http://www.egejfas.org/tr/
Ege Journal of Fisheries and Aquatic Sciences, 38(1), 63-68 (2021)

Su Ürünleri Dergisi (2021)
DOI: 10.12714/egejfas.38.1.07

RESEARCH ARTICLE ARAŞTIRMA MAKALESİ

Bioinformatics of Nile tilapia (*Oreochromis niloticus*) lymphocyte cytosolic protein 1 (*lcp1*) gene

Nil tilapiası (*Oreochromis niloticus*) lenfosit sitosolik protein 1 (*lcp1*) geninin biyoinformatiği

Mehtap Bayır[1]* ● Gökhan Arslan[2]

[1] Atatürk University Agricultural Faculty, Department of Agricultural Biotechnology, Erzurum https://orcid.org/0000-0002-7794-1058
[2] Atatürk University Aquaculture Faculty, Erzurum https://orcid.org/0000-0002-8634-8598

*Corresponding author: mehtap.bayir@atauni.edu.tr

Received date: 18.02.2020 Accepted date: 06.10.2020

How to cite this paper:
Bayır, M. & Arslan, G. (2021). Bioinformatics of Nile tilapia (Oreochromis niloticus) lymphocyte cytosolic protein 1 (lcp1) gene. *Ege Journal of Fisheries and Aquatic Sciences,* 38(1), 63-68. DOI: 10.12714/egejfas.38.1.07

Abstract: Bioinformatics analysis of lymphocyte cytosolic protein 1 (*lcp1*) gene in tilapia (*Oreochromis niloticus*) which is a model organism in experimental studies were completed in this study. For this purpose, characterization and identification of *lcp1* has been completed and ensembl database has been used to design the structure of *lcp1* gene. In addition, the chromosome region of tilapia *lcp1* and other genes in the same region with *lcp*1were determined. The chromosome of these genes were detected in zebrafish and human which are identical orthologos of tilapia. Conserved gene synteny designed manually according to these chromosomal regions. In addition, amino acid sequences synthesized by *lcp1* gene in some vertebrates were determined using some bioinformatics databases such as UNIPROT, ENSEMBL and NCBI before determine the phylogenetic relationship between these organisms and tilapia. Sequence similarity-identity rate of tilapia *lcp1* gene with zebrafish, rainbow trout, human, mouse and platyfish *lcp1/Lcp1* was calculated using BLOSUM62 matrix algorithm. This study is of great importance for the completion of in silico analysis of *lcp1* gene in tilapia because it is an aquatic model organism and it has an important place among economic aquaculture species. However this study provides the basic pioneering information for the future studies on molecular stress response in fish.

Keywords: Tilapia, genomic organisation, lcp1, phylogeny

Öz: Bu çalışmada, deneysel çalışmalarda model organizma olarak kullanılan tilapia (*Oreochromis niloticus*) 'da lenfosit sitozolik protein 1 (*lcp1*) geninin biyoinformatik çalışmaları yapılmıştır. Bu amaçla, *lcp1*'in karakterizasyonu ve tanımlaması yapılmış ve *lcp1* geninin yapısını tasarlamak için ensembl veri tabanı kullanılmıştır. Ayrıca tilapia *lcpl* geninin üzerinde yer aldığı kromozom bölgesi ve *lcpl* geni ile aynı bölgedeki diğer genler ile bu genlerin de yerleri belirlenmiştir. Ayrıca belirlenen bu genlerin kromozom bölgeleri, tilapianın ortologları olan zebra balığı ve insanda da tespit edilmiştir. Bu kromozomal bölgelere göre manuel olarak korunmuş gen yapısı dizayn edilmiştir. Bunlara ilaveten, tilapia ile diğer omurgalılar arasındaki filogenetik ilişkinin belirlenmesi amacıyla UNIPROT, ENSEMBL ve NCBI gibi bazı biyoinformatik veritabanları kullanılarak *lcp1* geni tarafından sentezlenen amino asit sekanslarına ulaşılmış ve bu seklanslar kullanılark filogenetik ağaç, Mega programı yardımıyla, Maksimum Olasılık Metodu'na göre oluşturulmuştur. Zebra balığı, gökkuşağı alabalığı, insan, fare ve plati balığı *lcp1/Lcp1* ile tilapia *lcp1* geninin sekans benzerlik-özdeşlik oranı BLOSUM62 matris algoritması kullanılarak hesaplanmıştır. Bu çalışma, tilapiada *lcp1* geninin in-siliko analizinin tamamlanması için büyük önem taşımaktadır. Çünkü tilapia, sucul bir model organizmadır ve ekonomik su ürünleri türleri arasında önemli bir yere sahiptir. Bu nedenlerle, yapılan bu çalışma, balıklarda moleküler stres yanıtı üzerine yapılacak çalışmalar için temel *oluşturacak ve her zaman başvurulacak bilgiler sağlamaktadır.*

