Designing PCR Primer of VEGFA Gene for CreatingDendrogram Using Bioinformatics Tools

*J. Raja Brindha,**Sreeya G. Nair, *N. C. J. Packia Lekshmi, *Ashiya A., *Raji M. R. *Department Of Allied Health Sciences, Noorul Islam Centre for Higher Education, Kumaracoil, Kanyakumari District, Tamil Nadu, India.

**Department of Zoology, Sree Ayyappa College for Women, Chunkankadai, Kanyakumari District, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India.

Abstract: Bioinformatics tools can be used for the organization and for analysis of biological data. The primers for a polymerase chain reaction can be designed by using bioinformatics tools. The parameters like length of the primer, GC content and melting temperature can be easily optimized thus enabling to implement the necessary changes to design the desired primers. In silico analysis can be employed to test the specificity of the primers. The phylogenetic tree of VEGFA gene revealed that the VEGFA gene in *Homo sapiens* is similar to that of *Xenopus laevis*. The gene variants in *Oxyura jamaicensis* and in *Mus musculus* are similar and they share the same evolutionary pattern with *Danio rerio*. VEGFA gene in *Bufo bufo* is similar to the gene in other organisms.

Key words: Homo sapiens, VEGFA gene, Xenopus laevis, Phylogenetic tree

1. Introduction

Polymerase chain reaction is a widely used molecular biology tool to amplify a specific DNA fragment. PCR was developed by an American biochemist, Kary Mullis in 1983 at the Cetus Corporation, for which he was awarded the Nobel Prize in Chemistry in 1993. The polymerase chain reaction involves rapid heating and cooling which is carried out in a thermal cycler (yourgenome.org).PCR has a broad range of applications in research, molecular diagnostics, forensics and agriculture. PCR is employed for genotyping as it enables to identify the sequence variations in alleles in specific cells or organisms.

Vascular Endothelial Growth Factor A Gene belongs to PDGF/VEGF growth factor family. It is heparin binding growth factors that are specific for vascular endothelial cells and induce angiogenesis in vivo (Leung *et al.*, 1989). Vascular endothelial growth factor A is the most prototypical and the most studied among the VEGF family (Alberto A'lvarez-Aznar *et al.*, 2017). It is a key regulator of both the physiological and pathological angiogenesis and belongs to a larger family of signaling ligands that exhibit variety of effects on different cell types (Rao and Lang, 2010). There are three tyrosine kinase receptors in VEGF family, but the receptors 1 and 2 mediate the signaling of VEGFA (Rao and Lang, 2010). The VEGF-A exhibit high vascular permeability and are derived from purified fluid secreted by a tumor (Ankit P. Laddha *etal.*, 2020; Senger *et al.*, 1983).

The Vascular Endothelial Growth Factor A gene induces the proliferation of vascular endothelial cells and promote the cell migration. The VEGFA have different biological functions and play a major role in angiogenesis (Neha Garg *et al.*,2019). VEGF-A exhibit various activities like inducing endothelial cell proliferation, cell migration, the proteolytic activity and stimulating the microvascular leakage (Regenfuss and Cursiefen,2010). It is a multifaceted cytokine that mainly influence certain endothelial pathways that promote the cell growth, survival and the vascular permeability (Kleinman and Ambati,2010).

The human VEGF gene was first cloned and characterized from the cultured vascular smooth muscle cells. The gene contains 8 exons. Due to the alternative splicing the regions encoded by 6^{th} exon lacks 165 amino acids from and the region encoded by 6^{th} and 7^{th} exons lacks 121 amino acids (OMIM, Tischer *et al.*, 1991).





Fig 1 – The chromosomal structure of Vascular Endothelial Growth Factor fromGeneCards.

(https://www.genecards.org)

Diseases associated with VEGFA gene

i) Poems Syndrome:

POEMS Syndrome is also known as crow- Fukase syndrome, which is a rare paraneoplastic syndrome characterized by the organomegaly, monoclonal gammopathy, endocrinopathy and also the changes in skin. The elevated serum or plasm levels of the vascular endothelial growth factor (VEGF) indicate the POEMS syndrome. The most commonly used therapies for the POEMS syndrome includes alkylators, steroids, chemotherapy with blood stem cell transplantation, bortezomib. and lenalidomide (MalaCards; Kyoto Encyclopedia of Genes and Genomes).

ii) Proliferative Diabetic Retinopathy

It is also known as the microvascular complication of Diabetes 1, characterized by sciatica and neuralgia. The ERK signaling and TGF-Beta pathway related to the VEGFA gene is responsible for the complications of the disease. Treatment include clopidogrel and glimepiride (MalaCards). **VEGFA gene in other animals**

Beside humans the VEGFA gene is also present in

- **Mus musculus* (house mouse)
- *Danio rerio (zebra fish)
- *Bufo (common toad)
- * Xenopus laevis (African clawed frog)

* Oxyura jamaicensis (ruddy duck)

Bioinformatics tools

i) GenBank

GenBank is a comprehensive database containing all the publicly available nucleotide sequences of about 4, 20,000 species (Eric W Sayers *et al.*,2019). The nucleotide sequences were submitted to GenBank using BankIt or using the tool tbl2asn or through the NCBI submission portal for individual laboratories as well as for batch submissions from large scale sequencing projects including whole genome shotgun(WGS) and the environmental sampling projects (Eric W Sayers *et al.*,2019). GenBank ensures the worldwide coverage through the daily data exchange with the DNA Data Bank of Japan (DDBJ) and with the European Nucleotide Archive (ENA). GenBank can be accessed through the NCBI database (www.ncbi.nlm.nih.gov/genbank/).

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

ii) ClustalW

ClustalW is a bioinformatics tool for aligning any number of the homologous nucleotide or protein sequences (Higgins *et al.*, 1994). ClustalW use the progressive alignment methods to obtain multiple-sequence alignment. The sequences having best alignment score or the most similar sequences are aligned first. Then, more distant groups of the sequences are aligned and finally a global alignment is obtained (Higgins *et al.*,1994). ClustalW algorithm first compute a rough distance matrix between each pair of the sequences based on their pairwise sequence alignment scores. The pairwise sequence alignment scores are computed using the pairwise alignment parameters for both the DNA and protein sequences. The next step is the neighbor-joining method creating a guide tree that generates a global alignment. The guide tree serves as a rough template that share the insertion and deletion features providing a close-to-optimal result (Higgins *et al.*,1994).ClustalW can be accessed through https://www.genome.jp.

iii) EMBOSS Cons

The European Molecular Biology Open Software Suite (EMBOSS) is a collection of the bioinformatics tools and software libraries that can be used individually. EMBOSS is programmed to run on UNIX platform (Lisa J. Mullan and Alan J. Bleasby,2002). EMBOSS Cons creates consensus sequence from a multiple sequence alignment. EMBOSS Cons can be accessed through https://www.ebi.ac.uk.

iv) BLAST

Basic Local Alignment Search Tool is a sequence similarity searching tool used to easily search a sequence database for the matches to a query sequence (Christiam Camacho *et al.*,2009; Altschul *et al* 1990, Altschul *et al.*,1997). BLAST contains several variants to compare the combinations of all the nucleotide or protein sequences against a nucleotide or protein database. It provides an expect value or the statistical information about the significance of each of the alignment. The Basic Local Alignment Search Tool (BLAST) can handle the databases having more than two billion letters and also it is possible to limit the search by a list of GenInfo Identifiers (GI) and can simultaneously search multiple databases (Christiam Camacho *et al.*, 2009). BLAST can be accessed by https://blast.ncbi.nlm.nih.gov.

v) Primer3 Plus

Primer3 Plus is a web interface to the Primer3 primer design program which is an alternative to the CGI-scripts that comes with Primer3. The Primer3 Plus combines the present-day web technologies to provide an intuitive user interface and has been developed in collaboration with the molecular biologists and technicians who regularly design the primers (Andreas Untergasser *et al.*,2007). Primer3 Plus supports many sequence formats including FASTA. The primers obtained from Primer3 Plus can be sent to an order form thus allowing tight integration into the laboratory ordering systems (Andreas Untergasser *et al.*,2007). Primer3 Plus can be accessed by http://www.bioinformatics.nl/primer3plus.

vi) In silico PCR

In silico PCR allows to map the locations of a pair of PCR primers within the genome of a model organism. It searches a sequence database with the help of PCR primers, using an indexing strategy for the fast performance. The minimum required length of the primers is 15 bases. The result is obtained in FASTA format, which contains all the sequences in the database that lie between and include the primer pair (Health Sciences Library System).

vii) ClustalW2 Phylogeny

ClustalW2 is one of the bioinformatics tool for the multiple sequence alignment of the DNA or protein sequences. It can easily align the sequences and help to create a phylogenetic tree online (Higgins *et al.*,1988). The phylogenetic tree formats available are CLUSTAL, Distance matrix and

