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 vascularpermeability 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 functionsand 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 6th exon lacks 165 amino acids from and the region encoded by 6th and 7th exons lacks 121 amino acids (OMIM, Tischer *et al*., 1991).
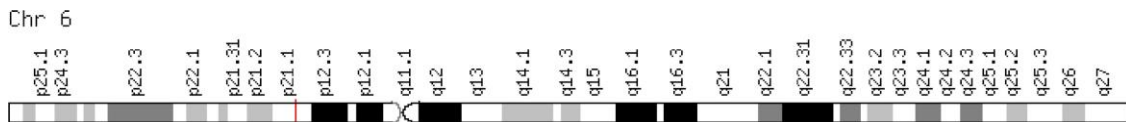
Fig 1 – The chromosomal structure of Vascular Endothelial Growth Factor fromGeneCards.
(https://www.genecards.org)

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/).

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

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'.
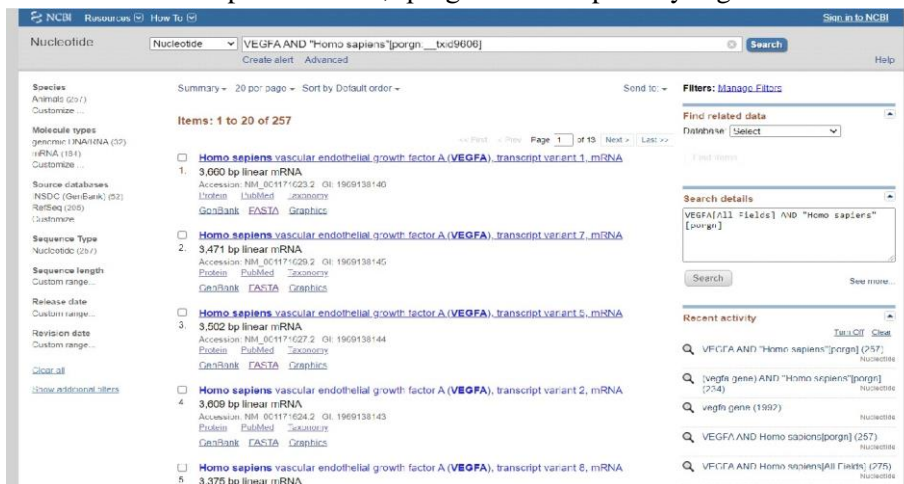- The taxonomic ID for Homo sapiens is 9606, 'porgn' indicate primary organism.

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 inthe 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 (URL-https://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.
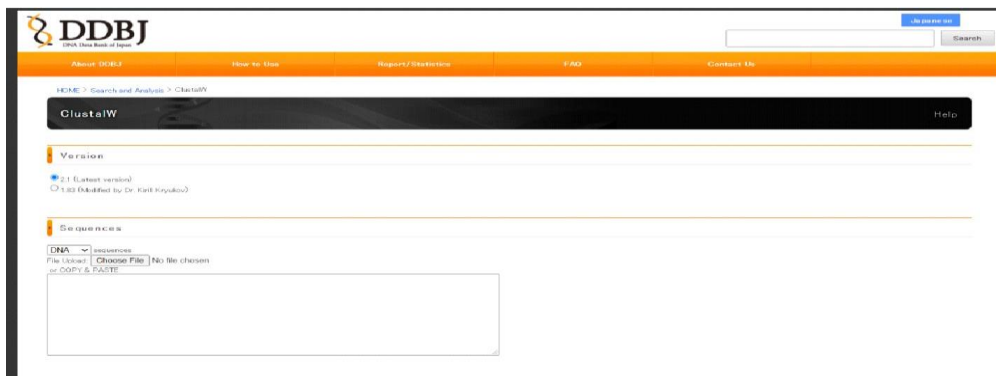

Fig 3- The ClustalW multiple sequence alignment program at DNA Data Bank of Japan (https://www.clustalw.ddbj.nig.ac.jp).