Anahtar kelimeler: Tilapia, genomik organizasyon, lcp1, filogenetik

INTRODUCTION

L-plastin which conserved from yeast to man belongs to the fimbrin family of actin-binding proteins consists of a head domain (10-kDa) and a core domain (60 kDa) (Bredscher, 1981; Lin et al., 1988; Adams et al., 1995). As it is known, cell movement is necessary for the immune system, but how it regulates the movement is much more important as it can be destructive in diseases such as cancer (Goldstein et al., 1985; Lin et al., 1988). *Lcp1* was first discovered in neoplastic human fibroblasts. In subsequent studies, this protein was found to increase regulation of many cancer cell lines (Park et al., 1994) and was also highly expressed in normal leukocytes, including macrophages, monocytes and neutrophils. Therefore, studies on *lcp1* have been focused on two branches: leukocyte biology of l-plastin and cancer biology of l-plastin. However, the main goal of these two groups is to investigate the effects of actin cell skeleton on regulation and cell mobility (Margaret et al., 2017). This protein has been the subject of many scientific researches (Otsuka et al., 2001; Chung and Deisseroth, 2004; Li and Zhao, 2011) especially because it is a cancer marker, but *lcp1* gene regulation, protein function and bioinformatics still need to be studied.

Tilapia is one of the most aqua cultured fish in tropical and subtropical regions due to its resistance to bad environmental conditions, easy breeding, high adaptability to salty and brackish waters, high plant and animal nutrient resources and rapid growth except from its sensitivity to low water temperature (Donaldson, 1979; Tekelioğlu et al., 1991). Tilapia is cultivated in 75 countries around the world and its most common farming species is *Oreochromis niloticus*. A large part of tilapia production is in Asia and China takes the first place in the production of tilapia and Philippines follows China (Urch, 1996). Tilapia which is widely farming in many countries, are preferred as a model organism in experimental and genetic studies due to its low chromosome number (2n = 44) and its whole genome sequence is completed (Guyon et al., 2012) as well as being an easily available species (Ergene et al., 1998).

Genetic similarities between species present in all organisms mean that studies on one organism can be used as a data source for other species (Collins et al., 1998). Therefore, in this study, the bioinformatics of *lcp1* gene in aquatic model organism, tilapia (*O. niloticus*) will be completed and leading data will be provided for molecular studies in other fish.

MATERIAL AND METHOD

Bioinformatics of lymphocyte cytosolic protein 1 (*lcp1*) gene

In order to investigate whether the lymphocyte cytosolic protein 1 (*lcp1*) gene is a functional or non-functional or pseudogene in tilapia, the cDNA sequence of this gene was reached from the ENSEMBL database and confirmed that the *lcp1* gene is a functional gene. Ensemble ID and Uniprot ID of tilapia *lcp1* gen was determined as ENSONIG00000016164 and I3KGW3, respevtively and its amino acid number was determined as 619.

The conserved gene synteny is designed to detect genes that are conserved in the same way as the orthologues of living organism. For this purpose, we designed the gene synteny by detecting conserved genes in tilapia, zebrafish and human. It was first determined which chromosomes and regions of *lcp1* gene were found in tilapia and then other genes in this chromosome were found and their locations were recorded. Then, the chromosomes and locations of these genes detected in zebrafish (*Danio rerio*) and human (*Homo sapiens*) which are orthologs of tilapia (Table1). Finaly the conserved gene synteny was designed manually using these datas (Figure 1).

Table 1. The genes used in conserved gene synteny and their location in tilapia, human and zebrafish

Gene	Gene symbol	Tilapia		Human		Zebrafish	
		Scaffold	Location	Chromosome	Location	Chromosome	Location
Integral membrane protein 2Bb	*itm2bb*	GL831210.1	0.81	13	48.23	9	25.33
Laccase domain containing 1	*lacc1*	GL831210.1	0.58	13	43.87	9	18.57
Leishmanolysin like peptidase	*lmln*	GL831210.1	0.47	3	197.96	9	38.50
Integrin subunit alpha	*itgav*	GL831210.1	0.51	2	186.59	9	11.59
5-hydroxytryptamine (serotonin) receptor 2A	*htr2aa*	GL831210.1	0.96	13	43.83	9	25.19
Succinate-CoA ligase ADP-forming beta subunit	*sucla2*	GL831210.1	0.87	13	47.74	6	10.03
Esterase D	*esd*	GL831210.1	1.13	13	46.77	9	25.17
Leucine rich repeats and calponin homology domain containing 1	*lrch1*	GL831210.1	1.12	13	46.55	9	25.09
Lymphocyte cytosolic protein 1	*lcp1*	GL831210.1	0.77	13	46.12	9	56.25
CCR4-NOT transcription complex subunit 11	*cnot11*	GL831210.1	0.79	2	101.25	9	56.23

For determining the phylogenetic relationship between the lcp1/LCP1 genes tilapia (*O. niloticus*) and zebrafish (*Danio rerio*), rainbow trout (*O. mykiss*), brown trout (*Salmo trutta*), gilthead seabream (*Sparus aurata*), yellow perch (*Perca flavescens*), fugu (*Takifugu rubripes*), pike perch (*Sander lucioperca*), northern pike (*Esox lucius*) Norway rat (*Rattus norvegicus*) mouse (*Mus musculus*), human (*Homo sapiens*) CLUSTALW (Thompson et al. 1994) BioEdit program (http://www.mbio.ncsu.edu/bioedit/page2.html) was used. Firstly, the neighbor joining method was applied using the MEGA6 (Tamura et al., 2013) program, then the phylogenetic tree was constructed according to the maximum likelihood method (Kell et al., 2018) (Figure 2). Zebrafish (*Danio rerio*) fascin actin-bundling protein 2a (*fscn2a*) gene was used as the outer group.

