerful hunters. *H. sapiens* began to spread in the world from Africa around 30,000 years ago [4]. Therefore, Neanderthals and early humans coexisted and mated. Modern genetic data show that Neanderthals mated with modern humans in Europe when they coexisted. As a result, almost 1-4% of the modern humans' genome consists of genes from Neanderthals. The genes that were inherited from Neanderthals help us fight deadly viruses such as Epstein-Barr. However, gene mutations have also resulted in diseases such as Crohn's disease, type-2 diabetes, lupus, heart diseases, and depression [1].

This study focused on the genomes that passed from Neanderthals to modern humans.

As a significant proportion of these, archaic-specific DNAs are found within the *TLR1-TLR6-TLR10* gene cluster that belong to toll-like receptors (TLRs). TLRs recognize the structure of

pathogens and provide natural immunity against many pathogens. Therefore, they are an important defense against pathogens. TLRs are known to respond to stimuli associated with various pathogens and to provide signal responses necessary for the activation of innate immune effector mechanisms and subsequent development of adaptive immunity [1].

A previous study indicated that modern humans carry three archaic-like haplotypes, and three TLRs that were inherited from archaic humans were identified. Two of these haplotypes resemble those of the Neanderthal genome, and the third haplotype resembles that of the Denisovan genome. The frequency of single nucleotide polymorphisms (SNPs) commonly shared in Neanderthal-like haplotypes varies among continents and populations. In Europe, allelic frequencies of Neanderthal-like core haplotypes are higher in Southern European populations [1].

First, we aimed to collect previously identified archaic-like SNPs that have clinical significance by meta-analysis. Then, we combined scientific knowledge and outcome from previous studies to determine diseases in modern humans, which were received genetically from Neanderthals. Second, a software program was developed to merge previously identified archaic-like SNPs and their clinical pathogenicity. Thus, this study and the developed software give us data regarding the origin of diseases in modern humans. Finally, an *in silico* model was designed for clinicians and researchers to trace the history of the archaic alleles and determine the possible correlation with the persons' phenotype, thus providing a better understanding to interpret the underlying mechanisms of the diseases.

Materials and methods

Recent data by Dannemann et al. [1] were used to determine the archaic-like SNPs that are represented within the human genome. The research group previously identified 79 archaic-like alleles within the *TLR6-TLR1-TLR10* gene cluster, which indicates repeated introgression from archaic humans. Meta-analysis was performed to find out the possible clinical significance of those genetic markers from 1000 genome populations.

Neanderthal introgression maps of Sankararaman et al. [5] and Vernot et al. [6] were used for the identification of archaic-like haplotypes that are potentially observed in modern human genomes. The introgression map presented by Sankararaman et al. [5] provides the possibility of the emergence of SNPs on the polymorphic positions of Neanderthals in modern humans [5]. Vernot et al. [6] detected introgressed regions of modern human reference sequence and compared these candidate regions with reference from Neanderthal genome [6]. They used introgression possibilities per SNP for all Asian and European individuals. They also calculated the difference between Neandertal probabilities from the distance between neighboring SNP pairs, including three *TLR* genes and an additional region of 50 kb (Chromosome 4:38.723.860-38.908.438) [1]. Potentially archaic-like SNPs in this region were identified in 109 Yoruba individuals in the genome dataset of Neanderthal or Denisovan genomes. Furthermore, Deamann et al. [6] agreed that this introgressed region covers chromosome 4 of 143 kb (Chromosome 4:38.760.338-38.905.731) and contains 61 archaic-like SNPs. This region overlaps with two haplotypes identified by Vernot et al. [6].

Microsoft visual studio C++ 2008 edition served as the integrated development environment and the C programming language was used to build software.

The software was generated in two parts. In the first part, software was created to allow the user to search information via the created database in two different ways. The user could conduct the search using SNP ID and chromosome location. The database was created in the second part.

The *in silico* genome browser was designed for the first time to show the data collected so far of all identified archaic-like SNPs and their clinical significance. Therefore, a program was created to generate a database that comprised all the data collected for 79 archaic-like SNPs. The SNP variation of ancestral nucleotides, the diseases caused by the SNPs, and allele frequencies and genotype frequencies according to 1000 genome populations were added to the program, which was created separately for each SNP ID.

The website has four main sections: *Homepage, About us, User guide,* and Contact.