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.
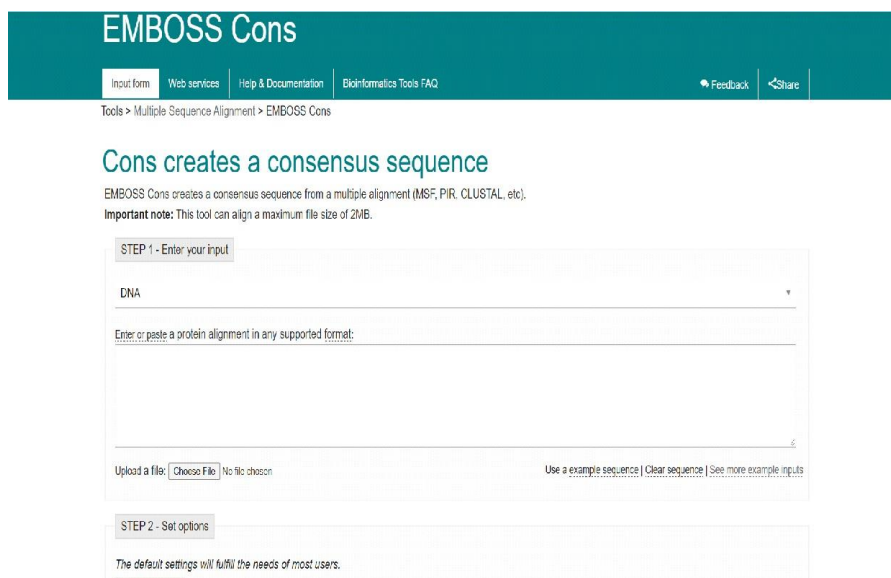- The result displayed on the new window can be downloaded by selecting 'Download outputfile'.

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

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.
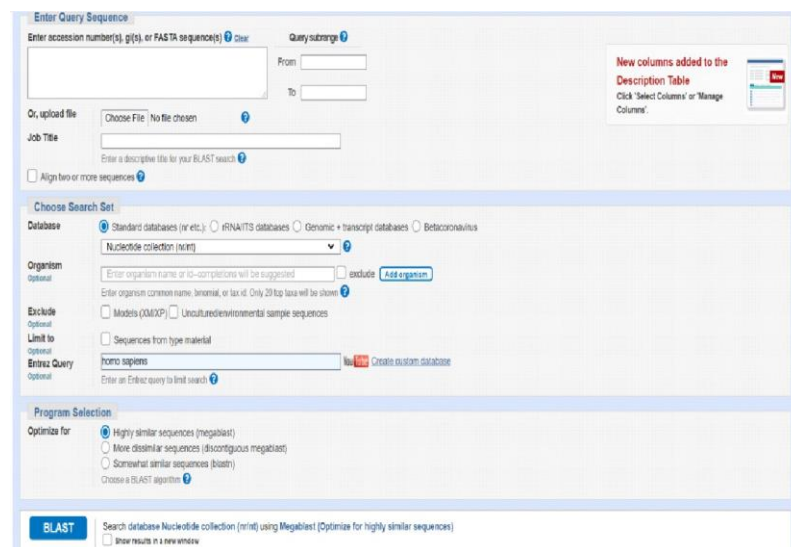

Fig 5 - The nucleotide BLAST (blast n) is used to compare the consensus sequence to the nucleotide sequences at GenBank. (https://blast.ncbi.nlm.nih.gov./)

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

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 tosecondary 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.

