BJMG 24 (1), 2021 • 35-40
DOI: 10.2478/bjmg-2021-0012
sciendo

ORIGINAL ARTICLE

COMPARATIVE ANALYSIS OF GENES ASSOCIATED WITH OBESITY IN HUMANS USING BIOINFORMATIC DATA AND TOOLS

Musliji ZS[1], Pollozhani AK[1], Lisichkov K[2], Deligios M[3], Popovski ZT[2,4,*]

[*]**Corresponding Author:** Professor Zoran T. Popovski, Ph.D., Department of Biochemistry and Genetic Engineering, Faculty of Agriculture and Food Sciences, Bld "Aleksandar Makedonski," bb PB 297, 1000 Skopje, Republic of North Macedonia. Tel: +389-70-252-731. Fax: +389-2-3134-310. E-mail: zoran_ popovski@yahoo.com

ABSTRACT

Obesity has become a serious global problem that still needs a solution. One of the factors that leads to obesity is genetic predisposition. The identity and characteristics of the genes involved have not yet been fully confirmed. Analyzing the genetic contribution to obesity is a major step towards the solution. In this *in silico* study, using online bioinformatics tools, we evaluate the role of four genes that are believed to contribute to obesity. Data were collected and analyzed for the sequences of four so-called obesity genes: *FTO* (fat mass and obesity-associated protein), *PPARG* (peroxisome proliferator activated receptor γ), *ADRB3* (adrenergic receptor β 3) and *FABP2* (fatty acid binding protein 2). In the first part of the research, information about the genes was collected and organized and data in FASTA, format are extracted from the National Center for Biotechnology Information (NCBI). In the second part, all genes were analyzed by comparing three species of organisms, *Homo sapiens* (human), *Mus musculus* (mouse) and *Gallus* (chicken). In the third part of this study, phylogenetic trees were constructed for each of the four genes, using blast local alignment search tool (BLAST) and molecular evolutionary genetics analysis (MEGA X) software. Our analysis reveals that the functions of all these genes are associated with overweight and obesity.

Keywords: Bioinformatics; Genes; Obesity; Phylogenetic trees.

[1] Department of Food Technology, Faculty of Technology, Mother Teresa University, Skopje, Republic of North Macedonia

[2] Department of Chemical and Control Engineering, Faculty of Technology and Metallurgy, University "St. Cyril and Methodius", Skopje, Republic of North Macedonia

[3] Department of Biomedical Science, University of Sassari, Sassari, Italy

[4] Department of Biochemistry and Genetic Engineering, Faculty of Agricultural Sciences and Food, University "St. Cyril and Methodius", Skopje, Republic of North Macedonia

INTRODUCTION

Obesity is becoming a major global challenge for humanity. It is expected that in one decade, 38.0% of adults around the world will be overweight, if the current growth rate continues [1]. Obesity occurs because of an unbalanced intake of energy. This imbalance contributes to the occurrence of many chronic diseases such as cardiovascular, diabetes, musculoskeletal disorders, and several types of malignant diseases [2-4].

Obesity is multifactorial. In addition to the non genetic factors, such as nutritional habits and physical inactivity, genetic factors and genetic predisposition play a significant role [2,5,6]. So far, 127 genetic loci have been studied that have a potential link to overweight and obesity [1,4].

Despite many attempts to find a solution to this phenomenon and to reduce the number of people suffering from these diseases, the long-term solution is still being investigated. The development of obesity as a phenomenon is complex [7] and has not been fully understood.

Prevention, as a promising strategy for dealing with this disease, can be achieved by better understanding and controlling of the factors that lead to its manifestation. The analysis and characterization of genetic factors associated with obesity is therefore particularly important.

In the last two decades, various tools have been developed to research, collect data, analyze, and better understand genetic factors. One way of gene analysis is through bioinformatic tools. Bioinformatics is a modern scientific discipline that combines computer science and molecular biology. Bioinformatic tools analyze proteins and nucleic acids, *i.e.,* genes and gene products using computer algorithms and appropriate databases [8].

Due to the ability to quickly analyze biological data, bioinformatics has become an immensely popular and useful field. Specifically, it enables the analysis of biological

data such as DNA, RNA, amino acid sequence of proteins, identification of various characteristics and molecular interactions, prediction of 3D structures, *etc.* All this can be done with tools that are widely available to potential users [9].

