iMedPub Journals http://www.imedpub.com

Comparative Genome Wide Identification, Gene Structure and Phylogeny Analysis of POU5F1 Transcription Gene in *Mus musculus, Mus caroli, Onychomys torridus* and *Arvicanthis niloticus* 

## Abstract

House mouse (Mus musculus) is a model organism and has role in understanding and studying human diseases. Mus musculus and human have many conserved domains. POU domain is one of those conserved domains and in this recent study, POU5F1 gene, part of POU domain is studying. POU5F1 is a POU domain, class 5, transcription factor 1. It has role in pluripotency development in embryo, embryonic and adult stem cell development, in protein binding, miRNA binding, chromatin DNA binding, RNA polymerase 2 cis-regulatory region sequence-specific DNA binding and involve in reprogramming of cells. In this study, the phylogenetic relation of Mus musculus with other subfamily members is found to use them as model alternate to Mus musculus. Initially the comparative genome wide analysis of POU5F1 gene was performed in Mus musculus specie, Mus caroli, Onychomys torridus and Arvicanthis nilotius. This study focused on the identification of POU5F1 gene in Mus musculus (common house mouse) among Mus caroli (Ryukyu mouse), Onychomys torridus (Southern grasshopper mouse), and Arvicanthis niloticus (African grass rat) based on the bioinformatics techniques and tools to explore evolutionary relationship. For the identification genome sequence retrieved from NCBI, for conserved domain analysis, PROSITE-Expasy, SMART Protein and Pfam were used. Multiple sequence alignment and phylogenetic analysis was performed by Clustal Omega and MEGA-X. The results showed that in Mus musculus POU5F1 gene is present on chromosome number 17 and has 6 exons, In Arvicanthis niloticus POU5F1 gene is present on chromosome number 20 and has 5 exons, Mus caroli POU5F1 gene is placed on chromosome number 17, and has 6 exon, and Onychomys torridus POU5F1 gene has 5 exons with unknown chromosome number. Finally, our results provide important genomic suggestions for upcoming studies of biochemical, physiological and phylogenetic understanding on POU5F1 gene among other species.

**Keywords:** POU5F1; *Mus musculus*; Multiple sequence alignment; Phylogenetic analysis; Oct-4; *Mus caroli; Onychomys torridus; Arvicanthis nilotius* 

**Received:** April 29, 2021; **Accepted:** May 13, 2021; **Published:** May 20, 2021

## Introduction

*Mus musculus* is the specie which has largest contribution in modern research. It could bread anywhere in the world as compared to previous fancy mice which could bread only in Europe and Asia. *Mus musculus* is a wild type and had commensalism to humans, therefore could be used in study of human diseases. It considered as the prime mammalian laboratory model [1]. The *Mus musculus* genome sequence is much related to human genome sequence and about 14% smaller than human genome with function homology.

#### **Mahwish Amin**\*

Department of Molecular Biology, Government College University Faisalabad, Faisalabad, Pakistan

\***Corresponding author:** Mahwish Amin, Department of Molecular Biology, Government College University Faisalabad, Faisalabad, Pakistan, Tel: 923247185617; E-mail: mahwishamin4111@gmail.com

**Citation:** Amin M (2021) Comparative Genome Wide Identification, Gene Structure and Phylogeny Analysis of POU5F1 Transcription Gene in *Mus musculus, Mus caroli, Onychomys torridus* and *Arvicanthis niloticus.* J Mol Biol Biotech Vol. 6 No.3: 04.

Most of genes in the mouse and human are conserved [2]. Out of that conserved genes, One POU gene family is decided for study. The *Mus caroli* is also a species of rodents and having habitat China, Cambodia, Indonesia, Japan, Laos, Malaysia, Taiwan, Thailand and Vietnam. The Onychmys torridus is also rodent family member and habitat in New Mexico, California and United States. The *Arvicanthis niloticus* has use as animal model for Type 2 Diabetes mellitus.

 $\label{eq:powerserv} POU \ domain \ is a structural motif of transcription factors which are composed of the POU-specific domain and the POU homeodomain.$ 

POU family has role in pluripotency development in embryo, embryonic and adult stem cell development, in protein binding, miRNA binding, chromatin DNA binding, RNA polymerase 2 cisregulatory region sequence-specific DNA binding and involve in reprogramming of cell's neurogenesis, brain formation and animal development. POU5F1 is the member of POU domain family. In mammals including humans, POU5F1 transcription factor belong to one of the largest gene families that play vital role in control of pluripotency. The pluripotent cells include primordial germ cells, embryonic germ cell and inner cell mass [3].

The genome wide analysis of POU5F1 gene in *Mus musculus* will help in future to analyse the undifferentiated cells at embryo level in humans. In this recent study, initially the genome wide analysis of POU5F1 gene in *Mus musculus* species was performed and based on the NCBI retrieved genome sequence. This study focused on the identification of POU5F1 gene in *Mus musculus* and perform phylogenetic analysis among *Mus caroli* (Ryukyu mouse), *Onychomys torridus* (Southern grasshopper mouse) and *Arvicanthis niloticus* (African grass rat) to explore evolutionary relationship.

