

A Brief Study on Analysis of Genes Important for Endometrial Cancer

Atreyee Majumder and Malavika Bhattacharya[#]

Department of Biotechnology
Techno India University, West Bengal
Kolkata, India

Abstract— Endometrial cancer begins in the layer of cells that form the lining (endometrium) of the uterus. The genetic causes, mainly hereditary causes, contribute to 2–10% of endometrial cancer. Since endometrial cancer is becoming one of the main cancers effecting women in eastern India, the genes important in its context need to be identified and studied thoroughly. This study was performed with the aim of understanding conservation of the genes and the proteins encoded by them which are important with reference to endometrial cancer. For this purpose, the genes that were analyzed included PTEN, Tp53, WNT4, BRAF, PIK3CA, MGMT and SMAD family. The GENE similarity between Human (*Homo sapiens*), Mice (*Mus musculus*) and Zebra fish (*Danio rerio*) were analyzed. The FASTA Nucleotide Sequences were taken from NCBI-GENE and the FASTA Protein sequences were taken from UNIPROT. The ‘Phylogenetic tree view’ of Nucleotide and Protein and ‘Cobalt Multiple Sequence Alignment’ of protein showed GENE similarity between the three organisms. Around 70%-85% identity was observed between genes of human, mice and zebra fish.

Keywords—Endometrial Cancer; Hereditary factors; Gene conservation; Sequence analysis

I. INTRODUCTION :

Endometrial cancers are responsible for about 5% and 2% of worldwide cancer incidence and mortality among women, respectively. In women, endometrial cell carcinomas (ECCs) are the most common malignancy of the female genital tract in the world and the fourth most common one after breast, lung, and colorectal cancer. Studies have indicated that although approximately 90% of cases of endometrial cancer are sporadic, the remaining 10% of cases are hereditary (Mikuta, 1993; Lentz, 1994; Rose, 1996; Murali *et al.*, 2014).

Based on reports of genes affected under endometrial cancer incidences, seven genes were selected for analysis. They included PTEN, Tp53, WNT4, BRAF, PIK3CA, MGMT and SMAD family (Doll *et al.*, 2008; Prat and Franceschi, 2014).

II. MATERIALS AND METHODS

The genes that were analyzed included PTEN, Tp53, WNT4, BRAF, PIK3CA, MGMT and SMAD family. The GENE similarity between Human (*Homo sapiens*), Mice (*Mus*

musculus) and Zebra fish (*Danio rerio*) were analyzed. The FASTA Nucleotide Sequences were taken from NCBI-Nucleotide and the FASTA Protein sequences were taken from UNIPROT. Basic Local Alignment Search Tool Analysis (BLAST) comparison between nucleotide or protein sequences was done and regions of local similarity between sequences were identified.

BIOINFORMATICS TOOLS IN GENETIC ANALYSIS (Zhang *et al.*, 2000; Morgulis *et al.*, 2008):

BLAST: The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

FASTA: FASTA format is a text-based DNA and protein sequence alignment software package format for representing either nucleotide sequences or amino acid sequences, in which nucleotides or amino acids are represented using single-letter codes. The format also allows for sequence names and comments to precede the sequences.

UNIPROT: UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literature. The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

III. RESULTS:

The ‘Phylogenetic tree view’ of Nucleotide and Protein and ‘Cobalt Multiple Sequence Alignment’ of protein showed GENE similarity between the three organisms.

FIGURES

A. Nucleotide Blast (Nucleotide—Nucleotide):

Optimized for Highly similar sequences (megablast)

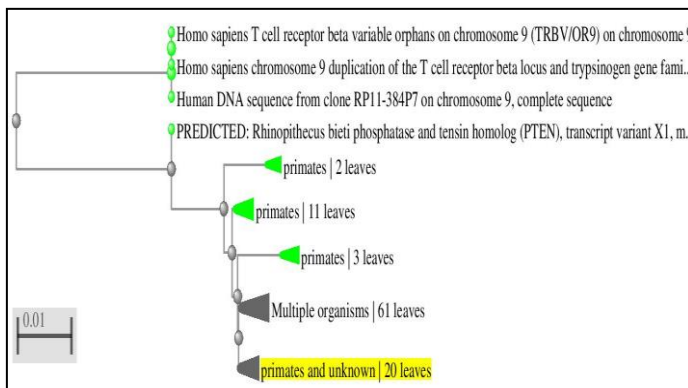
Nucleotide Sequence: The nucleotide collection consists of GenBank+EMBL+DDBJ+PDB+RefSeq sequences, but excludes EST, STS, GSS, WGS, TSA, patent sequences as well as phase 0, 1, and 2 HTGS sequences and sequences longer than 100Mb. The database is non-redundant. Identical sequences have been merged into one entry, while preserving the accession, GI, title and taxonomy information for each entry.