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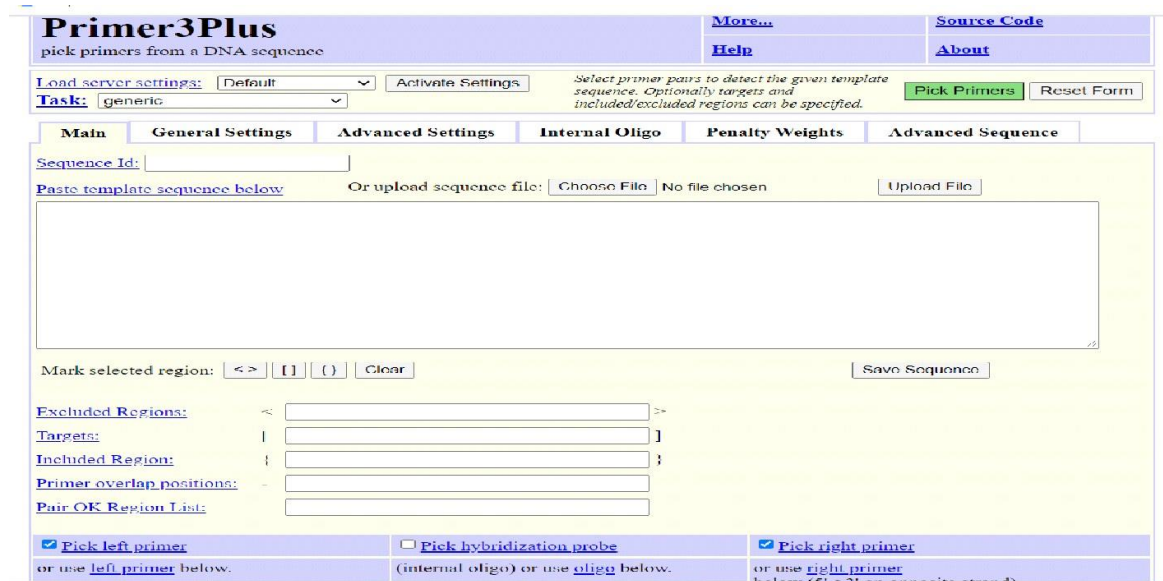
Fig 7- Primer3 Plus homepage. The nucleotide sequences are uploaded after setting the corresponding parameters.(https://www.primer3plus.com)

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 selectedfrom 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.
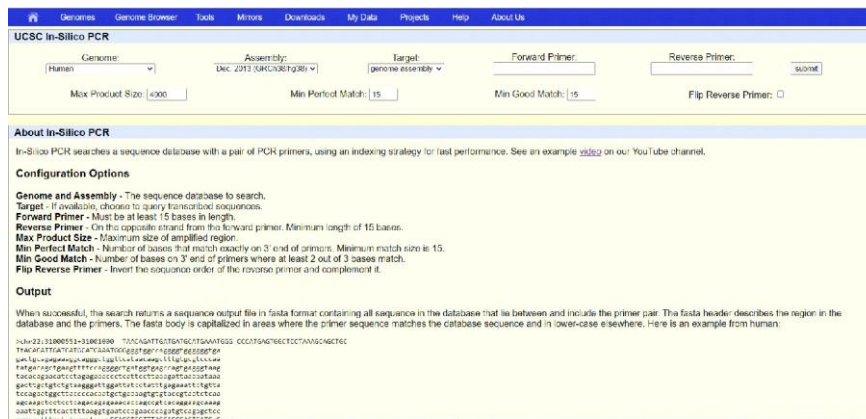

Fig 8- UCSC Genome Browser homepage.(https://www.genome.ucsc.edu/)