# **Materials and Methods**

#### Identification of POU5F1 gene in Mus musculus

Nucleotide sequences of POU5F1 gene in *Mus musculus, Mus caroli, Onychomys torridus and Arvicanthis niloticus* were retrieved from NCBI in FASTA format (NC\_000083.7, NC\_047677.1, NC\_034586.1, NW\_02341287.0). Homologous sequences were then obtained from BLASTN. The variants which have E-value less than 0 were considered for the homology analysis. All the variants were checked for sequence homology by the graphic summery and alignment. Conserved domains of the nucleotide sequences were found by Pfam.

## **Conserved domains analysis**

Conserved domains are recurring units in the molecular evolution that can be determined by sequence analysis. It informs about pattern or motif to see its similarity in other polypeptide sequences.

The protein sequence of POU5F1 gene in *Mus musculus* was retrieved in FASTA format by NCBI (NP\_001239381.1, NP\_038661.2) *Mus caroli* (XP\_021005284.1, XP\_021005285.1), *Onychomys torridus* (XP\_036032305.1), and *Arvicanthis niloticus* (XP\_034379773.1). Conserved domains in the protein sequence were determined through PROSITE-Expasy, SMART Protein and Pfam [4-6]. Genome information related to cDNA sequence, length of protein, ORF (Open Reading Frame), chromosome location was also retrieved from NCBI. The NCBI BLAST retrieved sequence of *Mus musculus* was used as query sequence to obtain nearly related matches and precise results. For the confirmation of presence of signature POU5F1 gene in *Mus musculus, Mus caroli, Onychomys torridus and Arvicanthis niloticus* were tested by SMART Protein database and Pfam, PROSITE-Expasy [4-6].

#### **Multiple sequence alignment**

Clustal Omega tool, MEGA-X and NCBI were used to see the related genes or proteins to find the evolutionary relationship between genes and to identify functionally or structurally related genes in other species. Sequences of *Mus caroli, Onychomys torridus,* and *Arvicanthis niloticus* were used. BLAST was used to obtain multiple sequences, followed by sequence alignment by ClusterOmega and MEGA-X respectively [7,8]. The evolutionary relationship of POU5F1 gene was found in *Mus musculus, Mus caroli, Onychomys torridus,* and *Arvicanthis niloticus*. First, multiple sequence alignment was performed by MEGA. X by using CLUSTAL W, and another database Cluster Omega that was used online.

#### Gene structure analysis

For the analysis of gene structure in *Mus musculus, Mus caroli, Onychomys torridus* and *Arvicanthis niloticus*. Gene Structure Display Server was used [9]. It informs about the gene features like exons, intron and conserved elements. It can be used for both genomic and cDNA sequence phylogenetic tree construction. For the construction of gene structure, cDNA sequence and genomic DNA sequence in FASTA format were put in Gene Structure Display Server database and the result showed in the form of gene structure.

# Phylogenetic analysis and phylogenetic tree construction

The phylogenetic tree shows the evolutionary relationship between the related organisms and also inform about differences between them. Phylogenetic analysis was performed by using MEGA-X software. To see how many times the branches repeat in phylogenetic tree bootstrap kept 1000 replicates and pairwise deletion and poison correction was used to construct phylogenetic tree. During phylogenetic tree construction Maximum Likelihood method was used. Phylogenetic tree was also generated by database Cluster Omega, MEGA-X and IQ-tree [7,10]. Then using three databases, rooted phylogenetic tree was built with *Mus musculus, Mus caroli*, Onychomystorridus, Arvicanthisniloticus and 13 other subfamily members.

## Results

## Identification of POU5F1 gene in *Mus musculus, Mus caroli, Onychomys torridus* and *Arvicanthis niloticus*

The POU5F1 protein sequence of *Mus musculus, Mus caroli,* Onychomystorridus, and Arvicanthisniloticus was retrieved from NCBI. The complete information is available in **Table 1**.

2021

Vol 6 No.3:04

Proposed names	Gene symbol	Protein accession	RNA accession	Exons	Chr	ORF length	Amino acid length	Start of genomic location	Conserved domains in protein sequence
<i>Mus musculus</i> POU domain	Pou5f1	NP_001239381.1	NM_001252452.1	6	17 B1;17 18.69cM	666	221	35816929	smart 00352
Arvicanthis niloticus POU domain	LOC117724197	XP_034379773.1	XM_034523882.1	5	20	1059	352	50621362	Smart00352 Pfam00046
<i>Mus caroli</i> POU domain	Pou5f1	XP_021005284.1	XM_021149625.1	6	17	1059	352	32277972	Smart00352 Pfam00046
<i>Mus caroli</i> POU domain	Pou5f1	XP_021005285.1	XM_021149626.1	6	17	666	221	32277972	Smart00352 Pfam00046
Onychomys torridus POU domain	Pou5f1	XP_036032305.1	XM_036176412.1	5	Unknown	1080	359	2412441	Smart00352 Pfam00046