1) PTEN:

>NC_000010.11:87863625-87971930 *Homo sapiens*
chromosome 10, GRCh38.p13 Primary Assembly

>NC_000085.6:32757577-32826160 *Mus musculus* strain
C57BL/6J chromosome 19, GRCm38.p6 C57BL/6J

>NC_007123.7:c17424162-17400669 *Danio rerio* strain
Tuebingen chromosome 12, GRCz11 Primary Assembly]



NB:-- In this sequence chromosome 9 complex DNA sequence of PTEN

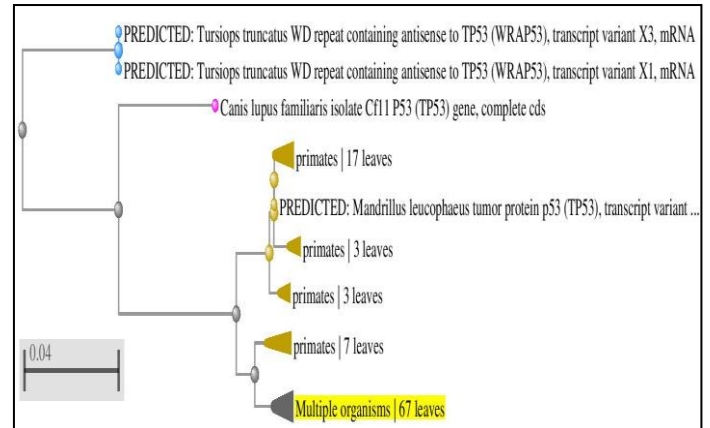
Figure 1: Nucleotide Blast results for PTEN

2) P53:

>NC_000017.11:c7687550-7668402 *Homo sapiens*
chromosome 17, GRCh38.p13 Primary Assembly

>NC_000077.6:69580348-69591873 *Mus musculus* strain
C57BL/6J chromosome 11, GRCm38.p6 C57BL/6J

>NC_007116.7:24086227-24097807 *Danio rerio* strain
Tuebingen chromosome 5, GRCz11 Primary Assembly]



NB: Antisense of TP53 transcript variant, mRNA WD repeats.

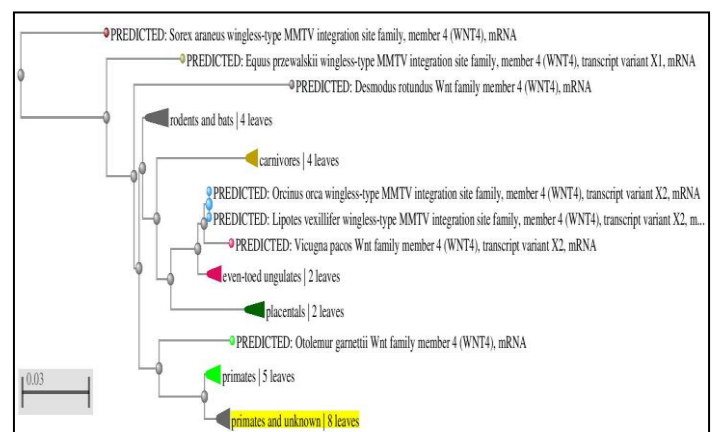
Figure 2: Nucleotide Blast results for P53

3) WNT4:

>NC_000001.11:c22143981-22117308 *Homo sapiens*
chromosome 1, GRCh38.p13 Primary Assembly

>NC_000070.6:137277635-137299501 *Mus musculus* strain
C57BL/6J chromosome 4, GRCm38.p6 C57BL/6J

>NC_007122.7:c39202915-39167846 *Danio rerio* strain
Tuebingen chromosome 11, GRCz11 Primary Assembly]



NB: MMTV integration site family WNT4 transcript variant.