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.
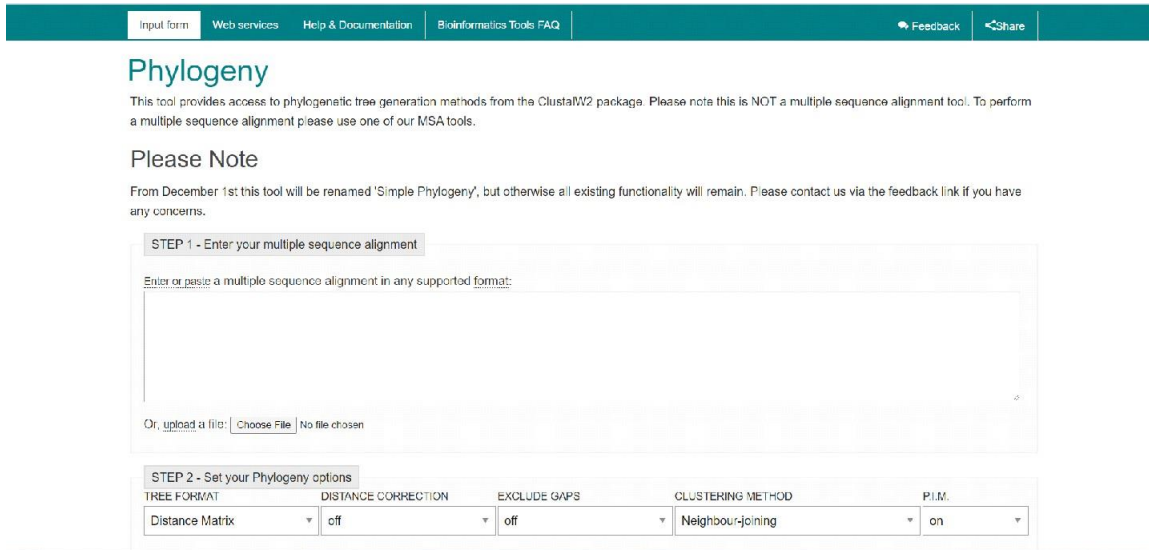- The phylogenetic tree will be displayed in a new window.

Fig 10- The phylogenetic tree is constructed using ClustalW2 phylogeny.
(**https://www.ebi.ac.uk.**) **The multiple sequence alignment of the nucleotide sequences fromthe 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.

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.

>gi|1969138146|ref|NM_001171623.2|_Homo_sapiens_vascular_endothelial_grow
th_factor_A__VEGFA___transcript_variant_1__mRNA
GCGGAGGCTTGGGGCAGCCGGGTAGCTCGGAGGTCGTGGCGCTGGGGGCTAGCACCAGCG
CTCTGTCGGGAGGCGCAGCGGTTAGGTGGACCGGTCAGCGGACTCACCGGCCAGGGCGCT
CGGTGCTGGAATTTGATATTCATTGATCCGGGTTTTATCCCTCTTCTTTTTTCTTAAACA
TTTTTTTTTAAAACTGTATTGTTTCTCGTTTTAATTTATTTTTGCTTGCCATTCCCCACT
TGAATCGGGCCGACGGCTTGGGGAGATTGCTCTACTTCCCCAAATCACTGTGGATTTTGG
AAACCAGCAGAAAGAGGAAAGAGGTAGCAAGAGCTCCAGAGAGAAGTCGAGGAAGAGAGA
GACGGGGTCAGAGAGAGCGCGCGGGCGTGCGAGCAGCGAAAGCGACAGGGGCAAAGTGAG
TGACCTGCTTTTGGGGGTGACCGCCGGAGCGCGGCGTGAGCCCTCCCCCTTGGGATCCCG
CAGCTGACCAGTCGCGCTGACGGACAGACAGACAGACACCGCCCCCAGCCCCAGCTACCA
CCTCCTCCCCGGCCGGCGGCGGACAGTGGACGCGGCGGCGAGCCGCGGGCAGGGGCCGGA
GCCCGCGCCCGGAGGCGGGGTGGAGGGGGTCGGGGCTCGCGGCGTCGCACTGAAACTTTT