**Table 1:** S1 Proposed nomenclature and important features of Pou5f1 genes.

#### **Conserved domain analysis**

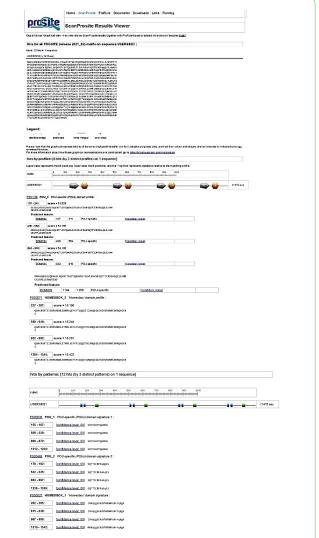
*Mus musculus, Mus caroli, Onychomys torridus,* and *Arvicanthis niloticus* have same two conserved domains smart00352 and Pfam00046. *Mus musculus* and *Mus caroli* have two isoforms, *Onychomys torridus,* and *Arvicanthis niloticus* have single isoform. The ORF is also obtained online by ORF finder, *Mus musculus* and *Mus caroli* have two ORF values 1059 and 666. *Arvicanthis niloticus* has single ORF that is 1059 and *Onychomys torridus* also has single ORF with value 1080. The Pfams, SMART Protein and PROSITE-Expasy results are attached below **(Figures 1-3)**.

Show the deta	aled description of this res fam-A matches to your se				licant	and 80		nificant	0	a the	-			•
	ch options and sequence earch form to look for Pfa				uence									
Significant Show or hide a	Pfam-A Matches													
Family	Description	Entry type	Clan			Align Start		HM From		HMM length	Bit score	E- value	Predicted active sites	Show/hide alignment
Pou	Pou domain - N-terminal to homeobox doma	Domain	CL0121	3	74	5	74	з	72	72	118.4	9.3e- 35	n/a	Show
Pou	Pou domain - N-terminal to homeobox doma	Domain	<u>CL0121</u>	355	426	357	426	з	72	72	118.4	9.3e- 35	n/a	Strow
Pou	Pou domain - N-terminal to homeobox doma	Domain	CL0123	707	778	709	778	з	72	72	118.4	9.3e- 35	n/a	Show
Pou	Pou domain - N-terminal to homeobox doma	Domain	<u>C10123</u>	1059	1130	1051	1130	з	72	72	118.4	9.3e- 35	n/a	Show
Pow	Pou domain - N-terminal to homeobox doma	Domain	<u>CL0121</u>	1280	1351	1292	1751	э	72	72	118.4	9.3e- 35	n/a	Show
Pou	Pou domain - N-terminal to homeobox doma	Domain	<u>CL0123</u>	1638	1709	1640	1709	з	72	72	115.3	8.5e- 34	n/a	Show
Homeodomain	Homeodomain	Domain	<u>CL0123</u>	93	149	93	149	1	57	57	62.4	2.6e- 17	n/a	Show
Homeodomain	Homeodomain	Domain	CL0123	445	501	445	501	1	57	57	62.4	2.6e- 17	n/a	Silvow
Homeodomain	Homeodomain	Domain	CL0121	797	853	797	853	1	57	57	61.2	6.2e- 17	n/a	Show
Homeodomain	Homeodomain	Domain	<u>CL0121</u>	1149	1205	1149	1205	1	57	\$7	61.4	5.2e- 17	n/a	Show
Homeodomain	Homeodomain	Domain	<u>CL0123</u>	1370	1426	1370	1426	1	57	57	61.4	5.2e- 17	n/a	Show
Homeodomain	Homeodomain	Domain	CL0123	1728	1784	1728	1784	1	57	57	60.9	7.7e-	n/a	Show

torridus and Arvicanthis niloticus by Pfam.

#### **Multiple sequence alignment**

Multiple sequence alignment was generated by NCBI and CLUSTAL Omega. They are aligned pair-wise. Multiple sequences are a phylogeny-aware alignment and that information use in the generation of phylogenetic tree (Figures 4-6) [11].



**Figure 2:** 2a-2b Conserved domains of *Mus musculus, Mus caroli, Onychomys torridus,* and *Arvicanthis niloticus* by PROSITE-Expasy.

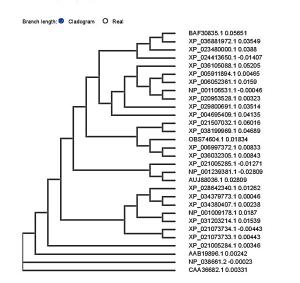
# Journal of Molecular Biology and Biotechnology

