

Human low-density lipoprotein receptor sequence analysis

Marwah Ali Zghair,¹ Fadhil Jawad Al-Tuma,² Hassan Mahmoud Abu Al-Maali³

Abstract

Objective: To introduce the local low-density lipoprotein receptor sequence to the GenBank database and to compare the differences with other sequences.

Method: The study was conducted from January to March 2020 at the College of Pharmacy, Karbala University, Iraq. The process of genotyping was done by the amplification of 150bp area of low-density lipoprotein gene by a polymerase chain reaction of deoxyribonucleic acid extracted from the blood samples of the local subjects. Data sequencing was analysed using the basic local alignment search tool software.

Results: The phylogenetic tree for low-density lipoprotein receptor showed its evolutionary relationship in human and other biological species that may have a common ancestor. The outcomes reflected the geographic association in the evolution process of low-density lipoprotein receptor.

Conclusion: Similarity between the studied region and another sequencing in the database explained that the analogous sequences were of the same genetic origin as the low-density lipoprotein receptor and may not necessarily have the same function.

Keywords: Low-density lipoprotein receptor, Sequencing, BLAST. (JPMA 71: S-12 [Suppl. 9]; 2021)

(DOI:)

Introduction

The Low-density lipoprotein receptor (LDLR) is a "protein rich" cell surface receptor playing a significant role in the cholesterol homeostasis.¹ The LDLR binds the Low-density lipoprotein (LDL) and transfers it to the cell lysosomes whereas the protein undergoes degradation, and the cholesterol is made available to repress the microsomal enzyme 3 hydroxy-3 methylglutaryl coenzyme A (HMG-CoA) reductase.²

In the LDLR protein, the most frequent amino acids are aspartate (8.70%), leucine (7.70%), serine (8.10%), glycine (7.20%), cysteine (7.30%) and valine (6.70%). There are also phenylalanine (3%) and tryptophan (2.30%) varieties. Such amino acids dispersal has been consistent with the one that has been generally reported for human proteins.³ Since LDLR is one of the mosaic proteins it is a member of numerous super-gene families and it has been built up of the exons shared with various proteins.⁴ It is known that 13 of the 18 exons sequences of the LDLR encoding gene are homologous to sequences in other proteins and as following:- 5 sequences similar to the C-9 complement's component arrangement; 3 translating a sequence identical to repeat sequences in the Epidermal growth factor (EGF) precursor, 3 other sequences identical to the blood clotting proteins system (the

clotting factors X&IX and protein C); and the last 5 exons encoding the non-repeated sequences which have been only shared with the EGF precursor.⁵

Molecular diagnosis has been known as the most specific approach for the identification of patients with numerous diseases of genetic origin.⁶ Even though there are several gene mutations of the LDLR identified in subjects of numerous countries, there is a very limited amount of genetic data for the Iraqi population and this study is within the first studies that introduce the LDLR gene sequencing of Iraqi subjects for the gene bank data.

The mutations of the LDLR may take place in a promoter, with exons or introns of the coding gene.⁷ Most variants fall in EGF precursor-like (48%) or ligand-binding domains (41%), with a maximal incidence that has been stated in exon 4 (20%).⁸ Mutations in the LDLR gene were classified into five functional groups, according to the mutant protein characteristics that have been formed and researched in the fibroblasts of the patients.⁹

Class1 mutations disturb the synthesis of the receptor in the endoplasmic reticulum; Class 2 mutations obstruct the LDLR adjustments by inhibiting the proper passage of the receptor; Class 3 mutations prevent the attachment of LDL to its receptor; Class 4 mutations stop the usage of the receptor-ligand compound; and Class 5 mutations result in blockage of correct LDLR recycling.¹⁰

The current study was planned to introduce the LDLR sequence of Iraqi individuals to the Gene Bank database, get the most similar global sequences and to analyse the

¹Department of Pharmaceutics, College of Pharmacy, ²Department of Chemistry and Biochemistry, College of Medicine, ³Department of Clinical Laboratory Sciences, College of Pharmacy, University of Kerbala, Kerbala, Iraq.

Correspondence: Fadhil Jawad Al-Tuma. Email: fadhil.jawad@uokerbala.edu.iq

sites of similarity and differences with other sequencings.