Gene structure of tilapia *lcp1* which consists of exon-intron organization, amino acids produced by exons, 5 'UTR (with TATA box located in this region) and 3'UTR (showing poly A tail in this region) region of the gene, the starting point (+1) of the transcription was designed using data from ensembl database (Table 2). Sequence similarity and identity rate of tilapia *lcp1* gene with zebrafish, rainbow trout, human, mouse and platyfish *lcp1/Lcp1* was calculated using BLOSUM62 matrix algorithm and Bioedit program, CLUSTALW (Thompson et al., 1994) (Table 3).

RESULTS AND DISCUSSION

Bioinformatics study

Bioinformatics studies should be completed before experimental studies to understand how the expression of genes changes with various stress factors, in molecular studies, Therefore, this study will provide important bioinformatics for both fish physiology studies and studies on other vertebrates since tilapia is an important model organism. Although there are several studies on tilapia (Urch, 1996; Ergene et al., 1998; Tekelioglu et al., 1991; Kaya and Akbulut, 2012), study on bioinformatics of this model organism are still very poor

Genomic sequences analysis of tilapia *lcp1* gene

Some algorithms and databases such as ENSEMBL, UNIPROT and NCBI databases and BioEdit software, BLOSUM62 matrix program and MEGA6 program were used for in silico analysis such as designing gene structure, phylogenetic analysis, determining similarity and identy between tilapia and some other vertebrates, designing conserved gene in this study. It was seen tilapia *lcp*1 gene has 15 exons and 14 introns (Table 2).