**2021** Vol 6 No.3:04

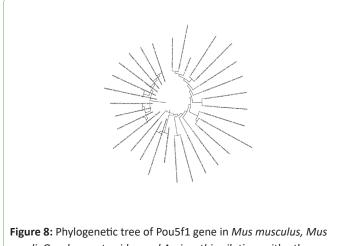
Arrithmatical     w is in in in in in in in	CLUSTAL 0(1.2.4) multiple sequence alignment
Domain architecture analysis Display all proteins with similar:	BAF30835.1 0 XP 825(65988.1 0 XP 825(65988.1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
Dumin magnitude: Provide having all the dominist to the party in the name offer. Additional densities are allowed     Tomin magnitude: The shows with the mechanic equivalents that a fact encore orget of the dominism of the array.     The DALVE degrane down requestes a summary of the truther down bodies. Densities: with a summary toring the struther down bodies. Densities: with a summary toring the struther down bodies are also been associated as a summary of the truther down bodies. Densities: with a summary toring the struther down bodies are also been associated as a summary of the truther down bodies are associated and down bodies are associated as a summary toring the struther down bodies are associated as a summary toring the struther down and and are associated as a summary toring the struther down and are associated as a summary toring the struther down and are associated as a summary toring the struther down and are associated as a summary toring the struther down and are associated as a summary toring the struther down and are associated as a struther down and and are associated as a struther down and a	XP         84695489.1         MAGHLTSDFAFSPPGGGGDGGFGGFEQGWODPRANLSFPGPGGSGIGPGVGGTEVMGT         68           XP         83681972.1
Confidently predicted domains, repeats, motifs and features: Peatures NOT shown in the diagram: 🕢	XP         0e66652361.1         ModeLLASCPAFSPPP6GGGDROPGPERAVDPRATURSFQGPPCSGGTGPGVPACALWGL         69           MP         0e110e531.1         ModeLLASCPAFSPPPGGGGDROPEPCAVDPRATURSFQGPCSGGTGPGVPACALWGL         69           XP         22093328.1         ModeLLASCPAFSPPFGGGGDROPEPCAVDPRATURSFQGPCSGGTGPGVPACALWGL         69           XP         22093328.1         ModeLLASCPAFSPPFGGGGDROPEPCAVDPRATURSFQGPCSGGGTGPCAVDPACALWGL         69
Sum that fail Labor	XP_0234880808.1        0           XP_0234880808.1         MAGHLASDFAFSPPPGG-GD0SAGLEPGAVDPRTMLSFQGPPGGPGIGPGSEVLGI         5           XP_02342349.1         MAGHLASDFAFSPPPGG-GD0SAGLEPGAVDPRTMLSFQGPPGGPGIGPGSEVLGI         55           XP_034379773.1         MAGHLASDFAFSPPPGG-GD0SAGLEPGAVDPRTMLSFQGPPGGPGIGPGSEVLGI         55
Instrumentic 207 200 VA         ELL         35 11 3 1010 meteold in the instrument of the instrum	xp         643884897.1         ModelLASSPARSPPP6G-COSCAGLEPGAVDPRFNLstPGGPESGPEIGPGSEVLGI         55           xp         621097373.1         ModelLASSPARSPPFGG-COSCAGLEPGAVDPRFNLstPGGPEGGPEGGPGSEVLGV         6           xp         621097373.1         ModelLASSPARSPPFGG-COSCAGLEPGAVDPRFNLstPGGPEGGPEGGPGSEVLGV         55           xp         621092173.1         ModelLASSPARSPPFGG-COSCAGLEPGAVDPRFNLstPGGPEGGFEGGPGSEVLGV         55           xp         621092173.1         ModelLASSPARSPPGG-COSCAGLEPGAVDPRFNLstPGGPEGGFEGGPGSEVLGI         55
Intermultiple 90         90         NA         AZZ         90         90         Precision           intermultiple 90         40         NA         EZI         63         75         10         Precision           intermultiple 90         40         NA         EZI         63         75         10         Precision           EXII.         464         1003         1004         F         10         Precision           EXII.         464         1003         1004         F         10         Precision           EXII.         1004         1004         F         1004         Precision         Precision           EXII.         1004         1004         F         1004         Precision         Precision           EXII.         1004         1004         Precision         Precision         Precision         Precision           EXII.         10041         Precision         Precision         Precision         Precision         Precision           EXII.         10041         Precision	XP_031203214.1         MAGHLASDFAFSPPPGG-GDGSAGLEAGWDPRTMLSFQ8PPGGPGIGPGSEVLGI         55           XP_021005285.1         0 <td< td=""></td<>
collad coll         12701397NA         CDC07_M         2088219599800         threshold           PCM         13091380 288-077         LA         246-62111330         threshold           PCM         13091380 288-077         LA         246-62111330         threshold	NP_038661.2         McMLMLSDFAFSPPGG-G0DSAGLEPGMVDPFTLSFQ0PPGGFQIGPGSEVLGI         55           AA819886.1         McMLASDFAFSPPGG-G0DSAGLEPGMVDSFTLLSFQ0PPGGFQIGPGSEVLGI         55           CCA36682.1
