DNA Barcoding of Cyprinid Fish *Chagunius chagunio* Hamilton,1822 from Phewa Lake, Nepal

Bhaba Amatya[1]

[1] Prithvi Narayan Campus, Pokhara, Tribhuvan University, Nepal

Correspondence: Bhaba Amatya, Prithvi Narayan Campus, Pokhara, Tribhuvan University, Nepal. Tel: 98-0-654-4253. E-mail: bhabaamatya@yahoo.com

Abstract

The present study is the first of its type that uses a technique of DNA barcoding to determine identification and relationship of a species of fish from Phewa lake, Nepal. The mitochondrial DNA from two ethanol-preserved samples of fish, randomly collected from Phewa lake, was extracted using Gene AllExgene ᵀᴹtissue extraction kit. 650 base pair of mitochondrial cytochrome c oxidase subunit 1 (CO1) was amplified using a cocktail of four primers and was sequenced bidirectionaly using Sanger sequence method. The DNA sequences were edited using AliView software. The sequences confirmed *Chagunius chagunio* as their alignment with 16 reference sequences belonging to *Chagunius chagunio* in the NCBI GenBank, scored highest percentage of Query Cover (75% to 100%) and Percentage Identity (97.29% to 100%). The MEGA software analysed the DNA sequences to obtain their corresponding protein sequences. The DNA sequences were submitted to the GenBank and accession numbers (MN087472 and MN087473) were obtained. Clustal Omega software analysed multiple sequence alignment among 19 homologous DNA sequences of *Chagunius chagunio* from India, Bangladesh and Phewa lake, Nepal. The percentage of similarity among the aligned sequences was calculated as 39.3%. Based on the neighbour joining tree, the *Chagunius chagunio* of Phewa lake is found closely related with *Chagunius chagunio* of Bangladesh.

Keywords: DNA Barcode, *Chagunius Chagunio*, Cyprinid Fish, CO1

1. Introduction

The cyprinid, *Chagunius chagunio,* commonly called 'rewa' in Nepal, is distributed in India, Bangladesh, Pakistan, Thailand, Myanmar and Nepal. It is a least concerned species in IUCN red list status (2014). In Nepal, it is a vulnerable species of fish and is distributed in different rivers of the country, including Phewa lake, Bagmati, Trishuli, Gandaki, Bheri, Karnali and Mahakali rivers of Nepal.

In Phewa lake, the cyprinid fishes are facing steep competition with the invasive species, mainly tilapia (*Oreochromis niloticus* and *O. mossambicus*). Therefore, conservation of the cyprinid species is a special concern in the recent years. One of the best ways to protect a species is to preserve its genetic resources and genetic diversity besides the preservation of the ecosystem (Hedrick, 2001). As there is complete lack of information on genetic analysis of fish species in the lake, I have barcoded a single species of cyprinid fish from the lake using partial sequence of mitochondrial CO1 gene. The reasons for the selection of mitochondrial gene are: it is a haploid genome; it shows high copy number and lacks introns; it exhibits low recombination and is maternally inherited. Species identification through genetic analysis is almost always efficiently solved by the use of a standardized molecular approach, such as, the DNA barcoding, using CO1 gene (Hebert et al., 2003; Hjibabei et al., 2007). The most commonly used gene that is used as the barcoding marker, is protein coding gene cytochrome-c oxidase 1 with base length of 648 bp (Zhang & Hewitt, 1997).

The sequence of 648 base pair of mitochondrial cytochrome c oxidase 1 gene (CO1) is used as the DNA barcode, which is highly reliable in identifying most of all animal groups. The advantage of using CO1 is that it is short enough to be sequenced quickly and cheaply yet long enough to identify variations among species. The suitability of CO1 gene for species identification is due to the fact that its mutation rate is often fast enough to distinguish closely related species and also because its sequences are conserved among conspecifics. High mutation rate in CO1 causes intraspecific variation which leads to species delimitation/identification (Hlaing et al., 2009; Wheat & Watt, 2008; Williams & Knowlton, 2001). Congeneric species of animals show more than 2% sequence

divergence (Hebert et al., 2003). Intraspecific divergences in mitochondrial genes in animal species are rarely higher than 2% and most are less than 1% (Avise, 2000). The higher divergence occurs when the species are geographically isolated. Many causes of high divergences are due to unclear status and taxonomic uncertainty (Avise & Walker, 1999).

Mitochondrial genomes are small (usually less than 20,000 bp), circular and maternally inherited (Boor, 1999). These genes are preferably used as universal markers in animal DNA barcoding as they are being maternal (Birky, 2001). The DNA genome is high in number per cell and are useful for population genetic and phylogenetic studies (Hu et al., 2004; McManus et al., 2004). Mitochondrial DNA is regarded as an important tool in the study of evolutionary relationships among various taxa owing to its conserved protein coding regions, high variability in non-coding sequences, and lack of recombination (Olive et al., 1983; Ingman et al., 2000). Sequence divergence accumulates more rapidly in mitochondrial DNA than in nuclear DNA owing to a faster mutation rate and lack of repair system, meaning that it often contains high levels of informative variation (Khan et al., 2008). The mitochondrial CO1 gene offers different observations in environmental science and systematic of fishes (Hubert et al., 2008) and permits researchers to receive ambiguous species to accomplish data practically and rapidly (Cowan et al., 2006).