Table 2. Tilapia (*Oreochromis niloticus*) *lcp1* gene nucleotide sequence

```
5'gacctcagagtgctctgatcttattacacgcccagcagtaatgaacagagaggctttt
gtctgtcaggcaggaagtgaagcacaagcaggaagttttacactggcagaagtgagaact
gagcagtttgcttcacagtagaaaatcaggataccatctttgagtttcctctttctttct
ctaccacatggtccccgcttacatgctactcggtgctgacaggctttgacttttaaatTT
                                        +1
AATgatgtcttgtttttttctcttgtgcCAGTTTCAGGTTCCTCTATTGATTCACGCTTC

AACATGGCCGCCCCACCAATCACTGCAGAGGAGCTGGAGGATCTTAGAGAAGCATTCACT
-N--M--A--A--P--P--I--T--A--E--E--L--E--D--L--R--E--A--F--T-
AAAATTGgtaag'N702'tccagATGTGGACAATGACGGCGTTATCAGCAAGGATGAGCT
-K--I--                 D--V--D--N--D--G--V--I--S--K--D--E--L
TGATGCTGTCCTCAAGGCCGCCAACCTTTCACTGCCCGGCTACAGAGTCAGAGAGATGAT
--D--A--V--L--K--A--A--N--L--S--L--P--G--Y--R--V--R--E--M--I
CCAGGAGCTGAGCAAGAGCAGCGAAGAGCTCAACTTTGATAAGTTTACTGAAgtggg'N127'
--Q--E--L--S--K--S--S--E--E--L--N--F--D--K--F--T--E-
tccagATTGTCCATGGACTAAAGAGTGCAGAGGTTGCAAAGACCTTCAAGAAAGCAATCA
     -I--V--H--G--L--K--S--A--E--V--A--K--T--F--K--K--A--I--
CCAAGAAGGAGGGCATCTGTAATGTGGCAGGAACCTCAGAGCAGACTGGCACTCAGCACT
T--K--K--E--G--I--C--N--V--A--G--T--S--E--Q--T--G--T--Q--H--
CTTACTCAGgtcag'N670'tgcagAGGAGGAGAAGGTAGCCTTTGTGAACTGGATCAAT
S--Y--S--                E--E--E--K--V--A--F--V--N--W--I--N-
AAAGCTCTGGAGAAGGATCCAGACTGCAAACACGTTCTGCCCATGGATCCTAACACCAAC
-K--A--L--E--K--D--P--D--C--K--H--V--L--P--M--D--P--N--T--N-
GACCTGTTCACCGCCATGGGAGATGGGATCGTTCTCTGgtacg'N1151'gttagTAAAA
-D--L--F--T--A--M--G--D--G--I--V--L--C                  --K-
TGATCAACCTGTCTGTAGCCGACACCATCGATGAGAGAACAATCAACAAGAAAAAGCTCA
-M--I--N--L--S--V--A--D--T--I--D--E--R--T--I--N--K--K--K--L-
CACCCTTCACCATCCAGgtgag'N98'aacagGAGAACCTGAACCTGGCCCTGAACTCGG
-T--P--F--T--I--Q          --E--N--L--N--L--A--L--N--S-
CATCAGCTATTGGCTGCCATGTGGTGAACATTGGAGCTGAGGACCTGAAGGAGGGCAGGC
-A--S--A--I--G--C--H--V--V--N--I--G--A--E--D--L--K--E--G--R-
AGCACCTGGTCCTGG GTCTGCTGTGGCAGGTCATCAAGATCGGGCTGTTCGCTGACATC
-Q--H--L--V--L--G--L--L--W--Q--V--I--K--I--G--L--F--A--D--I-
GAGCTCAGCAGGAATGAAGgtgtg'N226'tgcagCTCTGATCGCTCTGCTGCGGGATGG
-E--L--S--R--N--E--                A--L--I--A--L--L--R--D--G
AGAGAGTCTTGAGGATCTGATGAAACTTTCCCCCGAGGAGCTGCTGTTGCGTTGGGCCAA
--E--S--L--E--D--L--M--K--L--S--P--E--E--L--L--L--R--W--A--N
CTATCACCTGGAGGAGGCGGGCTGTGGCAAGATCAACAACTTCAGCAACGACATCAAGgt
--Y--H--L--E--E--A--G--C--G--K--I--N--N--F--S--N--D--I--K-
gat'N334'tccagGATTCGAAGGCGTACTACAACCTGCTGAACCAGGTGGCACCCAAAG
            -D--S--K--A--Y--Y--N--L--L--N--Q--V--A--P--K--
GAGACGAAGAGGGAATTCCCCCCATTGCCGTTGACATGTCAGGACTCAGGgtaag'N349'
G--D--E--E--G--I--P--P--I--A--V--D--M--S--G--L--R-
tgcagGAGAAAGACGACCTGAAGCGAGCGGAGCTCATGCTGGACCAGGCCGAAAGGCTCG
     -E--K--D--D--L--K--R--A--E--L--M--L--D--Q--A--E--R--L--
GGTGCAGACAGTTCGTTATGCCTACAGATGTCGTCCGTGGCAACCCTAAGCTCAACTTGG
G--C--R--Q--F--V--M--P--T--D--V--V--R--G--N--P--K--L--N--L--
CTTTTGTCGCCAATCTGTTCAACAAATACCCGGCTCTGAAGAAGCCAGAGAACCAGGACA
A--F--V--A--N--L--F--N--K--Y--P--A--L--K--K--P--E--N--Q--D--
TCGACTGGAGCTCCATCGAAGgtggg'N95'tccagGTGAAACCAGAGAGGAGCGCACCT
I--D--W--S--S--I--E--              G--E--T--R--E--E--R--T--
TCAGGAACTGGATGAACTCACTGGGAGTCAACCCTCGTGTAAACCATCTCTACGCgtaag
F--R--N--W--M--N--S--L--G--V--N--P--R--V--N--H--L--Y--A
'N123tgcagAGACACTCGACGACGCCTTGGTGATTTCCAGCTGTACGAGAAGATCAAAG
          --D--I--D--D--A--L--V--I--F--Q--L--Y--E--K--I--K--
TACCAGTGGACTGGGACAGAGTCAACAAACCTCCCTACCCCAAACTGAGCAGCAACATGA
V--P--V--D--W--D--R--V--N--K--P--P--Y--P--K--L--S--S--N--M--
AGAAGgtaca'N563'tccagCTGGAGAACTGTAACTACGCTGTGGAGCTGGGAAAGAAG
K--K-                -L--E--N--C--N--Y--A--V--E--L--G--K--K-
GAGGCCAAGTTCTCTCTGGTCGGCATTGCGGGTCAGGATCTGAACGCAGGGAATCGAACC
-E--A--K--F--S--L--V--G--I--A--G--Q--D--L--N--A--G--N--R--T-
CTCACCCTCGCTCTGCTCTGGCAGCTCATGAGGAGgtaaa'N245'ctcagGTACACCCT
-L--T--L--A--L--L--W--Q--L--M--R--R                --Y--T--L
GAATATTTTGGAGGACCTGGGCGATGGGCAAAAGGTGATCGATGACACCATCGTGTCCTG
--N--I--L--E--D--L--G--D--G--Q--K--V--I--D--D--T--I--V--S--W
GGTCAACGACAACCTGACAAGGGCCGGAAAATCCACAATCTCCAGCTTTAAGgtaac'N362'
--V--N--D--N--L--T--R--A--G--K--S--T--I--S--S--F--K-
gttagGACGGCTCCATCAGCACCAGCATGCCGGTTCTTGACCTGATCGATGCCATCCAGC
     -D--G--S--I--S--T--S--M--P--V--L--D--L--I--D--A--I--Q--
CTGGATCCATCCGGTACGACCTGCTGAAGACCGAAGACCTGACTGAGGAGGAGAAACTCA
P--G--S--I--R--Y--D--L--L--K--T--E--D--L--T--E--E--E--K--L--
ACAACgtacg'N93'tttagGCAAAGTACGCTATCTCCATGGCCAGGAAGATCGGCGCCC
N--N-                -A--K--Y--A--I--S--M--A--R--K--I--G--A--
GGGTGTATGCACTCCCCGAGGACCTGGTGGAGGTCAAGCCTAAGATGGTGATGACGGTGT
R--V--Y--A--L--P--E--D--L--V--E--V--K--P--K--M--V--M--T--V--
TCGCCTGCCTCATGGCTCGTGGCCTGAGGGCCTAAgtcccgtctgtgaagaggagacctc
F--A--C--L--M--A--R--G--L--R--A--*-
taaagctagcacccaaaggAAATAAgttgctaatgcagaggctccactcacaagttgtggc
tttagactgtaaccatttaatgaccagaaatattggatttatgatggcgatacagctgtgg
gacacatttcgtaacatctgctggtgattgatgtcttttactcgtgtttctttaaatgta
caatagaaaactttgttttattcaaaaccaagaagctgaaataatcccaagggtcaagtta
aacctcacctaaagagttg 3'
```