Inscrimulation 1999 1659 No.         Zad. Ent.         2007 1973 1970 Mercheld           Inscrimulation 1999 1659 No.         Zad. 2002         2605 2073 400 Mercheld           Inscrimulation 1990 1659 No.         Zad. 2002         2605 2074 400 Mercheld           Inscrimulation 1990 1659 No.         Zad. 2002         2010 2010 Mercheld           Inscrimulation 1990 1659 No.         Zad. 2002         2010 2010 Mercheld           Inscrimulation 1990 1659 No.         Zad. 2002         2010 2010 Mercheld	NP_00123938.1         0           AUJB8986.1         0           yp 021507837.1         MAGPLAPDFAFSPPPGG-CBGSGGLEPGANDPRTNLSFQCPPGVPGVFQVPQVAPGSEVLGT           yp 031598787.1         MAGPLAPDFAFSPPPGG-CBGSGGLEPGANDPRTNLSFQCPPGVPGVFQVPQVAPGSEVLGT           yp 031599787.1         MAGPLAPDFAFSPPPGG-CBGSGGLEPGANDPRTNLSFQCPPGVFQVFQVAPGSEVLGT
DQL         1141-1083 LBM-dF         DQL         1259-3041 LBM         Benedity           EXEC         1164 LBM         116	xp         e3819996.1         ModelLASDPAFSPPPGG-COSSGLEPGANDPRTLISTGOPPGGETGEPVGROSSTVLGT         59           OBS74649.1         ModelLASDPAFSPPGG-COSSGLEPGANDPRTLISTGOPPGGETGEPVGROSSTVLGT         59           xp         e8699737.1         ModelLASDPAFSPPPGG-COSSGLEPGANDPRTLISTGOPPGROSTPVLGT         59           xp         e86929737.1         ModelLASDPAFSPPPGG-COSSGLEPGANDPRTLISTGOPAGETGEPVGROSSTVLGT         59           xp         e3692387.3         ModelLASDPAFSPPPGG-COSSGLEPGANDPRTLISTGOPAGETGEPVGROSSTVLGT         59
Diff.         Diff.         Diff.         Diff.         Diff.           CLL         10110121200-04-0         ZZL         20100101000-06-04         Diff.         Diff. </td <td>BAF30835.1</td>	BAF30835.1
Figure 3: Conserved domains of Mus musculus, Mus caroli,	XP         936165888.1
Onychomys torridus and Arvicanthis niloticus by SMART-Protein.	XP         080682361.1         PSCPPPYDLCGGM-AYCAPQVGVGPVPPGGLETPQPEGEGA-AQVESNSEGASPIP-CAA         49           XP         080682361.1         PSCPPPYDLCGGM-AYCAPQVGVGVPVPGGLETPQPEGEGA-AQVESNSEGASPIP-CAA         117           NP         081186531.1         PACPPPYDCGGM-AYCAPQVGVGVPVPGGLETPQPEGEGA-AQVESNSEGASPIP-CAA         117
	XP_02095328.1         PACPPPYDFC5GM-ArCAPQVGVGLVPQ5GELETPQF6EAG-ARVESNSEGASPEP-CAA         117           XP_024413659.1
M.muzoukasmRNA for Oct-4 protein Sequence ID: XS2837 / Length: 1248 Number of Matcher: 5	XP_025672340.1         SPCPDAYEFCG0H-AYCGPQVGLGUVPQVWETLQPEGQM-ARVESNEEESSEPCAAR         113           XP_0434379773.1         SPCAPAYEFCG0H-AYCGPQVGLGUVPQQAVETLQPEGQM-ARVESNEETSESECAAS         113           XP_043438497.1         SPCAPAYEFCG0H-AYCGPQVGLGUVPQAVETLQPEGQM-ARVESNEETSESECAAS         113           XP_043438497.1         SPCAPAYEFCG0H-AYCGPVGLGUVPQAVETLQPEGQM-ARVESNEETSESECAAS         113           XP_024973734.1         SPCAPAYEFCG0H-AYCGPVGLGUVPQAVETLQPEGQM-ARVESNEETSESECAAS         10
Range 1: 743 to 1225 Score Expect Identities Gaps Strand Frame	XP_02187373.1         SPCPPAYEFC6GWAAYCGROVGLGLVPQVKVELUPPEGQG-ARVESIKEGASSEPCADR         114           NP_001009178.1         SPCPPAYEFC6GW-AYCGPVGLGLVPQVKVETLQPEGQM-ARVESIKEGASSEPCADR         113           XP_031283214.1         SPCPPAYEFC6GW-AYCGPVGLGLVPQVKVETLQPEGQM-ARVESIKEGASSEPSAR         113
885 bin(483) 0.0() 482493(100%) 04430(9) Pilus <sup>Pilus</sup> Query 3582185 Stroff Constant Factor (1997) 1997 (2007) 1997 (2007) 1997 (2007) 1997 (2007) 1997 (2007) 1997 (2007) Sigt: 743 distributing and a strong the strong and a strong	XP_02108238.1         0           XP_02108238.1         SPCPPAYEFC60M-AXC6PQ/05LGLVPQ/0VFTLQPE6QAG-ARVESINSEGASSEPCADR         113           NP_03866.1.2         SPCPPAYEFC60M-AXC6PQ/05LGLVPQ/0VFTLQPE6QAG-ARVESINSEGASSEPCADR         113           AM819866.1         SPCPPAYEFC60M-AXC6PQ/05LGLVPQ/0VFTLQPE6QAG-ARVESINSEGASSEPCADR         113
Open-rg         35821284         TOCCMERGEARMENT ATTACKET TOCCMERGEARGE         55821284           S5yte         883         TOCCMERGEARGEARGEAT TOCCMERGEARGEAT         5612164           Open-rg         35821284         TOCCMERGEARGEARGEAT         5612164           Open-rg         35821284         TOCCMERGEARGEARGEAT         5621264	ABL25000.1 SPCPATECX0MALCVQVQCLVPQVVELVPEQAGAAAXESNESISSFCADB 13 CA35652.1 SPCPATECX0AAXCCQVQCLUPQVQVELVPEQAGAAAXESNESTSSFCADB 85 NP 001239381.1 0
55/rt 863 CCHTANICIANTHICIANTHICIANTHICIANTHICIANTHICIANTHICIAN 922 Ourry 35821865 CTCHTANTICCTURTICCTURTICCTURTICCTURTICTUR 55821824 Sector 923 HTMLINICTURTICCTURTICCTURTICCTURTICTURTICTUR 982	Figure 5: Multiple sequence alignment of Mus musculus, Mus
sejete 932 CHARTMANDELTHICHMAAAAAAACHUUCHUUTAHAAAAAAAAAAAAAAAAAAAAAA	caroli, Onychomys torridus and Arvicanthis niloticus by CLUSTAL-
Query 35821485 CARGEGERANDAL CHARGENCTINGCT TRACT TRACT TO CONTRACT A SECONDAL SECON	Omega.