Materials and Methods

The research was done according to the International ethical guidelines and approved by an independent ethics review committee (ERC). The participants voluntarily agreed to participate in the study, which was documented.

The study was conducted from January to March 2020 at the College of Pharmacy, Karbala University, Iraq. The inclusion criteria were subjects between 25-50 years age and generally in good health as judged by the participating physicians. They had no history of genetic or inherited diseases and had no genetic tests done earlier inclusive of the family members also. Five mL venous blood was drawn from each candidate by direct puncture to the antecubital vein. Genomic deoxyribonucleic acid (DNA) was extracted from the blood samples using Promega kit and according to the labelled instructions, and the concentration and purity of the collected DNA were measured using the nano-drop technique.¹¹ The process of genotyping was done by the amplification of 150bp region by polymerase chain reaction (PCR) using two primers, P1: 5'-GAG TGT CAG GAT CCC GAC ACC TGC GCC-3' and P2: 5'-AAG TCG ACC CAC CCG CCT GCC TCC CGT-3' and 35 thermal cycles.¹² The data obtained from DNA sequencing for the studied region of the LDLR gene was analysed with the basic local alignment search tool (BLAST) software to get the most similar sequences to LDLR sequence from the GenBank database, and to analyse the sites of similarity and differences with other sequences.

Results

The data obtained from the DNA sequences analysis for

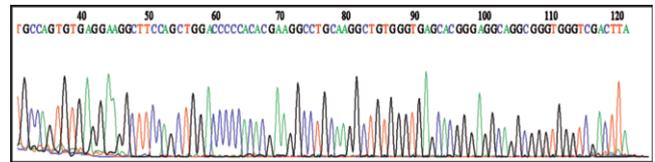


Figure-1: The amplification region of the low-density lipoprotein receptor (LDLR).

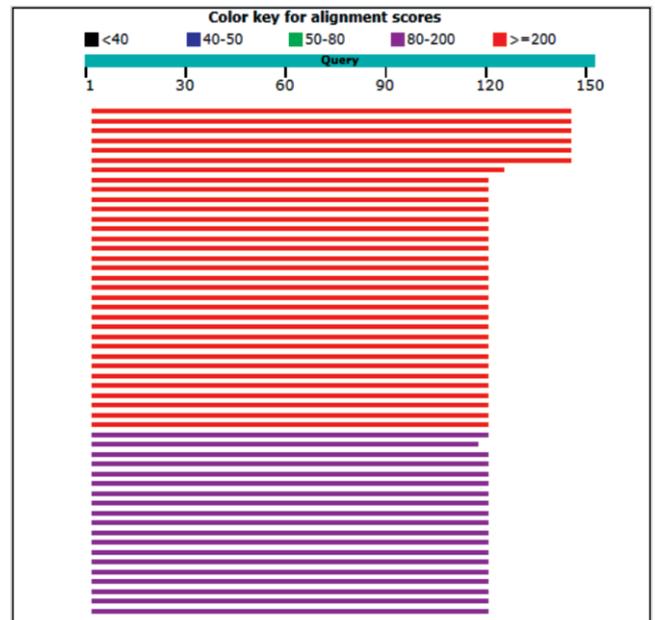


Figure-2: Graphic summary of sequences similar to the low-density lipoprotein receptor (LDLR) gene taken from local patients. The bar is coloured according to the obtained alignment scores. Colour codes and the bar length indicate the similarity area with the sequence of the query.

the studied region of the LDLR gene was analysed and recorded for both forward and reverse primers as nucleotides sequencing (Figure-1).

The graphical output provided by BLAST-n provided a

Table-1: Descriptions of similar sequences with low-density lipoprotein receptor (LDLR) gene.