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

NEXUS. ClustalW2 Phylogeny can be accessed by http://www.ebi.ac.uk. Phylogenetic tree

The phylogenetic tree also known as a dendrogram is a diagram that describes the evolutionary relationship among organisms or a group of organisms. The branches in a phylogenetic tree indicate the evolutionary relationship. The split led point is called a branch point and it represents the single lineage, which evolved into a distinct new one. A lineage that evolved early from the root and which remains unbranched is called a basal taxon. The two lineages arising from the same branch point are called sister taxa. A branch having more than two lineages is called a polytomy (courses.lumenlearning.com)

2. Materials and Methods

2.1 Search string establishment

- The first step in designing a PCR primer is the construction of a specific database consisting all the nucleotide sequence that are available at GenBank from national center for biotechnology Information (NCBI) (André O.S Lima and Sérgio P.S Garcês,2006).
- A good search string is established by using correct keywords and the Boolean operators.
- The Boolean operators 'OR' and 'AND' are used to connect the word to be obtained in the result whereas the Boolean operator 'NOT' is used to exclude the entries from the results.
- The taxonomic ID of the organism helps to retrieve the nucleotide sequences of the gene present in that particular organism.
- A search string is established to obtain the nucleotide sequences of Vascular Endothelial Growth Factor A (VEGFA) gene from the GenBank. (URL-https://www.ncbi.nlm.nih.gov).
- The search string created is:

"VEGFA AND "Homo sapiens" [porgn: -txid9606]".

- The 'VEGFA' and 'Homo sapiens' are the key words, connected by the Boolean operator'AND'.
- The taxonomic ID for Homo sapiens is 9606, 'porgn' indicate primary organism.

Kusources E				SIGN IN TO NUBI
Nucleotide	Nucleotide VEGFA AND "Homo sapiens"[porgn_txid9606]		Search	
	Create alert Advanced			Help
Species	Summary + 20 por page + Sort by Default order +	Send to: -	Filters: Manago Filtors	
Animals (207)				
Gustomize	Items: 1 to 20 of 257		Find related data	-
Molecule types genomic DNA/RNA (32)	se First is Prov. Page 1 of 13. Ne	ot > Last >>	Database Select	*
(IRNA (184) Customize	Homo sepiens vascular endothelial growth factor A (VEGFA), transcript varient 1, mRN 3,660 bb linear mRNA	A		
Source databases	Accession: NM_001171623.2 (3): 1969138140			
NSDC (GenBank) (52)	Litotein LiubMed Laxonomy		Search details	
RetSeq (205) (Sustamize	GonBank FASTA Graphics		VEGFA[All Fields] AND '	Homo sapiers"
Sequence Type	Homo sepiens vascular endothelial growth factor A (VEGFA), transcript variant 7, mRN	A	61 03	
Nucleotide (267)	2. 3,471 bp linear mRNA			
200000000000000	Accession: NM_001171629.2 GI: 1969138145			
Sequence length	Protein PubMed Taxonomy		Search	Same revenue
Gustom range	GenBank LASTA Graphics		And a second sec	Sed more.
Release date				
Custom range	Homo sapiens vascular endothelial growth factor A (VEGFA), transcript variant 5, mRN	A	Recent activity	
Revision date	3 3,502 bp linear mRNA			Tura Off Clear
Custom range	Accession: NM_001171027.2 GI: 1969138144			ions Trorant (357)
	Fidem Fidered _aconomy		C VEGLAMANT FILMED SAL	Nucleotide
Clear all	CRAHONK LASIA LZAPRICS		-	Contractor and the second second
ichow addroonal others	Home canings years lay and the list arouth (actor & (///CEA)) transactor years of a RN	A	(vegta gene) AND "Hom (234)	o sepiens (porgn) Nucleation
LINE PARAMETER	4 2 600 ks lister s DNA		(FOA)	
	 3,009 Dp inteal miking Automatica MM, 001171624 2 (20. 1060138149) 		Q vegfa gene (1992)	Birettestia
	Protein PubMed Taxonoriw			NUClection
	GenBank FASTA Graphics		Q VEGFA AND Homo sape	ons[porgn] (257) Nucleating
	Homo sapiens vascular endothelial growth factor A (VEOFA), transcript variant 8, mRN	A	Q VEGEA AND Homo sapi	ens[All Fields] (275) Nucleation

Fig 2- The nucleotide sequence search result from GenBank obtained by using the search string.

2.1.2 Data bank construction using the sequences retrieved from GenBank

• To retrieve the nucleotide sequence of the VEGFA gene, the search string is entered in the box, at the NCBI homepage.

- The 'nucleotide' option is selected from the search list.
- The nucleotide sequence considered of interest is related to the chosen gene that belongsto the desired genus and has the complete DNA Sequence (André O.S Lima and Sérgio P.O Garcês, 2006).
- The Accession number helps to identify the latest submission.
- The nucleotide sequences can be selected by filling the text box next to the entries.
- After selecting the desired nucleotide sequences, the sequences need to be copied in fastaformat into a text document, which forms the specific data bank.
- The fasta sequence can be obtained by selecting 'FASTA' Option from the display listlocated at bottom of the specific entries.
- The 'send to text' option will provide the fasta sequence in a ready to copy format.
- Four nucleotide sequence of VEGFA gene are selected for data bank construction. Theyare,
- Homo sapiens vascular endothelial growth factorA (VEGFA), transcript variant 1, mRNA.
- Homo sapiens vascular endothelial growth factorA (VEGFA), transcript variant 4, mRNA.
- Homo sapiens vascular endothelial growth factorA (VEGFA), transcript variant 3, mRNA
- Homo sapiens vascular endothelial growth factorA (VEGFA), transcript variant 10, mRNA
- All the nucleotide sequence is obtained in FASTA format.

2.1.3 Multiple Sequence Alignment using ClustalW

- The multiple sequence alignment of the nucleotide sequences is obtained by using the bioinformatics tool ClustalW.
- Multiple sequence alignment is the method of aligning three or more biologicalsequences having similar length.
- Clustal is a series of computer programs used in bioinformatics.
- ClustalW is a multiple sequence alignment program for DNA or protein sequence.
- The Clustal W can be accessed via the web server of the DNA Data Bank of Japan (URLhttps://www.clustalw.ddbj.nig.ac.jp).
- 'FASTA' is chosen as the output format. The 'DNA' is selected for sequences.
- The nucleotide sequence was uploaded in the box provided in the homepage.
- The data bank construction of the desired nucleotide sequence makes it easier to uploadall the sequences.
- The 'send to clustal w' option was selected and the results are displayed in a new windowfrom which it is downloaded.

DDBJ					Ja pane su Search
About DDB3	How to Use	Report/Statistics	FAD	Contact Us	
HDME 2 Search and Analysis 2 Clu	stallV				
ClustalW					Help
Version					
 2.1 (Latest version) 1.83 (Modified by Dr. Kirll Kryukov) 					
Sequences					
DNA vesquences Tile Ublad: Choose File No file of or COPY & PASTE	10500				

Fig 3- The ClustalW multiple sequence alignment program at DNA Data Bank of Japan (https://www.clustalw.ddbj.nig.ac.jp).

2.1.4 Consensus Sequence Determination using EMBOSS Cons

- The consensus sequence is a unique nucleotide sequence representing the most conserved nucleotides among the aligned genes (André O.S Lima and Sérgio P.S Garcês, 2006).
- The consensus sequence obtained will be used for further BLAST analysis and for thepriming position selection.
- The gaps will be present in the consensus sequence, if the aligned nucleotide sequences donot attain an identity threshold
- The pattern of the consensus sequence is determined by the degree of identity of the alignedgenes.
- The nucleotide sequences having a high identity will give a representative and consistent consensus sequence.
- The sections of the consensus sequence having high identity are called a conserved domainor region.
- The conserved regions or domains are sections where the DNA is least modified during evolution.
- The consensus sequence is obtained from the multiple sequence alignment of VEGFA geneusing EMBOSS cons.
- EMBOSS cons is a bioinformatics tool from The European Bioinformatics Institute thatcreates a consensus sequence from a multiple alignment.
- EMBOSS cons homepage is accessed using the URL-https:// www.ebi.ac.uk.
- 'DNA' is selected from the drop down, list for the sequences.
- The multiple sequence alignment from clustal W was uploaded in the box entitled 'Enter orpaste a nucleotide alignment in any supported format.
- Choose submit, under the 'submit your job' on step 3 to obtain the results.
- The result displayed on the new window can be downloaded by selecting 'Download outputfile'.

	Help & Documentation	Bioinformatics Tools FAQ		🗭 Feedback	<share< th=""></share<>
ols > Multiple Sequence Alig	nment > EMBOSS Cons				
Jons create	s a conse	nsus sequence			
MBOSS Cons creates a cor	sensus sequence from a	a multiple alignment (MSF, PIR, CLUS	TAL, etc).		
nportant note: This tool car	i align a maximum file si: 	Ze of ZIVIB.			
STEP 1 - Enter your input					
DNA					*
Enter or paste a protein align	ment in any supported for	ormat:			
			line and a second of the	10	<u>6</u>
Upload a Tile: Choose File N	o filo choson		use a example sequence Clear	sequence see more ex	cample inplits

Fig 4 - The EMBOSS Cons tool from The European Bioinformatics Institute for finding consensus sequence.(https://www.ebi.ac.uk.)