George Statistics Human Constitution Constitution Constitution (Constitution Constitution) Statistics Softer 1989 Human Constitution Constitution Constitution (Constitution) Statistics George Statistics Constitution Constitution (Constitution) Statistics Softer 1983 Human Constitution (Constitution) Statistics (Constitution) Statistics Softer 1983 Human Constitution (Constitution) Statistics (Constitution) Statistics Statistics (Constitution) Statistics (Constitution) Statistics (Constitution) Statistics Statistics (Constitution) Statistics (Constitution) Statistics (Constitution) Statistics Statistics (Constitution) Statistics	
Query 35821665 GTA 35821667 Sbjct 1223 GTA 1225	
PREDOTED-ArActanthia holdioux POU domain, class 5. Instruction factor 1-like (LOC11724167), mRNA Sequence ID:XM, GM223882.1 Length: 1059 Number of Matches: 4 Range 1: 1o 384	
Store         Broyset         Mendles         Gray         Brendle         Transe           6V6.bit/2(27)         24:691         356534(956)         DOR4(956)         DVRPF/ss           0ummy         355134(956)         356534(956)         DVRPF/ss         DVRPF/ss           0ummy         355134(956)         ATGGC GROUP ACCESSION CONCELLATION CONCELATION CONCELLATION CONCELATION CONCELLATION CONCELATION	
Query         35817057         TORCHARGETURANTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	
Owery 38437317 characterizationancherrinaworthreaderth	
SBjet 1m. ddfhafdaff Undaffadaffadaffadaffadaffadaffadaffadaf	
Query         35817287         Additional Control Contron Control Control Contron Control Control Contron Con	Mit Augment Taplene Phyloperies and phyloperies (and
Query 3931237 AND/DEMOLECULE (TURNING) 59612500 594-te 141 AMARTINI TURNING 1546 PREDICTED: Adventishe politica POUL (oncar:, class 5, transcriptica for 548 e.g.OC117724197), mRNA Despusee DJ: MUI (3452421 Length: 105) MUINT (4414) Despusee DJ: MUI (3452421 Length: 105) MUINT (4414)	
Range 1: Ito 384 Score Expect Identifies Gaps Strand Frame	Expanding     Constraints     Constraint     Constraint     Constraint     Constraint     Constraints
604.bb/(227) 2e-1670) 3555.94(95%) 0:034.90%) PumPlus Query 3551.9997 Affectingsackertingertingsachertings	L CLORENCE (Transferrence) en la provinción de la provinci de la provinción de la provinción de la provinción de la provinció
Query         383/372         Topological product intervention of the control of the	14 AP_04009/73 100 demait cass Stranogene factor flatin envypen 17.90 (2017) 23 100 demait cass Stranogene factor flatin envypen 19.00 H LASSIA FS PP CODDISACT P WODP TWLSFOLPP OF CODDISACT P WODP TWLSFOLPP OF CODDISACT P WODP TWLSFOLPP OF CODDISACT P AT CODD
Sejet 121 (Clashidder Challen Call Chalden Hander Herner Challen Hander Herner Hander	1 2 2 5 2012 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1
Query         35837237         CICIMANT EXECUTION CONCERNMENT AND CONCERNA CONCERNMENT AND CONCERNA C	Figure 6: Multiple Sequence Alignment of Mus musculus, Mus
Owery Statistics And Antonio Statistics (Statistics Statistics Sta	caroli, Onychomys torridus and Arvicanthis niloticus By MEGA-X.
CHTCP 4	·
Score         Expect         Identities         Days         Brand         Frame           477 bits(258)         3e-1220         4014468(25%)         104450(25%)         Puol*Pus           0evry         3981693         014460(25%)         104450(25%)         Puol*Pus	
SB5ct 1 det inder den für frageringen för den för state i den för som	Phylogenetic analysis and phylogenetic tr
Query 35817851 eAFGGRY.McGaeGGFTGGGYCGGOCTGGGYCGCTCGGACTCTCGACTCCAA, 35817110	construction
0.0077 33837313 000711705051050715000	The characterization of the evolutionary relationship for POUS
Query 35837111 GOLTITUGEGETTEGGA	
Querry         39317131         GOOD CONSTRUCTION         1000 CONSTRUCTION         2000 CONSTRUCTION         2000 CONSTRUCTION           Querry         39317139         CONSTRUCTION         1000 CONSTRUCTION         3000 CONSTRUCTION </td <td>gene among Mus musculus, Mus caroli, Onychomys torridus a Arvicanthis niloticus was performed by generating the root</td>	gene among Mus musculus, Mus caroli, Onychomys torridus a Arvicanthis niloticus was performed by generating the root
Query         1981/7111         Open (1)         1981/7113         500         1981/7134         500         1981/7134         500         1981/7134         500         1981/7134         500         1981/7134         500         1981/7134         500         1981/7134         500         1981/7134         1981/7134         500         <	The characterization of the evolutionary relationship for POU5 gene among <i>Mus musculus, Mus caroli, Onychomys torridus a</i> <i>Arvicanthis niloticus</i> was performed by generating the root tree of four species by the maximum likelihood (ML) method. T figure showed that 31 subfamily members have homology ge