Osman *et al.* [4] has recently performed a bioinformatic analysis of the single nucleotide polymorphisms (SNP) of the human *FTO* gene (fat and obesity gene) and suggested that the use of *in silico* analysis may be a good approach to targeting SNPs in other genes associated with the appearance of overweight and obesity. Appa Rao *et al.* [7] has also used bioinformatic tools to analyze the genes involved in diabetes-related obesity. A similar study was conducted by Abdella [10], which concluded that this method of analysis was useful for further studies related to therapeutic and preventive findings for certain diseases.

In this study, using online bioinformatic tools, data related to the *FTO*, *PPARG* (peroxisome proliferator activated receptor γ), *ADRB3* (adrenergic receptor β 3) and *FABP2* (fatty acid binding protein 2) genes, which have been associated with obesity in humans, were collected and analyzed.

MATERIALS AND METHODS

The data for the sequences of the four genes (and corresponding proteins) have been extracted from the National Center for Biotechnology Information (NCBI) [11]. Using these data, we have summarized the general information about the genes in humans and based on the homology information, we have compared the location of the genes in three species.

Next, a phylogenetic tree was contracted using basic local alignment search tool (BLAST) and multiple sequence alignment (MEGA X) software. The BLAST is an online tool that enables study of the evolutionary history of a gene or protein by comparing the homologous [10]; MEGA X is a package for performing fast and accurate multiple sequence alignment of potentially multiple large sequences of large number of proteins or DNA/RNA sequences [12]. The method of phylogenetic inference that is used for constructing the phylogenetic trees is distance-matrix methods neighbor joining (NJ) [13]. We have chosen this method because it provides the best trade-off between accuracy and complexity (computation time) [10]. The evolutionary distances were computed using the Maximum Composite Likelihood method [13,14].

RESULTS

Using the data obtained through NCBI, the *FTO*, *PPARG*, *ADRB3* and *FABP2* genes are characterized and an ontological table is created (Table 1). This table summarizes the key information about these genes including name, location, function, *etc.*

As seen in Table 1, the four genes are in different loci. Homology is evident among similar species, with minor differences. These genes have various functions, however what they have in common is their contribution to the increase in energy intake, *i.e.,* their contribution to overweight and obesity. We also see a link in the last column where all genes, in addition to other diseases, play a role in obesity-related disorders, such as metabolic disorders, weight-related disorders, and others.

In the second part of this research, all genes are analyzed by comparing three species of organisms, *Homo*

Table 1. Data on genetic ontology of the genes (proteins) investigated in this study.

Symbol	Name	Location	Tissue/Expression	Homologs	Function	Pathology of Diseases
FTO	*FTO* α-ketoglutarate-dependent dioxygenase	16q12.2	brain, adrenal glands and 25 other tissues	chimpanzee; rhesus monkey; dog; cow; mouse; rat; chicken; zebrafish; frog	(exact function of this gene is not known); reversing alkylated DNA and RNA damage by oxidative demethylation	strong association with BMI, obesity risk and T2DM
PPARG	peroxisome proliferator activated receptor γ	3p25.2	biased expression in fat, urinary bladder, colon, stomach and nine other tissues	chimpanzee; rhesus monkey; dog; cow; mouse; rat; chicken; zebrafish	helps in regulating transcription of various genes; regulator of adipocyte differentiation	obesity, DM, ather oscelrosis, cancer
ADRB3	adrenergic receptor β 3	8p11.23	ovary, urinary bladder, placenta and two other tissues	chimpanzee; rhesus monkey; dog; cow; mouse; rat; chicken; zebrafish; frog	mediate catecholamine-induced activations of adenylate cyclase through the action of G proteins; involved in the regulation of lipolysis and thermogenesis	obesity and body weight-related disorders
FABP2	fatty acid-binding protein 2	4q26	small intestine, duodenum, colon	chimpanzee; rhesus monkey; dog; cow; mouse; rat; chicken; zebrafish; frog	fatty acid-binding protein, uptake, intracellar metabolism, transport of long-chain fatty acids; may act as a lipid sensor to maintain energy homeostasis	obesity, kidney diseases, metabolic disorder

BMI: body mass index; T2DM: type 2 diabetes mellitus; DM: diabetes mellitus.