2.1.5 Basic Local Alignment Search Tool(BLAST)

- The best sections for the PCR primer design are located at the conserved region in the consensus sequence (André O.S Lima and Sergio P.S Garcês.2006).
- To determine the conserved region, the whole length consensus sequences is compared to those

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

nucleotide sequences deposited at the GenBank.

- The regions having highly aligned sequences are considered as the best regions for the primer design.
- The Basic Local Alignment Search Tool (BLAST) can be accessed from the NCBI website using the URL https://blast.ncbi.nlm.nih.gov./.
- The nucleotide blast or 'blast n' option is selected as the nucleotide sequences are compared.
- The consensus sequence from EMBOSS cons are uploaded in 'Enter Query Sequence'.
- The primer is designed for VEGFA gene in Homo sapiens, so that the search is limited to
- Homo sapiens by entering 'Homo sapiens 'in Entrez Query box.
- The 'BLAST' option at the bottom page is selected to receive the results.
- The results will be displayed within seconds to minutes, depending on how busy the NCBI server is at that time.
- The results will be displayed graphically and as a text.
- The graphical Illustration facilitate the 'macro' view of the alignment.
- The 'macro' view gives a clear picture of the location of the sequences deposited at the GenBank, which exhibit a homology to the consensus sequence.
- In the BLAST result the query sequence or the consensus sequence appear as a wider red bar with the nucleotide numbers and the GenBank sequences are represented ads narrow bars.
- The color of bars depends on the size and degree of identity.
- From the BLAST result, two sections of the consensus sequence (query sequence), where the largest number of high identity sequences aligned are selected.
- The selected sections should be at least 30-40 bp long and must be 150-200 bp apart from each other.
- After determining the best sections, the nucleotide sequences from that specific consensus block is selected for primer design.

	From	New columns added to the Description Table Citid Static Column' or Manage
Or, upload file	Choose File No file chosen	Columns'.
Job Title		
	Enter a descriptive title for your BLAST search 😧	
Aign two or m	cne sequences 😧	
Choose Sear	rch Set	
Database	Standard databases (nr etc.): RNA/ITS databases Genomic + transcript databases Betacoronav	itus
	Nucleotide collection (nrimt) v 💡	
Organism	Friter contarium name or it_controletions will be successed	
Optional	Enter organism common name, binomial, or tax id. Only 20 too taxa will be shown 0	
Exclude	Models (XMXP) Unculturedienvironmental sample sequences	
Limit to	Sequences from type material	
Entrez Query	homo sapiens Bauting Create custom database	
Optional	Entire an Entirez query to limit search 🕖	
Program Sel	ection	
Optimize for	Highly similar sequences (megublast) Highly similar sequences (discordiguous megatilast) Somewhat similar sequences (bitam) Torone BLKS* equivative @	

Fig 5 - The nucleotide BLAST (blast n) is used to compare the consensus sequence to the nucleotide sequences at GenBank. (https://blast.ncbi.nlm.nih.gov./)

2.1.6 Primer Designing using Primer3 Plus

- The efficacy and the sensitivity of a polymerase chain reaction depends on the efficiency of the primers used (Kamel A. Abd-Elsalam,2003; He et al.,1994).
- The capability of an oligonucleotide primer to serve as a PCR primer depends on a) the kinetics of the association and dissociation regarding the primer-template duplexes at the annealing and extension temperatures in the reaction b) the location of the mismatched nucleotides c) the efficiency of the

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

DNA polymerase enzyme which there by identify and extend a mismatched duplex.

- The primers that are complementary to the target sequence must fulfill certain criteria like length of primer, the GC content, melting and annealing temperature, the 3' and 5' end specificity.
- Primers must be of at least 18 nucleotides in length in order to minimize the chances of having problems with the secondary hybridization site present on the vector or insert.
- Primers having long runs of a single base is avoided.
- The primers with the optimal melting temperature in the range of 52-58°C yield better resultas that of primers with the lower melting temperature.
- Primers having the melting temperature, greater than 65°C are avoided as it leads to secondary annealing.
- GC content in the primers provide the information about the strength of annealing.
- The primers should have the GC content in the range 45-60 percent.
- The primer of VEGFA gene is designed using the Primer3 Plus. The Primer3 Plus tool can beaccessed using the URL- https://www.primer3plus.com.
- The parameters like primer size primer melting temperature, GC content are adjusted to optimum values at the 'General Settings' tab from the homepage.
 Primer Size: Min-18 Opt- 20 Max-30
 Primer Tm: Min-52°C Opt-60°C Max-63°CPrimer GC %: Min-45 Opt-50 Max-60
- The 'Load Settings' option is selected to implement changes in default values.
- The selected nucleotide sequence from the conserved region is copy pasted to the boxentitled 'Paste template sequence below' in the homepage.
- The 'Pick Primers' button on the top right corner is selected to obtain the Forward andReverse primer from the submitted sequence.

Primer3Plus			More	Source Code
pick primers from a DNA sequence	Help	About		
Load server settings: Default Task: generic v	 Activate Settings 	Select primer po sequence. Optio included/exclude	irs to detect the given temp nally targets and ed regions can be specified.	Pick Primers Reset Fo
Main General Settings A	dvanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence
Primer GC% <u>Mint</u> 45 Concentration of monovalent cations:	<u>Opt</u> 50.0 50.0 <u>Anna</u>	Max: 80 ealing Oligo Concen	tration: 50.0	
Aispriming/Repeat Library: NONE	Ins [<u>con</u>	simator of diviria		
Load and Save To upload or save a settings file from '	our local computer, c	hoose here:		
		Cause Cattingen		

Fig 6- The General Settings tab at Primer3 Plus. The parameters like primer size, melting temperature and GC content are adjusted to their optimum range to obtain highly efficient primers.

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

			<u>intorem</u>	<u>Source Code</u>	
ick primers from a DNA sequen	ee		Help	About	
oad server settings: Default	Activate Setting:	S Select primer pa sequence. Optic included/exclud	urs to detect the given temp nally targets and ed regions can be specifica	Pick Primers Re	set Forr
Main General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Advanced Sequence	
equence Id:					
aste template sequence below	Or upload sequence :	file: Choose File No	file chosen	Upload Filo	
Aark selected region: <> []]	() Clear		Ĺ	Save Sequence	-19
Mark selected region: <> [1]	() Cloar		L	Save Sequence	-78
Mark selected region: <> []	() Cloar	> 1	L	Save Sequence	-78
Mark selected region: <> [1] xeluded Regiona: < argets:	() Cloar	>> 1	L	Save Sequence	
Mark selected region: <> [1] xcluded Regiona: < argets: included Region: rimer overlap positions: -	() Cloar	>>]	L	Save Sequence	
Mark selected region: <> [1] xcluded Regions: < argets: 1 heluded Region: 1 rimer overlap positions: - air OK Region List:	() Cloar	>=]]]	l	Save Sequence	18
Mark selected region: <> [1] ixcluded Regions: <	() Cloar	>>]] ; ization probe	Pick right	Savo Soquonco	

Fig 7- Primer3 Plus homepage. The nucleotide sequences are uploaded after setting the corresponding parameters.(https://www.primer3plus.com)

2.1.7 In silico testing

- The term 'in silico' refers to the biological experiments that are conducted on a computer orby means of computer stimulation.
- The polymerase chain reaction of the primer designed is done virtually by using the 'In silicoPCR' tool from the UCSC Genome Browser.
- The UCSC Genome Browser can be accessed by the URL- https://www.genome.ucsc.edu/.
- The 'In silico PCR' tool can be selected from the home page itself.
- The primer was designed for the VEGFA gene in Homo sapiens, so 'Human' is selected from the drop-down list for Genome.
- The latest assembly [Dec.2013 (GRCh38/hg38)] is selected.
- The Forward and Reverse primer obtained from the Primer3 Plus are entered into respectiveboxes.
- On clicking the 'Submit' button, the amplified sequence will be displayed in a new window.