Results for job clustalo-l20210418-125020-0955-36365813-p2m

Phylogenetic Tree This is a Neighbour-joining tree without distance corrections.



**Figure 7:** Phylogenetic tree of Pou5f1 gene in *Mus musculus, Mus caroli, Onychomys torridus and Arvicanthis niloticus* with other subfamily members having POU domain by CLUSTAL-Omega.



**Figure 8:** Phylogenetic tree of Pou5f1 gene in *Mus musculus, Mus caroli, Onychomys torridus and Arvicanthis niloticus* with other subfamily members having POU domain by MEGA-X.

The phylogenetic analysis results showed that CAA36682.1 octamer-binding protein 4 *Mus musculus*, NP 038661.2 POU domain class 5 transcription factor 1 isoform 1 *Mus musculus*, AAB19896.1 Oct3 Mus sp., NP 001239381.1 POU domain class 5 transcription factor 1 isoform 2 *Mus musculus*, AUJ88036.1 POU domain class 5 transcription factor 1 transcript variant 3 oct4b *Mus musculus*, XP 021005284.1 POU domain class 5 transcription factor 1 isoform X1 *Mus caroli* XP, 021005285.1 POU domain class 5 transcription factor 1 isoform X2 *Mus caroli*, XP 021073734.1 POU domain class 5 transcription factor 1 isoform X2 *Mus pahari*, XP 021073733.1 POU domain class 5 transcription factor 1 isoform X1 *Mus pahari*, XP 028642340.1 POU domain class 5 transcription factor 1 *Grammomys surdaster*, XP 034379773.1

LOW QUALITY PROTEIN: POU domain class 5 transcription factor 1-like Arvicanthis niloticus, XP 034380407.1 POU domain class 5 transcription factor 1 Arvicanthis niloticus, XP 031203214.1 POU domain class 5 transcription factor 1 Mastomys coucha, NP 001009178.1 POU domain class 5 transcription factor 1 Rattus norvegicus, XP 021507032.1 POU domain class 5 transcription factor 1 Merionesung uiculatus, XP 038199969.1 POU domain class 5 transcription factor 1 Arvicola amphibius, OBS74604.1 hypothetical protein A6R68 14882 Neotoma lepida, XP 006997372.1 PREDICTED: POU domain class 5 transcription factor 1 Peromyscus maniculatus bairdii , XP 036032305.1 POU domain class 5 transcription factor 1 Onychomystorridus, XP 029800691.1 POU domain class 5 transcription factor 1 Suricata suricata, XP 005911894.1 PREDICTED: LOW QUALITY PROTEIN: POU domain class 5 transcription factor 1 Bos mutus, XP 006052361.1 POU domain class 5 transcription factor 1 Bualus bubalis, NP 001106531.1 POU domain class 5 transcription factor 1 Susscrofa, XP 020953528.1 POU domain class 5 transcription factor 1 isoform X1 Susscrofa, XP 004695409.1 PREDICTED: POU domain class 5 transcription factor 1 isoform X1 Condylura cristata, XP 036105088.1 POU domain class 5 transcription factor 1 isoform X2 Molossus molossus, XP 024413650.1 POU domain class 5 transcription factor 1 isoform X3 Desmodus rotundus, XP 023480000.1 POU domain class 5 transcription factor 1 isoform X2 Equus caballus, XP 036881972.1 POU domain class 5 transcription factor 1-like Manis javanica, BAF30835.1 octamerbinding transcription factor-3B Macaca fascicularis have POU domain homology and they have phylogenetic relationship with each other.

It signifies that out of these 31 subfamily members, anyone can be used for further pluripotency research work due to their phylogenetic relationship. The phylogenetic trees are presented below.

## Gene structure analysis of POU5F1 gene

The gene structure of POU5F1 gene was obtained online by Gene Structure Display Server. It provides the information about the intron and exon position and number in evolutionary times. The Gene Structure Display Server showed that the genome range of 35816929-35821674 of *Mus musculus* shows 6 exons and 5 introns, the genome range of *Arvicanthis niloticus* c50625435-50621362 show 5 exons and 4 introns, the genome range of *Mus caroli* 32277972-32282748 show 6 exons and 5 introns and the genome range of *Onychomys torridus* c2417075-2412441 show 5 exons and 4 introns, the positions are showed in the result diagram below (**Figures 9a-9d**).