*Tilapia (*Oreochromis niloticus*) *lcp1* gene structure. The exons of the *lcp1* are shown in capital letters. The starting site of transcription is +1, 5 UTR sequence and 3' UTR sequence are shown in lower case. The first 5 nucleotides and last 5 nucleotides of the introns and the rest of the nculeotides number are given red and lower case. The TATA box and the poly adenylation signal (AATAAAA) are shown in capital letters and painted in blue. Amino acids are shown in capital letters which are placed under exons.Stop codon (TGA) is specified asterisk.

Orthology of tilapia *lcp1* gene with *lcp1* genes from other vertebrates

Protein sequence alignment of each tilapia *lcp1* gene with *lcp1*s from platyfish (*Xiphophorus maculatus*), zebrafish (*Danio rerio*), rainbow trout (*Oncorhynchus mykiss*), human (*Homo sapiens*) and mouse (*Mus musculus*) was performed using CLUSTAL W (Thompson et al., 1994), and sequence identity and similarity of tilapia Lcp1 with Lcp1s from other teleost fishes and tetrapods were detected. Tilapia Lcp1 shared highest percentage sequence identity and similarity with fish Lcp1 sequences from tetrapods (Table 3). According to the Table 3, identity-similarity rates of tilapia *lcp1* gene were found as 89-95% with platyfish (Pf), 87-94% with zebrafish (Zf), 85-91% with rainbow trout (Rt), 79-88% with human (Hu) and 79-88% mouse (Mo).