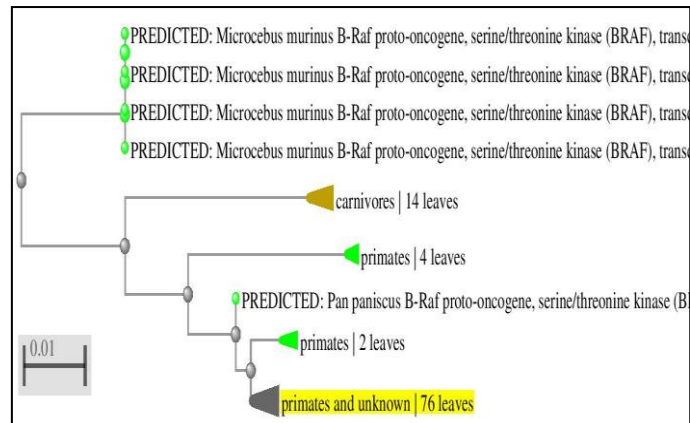
Figure 3: Nucleotide Blast results for WNT4

4) B-Raf:

>NC_000007.14:c140924929-140713328 *Homo sapiens*
chromosome 7, GRCh38.p13 Primary Assembly

>NC_000072.6:c39725658-39603231 *Mus musculus* strain
C57BL/6J chromosome 6, GRCm38.p6 C57BL/6J

>NC_007115.7:c12102123-12072421 *Danio rerio* strain Tuebingen chromosome 4, GRCz11 Primary Assembly]



NB: Braf proto Oncogene is similar in Microcebus murinus

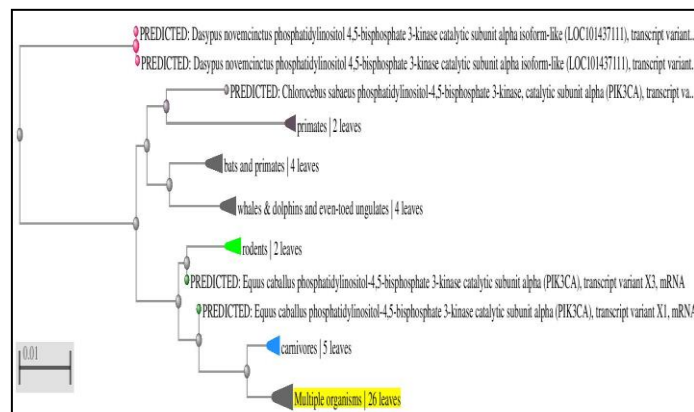
Figure 4: Nucleotide Blast results for B-Raf

5) PIK3CA:

>NC_000003.12:179148114-179240093 *Homo sapiens* chromosome 3, GRCh38.p13 Primary Assembly

>NC_000069.6:32397059-32468486 *Mus musculus* strain C57BL/6J chromosome 3, GRCm38.p6 C57BL/6J

>NC_007122.7:c34522237-34484780 *Danio rerio* strain Tuebingen chromosome 11, GRCz11 Primary Assembly]



NB: PIK3CA transcript variant for the different primates like alpha and beta.

Figure 5: Nucleotide Blast results for PIK3CA

6) MGMT:

NC_000010.11:129467241-129770983 *Homo sapiens* chromosome 10, GRCh38.p13 Primary Assembly

>NC_000073.6:136894463-137128193 *Mus musculus* strain C57BL/6J chromosome 7, GRCm38.p6 C57BL/6J

>NC_007128.7:c20218391-20214215 *Danio rerio* strain Tuebingen chromosome 17, GRCz11 Primary Assembly



NB: MGMT transcript variant for the ncRNA where intron 1.