Descriptions	Maximum score	Query cover	E value	Identity	Accessions
Homo sapiens LDLR gene, exon 8 and partial cds	248	94%	3.00E-62	98%	HQ190917.1
Homo sapiens LDLR gene, complete cds	248	94%	3.00E-62	98%	FJ525879.1
Homo sapiens LDLR, Ref SeqGene (LRG274) on the 19th chromosome	248	94%	3.00E-62	98%	NG_009060.1
Homo sapiens chromosome 19 clone CTB-164019, full sequence	248	94%	3.00E-62	98%	AC011485.6
Homo sapiens LDLR gene, partial cds	248	94%	3.00E-62	98%	AF217403.1
Homo sapiens LDLR gene, full cds	248	94%	3.00E-62	98%	AY324609.1
Homo sapiens LDLR gene, full cds	222	80%	2.00E-54	99%	AH002776.2
Homo sapiens LDLR, transcript variant 6, m-RNA	213	77%	1.00E-51	99%	NM_001195803.1
Homo sapiens LDLR, transcript variant 4, m-RNA	213	77%	1.00E-51	99%	NM001195800.1
Homo sapiens LDLR, transcript variant 3, m-RNA	213	77%	1.00E-51	99%	NM001195799.1

Query cover: Coverage percentage for complete sequent, E value: Coincidence similarity percentage, Identity: Percentage. The six columns to the right in the description table show the maximum score (the summation of the similarities and identities that have been dictated by the default matrix of the comparison), the Total Score, the percent query coverage (percentage of the query the subject hits cover), the E-value score, and the percent Identity between the query and the subject sequences. The last column contains the National Center for Biotechnology Information (NCBI) accession number for the corresponding hit sequence. The E-value (Expected value) is the amount of the alignments expected to take place by chance in a search of the data-base and is associated with a value of probability (p).

Table-2: Descriptions of similar sequences with low-density lipoprotein receptor (LDLR) protein.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Chain L, The Structure Of The LDLR /pcsk9 Compound Reveals LDLR In An Extended Conformation	72.4	72.4	100%	5.00E-14	100%	3P5B_L
LDLR isoform CRA_c "Homo sapiens"	72.4	72.4	100%	6.00E-14	100%	EAW84170.1
Chain L, The Structure Of The LDLR/pcsk9 Compound show The Receptor In An Extended Conformation	70.1	70.1	100%	3.00E-13	100%	3P5C_L
LDLR "Homo sapiens"	64.3	64.3	96%	3.00E-13	100%	ADM85889.1
LDLR isoform 4 precursor	69.7	69.7	100%	4.00E-13	100%	NP_001182729.1

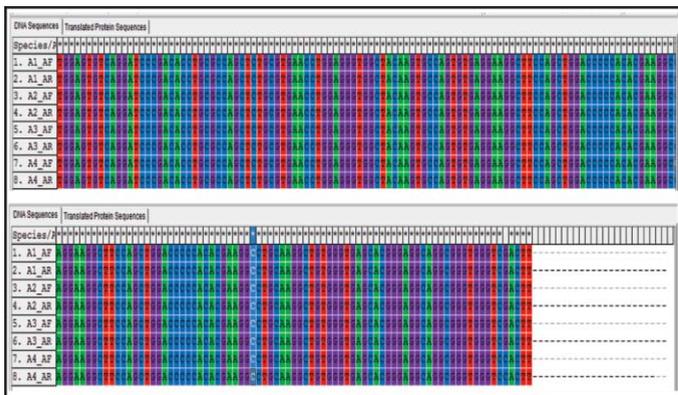


Figure-3: Low-density lipoprotein receptor (LDLR) deoxyribonucleic acid (DNA) sequences.

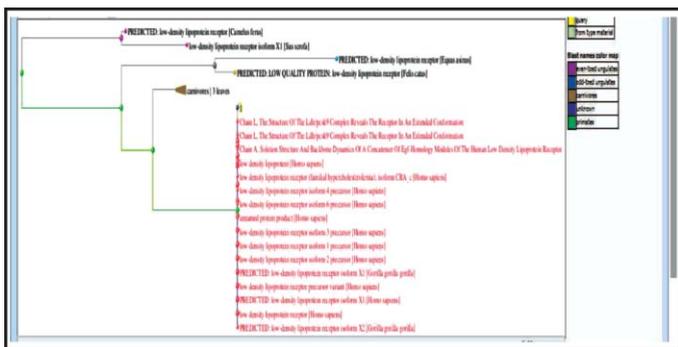


Figure-4: Rectangle tree view for low-density lipoprotein receptor (LDLR).

quick overview of the LDLR, or the query, sequence and resulted in the location of similar hit sequences (Figure-2).