Table 3. Identity-similarity rate between Tilapia (Ti) and, Platyfish (Pf) Zebrafish (Zf), Rainbow trout (Rt), Human (Hu) and Mouse (Mo)

```
Ti lcp1   1  -MAAPPITAEELEDLREAFTKIDVDNDGVISKDELDAVLKAANLSLPGYRVREMIQELSK
Pf lcp1   1  -...TK.S.....E.....A.....GN.F......NELFR....A....K....V...TR
Zf lcp1   1  -...AQ.S...M.E.......V...GN.H..T...N.LF.....P.........I...I.R
Rt lcp1   1  MA.PAQ.SQD...E.....A.....SH.H.GT...NDLF.....P.........I..D.T.
Hu Lcp1   1  -..RGSVSD..MME.....A.V.T.GN.Y..FN..NDLF...C.P........ITEN.MA
Mo Lcp1   1  -..RGSVSD..MME.....A.V.T.GN.Y..CN..NDLF...C.P........ITEN.MA

Ti lcp1  60  S-----SEELNFDKFTEIVHGLKSAEVAKTFKKAITKKEGICNVAGTSEQT--GTQHSYS
Pf lcp1  60  T-----.NS.T.EE...M......T......R...N..............S--.......
Zf lcp1  60  TMDLNQDGKIT..E.AKV..D...S......R...N......S.......S--.......
Rt lcp1  61  TGDLH-DGKVT.NE.ANV......T..........N.....NA.......SSS.......
Hu Lcp1  60  TGDLDQDGRIS..E.IK.F.....TD.....R...N......AIG.....SSV.......
Mo Lcp1  60  TGDLDQDGKIS..E.IKVF.....T......R...N......AIG.....SSV.......

Ti lcp1 113  EEEKVAFVNWINKALEKDPDCKHVLPMDPNTNDLFTAMGDGIVLCKMINLSVADTIDERT
Pf lcp1 113  .........................I.....N.....................P.......
Zf lcp1 118  ..........V..........Q.......S.D.....V..............P.......
Rt lcp1 120  ..........V...................T.......V...........Q..P.......
Hu Lcp1 120  ....Y...........N....R..I..N.......N.V..............P.......
Mo Lcp1 120  ....Y...........N....R..I..N...D...N.V..............P.......

Ti lcp1 173  INKKKLTPFTIQENLNLALNSASAIGCHVVNIGAEDLKEGRQHLVLGLLWQVIKIGLFAD
Pf lcp1 173  ..........V.....................................................
Zf lcp1 178  .....................................................................
Rt lcp1 180  .....................................................................
Hu Lcp1 180  ..........................................KPY.....................
Mo Lcp1 180  ..........................................KPY.....................

Ti lcp1 233  IELSRNEALIALLRDGESLEDLMKLSPEELLLRWANYHLEEAGCGKINNFSNDIKDSKAY
Pf lcp1 233  ..I.....................................................NS........
Zf lcp1 238  ..I.....................V.....................P......S........
Rt lcp1 240  ..I.....................................S......S........
Hu Lcp1 240  ..............E.........................N...N..G...T........
Mo Lcp1 240  ..............E.........................N...T..T...T........

Ti lcp1 293  YNLLNQVAPKGDEEGIPPIAVDMSGLREKDDLKRAELMLDQAERLGCRQFVMPTDVVRGN
Pf lcp1 293  ..I..........D.V.A..I....I...E......C.....DK..........A......
Zf lcp1 298  ..I..............A.PI.I..I..........C..E..D........TA.......
Rt lcp1 300  ..I..............L..I.I..I...E.I....C..E..D........TA.......
Hu Lcp1 300  .H..E...........V.AVVI..........IQ...C..Q...........TA.......
Mo Lcp1 300  .H..E............AVVI..........IQ...C..Q...........TA.......

Ti lcp1 353  PKLNLAFVANLFNKYPALKKPENQDIDWSSIEGETREERTFRNWMNSLGVNPRVNHLYAD
Pf lcp1 353  .....................................................................
Zf lcp1 358  ......Y.....................................................V.
Rt lcp1 360  ......YI....................................................V.
Hu Lcp1 360  .......I.....R....H.........GAL.............................S.
Mo Lcp1 360  .......I..........H.........GAL.............................S.

Ti lcp1 413  IDDALVIFQLYEKIKVPVDWDRVNKPPYPKLSSNMKKLENCNYAVELGKKEAKFSLVGIA
Pf lcp1 413  ..............................G..............................
Zf lcp1 418  LA.....................K.........G..............................
Rt lcp1 420  ..............N.............S..G.................N..........
Hu Lcp1 420  LS..................N..........GG.................NQ........G
Mo Lcp1 420  LS..................N..........GG............D...NQ.........

Ti lcp1 473  GQDLNAGNRTLTLALLWQLMRRYTLNILEDLGDGQKVIDDTIVSWVNDNLTRAGKS-TIS
Pf lcp1 473  .....E................................V.E.......GT..E...G-...
Zf lcp1 478  .....E.................................I..E...Q...ET..Q...G-...
Rt lcp1 480  .....E...K..Q.................E.......N.....T....T..Q...G-...
Hu Lcp1 480  .....E.........I..............EI.G....N..I..N...ET.RE.K..SS..
Mo Lcp1 480  .....E.........V..............I.G....N..I..N...TT.KE.Q..SS.A

Ti lcp1 532  SFKDGSISTSMPVLDLIDAIQPGSIRYDLLKTEDLTEEEKLNNAKYAISMARKIGARVYA
Pf lcp1 532  ....KC..S.....................M.....................................
Zf lcp1 537  G.......S.....................A....D.....................................
Rt lcp1 539  G......A.....................I.V.....................................
Hu Lcp1 540  ....PK....L..............N.......N.NDD.....................................
Mo Lcp1 540  ....PK....L..............N.......N.DD.....................................

                                                  Identity (%)   Similarity (%)

 Ti lcp1  592 LPEDLVEVKPKMVMTVFACLMARGLRA-          100            100
 Pf lcp1  592 .........................M.RA           89             95
 Zf lcp1  597 .........................M.RI           87             94
 Rt lcp1  599 .........................MKRV           85             91
 Hu Lcp1  600 ........N.............GK.MKRV           79              88
 Mo Lcp1  600 ........N.............GK.MKRV           79              88
```

*The dots refer to repeating amino acids and lines represent undetectable amino acids