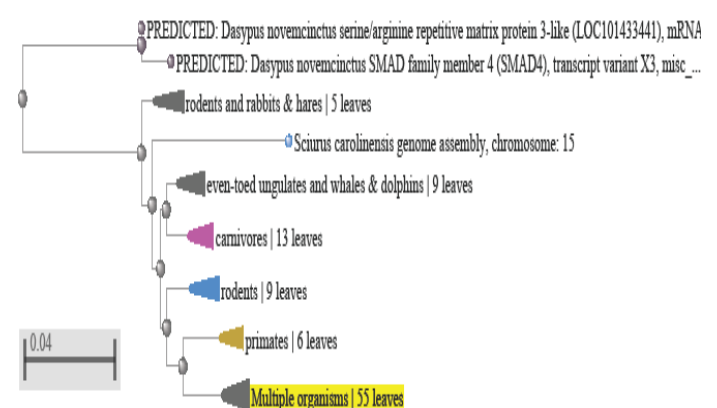
Figure 6: Nucleotide Blast results for MGMT

7) SMAD4:

>NC_000018.10:51030213-51085042 *Homo sapiens* chromosome 18, GRCh38.p13 Primary Assembly;

>NC_000084.6:c73703791-73634790 *Mus musculus* strain C57BL/6J chromosome 18, GRCm38.p6 C57BL/6J;

>NC_007116.7:6870899-6922681 *Danio rerio* strain Tuebingen chromosome 5, GRCz11 Primary Assembly]



NB: SMAD4 transcript variant of mRNA

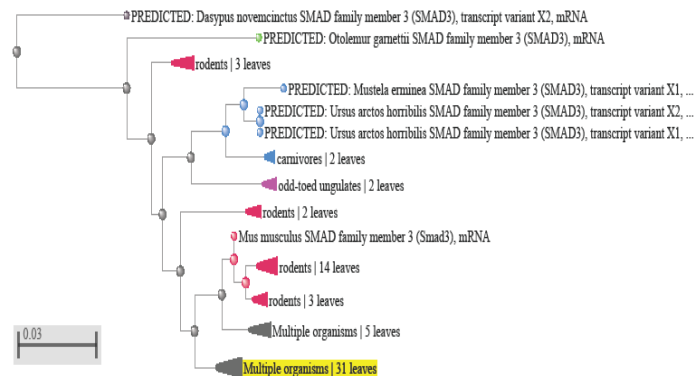
Figure 7: Nucleotide Blast results for SMAD4

8) SMAD3:

>NC_000015.10:67065602-67195195 *Homo sapiens* chromosome 15, GRCh38.p13 Primary Assembly;

>NC_000075.6:c63757994-63646766 *Mus musculus* strain C57BL/6J chromosome9, GRCm38.p6 C57BL/6J ;

> NC_007118.7:c34149234-34106111 *Danio rerio* strain Tuebingen chromosome 7, GRCz11 Primary Assembly]



NB:SMAD3 family transcript variant mRNA

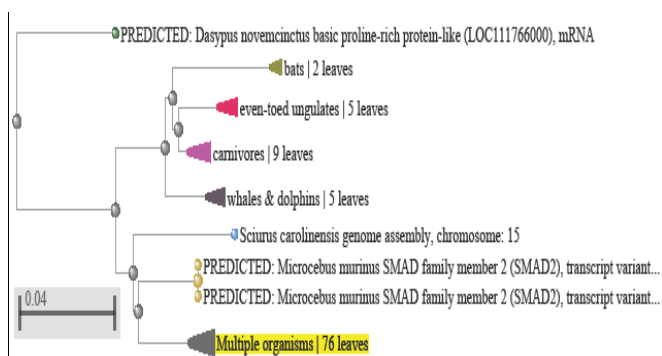
Figure 8: Nucleotide Blast results for SMAD3

9) SMAD2:

>NC_000018.10:c47931188-47808957 *Homo sapiens* chromosome 18, GRCh38.p13 Primary Assembly;

>NC_000084.6:76241090-76311748 *Mus musculus* strain C57BL/6J chromosome 18, GRCm38.p6 C57BL/6J;

>NC_007121.7:c14929659-14871136 *Danio rerio* strain Tuebingen chromosome10, GRCz11, Primary Assembly]



NB: SMAD2 transcript variant proline rich protein.