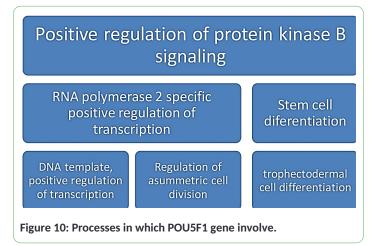
The POU5F1 gene is present in chromatin, cytoplasm, mitochondria, nucleolus, nucleoplasm, nucleus and transcription regulator complex. The POU5F1 gene was studied to see its phylogenetic relation with other subfamilies. In comparison to other subfamily members, most of them have same motifs and domains which show that they are closely related to each other in evolutionary time and have less than 0 E values. That conserved domains are involved in multiple processes like positive regulation of protein kinase B signalling, positive regulation of transcription

**2021** Vol 6 No.3:04

by RNA polymerase 2, DNA template positive regulation of transcription, regulation of asymmetric cell division, Stem cell differentiation, and trophectodermal cell differentiation (Figure 10) [12-16].



**Figure 9:** Figures 9a-9d exons positions of *Mus musculus, Mus caroli, Onychomys torridus and Arvicanthis niloticus* on the Gene Structure Display Server tool.



## Discussion

The genome identification of POU5F1 gene showed that in Mus musculus POU5F1 gene is present on chromosome number 17 and has 6 exons and 2 isoforms. The isoform 1 has 352 amino acids and maximum ORF length of 1059. The isoform 2 has 221 amino acids, with ORF length of 666. In Arvicanthis niloticus POU5F1 gene is present on chromosome number 20 and has 5 exons, maximum ORF length of 1059, and 352 amino acid. In Mus caroli POU5F1 gene is placed on chromosome number 17, and has 6 exon and 2 isoforms. The isoform 1 has 352 amino acids and maximum ORF of 1059. Its isoform 2 has 221 amino acids, and maximum ORF length of 666. In Onychomys torridus POU5F1 gene has 5 exons with unknown chromosome number, with 359 amino acids and 1080 maximum ORF length. All the species have same conserved domains which are smart 00352 and pfam 00046. It signifies that out of these subfamily members, anyone can be used for further pluripotency research work due to their phylogenetic relationship. The phylogenetic tree which was built with 31 subfamily members showed that they have common ancestors.

# Conclusion

In the present study, genome wide identification, conserved domain and phylogenetic analysis of POU5F1 gene were conducted in *Mus musculus*. *Mus caroli, Onychomys torridus* and *Arvicanthis niloticus*. This study revealed a high similarity in exon-intron structure of POU gene family in *Mus musculus*. *Mus caroli, Onychomys torridus* and *Arvicanthis niloticus*. This study supports the close phylogenetic relationship between M. caroli, *M. musculus*, *A. niloicus* and *O. torridus*. This identified gene can be a rich informational resource to specify diseases at embryo level in humans.

## References

- 1 Holsinger K E. (2013) Brenner's encyclopedia of genetics. Cytogenetics 268:270.
- 2 Maloy S R, Hughes K. (2013) Brenner's encyclopedia of genetics. Academic Press.
- 3 Parvin M S, Okuyama N, Inoue F, Islam M E, Kawakami A, et al. (2008) Autoregulatory loop and retinoic acid repression regulate pou2/pou5f1 gene expression in the zebrafish embryonic brain. Dev Dyn 237(5):1373-1388.
- 4 Sigrist C J A, Castro E, Cerutti L, Cuche B A, Hulo N, et al. (2012) New and continuing developments at PROSITE. Nucleic Acids Res 41(Database issue):D344-7.
- 5 Schultz J, Milpetz F, Bork P, Ponting C P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95(11):5857-5864.
- Finn R D, Bateman A, Clements J, Coggill P, Eberhardt R Y, et al.
   (2014) Pfam: the protein families database. Nucleic Acids Res.
   42(Database issue): D222-D230.
- 7 Sievers F, Higgins D G. (2014) Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105-116.

- 8 Kumar S, Stecher G, Li M, Knyaz C, Koichiro Tamura. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6): 1547-1549.
- 9 Guo A Y, Zhu Q H, Chen X, Luo J C. (2007) GSDS: a gene structure display server. Yi chuan 29(8):1023-1026.
- 10 Trifinopoulos J, Nguyen L T, Haeseler A V, Minh B Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44(W1):W232-W235.
- 11 Russell D J. (2014) Multiple sequence alignment methods. Springer 1079.
- 12 Campbell P A, Rudnicki M A. (2013) Oct4 interaction with Hmgb2 regulates Akt signaling and pluripotency. Stem Cells 31(6):1107-1120.

- 13 Rosner M H, Vigano M A, Ozato K, Timmons P M, Poirier F, et al. (1990) A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature 345(6277): 686-692.
- 14 Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M, et al. (1990) A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 60(3):461-472.
- 15 Plachta N, Bollenbach T, Pease S, Fraser S E, Pantazis P. (2011) Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. Nat Cell Biol 13(2):117-123.
- 16 Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, et al. (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123(5):917-929.