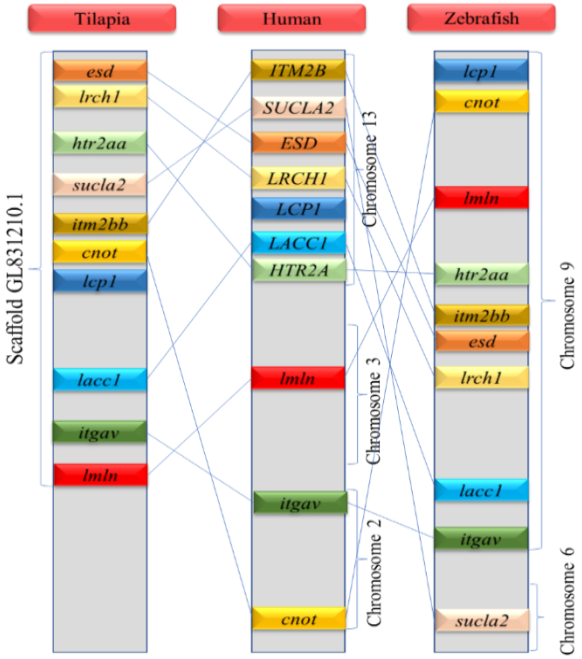

Figure 1. Conserved Gene Synteny of Tilapia *lcp1*gene

In order to detect the conserved genes of tilapia with zebrafish and human, it was first determined that the lcp1 gene was on the schaffold GL831210.1 and the other genes found in this schaffold were determined from the Ensembl genome database (Table 1). Then, conserved gene synteny was generated manually by region (Chromosome 2, 3 and 13 in human genome and Chromosome 6 and 9 in zebrafish genome) of these detected gene (Figure 1) As it is known, teleost fish have conserved regions for the gene structure of the same gene family, and the conserved gene synteny clearly demonstrates this (Figure 1). It's known that, teleost fish may have two copies of the genes as a result of duplication of the whole genome which are found as a single copy in the other organisms (Amores et al. 1998; Meyer and Schartl 1999; Postlethwait et al. 2000; Braasch and Postlethwait, 2012; Çapan, 2019; Bayır et al., 2020). The results indicate that the *lcp1* gene appeared as a result of teleost genome duplication (TGD) in bony fish, but one copy was lost after the TGD. It was seen sometimes in teleost fish such as Bayır et al., (2020) reported that of the medaka, tilapia, stickleback, puffer fish, platyfish, Makobe island cichlid, Midas cichlid, Amazon molly and fugu have a copy of creatine kinase gene (*ckma*), while zebrafish has two copies of creatin kinase gene, *ckma* and *ckm*

A neighbor-joining phylogenetic tree, constructed using sequences of Lcp1/LCP1 from zebrafish, rainbow trout, brown trout, gilthead seabream, yellow perch, fugu, pike, northern pike, Norway rat, mouse and human further supported this orthology (Figure 2). The phylogenetic tree which generated by maximum likelihood method showed that specific fish Lcp1 proteins clustered in distinct clades from tetrapods. The reliability of the tree was evaluated with a 1000 bootstrap replicates (Felsenstein, 1985).

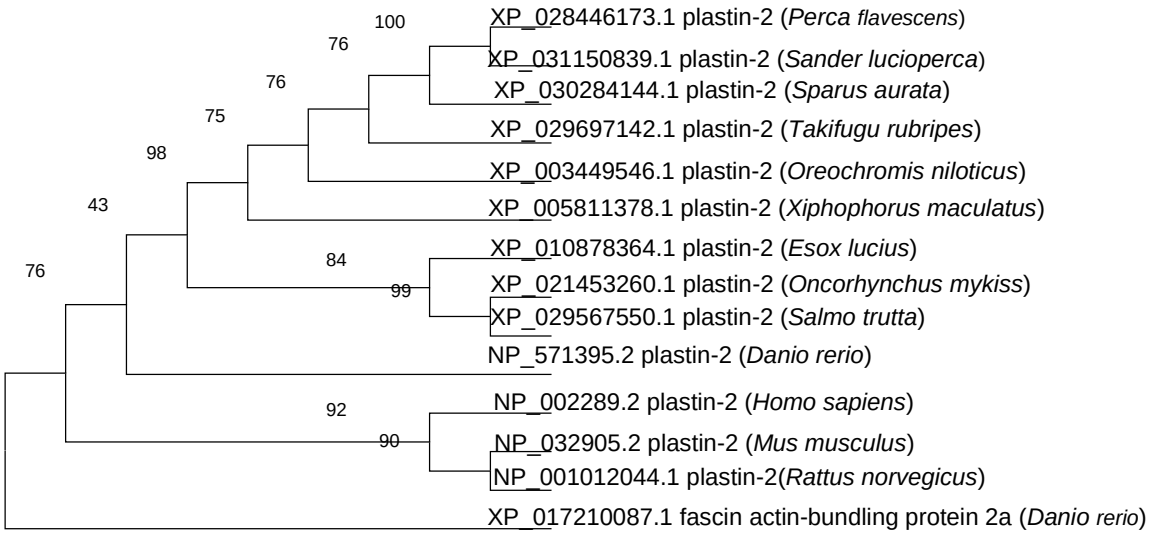

Figure 2. Phylogenetic tree of tilapia *lcp1* gene. Phylogenetic relationships between *lcp1* sequence from tilapia and the other fish and tetrapods. Tree was constructed using Maximum Likelihood method (Felsenstein, 1989). NCBI accession IDs of the sequences used for phylogenetic tree are given in phylogenetic tree

ACKNOWLEDGEMENTS

This article is based on bioinformatics in aquaculture. I would like to thank Prof. Jonothan M. Wright and Prof. Roy G. Danzmann who shared their bioinformatics knowledge with me as long as I worked in their laboratories as a post-doc researcher during two years.

REFERENCES

Adams, A.E., Shen, W., Lin, C. S., Leavitt, J. & Matsudaira, P. (1995). Isoform-specific complementation of the yeast sac6 null mutation by human fimbrin. *Molecular and Cellular Biology*, 15, 69-75.

Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho R.K., Langeland, J., PrinceV., Wang, Y.L., Westerfield, M., Wang, M.E., Y.L. Postlethwait, J. H. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science,* 282 (5394), 1711-1714. DOI: 10.1126/science.282.5394.1711

Bayır, M., Arslan, G., Yıldız Oğuzhan, P. (2020). Characterization, identification and phylogeny of the cretaine kinase (*ckma*) gene in medaka (*Oryzias latipes*). *Marine Science and Technology Bulletin,* 9(1), 15-22.