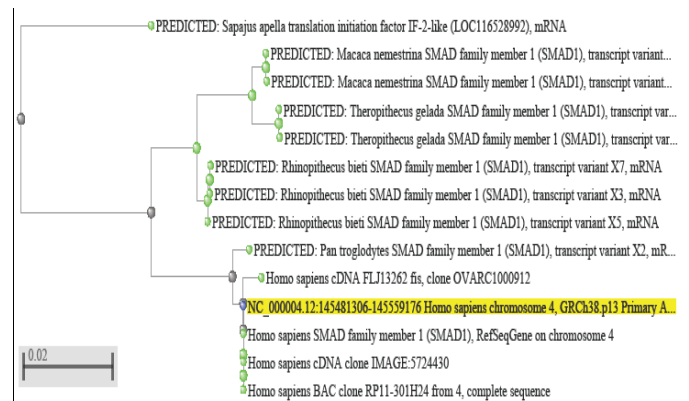
Figure 9: Nucleotide Blast results for SMAD2

10) SMAD1:

>NC_000004.12:145481306-145559176 *Homo sapiens* chromosome 4, GRCh38.p13 Primary Assembly;

>NC_000074.6:c79399428-79338395 *Mus musculus* strain C57BL/6J chromosome 8, GRCm38.p6 C57BL/6J;

>NC_007112.7:c35929002-35897043 *Danio rerio* strain Tuebingen chromosome 1, GRCz11 Primary Assembly]



NB: SMAD1 transcript of *Homo sapiens* cDNA member.

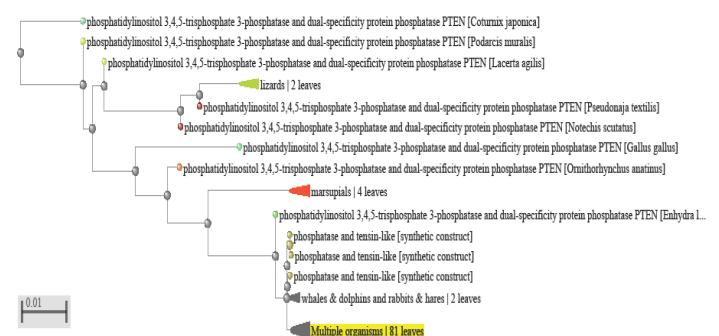
Figure 10: Nucleotide Blast results for SMAD1

B. Blastp (protein-protein BLAST)

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

1) PTEN:>sp|P60484|PTEN_HUMAN

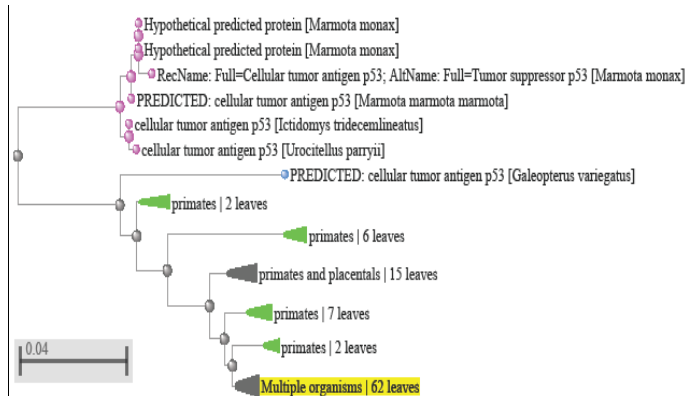
Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN OS=Homo sapiens OX=9606 GN=PTEN PE=1 SV=1;
>sp|O08586|PTEN_MOUSE Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN OS=Mus musculus OX=10090 GN=Pten PE=1 SV=1;
>tr|Q6TGR5|Q6TGR5_DANRE Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN OS=Danio rerio OX=7955 GN=ptenb PE=2 SV=1