In the *Homepage* section, a search can be conducted using SNP IDs or chromosome locations. If the given ID or location matches any on the database, the result will be visible on the screen. In the *About us* section, the user can get general information about the website. The *User guide* section is designed to guide the users on the use of the website. The *Contact section* is designed to allow the user to communicate with the website administrator regarding any queries they may have about the website.

Results

Before we conducted our study, all significant information about archaic SNPs was scattered at different places and various genome browsers. Therefore, we aimed to merge all information as the first step. Our merged meta-analysis data provided a better understanding of the mechanism and background of diseases.

Second, the *in silico* genome browser was created and transferred to the online platform. This generated genome browser provides online access to researchers and clinicians. After separately creating the domain name and hosting service, they were merged to create the publically free website http://archaics2phenotype.xxx.edu.tr/.

This website was generated for researchers and clinicians. The created database will facilitate the work of researchers because they can obtain all data with references via our browser. Our developed *in silico* model provides better understanding of the origin of genetic diseases or traits associated with archaic genomes. Moreover, it provides quick access to data for researchers and clinicians through genome browser.

A meta-analysis, which combines the results of multiple independent studies in a given subject, was performed to collect all the identified archaic-like SNPs. We used three international genome browsers and scientific articles for meta-analysis. We determined 79 archaic-like SNPs from the study of Dannemann et al. [1]. Then, 1000 genomes were used to check for SNP registration and identification in the 1000 genome populations.

The clinical significance of the identified SNPs was determined using the genetic browsers. In this study, three different international databases were used to collect data: Ensembl genome, 1000 genome, and dbSNP. Additionally, population genetic information was collected from 1000 genome data by Ensemble. Thus, for each population, allele frequencies and genotype frequencies were obtained for each determined SNP.

Main Points

- Our developed *in silico* model provides better understanding of the origin of genetic diseases or traits associated with archaic genomes.
- It provides quick access to data for researchers and clinicians through genome browser.
- The developed software was designed to help individuals and their belong populations to receive the best treatment in the future.

Allele frequency is the frequency of occurrence of a specific allele in a population. For example, if A is dominant allele and T is recessive allele, we have three different possibilities for allele combination. These would be AA, AT, and TT. Genotype frequency will be how often we see each allele combination in the population. But, allele frequency is how often we see each allele (A or T) in the population. Thus, allele frequency is the number of A alleles divided by the total number of alleles (A+T) or the number of T alleles divided by the total number of alleles in the population.

Minor allele frequency (MAF) is the less common allele frequency in the populations for each identified SNPs. MAF was selected 0.005 or more at the HapMap project. But, it was selected less than 0.005 at 1000 genome Project (http://www.internationalgenome.org/). Thus, researchers aimed to investigate low and rare variants for different populations.

We used the genetic information of 31 populations. The SNPs of these populations are registered in the three genome browsers mentioned earlier. For each SNP, allele and genotype frequencies, MAF, ancestral SNP information, chromosomal location, and importantly clinical significance was added.

Five major populations, namely African, American, Eastern Asia, European, and Southern Asia, and 26 subpopulations were used in this study. The African subpopulations studied were Yoruba in Ibadan, African Caribbean in Barbados, Mende in Sierra Leone, African Ancestry in Southwest US, Gambian in western division, Esan in Nigeria, and Luhya in Webuye. The American subpopulations studied were Colombian in Medellin, Peruvian in Lima, Mexican ancestry in Los Angeles, and Puerto Rican in Puerto Rico. The East Asian subpopulations studied were Chinese Dai in Xishuangbanna in Ho Chi Minh City, Han Chinese in Beijing, Japanese in Tokyo, and Southern Han Chinese . The European subpopulations studied were Toscani in Italy, Finnish in Finland, Utah residents with Northern and Western European ancestry, Iberian populations in Spain, and British in England and Scotland. Finally, the South Asian subpopulations studied were Gujarati Indian in Houston, Bengali in Bangladesh, Indian Telugu in the UK, Sri Lankan Tamil in the UK, and Punjabi in Lahore. All these population genetics data were merged with allele and genotype frequencies for each determined archaic-like SNPs (Table 1).