Details of the sequences similar to the LDLR gene were noted (Table-1). The DNA sequencing for LDLR was aligned by Molecular Evolutionary Genetics Analysis version 5 (MEGA-5) programme with ClustalW (Figure-3).

The phylogenetic tree for the LDLR achieved by BLAST showed the evolutionary relationships between the LDLR in human and other biological species that may have a common ancestor (Figure-4).

The protein sequences as a result of LDLR nucleotide translation to amino acids were compared with the most similar proteins by protein alignment according to GenBank database using BLAST-P software (Figure-5).

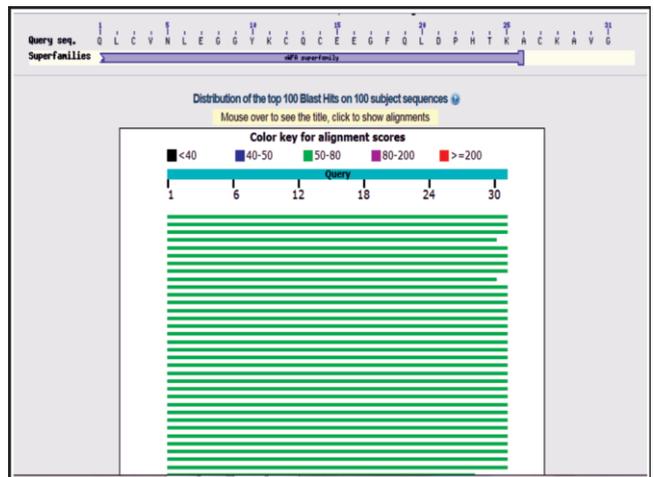


Figure-5: Summary of similar sequences to low-density lipoprotein receptor (LDLR) protein from local patients.

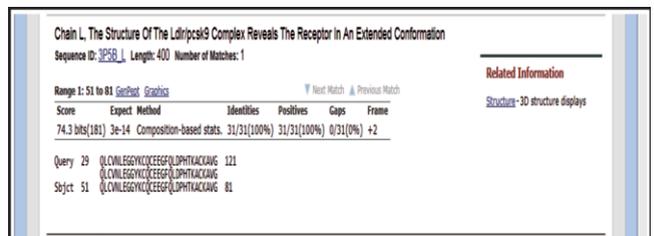


Figure-6: Alignments view of the low-density lipoprotein receptor (LDLR) sequence and its most similar hits.

The view of the BLAST-P alignment results provided more exhaustive information about every one of the hits and additionally played the role of a hyperlink to the corresponding GenBank sequence (Table-2).

In the alignment view for LDLR protein, the individual alignments were represented together with a region of LDLR amino acid, according to the site of homogeneity (Figure-6).

The main differences between the LDLR amino acids in humans and other living organisms were noted (Figure-7).

The implicated studied region of the LDLR protein was located on chain L of the LDLR structure (Figure-8).

The selected nucleotide sequencing for the LDLR gene



Figure-7: Comparison between the low-density lipoprotein receptor (LDLR) amino acids in humans and other living organisms.