BLASTP

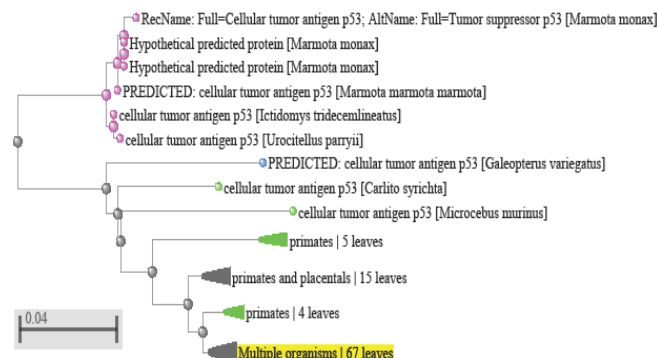
Figure 11: Blastp results for PTEN

2) **P53**: >sp|P04637|P53_HUMAN Cellular tumor antigen p53 OS=Homo sapiens OX=9606 GN=TP53 PE=1 SV=4; >sp|P02340|P53_MOUSE Cellular tumor antigen p53 OS=Mus musculus OX=10090 GN=TP53 PE=1 SV=4; >sp|P79734|P53_DANRE Cellular tumor antigen p53 OS=Danio rerio OX=7955 GN=tp53 PE=1 SV=1



BLASTP

Figure 12: Blastp results for P53



Multiple Sequence Alignment

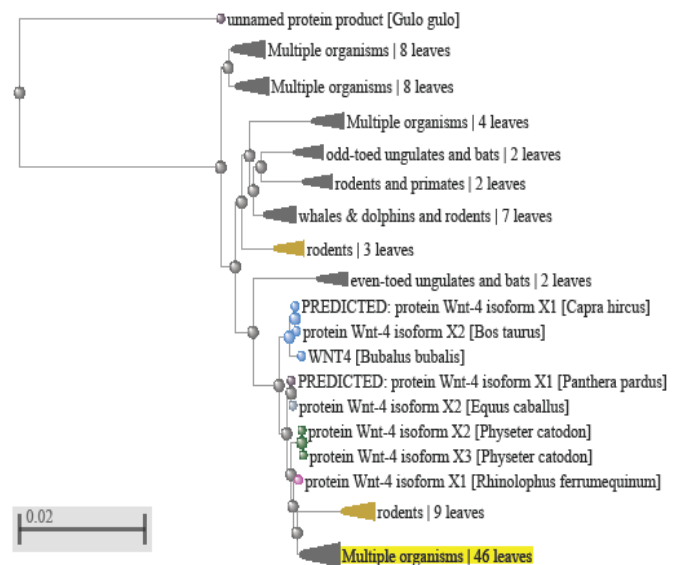
Figure 13: MSA results for P53

3) **WNT4**: >sp|P56705|WNT4_HUMAN Protein Wnt-4 OS=Homo sapiens OX=9606 GN=WNT4 PE=1 SV=4; >sp|P22724|WNT4_MOUSE Protein Wnt-4 OS=Mus musculus OX=10090 GN=Wnt4 PE=1 SV=1; >sp|P47793|WNT4A_DANRE Protein Wnt-4a OS=Danio rerio OX=7955 GN=wnt4a PE=2 SV=1