Mating between archaic humans and modern humans resulted in the adaptation to their envi-

Table 1. Shows a detailed information for the SNP rs5743557. Detailed report was created for each SNP. MAF, chromosomal location, clinical significance, allele frequencies for each population including sub-populations were reported

Alleles G/A [2] Ancestral:G [2] Ambiguity code: R[2] MAF: 0.18 [2]

Location: 4:38804475 [2]
SNP ID: rs5743557 [1]
GWAS Trait: *Helicobacter pylori* serologic status, Self-reported allergy [1]
Archaic-like haplotype cluster: The GWAS-identified archaic like SNP, seen in the archaic-like haplotypes III, IV, VII and differs from the other modern-human core haplotypes II, V, VI, VIII, IX, And The SNP differs between core haplotype III and other modern human core haplotypes II, V, VI, VIII, IX. [1]
Clinical Assertion Description: Not Known

Population	Allele Frequency [2]		Genotype Frequency [2]		
ALL	G: 0.821	A: 0.179	G|G: 0.698	A|A: 0.056	A|G: 0.246
AFRICAN	G: 0.988	A: 0.012	G|G: 0.976	A|A: 0.000	A|G: 0.024
ACB	G: 0.974	A: 0.026	G|G: 0.948	A|A: 0.000	A|G: 0.052
ASW	G: 0.926	A: 0.074	G|G: 0.852	A|A: 0.000	A|G: 0.148
ESN	G: 1.000	A: 0.000	G|G: 1.000	A|A: 0.000	A|G: 0.000
GWD	G: 0.991	A: 0.009	G|G: 0.982	A|A: 0.000	A|G: 0.018
LWK	G: 1.000	A: 0.000	G|G: 1.000	A|A: 0.000	A|G: 0.000
MSL	G: 1.000	A: 0.000	G|G: 1.000	A|A: 0.000	A|G: 0.000
YRI	G: 1.000	A: 0.000	G|G: 1.000	A|A: 0.000	A|G: 0.000
AMERICAN	G: 0.852	A: 0.148	G|G: 0.741	A|A: 0.037	A|G: 0.222
CLM	G: 0.840	A: 0.160	G|G: 0.713	A|A: 0.032	A|G: 0.255
MXL	G: 0.891	A: 0.109	G|G: 0.797	A|A: 0.016	A|G: 0.188
PEL	G: 0.941	A: 0.059	G|G: 0.882	A|A: 0.000	A|G: 0.118
PUR	G: 0.764	A: 0.236	G|G: 0.615	A|A: 0.087	A|G: 0.298
EAST ASIAN	G: 0.607	A: 0.393	G|G: 0.385	A|A: 0.171	A|G: 0.444
CDX	G: 0.796	A: 0.204	G|G: 0.634	A|A: 0.043	A|G: 0.323
CHB	G: 0.466	A: 0.534	G|G: 0.223	A|A: 0.291	A|G: 0.485
CHS	G: 0.600	A: 0.400	G|G: 0.352	A|A: 0.152	A|G: 0.495
JPT	G: 0.442	A: 0.558	G|G: 0.183	A|A: 0.298	A|G: 0.519
KHV	G: 0.758	A: 0.242	G|G: 0.566	A|A: 0.051	A|G: 0.384
EUROPEAN	G: 0.779	A: 0.221	G|G: 0.622	A|A: 0.064	A|G: 0.314
CEU	G: 0.813	A: 0.187	G|G: 0.667	A|A: 0.040	A|G: 0.293
FIN	G: 0.904	A: 0.096	G|G: 0.818	A|A: 0.010	A|G: 0.172
GBR	G: 0.830	A: 0.170	G|G: 0.681	A|A: 0.022	A|G: 0.297
IBS	G: 0.692	A: 0.308	G|G: 0.495	A|A: 0.112	A|G: 0.393
TSI	G: 0.678	A: 0.322	G|G: 0.477	A|A: 0.121	A|G: 0.402
SOUTH ASIAN	G: 0.837	A: 0.163	G|G: 0.693	A|A: 0.018	A|G: 0.288
BEB	G: 0.779	A: 0.221	G|G: 0.581	A|A: 0.023	A|G: 0.395
GIH	G: 0.811	A: 0.189	G|G: 0.660	A|A: 0.039	A|G: 0.301
ITU	G: 0.882	A: 0.118	G|G: 0.775	A|A: 0.010	A|G: 0.216
PJL	G: 0.865	A: 0.135	G|G: 0.740	A|A: 0.010	A|G: 0.250
STU	G: 0.843	A: 0.157	G|G: 0.696	A|A: 0.010	A|G: 0.294