SANTA CAUZ Genomics	Genome Browser
👘 Genomes Genome Browser Tools Mirrors	Downloads My Data Projects Help About Us
	Our tools
	Genome Browser Instanctively valuating periodic data COVID-19 Research use the SARS-GW 2 genome tomore and explane comovins distancts use the SARS-GW 2 genome Browser and explane comovins distancts BLT replay alignments download data from the Genome Browser distances Valiaria Annotation Integration get functional differentiations Data Integrator construct agence are from the Genome Browser distances Genome Browser in a Box (SBBS) run the Genome Browser annot large or server Insilice GR2 replay align CR2 replay

Fig 8- UCSC Genome Browser homepage.(https://www.genome.ucsc.edu/)

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

Container Containe aromaen ins		my crass intojouse intoj	A PROCESS OF	
UCSC In-Silico PCR				
Genome:	Assembly:	Target	Forward Primer.	Reverse Primer.
Humen 👻	Dec. 2013 (GRCh38/hg38) V	genome assembly v		suomt
Max Product Size: 4000	Min Porfect	Match: 15	Min Good Match: 15	Flip Reverse Primer:
About In-Silico PCR				
In-Silico PCR searches a sequence database	with a pair of PCR primers, using a	an indexing strategy for fast per	formance. See an example video on ou	ar YouTube channel.
Configuration Options				
Reverse Primer - On the opposite strand from Max Product Size - Maximum size of amplifie Min Perfect Match - Number of bases that ma Min Good Match - Number of bases on 3° and Fije Reverse Primer - Invert the seguence or Output When successful the search returns a seguence	the low and primer. Minimum long d region. Ich exactly on 3' and of primers. N i of primers where at least 2 out of for of the reverse primer and comp on output file in fasta format conta	ph of 15 bason. Imimum match sizo is 15. 3 bases match. Jemont it. Ining all sequence in the datab	see that iss between and include the notice	inser nair. The first a header describes the renion in the
database and the primers. The fasta body is cr >chr22i31000551+31001600 TAACASATTGATGATGATG	apitalized in areas where the prime	er sequence matches the catab	ase sequence and in lower-case elsew	here. Here is an example from human:
THACAPATTCATCATCATCAAATOOCeggtserrassesteesse	eten			
tatearascteasettttcraeseerteatesteascraetease	tang			
gattigtigtotgtaagggettiggattatictattigagaaattit	gtta			
ageaapctcctcct:agacagagaaacas:agccgtcacaggasgc	ааад			
auattgetttacttttaagstgaatcc3gaaccc3gatgtc2g5g	TC-C			

Fig 9- The In-silico PCR tool amplify the nucleotide sequence using the primers designed.

2.1.8 Construction of the phylogenetic tree using ClustalW2 phylogeny

- A phylogenetic tree also known as a dendrogram describes the evolutionary relationship among organisms.
- The phylogenetic tree can be constructed by two methods: distance and character based methods (Murat KASAP, 2005).
- In distance based method, Neighbor Joining method is employed. In neighbor joining method, each pair is evaluated for joining and the corresponding sum of all branches length is calculated.
- The pair having the smallest sum is considered as the closest neighbors and joined.
- A new branch is inserted between them and the rest of the tree and the branch length is recalculated.
- The Vascular Endothelial Growth Factor A(VEGFA) gene is present in Homo sapiens as well as in organisms like Mus musculus (house mouse), Daniorerio (Zebra fish), Bufobufo(Common toad), Xenopus laevis (African clawed frog) and Oxyura jamaicensis (Ruddy duck).
- The phylogenetic tree of VEGFA gene is constructed among these six organisms.
- The multiple sequence alignment of the nucleotide sequences from the GenBank was obtained using ClustalW.
- In case of Homo sapiens, the PCR amplified sequence is used.
- The phylogenetic tree is constructed using ClustalW2 phylogeny.
- The bioinformatics tool can be accessed by the URL-https://www.ebi.ac.uk.
- The multiple sequence alignment from the ClustalW is uploaded in the box entitled 'Enter your multiple sequence alignment'.
- 'Distance Matrix' is selected as tree format and 'Neighbor-joining' for the clustering method.
- The 'Submit' button at the bottom of the page is chosen to obtain the result.
- The phylogenetic tree will be displayed in a new window.

a multiple seq	tes access to pl uence alignmen	hylogenetic tree generati It please use one of our N	ion methods from MSA tools.	the ClustaW2 package. Ple	ase note this is NOT a multiple sequ	uence alignment tool.	To perform
Please	Note						
From Decemb any concerns.	er 1st this tool w	will be renamed 'Simple F	Phylogeny', but of	therwise all existing function	ality will remain. Please contact us vi	ia the feedback link if	you have
STEP 1 -	E <mark>n</mark> ter your multii	ple sequence alignment					
Enter or past	e a multiple sequ	uence alignment in any s	supported format:				
							4

Fig 10- The phylogenetic tree is constructed using ClustalW2 phylogeny. (https://www.ebi.ac.uk.) The multiple sequence alignment of the nucleotide sequences from the six organisms are used in distance based method.

3. Result and Discussion

3.1 Data Bank construction using the sequences retrieved from GenBank

The VEGFA gene nucleotide sequences were retrieved from GenBank using the search string. Adata bank containing the four nucleotide sequences in fasta format was created.

	>NM 001171626 2 Homo samiens vascular endothelial growth factor & (VEGEA)
	transcript variant 4 mDNA
	TGALLTGUTTTTTGGGGGTGALLGULGGAGLGUGGGGGGGGGG
	GTCGCGCTGACGGACAGACAGACAGACAGACACACCGCCCCAGCTACCACCTCCTCCCCGGCGGC
	GGACAGTGGACGCGGCGGCGAGCCGCGGGCAGGGGCCGGACCCCGCGCGCGGGGGG
	CGGGGCTCGCGCGTCGCACTGAAACTTTTCGTCCAACTTCTGGGCTGTTCTCGCTTCGGAGGAGCCGTG
0	STCCGCGCGGGGAAGCCGAGCCGAGCGGAGCCGCGAGAAGTGCTAGCTCGGGCCGGGAGGAGCCGCAGC
	CGGAGGAGGGGGGGGGGGAGGAGAGAGAGGAGGGGGGGG
	GCCGGGCTCATGGACGGGTGAGGCGGCGGTGTGCGCAGACAGTGCTCCAGCCGCGCGCCCCCAGGCC
	CTGGCCCGGGCCTCGGGCCGGGGA3GAAGAGTAGCTCGCCGAGGCGCCGAGGAGAGCGGGCCGCCCACA
	GCCCGAGCCGGAGAGGGAGCGCGGCCCGGCCCCGGTCGGGCCTCCGAAACCATGAACTITCTGCT
	GTCTTGGGTGCATTGGAGCCTTGCCTGCTGCTCCACCCCCACCAAGTGGTCCCAGGCTGCACCC
1	ATGGCAGAAGGAGGAGGGCAGAATCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGCGCAGCTACT
	SCCATCCAATCGAGACCCTGGTGGACATCTTCCAGGAGTACCCTGATGAGATCGAGTACATCTTCAAGCC
2	ATCCTGTGTGCCCCTGATGCGATGCGGGGGCTGCTGCAATGACGAGGGCCTGGAGTGTGTGCCCACTGAG
	GAGTCCAACATCACCATGCAGATTATGCGGATCAAACCTCACCAAGGCCAGCACATAGGAGAGAGA
	TCCTACAGCACAAAAATGTGAATGCAGACCAAAGAAAGATAGAGCAAGACAAGAAAATCCCTGTGGGCC
	TTGCTCAGAGCGGAGAAAGCATTTGTTTGTACAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACACA
	GACTCGCGTTGCAAGGCGAGGCAGCTTGAGTTAAACGAACG
	GAGCCGGGCAGGAAGGAAGGAGCCTCCCTCAGGGTTTCGGGAACCAGATCTCTCACCAGGAAAGACTGATA
	CAGAACGATCGATACAGAAACCACGCTGCCGCCACCACCATCACCATCGACAGAACAGTCCTTAATCC
1	AGAAACCTGAAATGAAGGAGAGAGAGACTCTGCGCAGAGCACTTTGGGTCCGGAGGCGAGACTCCGGCG
	GAAGCATTCCCGGGCGGGTGACCCAGCACGGTCCCTCTTGGAATTGGATTCGCCATTTTATTTTTCTTGC
	TGCTAAATCACCGAGCCCGGAAGAFTAGAGAGTTTTATTTCTGGGATTCCTGTAGACACCCCCCCC
	TACATACATTTATATATATATATATATATATATAAAAAATAAATATCTCTATTTTATATATAAAAAT
	ATATATATTCTTTTTTTTTTTTTTTTTTTTTTTTTTTT
	TGGACTTGAGTTGGGAGGGGAATGITCCCACTCAGATCCTGACAGGGAAGAGGAGGAGATGAGAGACTCT
	GCCATGATCTTTTTTTGTCCCACITGGTGGGGCCAGGSTCCTCTCCCCTGCCCAGGAATGTGCAAGGCC
	AGGCATGGGGGCAAATATGACCCAGTTTTGGGAACACCCACC
	ACCCCAGGTCAGACGGACAGAAAGACAGATCACAGGTACAGGATGAGGACACCGGCTCTGACCAGGAGT
	TTGGGGAGCTTCAGGACATTGCTGTGCTTTGGGGATTCCCTCCACATGCTGCACGCGCATCTCBCCCCCA

Fig 11- Data bank containing VEGFA gene nucleotide sequences in fasta format

3.2 Multiple sequence alignment using ClustalW

The bioinformatics tool ClustalW was used to obtain the multiple sequence alignment. The output format was chosen as FASTA.