CGTCCAACTTCTGGGCTGTTCTCGCTTCGGAGGAGCCGGTGGTCCGCGCGGGGGAAGCCGA
GCCGAGCGGAGCCGCGAGAAGTGCTAGCTCGGGCCGGGAGGAGCCGCAGCCGGAGGAGGG
GGAGGAGGAAGAAGAGAAGGAAGAGGAGAGGGGGCCGCAGTGGCGACTCGGCGCTCGGAA
GCCGGGCTCATGGACGGGTGAGGCGGCGGTGTGCGCAGACAGTGCTCCAGCCGCGCGCGC
TCCCCAGGCCCTGGCCCGGGCCTCGGGCCGGGGAGGAAGAGTAGCTCGCCGAGGCGCCGA
GGAGAGCGGGCCGCCCCACAGCCCGAGCCGGAGAGGGAGCCGAGCCGCGCCGGCCCCGG
TCGGGCCTCCGAAACCATGAACTTTCTGCTGTCTTGGGTGCATTGGAGCCTTGCCTTGCT
GCTCTACCTCCACCATGCCAAGTGGTCCCAGGCTGCACCCATGGCAGAAGGAGGAGGGCA
GAATCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGCGCAGCTACTGCCATCCAAT
CGAGACCCTGGTGGACATCTTCCAGGAGTACCCTGATGAGATCGAGTACATCTTCAAGCC
ATCCTGTGTGCCCCTGATGCGATGCGGGGGCTGCTGCAATGACGAGGGCCTGGAGTGTGT
GCCCACTGAGGAGTCCAACATCACCATGCAGATTATGCGGATCAAACCTCACCAAGGCCA
CCACATACCGACAGATCAGCTTCCTACAGCACAACAAATGTGAATGCAGACCAAACAAACA
TAGAGCAAGACAAGAAAAAAAATCAGTTCGAGGAAAGGGAAAGGGGCAAAAACGAAAGCG
CAAGAAATCCCGGTATAAGTCCTGGAGCGTGTACGTTGGTGCCCGCTGCTGTCTAATGCC
CTGGAGCCTCCCTGGCCCCCATCCCTGTGGGCCTTGCTCAGAGCGGAGAAAGCATTTGTT
TGTACAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACACAGACTCGCGTTGCAAGGC
GAGGCAGCTTGAGTTAAACGAACGTACTTGCAGATGTGACAAGCCGAGGCGGTGAGCCGG
GCAGGAGGAAGGAGCCTCCCTCAGGGTTTCGGGAACCAGATCTCTCACCAGGAAAGACTG
ATACAGAACGATCGATACAGAAACCACGCTGCCGCCACCACACCATCACCATCGACAGAA