(SNP: Single Nucleotide Polymorphism; GWAS: Genome-wide association studies; MAF: Minor allele freaquency; ACB: African Caribbeans in Barbados; ASW: Americans of African Ancestry in SW USA; ESN: Esan in Nigeria; GWD: Gambian in Western Divisions in the Gambia; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria; CLM: Colombians from Medellin, Colombia; MXL: Mexican Ancestry from Los Angeles USA; PEL: Peruvians from Lima, Peru; PUR: Puerto Ricans from Puerto Rico; CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese; JPT: Han Chinese in Beijing, China; KHV: Kinh in Ho Chi Minh City, Vietnam; CEU: Utah Residents (CEPH) with Northern and Western European Ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; TSI: Toscani in Italia; IBS: Iberian Population in Spain; BEB: Bengali from Bangladesh; STU: Sri Lankan Tamil from the UK; ITU: Indian Telugu from the UK; PJL: Punjabi from Lahore, Pakistan; GIH: Gujarati Indian from Houston, Texas).
Dannemann M, Andres AM, Kelso J. Introgression of Neandertal-and Denisovan-like Haplotypes contributes to adaptive variation in human toll-like receptors. Am J Hum Genet 2016; 98: 22-33.
Ensembl. Browse a Genome. 2016. Avalailable From: URL: http://www.ensembl.org/index.html

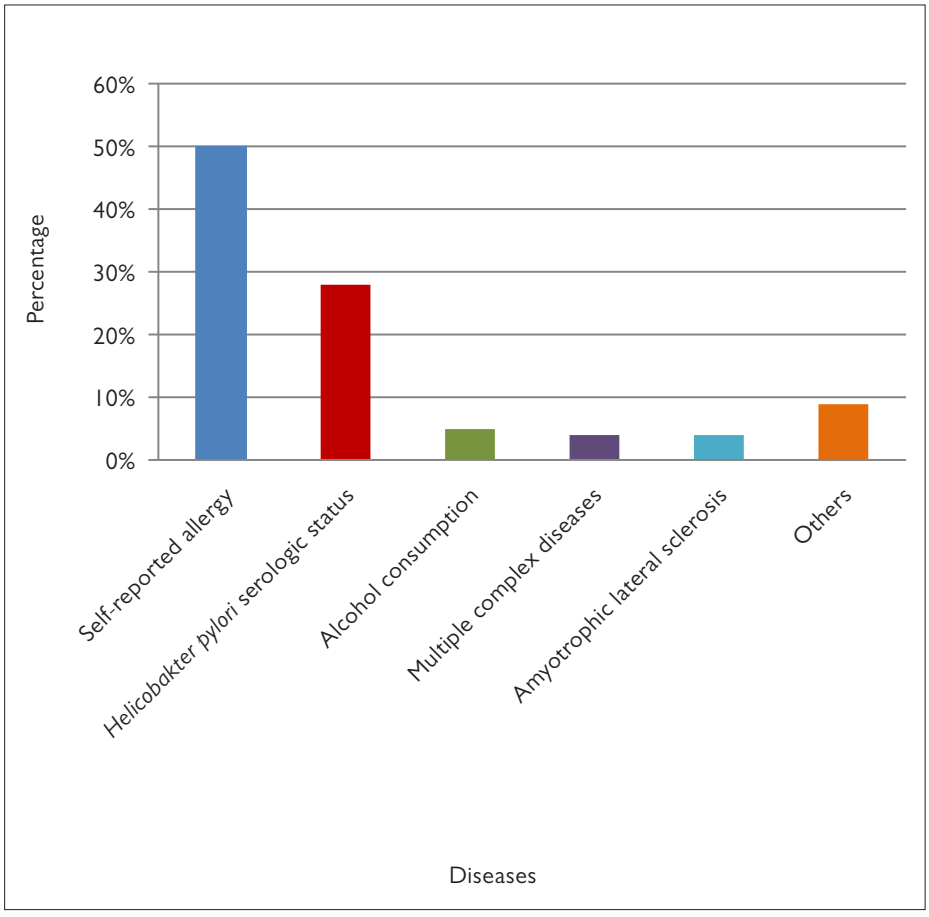

Figure 1. Illustration of the statistical calculation of the most common diseases or traits that might have been caused by archaic-like SNPs. The horizontal axis represents the most common diseases or traits and the vertical axis illustrates the frequency of the disease. Self-reported allergy is the most seen disease followed by *Helicobacter pylori* infection . Interestingly, alcohol consumption and amyotrophic lateral sclerosis had an association with archaic-like SNPs (5% and 4%, respectively). Other traits that were found <1% are endometriosis, blood pressure, coronary artery disease, abnormal lymphocyte counts, Paget's disease, height, allergic sensitization, breast cancer, and suicide attempts in bipolar and panic disorders.

ronment and local pathogens. Admixture of both hominins caused the introgression of common alleles; thus, modern humans gained adaptive immunity from archaic ones and survived as a result of natural selection. Archaic-like alleles regulate the gene expression of *TLR* genes [7]. Additionally, these alleles are also associated with resistance to several microorganisms and allergic diseases. Different alleles together with variety of gene expressions cause different disease phenotypes in modern humans. These archaic-like SNPs are responsible for some disease and disease susceptibility in the human genome (Figure 1). Our meta-analysis report lists each archaic-like SNP and its association with pathogenic diseases (Table 2).