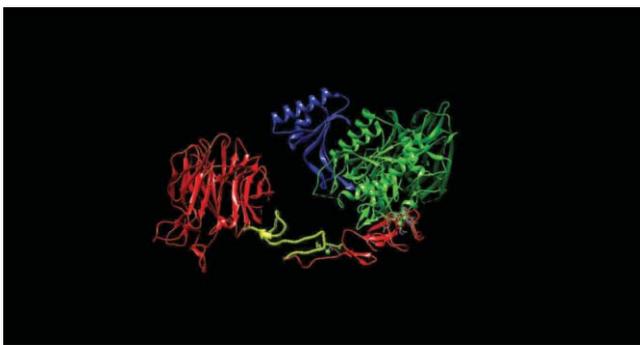


Figure-8: Chain L in the low-density lipoprotein receptor (LDLR) protein structure.

```

/ gene= LDLR
/ codon_start=1
/ product="low density lipoprotein receptor"
/ protein_id="AA056036.1"
/ translation="MGPNGWKL RNTVALLAAAGTAVGRCERNEFQCQDGKCSISKYK
VCDGSAECQDGSDESQETCLSVTCKSGDFSCGGRVNRICPQFHRCDGQVDCDNGSDEQ
GCPKTKCSQDFRCHDQKICISRFVCDSDRDCLDGSDSEACPFVLTGCPASFCNHSSTC
IPQLNACDNDPDCEDGSDENPQRGRLVYFQDSSPSCAFEFHCLSGECIHSNRCDG
GPDCKSDSEINC AVATCRPDEFQCSGDNCTHGSRCQDREYDCKDMSDEVGCVNVTLC
EGPKFKSGHCEITLDKVCNMRDCRDNSEPIKECGTNECLDINGGCSHVCHDLKIC
GYECLCPDGFQVAQRCEIDDECQDPDTCQLCVNLEGGYKCCQEEGFQLDPIKAC
KAVQSIAYLFFTRNREVRKHTLDRSEYTSLIPNLNINVVADTEVASNRIYNSDLQRH
ICSTQLDRAHGVSSYDVTISRDIAQPDGLAVDWIHSIYWTDSVLTGTVSVADTKGVKR
KTLFRENGSKPRAIVDPVHGFHYWTDNGTPAKIKKGLNGVDIYSLVTENIQNNGI
TLDLISGRLYWVDSKLSHSISIDVNGNRKTLLEDEKRLAHFSLAVFEDKVFHTDII
NEAIFSANRLTSSDNLAEENLSPEDNRFHNLTOPRGNVWCERTTSLNNGCQVLCCL
PAPQINPHSPKFTCACPDGHLARDMRSLCTEAEAAVATQETSIVRLKVSVAVTQTH
TTTTRPVDTSRLLPGATPGLTTVEIVTISHQALGDVAGRGNKPKPSSVRALSIPLIVL
LVFLCLGVFLLNKNRRLKININSINFNIPVYQKTEDEVIEHCHIQGYSYSPRQVLSLE
DDVA"
    
```

Figure-9: The amino acids sequencing for the studied area of the low-density lipoprotein receptor (LDLR) protein.

was translated into amino acids sequencing using BLAST (Figure-9).

Discussion

The results obtained from LDLR gene sequencing were aligned by BLAST, which compares the main information

of the biological sequences¹³ for finding the sequence similarity regions which provide evolutionary and functional clues on the function and the structure of LDLR. The results of similarity explained that these analogous sequences were of the same genetic origin as LDLR which may not necessarily have the same functions.¹⁴

The hits on the graphic summary (Figure-2) were red (S>200) and pink (S=80-200) in colour due to the high alignment score (S), and, consequently, the low significance of E-value, while the hits yielded from the other graphic summary (Figure-5) were green in colour (S=50-80) and this showed a lower value of alignment score (S) and a higher level of E-value. This outcome may have been related to the larger number of LDLR nucleotides aligned compared to the LDLR amino acids aligned. The piece of the aligned query increased the length and decreased the degree of homogeneity caused unexpectedly, explaining the lower E-value, the higher S-value and the colour of hits directed from black to red.

Besides, the BLAST-P graphic summary showed that the amplified region of the LDLR protein sequences may be homologous with Von Willebrand factor type A (vWA) super-family domain, which was found originally in blood coagulation protein von Willebrand's factor (vWF).¹⁵ The domain of the vWA, since its discovery, has drawn much interest due to its widespread occurrences in addition to its involvements in many different significant functions of the cells. These include the creation of the basal membranes, cell differentiation, cell migration, haemostasis, adhesion and signalling.¹⁶

On the other hand, the first sequence that aligned with LDLR gene with maximum coverage was the 429bp Homo-sapiens LDLR gene, exon 8 and partial coding DNA sequence (CDs) with c.1186+41T>A single nucleotide polymorphisms (SNP).¹⁷ The 2nd aligned sequence has been presented to the Gene Bank by two studies; the first one¹⁸ defined the genes that involved in the cholesterol metabolism in human and made chromosomal loci mapping for the LDLR, while the second group¹⁹ stated that the LDLR is a mosaic protein built up of the exons shared with various proteins, and, that is why, it belongs to numerous super-gene families.