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.

```
>EMBOSS0001
GCGGAGGCTTGGGGCAGCCGGGTAGCTCGGAGGTCGTGGCGCTGGGGGCTAGCACCAGCG
CTCTGTCGGGAGGCGCAGCGGTTAGGTGGACCGGTCAGCGGACTCACCGGCCAGGGCGCT
CGGTGCTGGAATTTGATATTCATTGATCCGGGTTTTATCCCTCTTCTTTTTTCTTAAACA
TTTTTTTTTAAAACTGTATTGTTTCTCGTTTTAATTTATTTTTGCTTGCCATTCCCCACT
TGAATCGGGCCGACGGCTTGGGGAGATTGCTCTACTTCCCCAAATCACTGTGGATTTTGG
AAACCAGCAGAAAGAGGAAAGAGGTAGCAAGAGCTCCAGAGAGAAGTCGAGGAAGAGAGA
GACGGGGTCAGAGAGAGCGCGCGGGCGTGCGAGCAGCGAAAGCGACAGGGGCAAAGTGAG
TGACCTGCTTTTGGGGGTGACCGCCGGAGCGCGGCGTGAGCCCTCCCCCTTGGGATCCCG
CAGCTGACCAGTCGCGCTGACGGACAGACAGACAGACACCGCCCCCAGCCCCAGCTACCA
CCTCCTCCCCGGCCGGCGGCGGACAGTGGACGCGGCGGCGAGCCGCGGGCAGGGGCCGGA
GCCCGCGCCCGGAGGCGGGGTGGAGGGGGTCGGGGCTCGCGGCGTCGCACTGAAACTTTT
CGTCCAACTTCTGGGCTGTTCTCGCTTCGGAGGAGCCGTGGTCCGCGCGGGGGAAGCCGA
GCCGAGCGGAGCCGCGAGAAGTGCTAGCTCGGGCCGGGAGGAGCCGCAGCCGGAGGAGGG
GGAGGAGGAAGAAGAGAAGGAAGAGGAGAGGGGGCCGCAGTGGCGACTCGGCGCTCGGAA
GCCGGGCTCATGGACGGGTGAGGCGGCGGTGTGCGCAGACAGTGCTCCAGCCGCGCGCGC
TCCCCAGGCCCTGGCCCGGGCCTCGGGCCGGGGAGGAAGAGTAGCTCGCCGAGGCGCCGA
GGAGAGCGGGCCGCCCCACAGCCCGAGCCGGAGAGGGAGCGCGAGCCGCGCCGGCCCCGG
TCGGGCCTCCGAAACCATGAACTTTCTGCTGTCTTGGGTGCATTGGAGCCTTGCCTTGCT
GCTCTACCTCCACCATGCCAAGTGGTCCCAGGCTGCACCCATGGCAGAAGGAGGAGGGCA
GAATCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGCGCAGCTACTGCCATCCAAT
CGAGACCCTGGTGGACATCTTCCAGGAGTACCCTGATGAGATCGAGTACATCTTCAAGCC
ATCCTGTGTGCCCCTGATGCGATGCGGGGGCTGCTGCAATGACGAGGGCCTGGAGTGTGT
GCCCACTGAGGAGTCCAACATCACCATGCAGATTATGCGGATCAAACCTCACCAAGGCCA
GCACATAGGAGAGATGAGCTTCCTACAGCACAACAAATGTGAATGCAGACCAAAGAAAGA
TAGAGCAAGACAAGAAAAaaaatcagttcgaggaaagggaaaggggcaaaaacgaaagcg
caagaaatcccgnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
nnnnnnnnnnnnnnnnnnnnnTCCCTGTGGGCCTTGCTCAGAGCGGAGAAAGCATTTGTT
TGTACAAGATCCGCAGACGTGTAAATGTTCCTGCAAAAACACAGACTCGCGTTGCAAGGC
GAGGCAGCTTGAGTTAAACGAACGTACTTGCAGATGTGACAAGCCGAGGCGGTGAGCCGG
GCAGGAGGAAGGAGCCTCCCTCAGGGTTTCGGGAACCAGATCTCTCACCAGGAAAGACTG
ATACAGAACGATCGATACAGAAACCACGCTGCCGCCACCACACCATCACCATCGACAGAA
CAGTCCTTAATCCAGAAACCTGAAATGAAGGAAGAGGAGACTCTGCGCAGAGCACTTTGG
GTCCGGAGGGCGAGACTCCGGCGGAAGCATTCCCGGGCGGGTGACCCAGCACGGTCCCTC
TTGGAATTGGATTCGCCATTTTATTTTTCTTGCTGCTAAATCACCGAGCCCGGAAGATTA
GAGAGTTTTATTTCTGGGATTCCTGTAGACACACCCACCCACATACATACATTTATATAT
ATATATATTATATATATATAAAAATAAATATCTCTATTTTATATATATAAAATATATATA
TTCTTTTTTTAAATTAACAGTGCTAATGTTATTGGTGTCTTCACTGGATGTATTTGACTG
CTGTGGACTTGAGTTGGGAGGGGAATGTTCCCACTCAGATCCTGACAGGGAAGAGGAGGA
GATGAGAGACTCTGGCATGATCTTTTTTTTTGTCCCACTTGGTGGGGCCAGGGTCCTCTCC
CCTGCCCAGGAATGTGCAAGGCCAGGGCATGGGGGCAAATATGACCCAGTTTTGGGAACA
CCGACAAACCCAGCCCTGGCGCTGAGCCTCTCTACCCCAGGTCAGACGGACAGAAAGACA
GATCACAGGTACAGGGATGAGGACACCGGCTCTGACCAGGAGTTTGGGGAGCTTCAGGAC
ATTGCTGTGCTTTGGGGATTCCCTCCACATGCTGCACGCGCATCTCGCCCCCAGGGGCAC
TGCCTGGAAGATTCAGGAGCCTGGGCGGCCTTCGCTTACTCTCACCTGCTTCTGAGTTGC
CCAGGAGACCACTGGCAGATGTCCCGGCGAAGAGAAGAGACACATTGTTGGAAGAAGCAG
CCCATGACAGCTCCCCTTCCTGGGACTCGCCCTCATCCTCTTCCTGCTCCCCTTCCTGGG
```