During the designing of the database, the program codes were written in C language in Visual Studio. The software of this study was basically divided into two parts.

In the first part, users could enter three different input options independently or simultaneously; the software was designed to allow users to search using SNP ID, or chromosome location of the interested SNP, or both. In the second part, the output of the searched input was displayed on the screen. In this part, the data acquired were used to create the *in silico* browser. After the creation of the necessary algorithms, all collected informative data about the 79 archaic-like SNPs were integrated with the new software. Thus, the archaics2phenotpe software was generated.

After creating the software and database, the website was generated and the database was posted on the website for online access. This website is an information sharing platform, which is available online to users.

Domain name and web hosting are required to create a website. First, the domain, archaics2phenotype.xxx.edu.tr, was created to setup the website for internet browsers. Second, the web hosting was created to activate the website. The users of the website could access all available data on the website through the web hosting. In addition, all data of the 79 archaic-like SNPs were stored in the hosting service. The database created using the software was transferred into the hosting service. Then, both the domain name and web hosting service were connected to each other and the website was activated eventually. As a final step, the interface of the website was designed. The appearance of the *in silico* genome browser is crucial for ease of use. The database was transferred to the website for online access and the data are at present freely available at http://archaics2phenotype.xxx.edu.tr/ to the public worldwide.

Discussion

Genetic and archeological studies showed that Neanderthals and modern humans interbred 50.000 years ago. The fossil findings revealed that the population of Neanderthals began to decline 40.000 years ago and the Neanderthal generations become extinct 39.000 years ago. There were many factors that contributed to their extinction and many hypotheses about their generation. First possibility is the rivalry for resources or direct warfare between Neanderthals and modern humans. Modern humans were more advanced technologically and were better hunters compared to the Neanderthals. Therefore, humans had better chances of survival. Second possibility is that the Neanderthals were adapted to cold climate. Their lives became difficult as the climate became warmer gradually. Another possibility could be the new pathogens and parasites found in the new environment [8].

Considerable genetic diversity occurs in humans by ancient polymorphisms. Thus, Neanderthal and modern haplotypes are not much diverged from modern human sequences. In Europe, the allelic frequencies of Neanderthal-like core haplotypes are higher in Southern European populations [1], for example, Tuscany in Italy and Iberian populations in Spain (TSI and IBS with frequencies of 39.3% and 38.3%, respectively). In Asia, Neanderthal-like allele frequency core haplotypes are higher in East Asian populations, such as Japanese in Tokyo (JPT, frequency 53.4%) and Han Chinese (CHB, frequency 53.6%). The frequencies of other Asian populations are between 21.7% and 41.9% [1].

In Neanderthal genome project, the genome was obtained from the bones found in the Vindija cave. The extracted Neanderthal DNAs were compared to those of five different modern humans (French, Chinese, Papua New

Table 2. Shows each listed archaic-like SNP and its associated disease. These archaic-like SNPs mainly cause self-reported allergies and *Helicobacter pylori* serologic status.