Table 2. Ontology: comparative data on the four genes/proteins in three species.

Gene	*Homo Sapiens* (humans)		*Mus Musculus* (mouse)		*Gallus* (chicken)	
	Description	**Location**	**Description**	**Location**	**Description**	**Location**
FTO	*FTO* α-ketoglutarate-ependent dioxygenase	chromosome 16; NC_000016.10 (53703963...54121941)	fat mass and obesity associated	chromosome 8; NC_00004.6 (91313367...91668433)	*FTO* α-ketoglutarate-dependent dioxygenase	chromosome 11; NC_006098.5
PPARG	peroxisome proliferator activated receptor γ	chromosome 3; NC_000003.12 (12287485...12434344)	perixome proliferator activated receptor γ	chromosome 6; NC_000072.6 (115360879...115490404)	perixome proliferator activated receptor γ	chromosome 12; NC_006099.5
ADRB3	adrenergic receptor β 3	chromosome 8; NC_000008.11 (37962990...37966599, complement)	adrenergic receptor β 3	chromosome 8; NC_00074.6 (27225776...27230845, complement)	adrenergic receptor β 3	chromosome 22; NC_006109.5 (2551442...2554479, complement)
FABP2	fatty acid-binding protein 2	chromosome 4; NC_000004.12 (119317250...119322138, complement)	fatty acid-binding protein 2, intestinal	chromosome 3; NC_000069.6 (122895072...122899506)	fatty acid-binding protein 2	chromosome 4; NC_006091.5

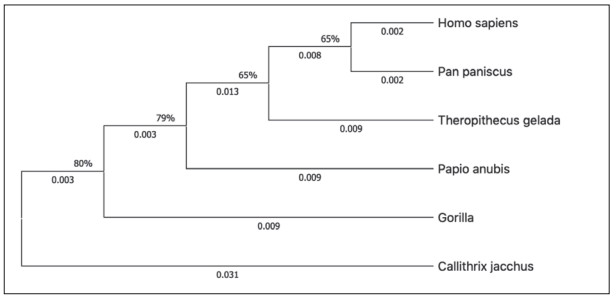

Figure 1. Phylogenetic tree constructed based on the alignment scores of *FTO* sequences.

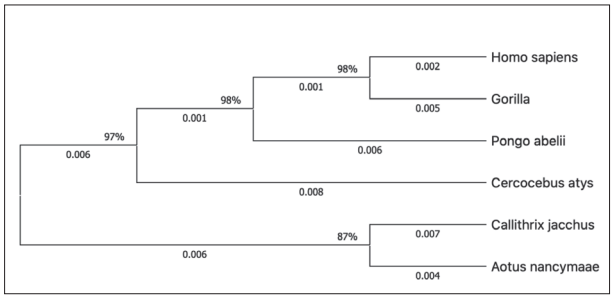

Figure 2. Phylogenetic tree constructed based on the alignment scores of *PPARG* sequences.

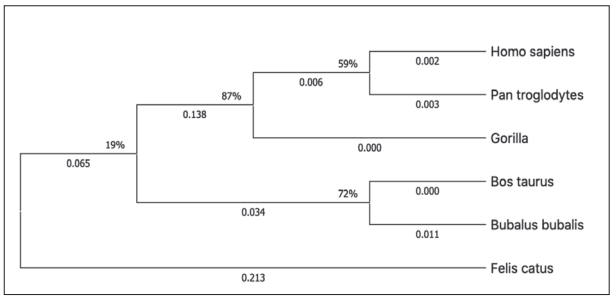

Figure 3. Phylogenetic tree constructed based on the alignment scores of *ADRB3* sequences.