Fig 12- The multiple sequence alignment from ClustalW.

3.3 Consensus sequence determination using EMBOSS Cons

The multiple sequence alignment file was uploaded in the EMBOSS Cons to determine the consensus sequence. The consensus sequence is highly aligned sequence.

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

>EMBOSS0001

GCGGAGGCTTGGGGCAGCCGGGTAGCTCGGAGGTCGTGGCGCTGGGGGCTAGCACCAGCG CTCTGTCGGGAGGCGCAGCGGTTAGGTGGACCGGTCAGCGGACTCACCGGCCAGGGCGCT CGGTGCTGGAATTTGATATTCATTGATCCGGGTTTTATCCCTCTTCTTTTTCTTAAACA TGAATCGGGCCGACGGCTTGGGGAGATTGCTCTACTTCCCCAAATCACTGTGGATTTTGG GACGGGGTCAGAGAGAGCGCGCGGGGCGTGCGAGCAGCGAAAGCGACAGGGGCAAAGTGAG TGACCTGCTTTTGGGGGTGACCGCCGGAGCGCGGCGTGAGCCCTCCCCCTTGGGATCCCG CAGCTGACCAGTCGCGCTGACGGACAGACAGACAGACACCGCCCCCAGCCCCAGCTACCA CGTCCAACTTCTGGGCTGTTCTCGCTTCGGAGGAGCCGTGGTCCGCGCGGGGGAAGCCGA GCCGAGCGGAGCCGCGAGAAGTGCTAGCTCGGGCCGGGAGGAGCCGCAGCCGGAGGAGGG GGAGGAGGAAGAAGAGAAGAAGAGAGAGAGGGGGCCGCAGTGGCGACTCGGCGCTCGGAA TOCCOAGGCCCTGGCCCGGGCCTCGGGCCGGGGAGGAGAGTAGCTCGCCGAGGCGCGA GGAGAGCGGGCCGCCCACAGCCCGAGCCGGAGAGGGAGCGCGAGCCGCGCCCCGG TCGGGCCTCCGAAACCATGAACTTTCTGCTGTCTTGGGTGCATTGGAGCCTTGCCTTGCT GCTCTACCTCCACCATGCCAAGTGGTCCCAGGCTGCACCCATGGCAGAAGGAGGAGGGCA GAATCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGCGCAGCTACTGCCATCCAAT CGAGACCCTGGTGGACATCTTCCAGGAGTACCCTGATGAGATCGAGTACATCTTCAAGCC ATCCTGTGTGCCCCTGATGCGATGCGGGGGGCTGCTGCAATGACGAGGGCCTGGAGTGTGT GCCCACTGAGGAGTCCAACATCACCATGCAGATTATGCGGATCAAACCTCACCAAGGCCA TAGAGCAAGACAAGAAAAaaaatcagttcgaggaaagggaaagggcaaaaacgaaagcg nnnnnnnnnnnnnnnnnnTCCCTGTGGGCCTTGCTCAGAGCGGAGAAAGCATTTGTT TGTACAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACACAGACTCGCGTTGCAAGGC GAGGCAGCTTGAGTTAAACGAACGTACTTGCAGATGTGACAAGCCGAGGCGGTGAGCCGG GCAGGAGGAAGGAGCCTCCCTCAGGGTTTCGGGAACCAGATCTCTCACCAGGAAAGACTG ATACAGAACGATCGATACAGAAAACCACGCTGCCGCCACCACCATCACCATCGACAGAA TTGGAATTGGATTCGCCATTTTATTTTTCTTGCTGCTAAATCACCGAGCCCGGAAGATTA TTCTTTTTTTAAATTAACAGTGCTAATGTTATTGGTGTCTTCACTGGATGTATTTGACTG CTGTGGACTTGAGTTGGGAGGGGAATGTTCCCACTCAGATCCTGACAGGGAAGAGGAGGA GATGAGAGACTCTGGCATGATCTTTTTTTTTTTTGTCCCACTTGGTGGGGCCAGGGTCCTCTCC CCTGCCCAGGAATGTGCAAGGCCAGGGCATGGGGGGCAAATATGACCCAGTTTTGGGAACA CCGACAAACCCAGCCCTGGCGCTGAGCCTCTCTACCCCAGGTCAGACGGACAGAAAGACA GATCACAGGTACAGGGATGAGGACACCGGCTCTGACCAGGAGTTTGGGGAGCTTCAGGAC ATTGCTGTGCTTTGGGGATTCCCTCCACATGCTGCACGCGCATCTCGCCCCCAGGGGCAC TGCCTGGAAGATTCAGGAGCCTGGGCGGCCTTCGCTTACTCTCACCTGCTTCTGAGTTGC

Fig 13- The consensus sequence from EMBOSS Cons.

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

3.4 Basic Local Alignment Search Tool (BLAST)

The consensus sequence was compared to that of other nucleotide sequences at GenBank using Nucleotide BLAST. The conserved domains were identified from the BLAST result was used to select the nucleotide sequence for primer design.

to hover to see the title 🔥 click to show alignments	Alignment Scores ■ < 40 ■ 40 - 50 ■ 50 - 80 ■ 80 - 200 ■ >= 200 3
100 sequences selected 🔞	Distribution of the top 221 Blast Hits on 100 subject sequences
	1 00000 1 0000 1 000 1000 1000 1000 10

Fig 14-The Graphic summary of BLAST result.

Douvolo	nd M	ConBonk C	rephice a .					Next + Brevieus @Descriptions
S Downic	<u>aa</u> ~	Gendank G	raphics Sort	by: E value		· (A) Amon o oui	int verient 1 m DNA	• Next • Previous • Descriptions
Sequence	ID: NM	001025366.3	Length: 366	0 Number of	f Matches: 2	AJ, transcri	pr variant 1, mixing	
See 1	more t	tle(s) ¥ See a	II Identical Pro	oteins(IPG)				2
Range 1:	1582 to	3660 GenBank	Graphics			Voxt N	Match A Provious Match	Related Information
Score 3840 bits	(2079)	Expect 0.0	Identities 2079/2079(1	.00%)	Gaps 0/2079(0%	Strand) Plus/P	lus	<u>Gene</u> - associated gene details <u>PubChem BioAssay</u> - bioactivity
Query 1	582 T	CCTGTGGGGCCTT	GCTCAGAGCGGAG	GAAAGCATTTG	TTTGTACAAGAT	CCGCAGACGTG	1641	Genome Data Viewer - aligned
Sbjct 1	582 T	CCTGTCGGCCTT	GCTCAGAGCGGAG	AAAGCATTTG	TTTGTACAACAT	CCGCAGACGTG	1641	genomic context
Query 10	642 T/	AATGTTCCTGCA.	AAAACACAGACTO	GCGTTGCAAG	GCGAGGCAGCTT	GAGTTAAACGA	1701	
Sbjct 10	542 Ť/	AATGTTCCTGCA	AAAACACAGACTO	GCGTTGCAAG	GCGAGGCAGCTT	GAGTTAAACGA	1701	
Query 1	702 A	GTACTTGCAGAT	STGACAAGCCGAG	GCGGTGAGCC	SGGCAGGAGGAA	GGAGCCTCCCT	1761	
Sbjct 17	702 A	GTACTTGCAGAT	STGACAAGCCGAG	GCGGTGAGCC	GGGCAGGAGGAA	GGAGCCTCCCT	1761	
Query 1.	762 G		CAGA CICICAG	CAGGAAAGAC	IGATACAGAACG		1821	
Sbjet 13	762 CI	\GGGTTTCGGG//\	CAGATCTCTCAC	CAGGAAAGAC	TGATACAGAACG	ATCGATACAGA	1821	
Query 18	822 A						1881	
Sbjct 18	822 AJ	ACCACGCTGCCGC	ACCACACCATCA	ACCATCGACAG	AACAGTCCTTAA	TCCAGAAACCT	1881	
	882 64	AATGAAGGAAGA	GAGACTCTGCGG	AGAGCACTTT	GGGTCCGGAGGG	CGAGACTCCGG	1941	

Fig 15- The highly aligned sequence selected based on BLAST result.

3.5 Primer designing using Primer3 Plus

Two blocks of conserved region, that are 30-40 bp long and 150-200 bp apart from each other are selected. The nucleotide sequence from the conserved region was chosen for primer designing. The selected nucleotide sequences, entered in Primer3 Plus after setting the parameters.