Braasch, I. & Postlethwait, J.H.I. (2012). Polyploidy in fish and the teleost genome duplication. In: Soltis, P.S., Soltis, D.E. (eds.), *Polyploidy and Genome Evolution*, 341-383. DOI:10.1007/9783-642-31442-1_17

Bretscher, A. (1981). Fimbrin is a cytoskeletal protein that crosslinks F-actin in vitro. *Proceeeding of the National Academy of Sciences*, 78,6849–6853. DOI:10.1073/pnas.78.11.6849

Chung, I. & Deisseroth, A.B. (2004). Recombinant adenoviral vector containing tumor-specific L-plastin promoter fused to cytosine deaminase gene as a transcription unit: Generation and functional test. *Archives of Pharmacal Research*, 27(6), 633-9.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R. & Walters, L. (1998). New Goals for the U.S. Human Genome Project: 1998-2003. *Science,* 282, 682-689. DOI:10.1126/science.282.5389.682

Çapan, E.C. (2019). Plati balığı (*Xiphophorus maculatus*)'nda katalaz enzim geninin biyoenformatiği ve doku spesifik dağılımı. Master Thesis. Atatürk University, Erzurum. 77 pp.

Donaldson, E. M. 1979. Hormonal enhancement of growth. *Fish Physiology*, 8, 456-597.

Ergene, S., Kaya, F., Pekcan, Y. & Oral, A. (1998). A Karyological Analysis of *Oreochromis niloticus* (L., 1758) (Pisces, Cichlidae) Used in Aquaculture, *First International Symposium on Fisheries and Ecology*, 1(10), 2-4, Trabzon.

Felsenstein, J. (1989). PHYLIP-Phylogeny inference package. *Cladistics*, 5, 164-166.

Goldstein, D., Djeu, J., Latter, G., Burbeck, S. & Leavitt, J. (1985). Abundant synthesis of the transformation-induced protein of neoplastic human fibroblasts, plastin, in normal lymphocytes. Cancer Res.; 45(11 Pt 2), 5643-7.

Guyon, R., Rakotomanga, M., Azzouzi, N., Coutanceau, J.P., Bonillo, C., D'Cotta, H., Pepey E, Soler, L., Rodier-Goud, M., D'Hont, A., Conte, M.A., van Bers, N. E.M., Penman, D.J., Hitte, C., Crooijmans, R.P.M.A., Kocher, T.D., Ozouf-Costaz, C., Baroiller, J.F. & Galibert F. (2012). A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics,*6(13), 222. DOI: 10.1186/1471-2164-13-222

Kaya, H., Akbulut, M. (2012). Kurşuna Maruz Bırakılan Tilapia *(Oreochromis mossambicus*)'nın Eritrosit Morfolojisinde Görülen Değişimler, 22(B), 10-15.

Kell, A.J.E., Yamins, D.L.K., Shook, E. N. & Norman-Haignere, S.V. (2018). A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neron*, 98(3), 630-644.

Li, J. & Zhao, R. (2011). Expression and clinical significance of L-plastin in colorectal carcinoma. *Journal of Gastrointestinal Surgery*, 15(11), 1982-1988. DOI: 10. 1007/s11605-011-1678-4

Lin, C.S., Aebersold, R. H., Kent, S. B., Varma, M. & Leavitt, J. (1988). Molecular cloning and characterization of plastin, a human leukocyteprotein expressed in transformed human fibroblasts. *Molecular and Cellular Biology*, 8, 4659-4668.

Margaret, J.K., Rachel, E.R., Emily A. B., Zachary J. C., Paul J.P., Taylor A. M., Jacek T. & Elizabeth, E.L. (2017). Targeted deletion of the zebrafish actinbundling protein L-plastin *(lcp1*). *PLoS ONE,* 13(1) . DOI:10.1371/journal. pone.0190353

Meyer, A. & Schartl, M. (1999). Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current Opinion in Cell Biology,* 11(6), 699-704. DOI:10.1016/s0955-0674(99)00039-3

Otsuka, M., Kato, M., Yoshikawa, T., Chen, H., Brown, E.J.Masuho, Y., Omata, M., Seki, N. (2001). Differential expression of the L-plastin gene in human colorectal cancer progression and metastasis. *Biochemical and Biophysical Research Communications*. 289(4), 876-81.

Park, T.S., Chen, Z.P. & Leavitt, J. (1994). Activation of the leukocyte plastin gene occurs in most human cancer cells. *Cancer Research*, 54(7), 1775-1781.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Force, A. H., Talbot, W.S. (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Research,* 10,1890-1902. DOI:10.1101/gr.164800

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar S, (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30 (12): 2725-2729. DOI: 10.1093/molbev/mst197

Tekelioğlu, N., Sarıhan, E., Polat, A. & Işık, O. (1991). Tilapyaların (*Tilapia zilli*) Çukurova Koşullarında Sera İçerisinde Kışlatılmaları ve Cinsiyetlerinin Hormonlu Yemlerle Değiştirilmesi. Ege Üni. Su Ür. Fak. Eğitiminin 10. Yılında Su Ürünleri Sempozyumu (12-14 Kasım), İzmir, 534-541 s.

Thompson, J.D., Higgins, D.G. & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673-4680. DOI:10.1093/nar/22.22.4673

Urch, M. (1996). Spotligt turns on Tilapia: 'the fish of the 21st Century' Seafood International June pp 25-27.