Fig 13- The consensus sequence from EMBOSS Cons.

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.

Fig 14-The Graphic summary of BLAST result.

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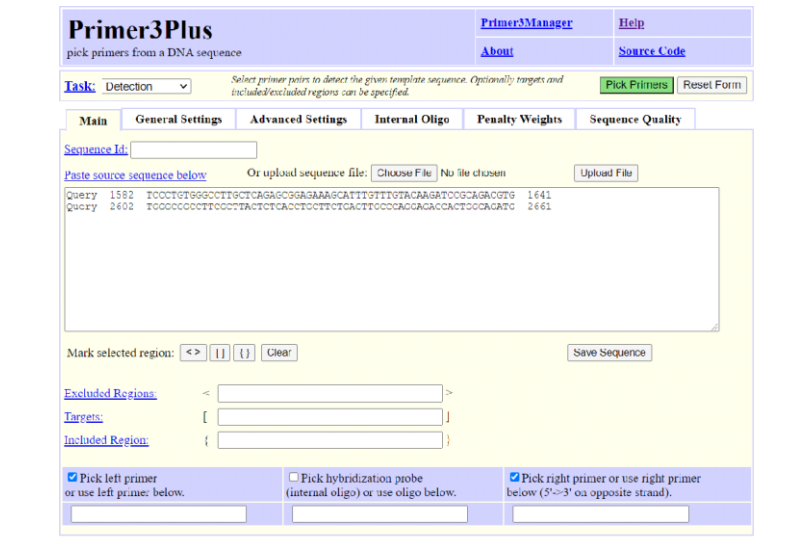
Fig 16- The nucleotide sequences are uploaded in Primer3 Plus.

Total five pairs of Forward and Reverse primer pairs was obtained, from the first pair was chosen for In silico testing.
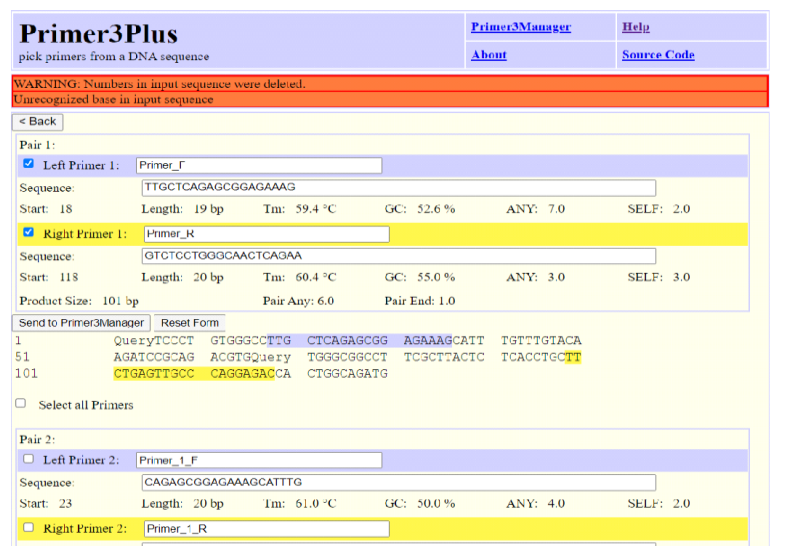
Fig 17- Five pairs of primers are obtained.