Prim	er3Plus				Prin	ier3Manage	r	Help	
pick primer	s from a DNA seque	ice			Abo	ut		Source Co	de
Task: Det	ection 🗸	Select prime included/exc	r pairs to detect th luded regions can	e given template sequen be specified.	ce. Optiona	lly targets and	P	lick Primers	Reset Form
Main	General Settings	Advan	ced Settings	Internal Oligo	Pena	lty Weights	Sequ	aence Quali	ty
Sequence Id									
Paste source	sequence below	Or uple	ad sequence fil	le: Choose File No	ile chosen		Upload	l File	
Mark select	ed region: <> []	{} Cle	ar				Save Se	quence	£
Mark select	ed region: <> [1] g <u>ions:</u> < [{} Cle	ar	>			Save Se	quence	æ
Mark select Excluded Re Targets:	ed region: <>)[] g <u>ions:</u> < [[]		ar	<pre></pre>			Save Se	quence	æ
Mark select Excluded Re Targets: Included Re	ed region:	() () () () () () () () () () () () () (ar	۱ ۱			Save Se	quence	£
Mark select Excluded Re Included Re Pick left j or use left p	ed region:	{} [] []	ar Pick hybrid (internal oligo)	>]]]] ization probe		☑ Pick righ below (53	Save Se t primer o	quence] or use right p ssite strand).	rimer

Fig 16- The nucleotide sequences are uploaded in Primer3 Plus.

Total five pairs of Forward and Reverse primer pairs was obtained, from the first pair was chosen for In silico testing.

About Source Code pick primers from a DNA sequence Source Code VARNING: Numbers in input sequence Introduction VARNING: Numbers in input sequence Source Code VARNING: Numbers in input sequence Source Code Statt 18 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 Statt 18 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 Statt 18 Length: 20 bp Tm: 60.4 °C GC: 55.0 % ANY: 3.0 SELF: 3.0 Preduct Size: 101 bp Pair Any: 6.0 Pair End: 1.0 SelF: 3.0 SelF: 3.0 Send to Primer@Manager] Reset Form	Primer3F	Plus	Primer3Manager	Help
SARNING: Numbers in input sequence were deleted. Intrecognized base in input sequence < Back Print: Left Primer 1: Print: Sequence: ITIGCTCAGAGEGGAGAAAG Start: 18 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 Skingt Primer 1: Primer_R Sequence: OTCOCTOGOGCAACTCAGAA	pick primers from a L	DNA sequence	About	Source Code
Intercognized base in input sequence <pre></pre>	ARNING: Numbers	in input sequence were deleted.		
E Back Pair 1: Pair 1: Pair 2:	nrecognized base in i	input sequence		
Pair 1: I cff Primer 1: Primer_T Sequence: ITGCTCAGAGGGGAGAAG Start: 18 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 I kight Primer 1: Primer_R	< Back			
Image: Sequence: Primer_f Sequence: ITGCTCAGAGCGGAGAAAG Start: I8 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 If Right Primer I: Primor_AR	Pair 1:			
Sequence: TTGCTCAGAGGGGAGAAAG Start: 18 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 © Kight Primer 1: Primer_R	Left Primer 1:	Primer_F		
Start: 18 Length: 19 bp Tm: 59.4 °C GC: 52.6 % ANY: 7.0 SELF: 2.0 Start: 18 Length: 20 bp Tm: 60.4 °C GC: 55.0 % ANY: 3.0 SELF: 3.0 Product Size: 101 bp Pair Any: 6.0 Pair End: 1.0 Self: 3.0 Pair End: 1.0 Sendit 0 Primer3Manaer Reset Form	Sequence:	TTGCTCAGAGCGGAGAAAG		
Image: Sequence: OTCTCCTGOGGAACTCAGAA Sequence: OTCTCCTGOGGAACTCAGAA Start: 118 Length: 20 bp Tm: 60.4 °C GC: 55.0 % ANY: 3.0 SELF: 3.0 Product Size: 101 bp Pair Any: 6.0 Pair End: 1.0 Send to Priver@Amager ResetForm	Start: 18	Length: 19 bp Tm: 59.4 °C GC: 52	.6 % ANY: 7.0	SELF: 2.0
Sequence: OTCTCCT000GAACTCA9AA Start: 118 Length: 20 bp Tm:: 60.4°C GC: 55.0% ANY: 3.0 SELF: 3.0 Product Size: 101 bp Pair Any: 6.0 Pair End: 1.0 SELF: 3.0 Send to PriverGMenage! ResetForm	Right Primer 1:	Primer_R		
Start: 118 Length: 20 bp Tm: 60.4 °C GC: 55.0 % ANY: 3.0 SELF: 3.0 Product Size: 101 bp Pair Any: 6.0 Pair Ead: 1.0 SELF: 3.0 Sendto PrimersManager Reset Form Sendto PrimersManager TGTTTGTTACA QueryTCCCT GCGSCCTTG CTCAGAGCGG AGAAGCATT TGTTTGTACA 1 AGATCCGCAG ACGTGQUery TGGCGGGCCT TGCCTTACTC TCACCTGCTT 01 CTGAGTTSCC CAGGAGAGCA CTGGCAGAGA CTGGCAGAGA SELF: 3.0 Pair 2:	Sequence:	GTCTCCTGGGCAACTCAGAA		
Product Size: 101 bp Pair Any: 6.0 Pair End: 1.0 Send to Primer/Manager: Reset Form: Image: Constraint of the send of the	Start: 118	Length: 20 bp Tm: 60.4 °C GC: 55	.0 % ANY: 3.0	SELF: 3.0
Send to Primer3Manager] Reset Form QueryTCCCT GTGGGCCTTG CTCAGAGCGG AGAAAGCATT TGTTTGTACA AGATCGCCAG ACGTGQuery TGGGCGGCCT TGCCTACTC TCACCTGCTT OI CTGAGTTGCC CAGGASACCA CTGGCAGATG Select all Primers Pair 2: Left Primer 2: Primer_1_F Sequence CAGACGCGAGAAGCATTTG Start: 23 Length: 20 bp Tm: 61.0 °C GC: 30.0 % ANY: 4.0 SELF: 2.0 Right Primer 2: Primer_1_R	Product Size: 101 bj	p Pair Any: 6.0 Pair End	: 1.0	
QueryTCCCT GTGGGCCTTG CTCAGAGGG AGAAAGCATT TGTTTGTACA AGATCCGCAG ACGTGQuery TGGGCGGCCT TCGCTACTC TCACCTGCTT CTGACTTGCC CAGGAGACCA CTGGCAGATG Select all Primers Pair 2: Left Primer 2: Primer_1_F Sequence: CAGAGCGGACAAGCATTG Sequence: CAGAGCGGACAAGCATTG Kight Primer 2: Primer_1_R	Send to Primer3Manag	er Reset Form		
1. AGATCCGCAG ACGTGQ:ery TGGCCGCCT TCSCTTACTC TCACCTCGTT 0.1 CTGAGTTGCC CAGGAGACCA CTGGCAGATG Select all Primer 2: Primer 1_F CfGAGCGGCGACAGCATTG Sequence: CACAGCGGCGACAGCATTG Start 23 Length: 20 bp Tm: 61.0 °C GC: 50.0 % ANY: 4.0 SELF: 2.0 Right Frimer 2: Primer_1_R	Que	ryTCCCT GTGGGCCTTG CTCAGAGCGG AGA	AAGCATT TGTTTGTACA	
OI Dispersive ChoseAsACA CloseAsACA CloseAsACA Select all Primers Pair 2: Left Primer 2: Primer_1_R	AGA	ATCCGCAG ACGTGQuery TGGGCGGCCT TCG	CTTACIC TCACCIGC <mark>IT</mark>	
Pair 2: Left Primer 2: Primer_1_F Sequence: CGAGGCGGAAAGCATTG Start: 23 Length: 20 bp Tm: 61.0 °C GC: 50.0 % ANY: 4.0 SELF: 2.0 Right Frimer 2: Primer_1_R Image: Section 2007 Section 2007 Section 2007	Select all Primers			
Left Primer 2: Primer_1_F Sequence: CAGAGCOGAGAAAGCATTTG Start: 23 Length: 20 bp Tm: 61.0 °C GC: 50.0 % ANY: 4.0 SELF: 2.0 Right Primer 2: Primer_1_R <t< td=""><td>Pair 2:</td><td></td><td></td><td></td></t<>	Pair 2:			
Sequence: CAGAGCOGAGAAAGCATTTG Start: 23 Length: 20 bp Tm: 61.0 °C GC: 50.0 % ANY: 4.0 SELF: 2.0 Right Primer 2: Primer_1_R	Left Primer 2:	Primer_1_F		
Start: 23 Length: 20 bp Tm: 61.0 °C GC: 50.0 % ANY: 4.0 SELF: 2.0 Right Primer 2: Primer_1_R <td< td=""><td>Sequence:</td><td>CAGAGCGGAGAAAGCATTTG</td><td></td><td></td></td<>	Sequence:	CAGAGCGGAGAAAGCATTTG		
Right Primer 2: Primer_1_R	Start: 23	Length: 20 bp Tm: 61.0 °C GC: 50	.0 % ANY: 4.0	SELF: 2.0
	Right Primer 2:	Primer_1_R		