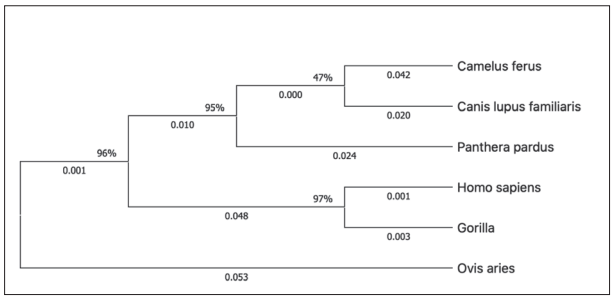

Figure 4. Phylogenetic tree constructed based on the alignment scores of *FABP2* sequences.

sapiens (human), *Mus musculus* (mouse) and *Gallus* (chicken). As a result, an ontological table has been created and is shown in Table 2. The same four genes are found in the three types of organisms. The location of these genes differs in all except for the *ADRB3* gene, where we see the same location (chromosome 8) in both *Homo sapiens* and *Mus musculus*.

The evolutionary relationship of organisms and genetic linkage for each gene is done separately by constructing a phylogenetic tree using MEGA X. The style of the trees we have used is the traditional and rectangular type.

Figure 1 shows the analysis of the *FTO* gene. The evolutionary history was inferred using the NJ method [13]. The evolutionary distances were computed using Maximum Composite Likelihood method [14] and are in units of the number of base substitutions per site. This analysis involved six nucleotide sequences. All ambiguous positions were removed for each sequence pair. There was a total of 1625 positions in the final dataset.

The six selected homologs are human (*Homo sapiens*), bonobo (*Pan paniscus*), gelada (*Theropithecus gelada),* olive baboon (*Papio anubis*), black snub-nosed monkey, gorilla (*Gorilla*), and the marmoset (*Callithrix jacchus*). We used BLAST analysis to select five homologs of homo sapiens. The six alignments were made using MEGA X.

From the constructed tree (Figure 1), we see that the more distantly related to human FTO gene is the marmoset *FTO* gene (*Callithrix jacchus*). Less distantly related (closer relative) to the human *FTO* gene is the bonobo *FTO* gene (*Pan paniscus*).

Figure 2 shows the phylogenetic tree of the *PPARG* gene. There was a total of 1199 positions in the final dataset. Six homologs were selected using BLAST, human (*Homo sapiens*), gorilla (*Gorilla*), orangutan (*Pongo abeii*), the sooty mangabey (*Cerocebus atys*), marmoset (*Callithrix jacchus*) and Nancy Ma's night monkey (*Aotus nancymaae*), which is a night monkey species from South America. From the results of the constructed tree (Figure 2), we see that the more distantly related to the human *PPARG* gene sequence are the marmoset (*Callithrix jacchus*) and Nancy Ma's night monkey *PPARG* genes (*Aotus nancymaae*). Less distantly related to the human *PPARG* gene is the *PPARG* gene of the gorilla.

Figure 3 shows the phylogenetic tree of the *ADRB3* gene. From the BLAST results, this gene was also present in more different species from which six homologs were selected. All ambiguous positions were removed for each sequence pair. There was a total of 1234 positions in the final dataset. The six selected homologs are the chimpanzee (*Pan troglodytes*), gorilla (*Gorilla*), human (*Homo sapiens*), cattle (*Bos taurus*), the water buffalo (*Bubalus bubalis*), and the cat (*Felis catus*). From the evolutionary analyses (Figure 3), we see that the more distantly related to the human *ADRB3* gene is the cat *ADRB3* gene (*Felis catus*). Less distantly related are the chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla*).

Figure 4 shows the phylogenetic tree of the *FABP2* gene. Six homologs were selected, the wild Bactrian camel (*Camelus ferus*), dog (*Canis lupus familiaris*), the leopard (*Panthera pardus*), human (*Homo sapiens*), gorilla (*Gorilla*) and the sheep (*Ovis aries*). This analysis involved six nucleotide sequences. All ambiguous positions were removed for each sequence pair. There was a total of 351 positions in the final dataset. From the results of the constructed tree, we see that the less distantly related to the human is the gorilla. More distantly related are the sheep (*Ovis aeirs)* and camel (*Camelus ferus)* species.

DISCUSSION

Bioinformatics is a relatively new discipline that has enormous potential for development. The use of bioinformatic tools allows testing and eventual validation of scientific hypotheses, which is of immense importance before starting with experimental work. Bioinformatics combined with other disciplines contribute to the diagnosis and prevention of various diseases with a proven genetic basis.