Besides, in the alignment view of BLAST for LDLR, the section of dissimilarity were indicated by gaps which may be the result of addition, deletion or substitution of a nucleotide or amino acid, whereas the identical part between the subject and query are known as the conserved area.²⁰ The absence of the substitutions, or the presence of only highly-conservative substitution cases in

a specific sequence region suggested that this region is structurally or functionally important.²¹

Moreover, in the phylogenetic tree, every one of the nodes with the descendants is representing the descendants' most recent common ancestor and the lengths of the edges are corresponding to the estimates of the time duration. Every one of the nodes has been a taxonomic unit, while taxonomy represents the classification of the organisms based on similarity.²²

The first taxonomic group in the LDLR evolutionary tree belongs to even-toed ungulates with two subgroups; one related to *Camelus ferus* LDLR predicted by automated computational analysis for male bactrian camel from Mongolian wild camel protection area, and the other member is related to LDLR precursor.²³ The LDLR gene has to be one of the potential causative genes for the LDL cholesterol and total cholesterol (TC) in the pigs, but there is heterogeneity in various populations. Other evolutionary ancestor for the LDLR in homo-sapiens is *Equus asinus* isolated LDLR from the blood sample of male Chinese Guanzhong donkey²⁴ in addition to the LDLR isoform X3 *Felis catus*²³ with all the organisms being devolved from Asia, which reflected the geographical association in the evolution process.

Similarly, the first items in the results of the BLAST-P alignments was mentioned by a study²⁵ which described Chain L in the structure of the LDLR-c with the sequence 100% identical to the studied region of the LDLR protein, the second identical sequence with 100% coverage is the LDLR protein sequence introduced to the sequences of the human genome database²⁴ as part of the sequence of the human genome research.

These results of multiple aligned sequences belonged to homo-sapiens LDLR protein with 100% coverage and identity to the studied sequence, confirm that this region is related to the LDLR gene and can be analysed for further researches.

Conclusion

The results of similarity between the studied region of the LDLR and another sequencing in the GenBank database explained that these analogous sequences are of the same genetic origin as the LDL and may not necessarily have the same function. Moreover, BLAST-P graphic summary showed that the amplified region of the LDLR protein sequences may be homologous with the vWA superfamily domain which has been discovered first in blood coagulation protein v-WF. The current study showed the LDLR sequence region to be structurally and functionally important and indicated by the absence of

the substitutions, or presence of only quite conservative substitution types in a specific region in comparison with sequences of other species.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: None.

References

1. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J* 2017;38:2459-72. doi: 10.1093/eurheartj/ehx144.
2. Grenkowitz T, Kassner U, Wühle-Demuth M, Salewsky B, Rosada A, Zemojtel T, et al. Clinical characterization and mutation spectrum of German patients with familial hypercholesterolemia. *Atherosclerosis* 2016;253:88-93. doi: 10.1016/j.atherosclerosis.2016.08.037.
3. Goldstein JL, Brown MS. A century of cholesterol and coronaries: from plaques to genes to statins. *Cell* 2015;161:161-72. doi: 10.1016/j.cell.2015.01.036.
4. Benito-Vicente A, Uribe KB, Jebari S, Galicia-Garcia U, Ostolaza H, Martin C. Validation of LDLr Activity as a Tool to Improve Genetic Diagnosis of Familial Hypercholesterolemia: A Retrospective on Functional Characterization of LDLr Variants. *Int J Mol Sci* 2018;19:e1676. doi: 10.3390/ijms19061676.
5. Do R, Stitzel NO, Won HH, Jørgensen AB, Duga S, Angelica Merlini P, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* 2015;518:102-6. doi: 10.1038/nature13917.
6. Al-Tu'ma FJ, Abu Al-Maali HM, Zghair MA. Correlation between anthropometric parameters with low density lipoprotein receptor (A370T) gene polymorphism in type II diabetes patients of Iraqi population. *World J Pharm Med Res* 2019;5:13-8.
7. Wang B, Tontonoz P. Liver X receptors in lipid signalling and membrane homeostasis. *Nat Rev Endocrinol* 2018;14:452-63. doi: 10.1038/s41574-018-0037-x.
8. Hsieh J, Koseki M, Molusky MM, Yakushiji E, Ichi I, Westerterp M, et al. TTC39B deficiency stabilizes LXR reducing both atherosclerosis and steatohepatitis. *Nature* 2016;535:303-7. doi: 10.1038/nature18628.
9. Ma S, Sun W, Gao L, Liu S. Therapeutic targets of hypercholesterolemia: HMGCR and LDLR. *Diabetes Metab Syndr Obes* 2019;12:1543-53. doi: 10.2147/DMSO.S219013.
10. Al-Khateeb A, Zahri MK, Mohamed MS, Sasongko TH, Ibrahim S, Yusof Z, et al. Analysis of sequence variations in low-density lipoprotein receptor gene among Malaysian patients with familial hypercholesterolemia. *BMC Med Genet* 2011;12:40. doi: 10.1186/1471-2350-12-40.
11. Muftin NQ, Jubair S, Hadi SM. Identification of MTHFR genetic polymorphism in Iraqi autistic children. *Gene Reports* 2020;18:e100585. DOI: 10.1016/j.genrep.2019.100585
12. Zghair MA, Al-Tu FJ, Abu Al-Maali HM. Association between the low density lipoprotein receptor (A370T) gene polymorphism with lipid profile in type 2 diabetes mellitus. *Karbala Journal of Pharmaceutical Sciences* 2018;9:103-12.
13. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 2004;32:20-5. doi: 10.1093/nar/gkh435.
14. Mizrahi I. GenBank. In: *The NCBI Handbook*, 2nd ed. Bethesda,