Fig 17- Five pairs of primers are obtained. _



3.6 In silico testing

The first pair of primers was selected and uploaded in In silico PCR tool from the UCSC Genome Browser. The amplified sequence has 3509 bp.

l	UCSC In-Silico PCR
	> <u>zbri:atzm194441z0447</u> P4090p TECTEAGAARANAAGAAAA ATTECTETAGAA TECTEAGAGGGGGAAAGAULULULULULUARAAAATTECTETAGAAAAAAAAAAAAAAAAAAAAAAAAAAA
	TIBBAGBOAGEDATUEGAGETEGETAGECABECABECABECABECABECABECABECABECABECAB
	ereeremergeeneersertemenessertemeterstatte
	plapped paragraphic fungtion compared that the compared to a second
	trigen 20 million 10 cm 20 million 20 mill
	ipcicpscullpauptelputtelputtelputtelisticitatiagt
	tgegeaagggtgtgtctcccactcagactaggttcctgagggtag pprogramtercectecgactaggttcctgagggttathere
	gc=cagactgaggctcctaggctaggttcctttaggttcctcttagat
	pppt-companying-mapping-companying-companying-a-companying-a-companying-a-companying-a-companying-a-companying-
	atcotcoccogotcolggigacaacccoageageccoisattiticage Terrcrengengergetgegegegegegittefperraititicitageatt
	ageccapingetcopingeocopggectegitingiacettigigagen
	<pre>ccaegicatipatipatiptgstcstcattcaggigtggstcagsc ccaegicacipatipatiptgstcstcattcaggigtggstcagsc</pre>
	ternarchitecritecticapetionessist
	pur-angurancempulskylingkylingkylingkylingkarancempulskylingkylin
	agtgritocctoccageaatgticttgtggtigtgtagtaggg grigorgggaoraggetgggtigtgtgtgtgtgttagttagg
	grrramtpanarcapagggrgtattattagtrtgsstsssstrtt
	ppersonal compared to page 11 (Long type 11 at 1 a to 1 at 1 at 1 at 1 at 1 at 1
	<pre>tp=:ppp://thetpapetgactopppet.com/tpapetcic/tagoot/c</pre>
	agtimic canaditge agent regiting significant as ABB and a second se
	τιμαρασμημιτική μεταγραφή μεταγραφή μεταγραφή μεταγραφή μεταγραφή μεταγραφή μεταγραφή μεταγραφή μεταγραφή μεταγ
	gccgaartgrcttctcctatctaagtgaagcatatttgaaggasa scaaactettuctsesettegaactatttgaaggasa
	ggcctacagccaccgggggacatggggatggtsitici
	contectededectectededectectededeasectectedec

Fig 18- The amplified sequence with the designed primers from In silico PCR tool.

In 2006, André O.S Lima and Sérgio Garcês in designed the PCR primer for *Streptomycesα*amylase. They established a search string consisting the key words and thetaxonomic ID. The search string established was ("txid 1883" AND alpha-amylase). The consensus sequence was analyzed using the BLAST tool and two blocks of the consensus sequence were selected for primer designing. They obtained the forward primer (5'-GCCGGTCAGCTACAAGAT-3') and reverse primer (5'-TCAACGACCTGCTCTCG-3')

using the Gene Runner v.3.05 software, tested In silico.

Amber L. Robertson and Allison R. Phillips, in 2017 designed the forward and reverse primers for *Pseudomonas putida* aroA gene. The quality of primers were assessed using the website http://www.sigma-genosys.com/calc/DNAcalc.asp and the specificity was determined using BLAST.

3.7 Construction of phylogenetic tree using ClustalW2 Phylogeny

The multiple sequence alignment of the amplified sequence and that of VEGFA gene nucleotide sequence from other animals are obtained using the ClustalW.

The VEGFA gene nucleotide sequences from other animals are

PREDICTED: *Oxyurajamaicensis* vascular endothelial growth factor A (VEGFA), transcriptvariant X13, mRNA

Musmusculus vascular endothelial growth factor A (Vegfa), transcript variant 2, mRNADanio rerio vascular endothelial growth factor A (vegfa), transcript variant 1, mRNA *Xenopus laevis* vascular endothelial growth factor A L homeolog (vegfa. L), mRNA

PREDICTED: *Bufo* vascular endothelial growth factor A (VEGFA), transcript variant X16,mRNA The multiple sequence alignment of the amplified sequence and that of VEGFA gene nucleotide sequence from other animals are obtained using the ClustalW.

GC.	-TAGCCATGTCCTTGA
CTGTGCAGGCT-CCCGG	CTCTGCT
CCTGCCT-	-CAGCTTTAGAGAAGGG
cc	CTTGTGTGCTCACAGCAGT
TT	TTACAATTTT
	GCGGCCCC-
CACACA	CAG
AAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	ACAGAAGTGAGTG
GTG-CCGTT	CATAAACGTCTACTAC
G	GC-AGC
ATG-TGCCAAGTA	-CGG-GAGACGTT
GGTGGA	TATCTACCAG
GAATAT	CCAGA
-CGAAGTGGAG	TTCATCTTCAAACCC
TCTTGT	GCGCTC
	ATGAGCTGTGCACG
ATG	CTGCAGC
GATGAGAC	TTTC
GAGTGCGTA-	GGAGTGG
TACAATGTCACCA	TGCAGATCTTG-AGGAT
GAGACCTC	TCAG
CTGTATGTGGAGAAG	AGTTTCCAGCAGC-ACAGTC
GCTGTGAATGCAGACA-A	AAGAAAGAAATTAAAGGCAAACAGGA
AAAAAAAGCAA	A-ACGAGGGA-AGGGCCATAAACGAAAGCGC
AAGAAAATCCGGTACAAAT	CG-CACCGCTTTCATTGTGAGCC
TTGCACAG	ACACAGAGAGGAGAAAACACTTGTTTGCCCAA
GACCCCCAG-ACCTGTAAATG	TTCCTGTAAAAACACAG-ACTCACGTTG-CAAAA
CGAAGCAGCTTGAGTTAAACG	AAAGGACTTGCAGGAGAGGGAAGAAGCCTA
GAACAGACCAGTCATATG-	CGGCCCCCTATTCTGCTGACCTC
GCTCACTAACAT	GCGCTCTGCCACCCCTGTGCAGCAAAATGTGAAAA
ACCGCGGCGGTGAGACGAG	GAGAGAGTCTCCGGAAAGAGACAAAATGGAGTGAAA
ATGA-AGGATCGAAA	CTCGGCACAGCACTTTGGTTCTCA-
TGCATTCCG	TGTCCCTGAGGGGGGGACAGAATAGCC-GTATTA
TTCAGTAAACGC-CATTTCTC	CTTCTCCTTTTCAAAATTGATGCTGCTACACATC
GAGAGCCAGA-ACA-GTGGGG	TCATA-ACTTTTTTTGGATTCTTATAAATATATT
ΔΤΔΑΤΔΑΤΤΑΔΑΤΑΤΑΓΑΤΑΓ	ΔΥΑΔΑΤΑΤΑΤΑΤΑ-ΤΑΤΤΤΤΤΤΤΤΤΤΑΤΑΤΑΤΑΤΑΤΑΤΑΤΑΤ

Fig 19- The multiple sequence alignment for constructing phylogenetic tree.

The phylogenetic tree was constructed using neighbour-joining method, which is a distance based method. The ClustalW2 phylogeny was used for phylogenetic tree construction.

Classifier Classifier <thclassifier< th=""> Classifier Classifi</thclassifier<>	ee			
View Phylogenetic Tree File	fo_vascular_sodothelial_growt			
((X1_046438819.1_08607CTFD_BL+6_bu M1_151488.3_lanin_menin_vescular_ :0.47778,	fo_vascular_endothelial_growth			
X1_94172166PRD1CIII)_CSyura M1_94174841_PR0_proventure_vacuu (8.848172 (cmv_d1/82969441/85476_02895p_116 M1_9918856991_kenopus_laevis_vas (8.22447);	endotheisal_growth_factor_Aa_ jamaicensis_vascular_endotheis lar_endotheisal_growth_lactor, cicasaaccasaase_aiciccciesaa cular_endotheisal_growth_lactor	Lacrop A_VSG2_transcript_veriaer_SG2 _vsgtaa_transcript_variaersM328U16/77 ial_growth_lactor_A_VIG92_transcript_varia_ _vsgta_transcript_veriaetsMak_0.19 / A_VSG2_transcript_veriaetsMak_0.19 / act(12668-0.12769), or_A_homeolog_vegtait_eM3429.09(22)	авда (8.1979), ••• _XI Iинад.(8.16/866), *9	
Phylogram				
Branch length: 🌒 Cladogram 🤅	C Real			
		XM_040430819.1_PREDICTED_E NM_131400.3_Danlo_rerio_vascula XM_035321361.1_PREDICTED0 NM_001317041.1_Mus_musculus_ chrfit_43781968+43785476_35009t_ NM_001085589.1_Xencpus_laevis	ulo_buio_vascular_endolheliai_growth_factor r_endotheliai_growth_factor_Aa_vegfaa_tu Svura_jamaicensis_vascular_endotheliai_gro vascular_endotheliai_growth_factor_A_Vegfa TTGGTGGAGGGGGAGAAG_GTGTCCGTG vascular_encotheliai_growth_factor_Ahon	A_VEGFA_transcrip ranscript_variant_1_mRi wth_factor_A_VEGFA_ transcript_variant_2 GGCAACTCAGAA 0.127/ neolog_vegfa.LmRN

Fig 20- The phylogenetic tree of VEGFA gene constructed using the bioinformatics tool ClustalW2 phylogeny.