From the analysis of these genes, we can see that greater similarities exist between human and some species of monkeys such as gorilla, chimpanzee and bonobo, also historically called the pygmy chimpanzee. We can note that the gorilla is more closely related in respect to the *FTO* and *ADRB3* genes, whereas for the other two genes, the chimpanzee species are the closest to humans.

Based on the Table 1 with data on genetic ontology of the genes (and corresponding proteins) investigated in this study, homology is evident. These genes have various functions, and what they have in common is their contribution to the increase in energy intake, *i.e.,* their contribution to overweight and obesity. They all play a role in obesity-related disorders, such as metabolic disorders, weight-related disorders, and others.

Furthermore, from the Table 2 ontological data, we see that the same four genes are found in the three types of organisms. The location of these genes differs in all except for the *ADRB3* gene in *Homo sapiens* and *Mus musculus* (chromosome 8).

Based on the analysis of the evolution of these genes, we can conclude that the closest homologs to humans are chimpanzees and gorillas. Less homology is observed between humans and other species included in the investigation such as the camel, cat, leopard, dog, the marmoset, *etc.*

Using bioinformatic tools to identify and characterize obesity-associated genes, we obtain valuable information about the underlying factors and causes of obesity and can contribute toward identifying solutions to this problem. The development of obesity is multifactorial and complex, and genetic predisposition itself depends on other factors such as gene expression. The possession of different variants of these genes is not always manifested with overweight or obesity. Few studies have found that the interaction between transcription factors and epigenetic modifications play a critical role in the expression of the obesity genes [15]. The pathogenesis in the metabolism and the regulation of the expression of these genes is still unclear. Systematic research and more data will be needed to understand the interactions and the effect of all these factors and eventually to identify treatments.

Declaration of Interest. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

REFERENCES

1. Singh R, Kumar P, Mahalingam K. Molecular genetics of human obesity: A comprehensive review. C R Biol. 2017; 340(2): 87-108.
2. Herrera BM, Lindgren CM. The genetics of obesity. Curr Diab Rep. 2010; 10(6): 498-505.
3. Shabana, Ullah Shahid S, Hasnain S. Identification of genetic basis of obesity and mechanistic link of genes and lipids in Pakistani population. Biosci Rep. 2018; 38(4): BSR20180281.
4. Osman MM, Khalifa AS, Mutasim AEY, Massaadd SO, Gasemelseed MM, Abdagader MA, *et al.* In silico analysis of single nucleotide polymorphisms (Snps) in human FTO gene. JSM Bioinform Genomics Proteomics. 2016; 1(1): 1-13.
5. Ali AT, Crowther NJ. Factors predisposing to obesity: A review of the literature. S Afr Fam Pract. 2010; 52(3): 193-197.
6. Zhao J, Grant SFA. Genetics of childhood obesity. J Obes. 2011; 2011: 845148.PAGE
7. Pevsner J. Bioinformatics and Functional Genomics, 3rd ed. Hoboken, NJ, USA: Wiley-Blackwell, Press; 2015.
8. Mehmood MA, Sehar U, Ahman N. Use of bioinformatics tools in different spheres of life sciences. J Data Mining Genomics Proteomics. 2014; 5(2): 158.
9. Appa Rao A, Tayaru N, Thota H, Babu Changalasetty S, Saroja Thota L, Gedela S. Bioinformatic analysis of functional proteins involved in obesity associated with diabetes. Int J Biomed Sci. 2008; 4(1): 70-73.
10. Abdella. Bioinformatics analyses of some functional genes and proteins involved in obesity-induces type 2 diabetes. Am J Bioinform. 2012; 1(2): 87-98.
11. National Center for Biotechnology Information. Available at: https://www.ncbi.nlm.nih.gov/. Accessed August 20, 2020.
12. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013; 30(5): 1229-1235.
13. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4(4): 406-425.
14. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA. 2004; 101(30): 11030-11035.
15. Huang Q, Ma C, Chen L, Luo D, Chen R, Liang F. Mechanistic insights into the interaction between transcription factors and epigenetic modifications and the contribution to the development of obesity. Front Endocrinol (Lausanne). 2018; 9: 370.