SNP ID	GWAS TRAIT
rs6841698	Self-reported allergy
rs10024216	Amyotrophic lateral sclerosis | *Helicobacter pylori* serologic status | Self-reported allergy
rs10008492	Endometriosis | *Helicobacter pylori* serologic status | Self-reported allergy
rs10470854	Self-reported allergy
rs4331786	Amyotrophic lateral sclerosis | Self-reported allergy | Blood Pressure
rs4513579	Self-reported allergy
rs10776482	*Helicobacter pylori* serologic status | Self-reported allergy
rs4129009	*Helicobacter pylori* serologic status | Self-reported allergy
rs10776483	*Helicobacter pylori* serologic status | Self-reported allergy
rs11466657	Self-reported allergy
rs11096955	Self-reported allergy
rs11096956	Helicobacter pylori serologic status | Self-reported allergy
rs11096957	Amyotrophic lateral sclerosis | Endometriosis | Self-reported allergy
rs4274855	*Helicobacter pylori* serologic status | Self-reported allergy
rs11466645	*Helicobacter pylori* serologic status | Self-reported allergy
rs11466640	*Helicobacter pylori* serologic status | Self-reported allergy
rs7694115	Self-reported allergy
rs11466617	*Helicobacter pylori* serologic status | Self-reported allergy
rs7653908	*Helicobacter pylori* serologic status | Self-reported allergy
rs7658893	v serologic status | Self-reported allergy
rs11725309	*Helicobacter pylori* serologic status | Self-reported allergy
rs10034903	Self-reported allergy
rs10004195	Helicobacter pylori serologic status | Self-reported allergy
rs12233670	Multiple complex diseases | Coronary Artery Disease | Lymphocyte counts | *Helicobacter pylori* serologic status | Self-reported allergy
rs6834581	Alcohol consumption
rs4833093	Alcohol consumption
rs6531663	Alcohol consumption
rs4543123	Amyotrophic lateral sclerosis | Paget`s disease | *Helicobacter pylori* serologic status | Alcohol consumption | Self-reported allergy
rs4624663	Panic disorder
rs4833095	Amyotrophic lateral sclerosis | Paget`s disease | *Helicobacter pylori* serologic status | Alcohol consumption | Self-reported allergy
rs5743604	*Helicobacter pylori* serologic status | Alcohol consumption | Self-reported allergy
rs5743596	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743595	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743594	Height
rs5743592	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743571	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743565	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743563	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743562	*Helicobacter pylori* serologic status | Self-reported allergy
rs5743557	*Helicobacter pylori* serologic status | Self-reported allergy
rs11722813	*Helicobacter pylori* serologic status
rs2101521	*Helicobacter pylori* serologic status | Self-reported allergy
rs17616434	*Helicobacter pylori* serologic status | Alcohol consumption | Self-reported allergy | Allergic sensitization

Table 2. Shows each listed archaic-like SNP and its associated disease. These archaic-like SNPs mainly cause self-reported allergies and *Helicobacter pylori* serologic status (Continue).

SNP ID	GWAS TRAIT
rs4833103	*Helicobacter pylori* serologic status | Self-reported allergy
rs6815814	*Helicobacter pylori* serologic status | Self-reported allergy
rs7696175	Breast cancer | Multiple complex diseases | Self-reported allergy
rs5743810	Self-reported allergy
rs1039559	Self-reported allergy
rs5743794	Suicide attempts in bipolar disorder | *Helicobacter pylori* serologic status | Self-reported allergy
rs5743788	Self-reported allergy
rs7665774	Self-reported allergy
rs7673348	*Helicobacter pylori* serologic status | Self-reported allergy
rs7687447	*Helicobacter pylori* serologic status | Self-reported allergy
rs6531672	Self-reported allergy
rs6531673	Self-reported allergy
rs7681628	Self-reported allergy
rs2174284	*Helicobacter pylori* serologic status | Self-reported allergy
rs3860069	*Helicobacter pylori* serologic status | Self-reported allergy
rs17582830	*Helicobacter pylori* serologic status | Self-reported allergy
rs2130296	Multiple complex diseases
rs721653	Self-reported allergy
rs902136	Self-reported allergy
rs11943027	Self-reported allergy
rs17582893	Suicide attempts in bipolar disorder | *Helicobacter pylori* serologic status | Self-reported allergy
rs2381345	Self-reported allergy
rs1873195	*Helicobacter pylori* serologic status | Self-reported allergy
rs17582921	*Helicobacter pylori* serologic status | Self-reported allergy
rs6851685	*Helicobacter pylori* serologic status | Self-reported allergy
rs6835514	Multiple complex diseases | Coronary Artery Disease | *Helicobacter pylori* serologic status | Self-reported allergy
rs1604834	Multiple complex diseases|Self-reported allergy
rs974734	Self-reported allergy
rs7688418	Self-reported allergy
rs7665932	Self-reported allergy
rs6531677	Self-reported allergy
rs12642243	Self-reported allergy
rs12641669	Self-reported allergy
rs1115259	Self-reported allergy
rs6824769	Multiple complex diseases | Self-reported allergy
rs7664107	Self-reported allergy