Murat KASAP, 2005 created a phylogenetic tree of Vascular Endothelial Growth Factor using distance based methods. The multiple sequence alignment was obtained using ClustalW. The pairwise evolutionary distances were calculated using the program PROTDIST. The analysis revealed that the VEGF family is composed of four major branches evolved from a common ancestor.

Conclusion

The PCR primer of Vascular Endothelial Growth Factor A (VEGFA) gene in *Homo sapiens* was designed with the help of bioinformatics tools. The nucleotide sequences of VEGFA gene was

Vol 12 Issue 03 2023 ISSN NO: 2230-5807

retrieved from GenBank of which the multiple sequence alignment and consensus sequence determined with ClustalW and EMBOSS Cons respectively. The sequences similar to the consensuses sequence are found by the sequence similarity searching tool BLAST. The conserved domains are identified and the nucleotide sequences from it are uploaded to Primer3 Plus to obtain the forward and reverse primers which are tested in silico. The phylogenetic tree of VEGFA gene was constructed using the amplified sequence and the sequence from organisms like *Mus musculus*, *Danio rerio*, *Bufo*, *Xenopus laevis* and *Oxyura jamaicensis*.

References

- Alberto A'Ivarez -Aznar, Lars Muhl, Konstantin Gaengel, 2017, Chapter Thirteen- VEGF Receptor Tyrosine Kinases: Key regulators of Vascular Function, Current Topics in Developmental Biology, Volume 123, Pages 433-482.
- [2] Altschul S, Gish W, Miller W, Myers E, Lipman D, 1990, Basic local alignment search tool, Journal of Molecular Biology, 215(3): 403-410.
- [3] Altschul S, Madden T, Sch"affer A, Zhang Z, Zhang J, Miller W, Lipman D, 1997, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Research, 25(17):3389-3402.
- [4] Amber L. Robertson and Allison R. Phlillips, 12 Oct 2017, Integrating PCR Theory and Bioinformatics into a Research-Oriented Primer Design Exercise, CBE-Life Sciences Education, Vol 7, No.1.
- [5] __André O.S Lima and Sérgio P.S. Garcês, 2006, Intrageneric Primer Design, Bringing Bioinformatics Tools to the Class, Biochemistry and Molecular Biology Education, Vol.34, No.5, pp 332-337.
- [6] Andreas Untergasser, Harm Nijveen, Xiangyu Rao, Ton Bisseling, Rene´ Geurts, Jack A.M. Leunissen, July 2007, Primer3 Plus an enhanced web interface to Primer3, Nucleic Acids Research, Volume 35, Issue suppl-2,1, pages W71-W74.
- [7] ___Ankit P. Laddha, Anil B. Gaikwad, Yogesh A. Kulkarni, 2020, Structural and Functional Insights of VEGF and its Receptors, Reference Module in Life Sciences.
- [8] Bing Yu, Changbin Zhang, 2011, In silico PCR analysis, Methods in Molecular Biology, 760:91-107.
- [9] Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, Thomas L Madden, 2009, BLAST+: Architecture and applications, BMC Bioinformatics 10, Article number:421.
- [10] Eric W Sayers, Mark Cavanaugh, Karen Clark, James Ostell, Kim D Pruitt, Ilene Karsch Mizrachi, 08 January 2019, GenBank, Nucleic Acids Research, Volume 47, Issue D1, Pages D94-D99.
- [11] Gang Hu, Lukasz Kurgan, 2019 Feb, Sequence Similarity Searching, Current Protocols in Protein Science, 95(1): e71.
- [12] Higgins D, Thompson J, Gibson D Thompson J.D, Higgins D.G., Gibson T.J, 1994, CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Research 22:4673-4680.
- [13] Higgins, D.G and Sharp, P.M, 1998, CLUSTAL: a package for performing multiple sequence alignment on a microcomputer, Gene, 73(1),273-244.
- [14] Jui-Hung Hung, Zhiping Weng, 2016 Sep 1, Designing Polymerase Chain Reaction primers using Primer3 Plus, Cold Spring Harbor Protocols,2016(9).
- [15] Kamel A. Abd-Elsalam, May 2003, Bioinformatic tools and guidelines for PCR primer design, African Journal of Biotechnology Vol.2(5), pp 91-95.
- [16] Kelli H Dyer, Paolo S Silva, Jennifer K Sun, Sep-Nov 2013, Vascular endothelial growth factor gene polymorphisms and vitreous proteome changes in diabetic retinopathy, Seminars in Ophthalmology, 28 (5-6): 347-54.

- [17] Klienman, J. Ambati, 2010, Injury and Repair: Neovascularization, Encyclopedia of the Eye, pages 400-407.
- [18] Kyoto Encyclopedia of Genes and Genomes.
- [19] Leung, D.W, Cachianes, G., Kuang, W-J., Goeddel, D.V, Ferrara, N.,1989, 'Vascular endothelial growth factor is a secreted mitogen'. Science 246:1306-1309.
- [20] Lisa J. Mullan, Alan J. Bleasby, March 2002, Short EMBOSS user guide, Briefings in Bioinformatics, Volume 3, Issue 1, Pages 92-94.
- [21] Marcin Bogusz, Simon Whelan, 2017 Mar 1, Phylogenetic Trees Estimation With and Without Alignment: New Distance Methods and Benchmarking, Systematic Biology,66(2): 218-231.
- [22] Murat KASAP, 2005, Phylogenetic Analysis of Vascular Endothelial Growth Factor Diversity, Turkish Journal of Biotechnology, Vol 29, pp 217-227.
- [23] Napoleone Ferrara, 2004 Aug, Vascular endothelial growth factor: basic science and clinical progress, Endocrine reviews,25(4):581-611.
- [24] ___Neha Garg, Prateek Kumar, Kundlik Gadhave, Rajanish Giri, 2019, Chapter Ten- The dark proteome of cancer: Intrinsic disorderedness and functionality of HIF- α along with its interacting proteins, Progress in Molecular Biology and Translational Science, Volume 166, pages 371-403.
- [25] Ramu Chenna, Hideaki Sugawara, Tadashi Koike, Rodrigo Lopez, Toby J Gibson, Desmond G Higgins, Julie D Thompson, 2003 Jul 1, Multiple Sequence Alignment with the Clustal series of programs, Nucleic Acids Research, 31(13):3497-500.
- [26] Rao and R.A Lang, 2010, Formation and Regression of the Primary Vitreous and Hyaloid Vascular Encyclopedia of the Eye, pages 151-156.
- [27] ___Regenfuss, C. Cursiefen, 2010, Concept of Angiogenic Privilege, Encyclopedia of the Eye, pages 334-338.
- [28] Ruslan Kalendar, Bekbolat Khassenov, Yerlan Ramankulov, Olga Samuilova, Konstantin I Ivanov, 2017 Jul, FastPCR: An in silico tool for fast primer and probe design and advanced sequence analysis, Genomics,109(3-4): 312-319.
- [29] Sagar Patel, Hetalkumar Panchal, Kalpesh Anjaria, 2012, Phylogenetic analysis of some leguminous trees using CLUSTALW2 bioinformatics tool, Institute of Electrical and Electronics Engineers.
- [30] Sourdis, M Nei, 1988 May, Relative efficiencies of maximum parisomy and distance-matrix methods in obtaining the correct phylogenetic tree, Molecular biology and evolution, 5(3): 298-311.
- [31] Timothy M Rose, Jorja G Henikoff, Steven Henikoff, 2003 Jul, CODEHOP (COnsensus- Degenerate Hybrid Oligonucleotide Primer) PCR Primer design, Nucleic Acids Research, 31(13): 3763-6.
- [32] Tischer, E., Mitchell, R., Hartman, T., Silva, M., Gospodarowicz, D., Fiddes, J. C., Abraham,
- [33] J. A, 1991, The human gene for vascular endothelial growth factor: multiple protein forms are encoded through alternative exon splicing. J. Biol. Chem. 266: 11947-11954.[PubMed: <u>1711045</u>]
- [34] William R Pearson,2014, BLAST and FASTA similarity searching for multiple sequence alignment, Methods in Molecular Biology, 1079:75-101.