BLAST P

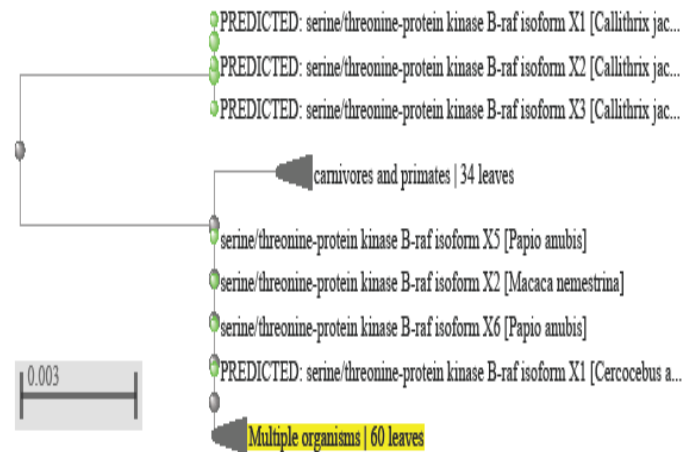
Figure 14: Blastp results for WNT4



Multiple Sequence Alignment

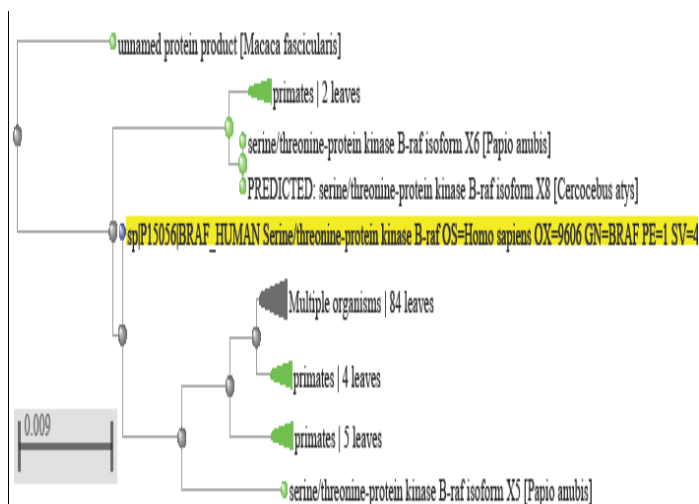
Figure 15: MSA results for WNT4

4) **BRAF**: >sp|P15056|BRAF_HUMAN Serine/threonine-protein kinase B-raf OS=Homo sapiens OX=9606 GN=BRAF PE=1 SV=4; >sp|P28028|BRAF_MOUSE Serine/threonine-protein kinase B-raf OS=Mus musculus OX=10090 GN=Braf PE=1 SV=3; >tr|Q75V91|Q75V91_DANRE Serine/threonine protein kinase BRAF OS=Danio rerio OX=7955 GN=braf PE=2 SV=1



BLAST P

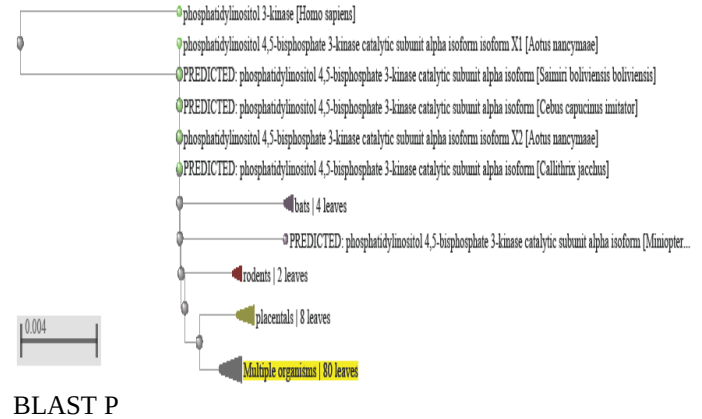
Figure 16: Blastp results for BRAF



Multiple Sequence Alignment

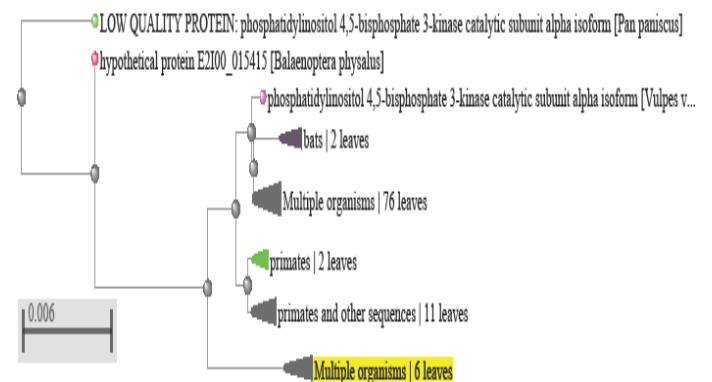
Figure 17: MSA results for BRAF

5) **PIK3CA**: >sp|P42336|PK3CA_HUMAN Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform OS=Homo sapiens OX=9606 GN=PIK3CA PE=1 SV=2; >sp|P42337|PK3CA_MOUSE Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform OS=Mus musculus OX=10090 GN=Pik3ca PE=1 SV=2; >tr|F1QAD7|F1QAD7_DANRE Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit OS=Danio rerio OX=7955 GN=pik3ca PE=3 SV=2