Guinea, and Africans from San and Yaruba groups) [2]. The results from the initial analyses showed that Neanderthal DNA was more similar to the non-African population's DNA than to the African one. The simplest explanation of this similarity was that there was a gene flow between Neanderthals and humans. There were significant differences between the modern humans and the Neanderthals in four genes: sperm-associated antigen 17 (*SPAG17*), which is responsible for sperm motility [8]; protocadherin-16 (*PCD16*), which is responsible for wound healing [9]; transcription termination factor *TTF1*, which is responsible for gene reading; and *RPTN* gene, which is highly expressed in hair follicles, skin, and sweat glands [10]. Apart from these, the mannose receptor C-type 1 (*MRC1*) gene, also found in Neanderthals and modern humans, played a role in cell communication. However, the Neanderthals carried a special mutation in the *MRC1* gene. This mutation did not appear in modern humans. It had led to the formation of a pale skin color and red hair in Neanderthals [11]. Another

difference was found in the forkhead box protein P2 (*FOXP2*) gene. In modern humans, the *FOXP2* gene does not have any effect. This gene is also called speech gene because speech disorders occur. Also, this gene was found in Neanderthals and chimpanzees [12]. Like these, there are differences in DNA levels among many genes. However, the results showed that 99.7% of the human and Neanderthal genomes are exactly the same while human and chimpanzee genomes showed 98.8% similarity.

The first encounter between *H. sapiens* and *H. neanderthalensis* was won by the Neanderthals. Approximately 100.000 years ago, *H. sapiens* migrated to the north and the east Mediterranean. These regions were the territory of Neanderthals and therefore modern humans could not settle there. This may be due to unfavorable climate, local parasites, and new diseases. Regardless of the cause, *H. sapiens* were driven out from these areas, and the Middle East remained in control of Neanderthals. About 70.000 years ago, the tribe came out of Africa for the second time. This time *H. sapiens* won and dominated the whole earth, not just the Middle East. They reached Europe and Eastern Asia within a short period of time. They passed through the open sea about 45.000 years ago and reached Australia, which was not reached by any other human-like species until that time [13]. The basis of these developments is the cognitive revolution that emerged 30.000-70.000 years ago. The cognitive revolution has added new thinking and new communication skills to *H. sapiens*. According to the most accepted theory of cognitive revolution, genetic mutations have altered the internal structure of the brain of *H. sapiens*. This change has allowed them to think in ways that have never been possible before and to communicate in new languages [13]. The reason for this mutation to occur in human DNA but not in Neanderthals is just a coincidence. According to this theory, the reason for the domination of *H. sapiens* in the world was caused only by a mutation that occurred in our genes by chance. Since cognitive revolution, *H. sapiens* have the ability to renew their behavior according to changing needs. This is the basis for *H. sapiens* to develop more than other *Homo* species and to dominate the world nowadays.

The biggest differences between Neanderthal and modern humans are strength and endurance [14]. Neanderthals were stronger and had more endurance than modern humans. The arms and thighs of modern humans are thinner than those of Neanderthals. It was important for Neanderthals to act quickly because they were hunter-gatherers [15]. The hands of modern humans are thought to have evolved for the delicate grip. The average height of Neanderthal men was 164-168 cm and that of women was 152-156 cm [15]. Introgressed Neanderthal sequences were identified in modern human autosomes and X chromosomes. Nevertheless, Neanderthal mitochondrial genome sequences were not reported within modern human genomes. In 2016, Mendez and colleagues stated that full mitochondrial DNA sequences were found in eight individuals. These individuals were from Spain, Germany, Croatia, and Russia. The Y chromosome was obtained from male Neanderthals in El Siron, Spain [16].

Mendez Conduzted a study [16] conducted a study of male Neanderthals who lived in Spain 49,000 years ago. The Y chromosome of these Neanderthals did not pass to modern humans. Europeans and Asians are missing chunks of Neanderthal DNA on their Y chromosomes. Thus, to conclude, female modern humans and male Neanderthals are not exactly compatible. Therefore, Mendez think that Neanderthals [16] think that Neanderthals may have problem in sperm production and they may have not produced many healthy male babies. As a result of this, the Neanderthal population might have declined rapidly [16].

As a result of the hybridization of early humans and Neanderthals between Europe and West Asia, non-African populations carry almost 1-4% Neanderthal DNA in their genome. Nevertheless, this Neanderthal DNA had both positive and negative effects on modern humans. Dannemann, et al. [17] agreed that introgressed genomes provide genetic adaptation to new environments. This could be a positive effect introgression of Neanderthal genome to humans. The introgression provides natural immunity to new environments and pathogens. Neanderthal alleles often are not adaptive to modern human genome [17].