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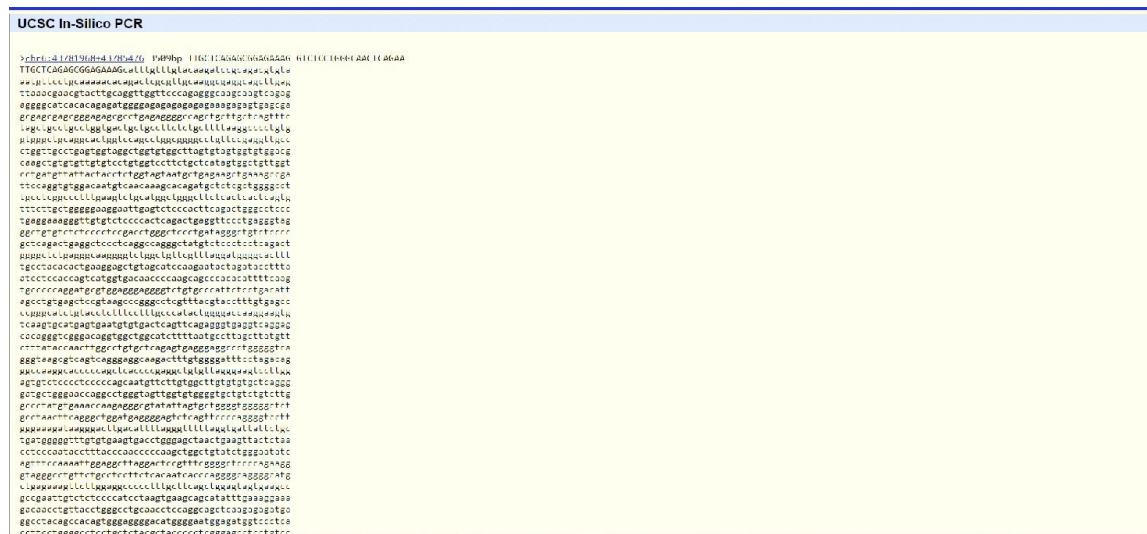
Fig 18- The amplified sequence with the designed primers from In silico PCR tool.

In 2006, André O.S Lima and Sérgio Garcês in designed the PCR primer for *Streptomyces* α-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, mRNA*Danio 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.

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.
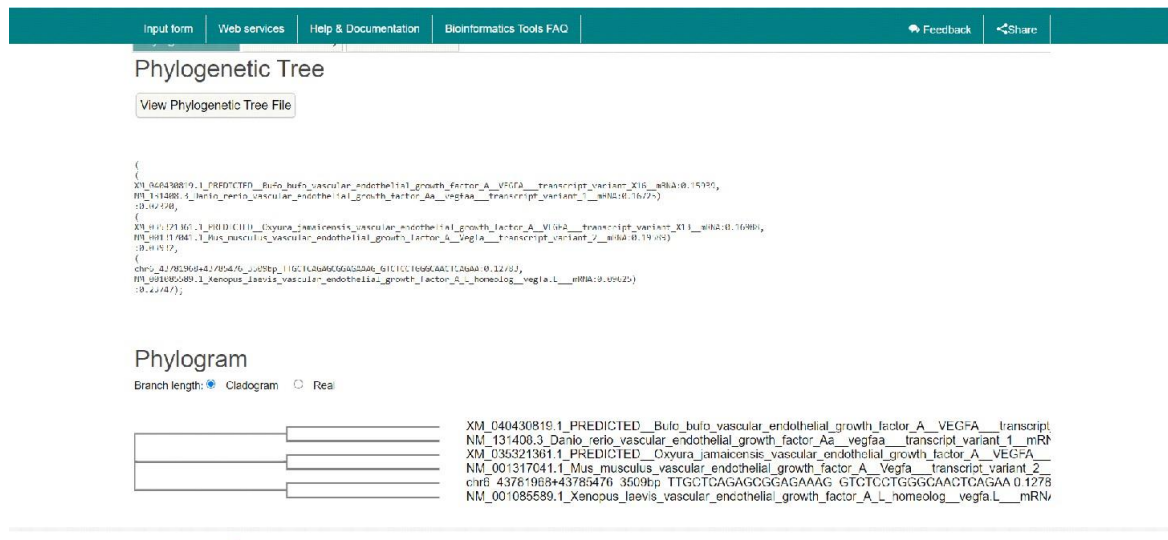

Fig 20- The phylogenetic tree of VEGFA gene constructed using the bioinformatics tool ClustalW2 phylogeny.

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

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

[1] Alberto A´lvarez -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: 1711045]

[34] William R Pearson,2014, BLAST and FASTA similarity searching for multiple sequence alignment, Methods in Molecular Biology, 1079:75-101.