BLAST P

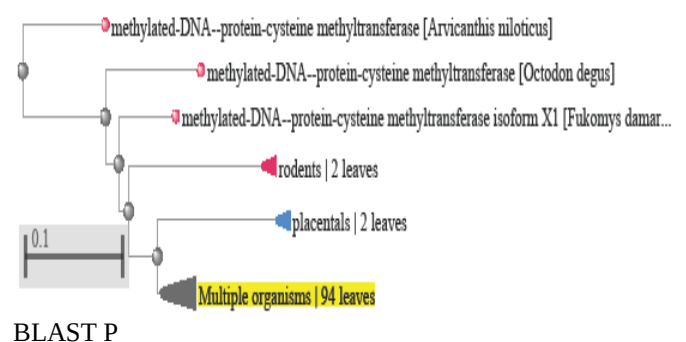
Figure 18: Blastp results for PIK3CA



Multiple Sequence Alignment

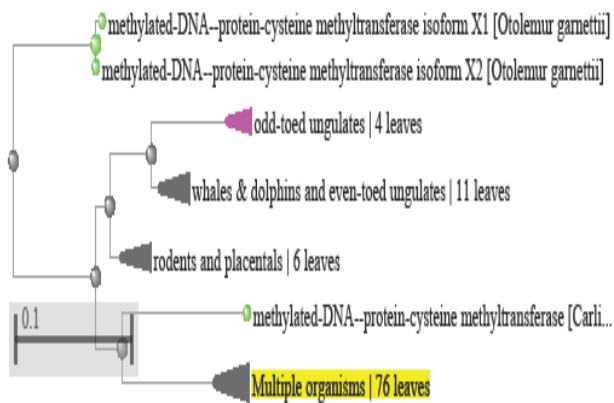
Figure 19: MSA results for PIK3CA

6) **MGMT** >sp|P16455|MGMT_HUMAN Methylated-DNA--protein-cysteine methyltransferase OS=Homo sapiens OX=9606 GN=MGMT PE=1 SV=1; >sp|P26187|MGMT_MOUSE Methylated-DNA--protein-cysteine methyltransferase OS=Mus musculus OX=10090 GN=Mgmt PE=1 SV=3; >tr|Q568B0|Q568B0_DANRE Mgmt protein (Fragment) OS=Danio rerio OX=7955 GN=mgmt PE=2 SV=1



BLAST P

Figure 20: Blastp results for MGMT



Multiple Sequence Alignment

Figure 21: MSA results for MGMT

IV. CONCLUSION

Endometrial cancer, the most common type of cancer affecting the female reproductive organs, starts within endometrium of the uterus. With reference to studies done so far, the genetic disorders can cause endometrial cancer with an overall hereditary contribution of around 2–10%. In this study, seven genes important with reference to endometrial cancer were analysed for conservation of both nucleotide as well as protein sequences across two model systems (mouse and zebra fish) which are widely used for studies on diseases occurring in human beings. The analyses indicated that around 70%-85%

identity is maintained between these genes in human beings (*Homo sapiens*), Mice (*Mus musculus*) and Zebra fish (*Danio rerio*).

ACKNOWLEDGEMENTS

The authors are thankful to Chancellor, Techno India University, West Bengal for providing the necessary infrastructural facilities.

REFERENCES

- [1] Lentz, S.S. Advanced and recurrent endometrial carcinoma: hormonal therapy. *Seminars in Oncology*, 1994, **21**, 100–106.
- [2] Mikuta, J.J. International Federation of Gynecology and Obstetrics staging of endometrial cancer. *Cancer*, 1993, **71**: 1460–1463.
- [3] Rose, P.G. Endometrial carcinoma. *New England Journal of Medicine*, 1996, **335**, 640–649.
- [4] Murali, R., Soslow, R.A. and Weigelt, B. Classification of endometrial carcinoma: more than two types. *Lancet Oncology*, 2014, **15**(7):e268-78.
- [5] Doll, A., Abal, M., Rigau, M., et al. Novel molecular profiles of endometrialcancer-new light through old windows. *Jourtnal of Steroid Biochemistry and Molecular Biology*, 2008, **108**(3-5):221-229. Epub 2007 Sep 15.
- [6] Prat, J., Franceschi, S.. Cancers of the female reproductive organs. In:StewartBW, Wild CP, editors. World Cancer Report 2014. Lyon, France: InternationalAgency for Research on Cancer; 2014.
- [7] Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 2000, **7**(1-2):203-14.
- [8] Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R. and Schäffer, A.A. Database Indexing for Production MegaBLAST Searches. *Bioinformatics*, 2008, **24**:1757-1764.