Our generated database will facilitate the work of researchers, providing all data with references via this website, and the developed *in silico* model provides a better understanding of the origin of the genetic diseases that are introgressed from archaic genomes. Furthermore, the genome browser provides quick online access to data for researchers, clinicians, or anyone who is interested in the history of early human life.

This computer software will aid the evaluation of the percentage of Neanderthal-derived sequences in modern humans, thus facilitating the assessment of the genetic diseases that originally came from Neanderthals.

The limitations of this study was the lack of comparative genomic data in the literature and genome browsers; it needs to be developed with the use of more genetic data. However, our *in silico* model provides better understanding of the origin of genetic diseases or traits that are associated with archaic genomes. Therefore, by better understanding the human genome make up, this precise medicine model will help individuals and their populations to receive precise treatment.

Ethics Committee Approval: Authors declared that the research was conducted according to the principles of the World Medical Association Declaration of Helsinki "Ethical Principles for Medical Research Involving Human Subjects", (amended in October 2013).

Informed Consent: Due to the metaanalysis design of the study, informed consent was not taken.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept - M.C.E., N.S.; Design- M.C.E., N.S.; Supervision - M.C.E.; Resources - M.C.E., N.S.; Materials - M.C.E., N.S.; Data Collection and/or Processing - M.C.E., N.S.; Analysis and/or Interpretation - N.S.; Literature Search - M.C.E., N.S.; Writing Manuscript - M.C.E., N.S.; Critical Review - M.C.E.

Acknowledgements: We would like to thank Near East University Faculty of Engineering, Department of Biomedical Engineering member Assoc. Prof. Terin Adalı for her support and great help.

Conflict of Interest: Authors have no conflicts of interest to declare.

Financial Disclosure: The authors declared that this study has received no financial support.

References

1. Dannemann M, Andres AM, Kelso J. Introgression of Neandertal and Denisovan like haplotypes contributes to adaptive variation in human toll-like receptors. Am J Hum Genet 2016; 98: 22-33. [Crossref]
2. Green RE, Krause J, Briggs AW, et al. A draft sequence of the Neandertal genome. Science 2010; 328: 710-22. [Crossref]
3. Villa P, Roebroeks W. Neandertal demise: An archaeological analysis of the modern human superiority complex. PLos ONE 2014; 9: e96424. [Crossref]
4. Groucutt S, Petraglia MD, Bailey G, et al. Rethinkink the dispersal of homo sapiens out of Africa. Evol Anthropol 2015; 24: 149-64. [Crossref]
5. Sankararaman S, Mallick S, Dannemann M, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 2014; 507: 354-7. [Crossref]

6. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. Science 2014; 343: 1017-21. [Crossref]
7. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell 2006; 124: 783-801. [Crossref]
8. Zhang Z, Jones BH, Tang W, et al. Dissecting the axoneme interactome: the mammalian orthologue of Chlamydomonas PF6 interacts with sperm-associated antigen 6, the mammalian orthologue of Chlamydomonas PF16. Mol Cell Proteomics 2005; 4: 914-23. [Crossref]
9. Matsuyoshi N, Imamura S. Multiple cadherins are expressed in human fibroblasts. Biochem Biophys Res Commun 1997; 235: 355-8. [Crossref]
10. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes Dev 2009; 23: 1247-69. [Crossref]
11. Kundu S, Rahman S, Thakur S, Ancient DNA and Neanderthals mystery. Int J Sci Eng Res 2014; 5: 1189.
12. Enard W, Przeworski M, Fisher SE, et al. Molecular evolution of FOXP2, a gene involved in speech and language. Nature 2002; 418: 869-72. [Crossref]
13. Harari YN. Sapiens: A Brief History of Humankind, Kolektif Kitap, Istanbul; 2015.
14. Papagianni D, Morse M. The Neanderthals Rediscovered, Thames & Hudson, London; 2013.
15. Helmuth H. Body height, body mass and surface area of the Neanderthals. Z. Morphol Anthropol 1998; 82: 1-12.
16. Mendez FL, Poznik GD, Castellano S, et al. The Divergence of Neandertal and Modern Human Y Chromosomes. Am J Hum Genet 2016; 98: 728-34. [Crossref]
17. Dannemann M, Kelso J. The contribution of Neanderthals to phenotypic variation in modern humans. Am J Hum Genet 2017; 101: 578-89. [Crossref]