- MD: National Center for Biotechnology Information (US); 2013.
15. Peyvandi F, Garagiola I, Baronciani L. Role of von Willebrand factor in the haemostasis. *Blood Transfus* 2011;9(Suppl 2):s3-8. doi: 10.2450/2011.002S.
 16. Paththinige CS, Rajapakse JRDK, Constantine GR, Sem KP, Singaraja RR, Jayasekara RW, et al. Spectrum of low-density lipoprotein receptor (LDLR) mutations in a cohort of Sri Lankan patients with familial hypercholesterolemia - a preliminary report. *Lipids Health Dis* 2018;17:100. doi: 10.1186/s12944-018-0763-z.
 17. Ye H, Zhao Q, Huang Y, Wang L, Liu H, Wang C, et al. Meta-analysis of low density lipoprotein receptor (LDLR) rs2228671 polymorphism and coronary heart disease. *Biomed Res Int* 2014;2014:e564940. doi: 10.1155/2014/564940.
 18. Ma R, Zhu X, Yan B. SCARB1 rs5888 gene polymorphisms in coronary heart disease: A systematic review and a meta-analysis. *Gene*. 2018;678:280-7. doi: 10.1016/j.gene.2018.08.024.
 19. Jha CK, Mir R, Banu S, Elfaki I, Chahal SMS. Heterozygosity in LDLR rs2228671 and rs72658855 Gene is Associated with Increased Risk of Developing Coronary Artery Disease in India -A Case-Control Study. *Endocr Metab Immune Disord Drug Targets* 2020;20:388-99. doi: 10.2174/1871530319666191015164505.
 20. Ma S, Sun W, Gao L, Liu S. Therapeutic targets of hypercholesterolemia: HMGCR and LDLR. *Diabetes Metab Syndr Obes* 2019;12:1543-53. doi: 10.2147/DMSO.S219013.
 21. Zhong J, Chen X, Wu N, Shen C, Cui H, Du W, et al. Catechol-O-methyltransferase promoter hypomethylation is associated with the risk of coronary heart disease. *Exp Ther Med* 2016;12:3445-9. doi: 10.3892/etm.2016.3757.
 22. Podan J. The Coral of Life. *Evol Biol* 2019;46:123-44. Doi: 10.1007/s11692-019-09474-w.
 23. Zeng Z, Chen R, Liu C, Yang H, Chen C, Huang L. Evaluation of the causality of the low-density lipoprotein receptor gene (LDLR) for serum lipids in pigs. *Anim Genet* 2014;45:665-73. doi: 10.1111/age.12183.
 24. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304-51. doi: 10.1126/science.1058040.
 25. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Ann N Y Acad Sci* 2013;1281:64-91. doi: 10.1111/nyas.12098.
-