The DNA barcodes are stored in an open-access digital library (Barcode of Life Data Systems and GenBank/National Center for Biotechnology Information) that can be used to compare the DNA barcode sequences of unidentified samples from the field by matching them to known sequences with associated species names in the database, so that users can recognize species and retrieve information about them quickly and cheaply.

The DNA barcoding of organism is effective in species identification at all stages of life, differentiating among phenotypically alike species (cryptic species) and identifying products in commerce, e.g. herbal supplements, wood, skin, bone and other animal parts .Based on the advantages of DNA barcoding, scientists all over the world have established the database and repositories of mitochondrial DNA sequences for all animals, including fish, and is adopted as a global bio-identifying system for animals in recent years.

2. Methods and Methodology

2.1 Study Area

Phewa lake is located at the south western part of kaski district about 200km west from Kathmandu Valley. The lake is at an altitude of 742m (2,434ft) and covers an area of about 4.43km^2 (1.7sq m) (Rai, 2000). It has an average depth of about 8.6 m (28ft) and a maximum depth of 24m (79ft) (Shrestha, 2003). The maximum water capacity of the lake is approximately 43,000,000 cubic meters (35,000 acre ft) (Pokharel, 2003). Cyprinid is the dominant fish species in Phewa lake in terms of species richness and abundance (Giri, 2013). DNA extraction, polymerase chain reaction and gel electrophoresis was carried out at the Centre for Molecular Dynamics Nepal (CMDN) in Kathmandu, Nepal.

2.2 Sample Collection

The fish specimens were caught in the wild and morphologically identified *in situ* by visual inspection. Two samples, each of approximately 100 mg of white muscle tissue, and fin clip from two individual fish from the lake were collected and preserved in 95% ethanol.

2.3 Mitochondrial DNA Extraction

The DNA was extracted by using Gene AllExgene ™tissue extraction kit. The following steps were involved in the purification of DNA:

1. 20mg of tissue was minced with a sharp scalpel as small as possible and put in a 2ml tube and mixed with 200ul of TL buffer and vortexed for 15 seconds**.** (This step was for lysis of cell of the sample).

2. 20ul of proteinase K solution was added to it and mixed by vortexing. Incubation was done at 56℃ for overnight for complete lysis. (This step was for breaking down of cell protein).

3. 400ul of buffer TB was added and mixed by vortexing. The tube was spinned down briefly to remove any drops from inside of the lid.

4. The mixture was transferred into the spin column for centrifugation for 1min at 6000xg and replaced the collection tube with new one.

5. 600ul of Buffer BW was added and centrifuged for 30sec at 6000xg above and replaced the collection tube with a new one.

6. 700ul of buffer TW was added and centrifuged for 30 sec at 6000xg above. The filtrate was discarded and the SV column was reinserted back into the collection tube.

7. The mixture was centrifuged at full speed (above 13000xg) for 1min to remove residual wash buffer and placed the SV column in a fresh 1.5ml tube.

8. 100ul of buffer AE was added and incubated for 2min at room temperature and centrifuged at full speed (>13000xg) for 1min. and stored at -20℃.

2.4 Amplification and Sequencing of CO1 Fragment

The partial CO1 segment of mitochondrial DNA was targeted for DNA barcoding using cocktail of four fish specific primers (Table 1) which amplified 650 bp region of the gene (Ivanova et al., 2007).

Table 1. Primers used for CO1 amplification

Target gene	Primer ID	Sequence 5'-3'	Band size (bp)	Reference
CO1	VF2_t1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	650	Ivanova et al. (2007)
	FishF2_t1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC		
	FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA		
	FR1d_t1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA		

A total of 25µl PCR final reaction was prepared containing 12.5µl of 2× Qiagen multiplex master mixes, 2.5µl of 5× Q-solution, 0.25µl 10pMol/µl fish CO1 cocktail primer sets and 2µl of extracted DNA. The thermocycling (MJ research Tetrad PTC -225 Thermal cycler, USA) condition was 95℃ for 15 min followed by touch down PCR of 5 cycles at 94℃ for 60 secs, 48℃ for 50 sec and 72℃ for 50 sec followed by 35 cycles at 94℃ for 60 secs, 50℃ for 50 sec and 72℃ for 50 sec with the final extension at 72℃ for 5 min. The amplified 650 bp target PCR product was visualized in Gel-Doc, (Maor Scientific TM) under 2% agarose gel electrophoresis. Both the samples amplified positive.
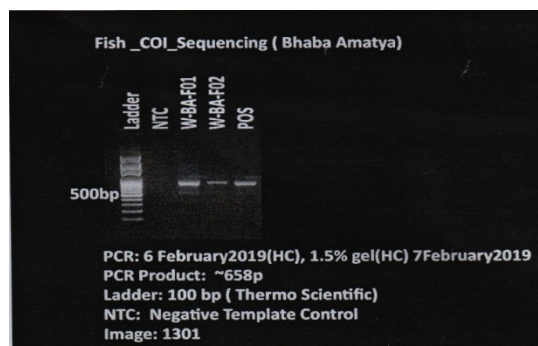

Figure 1. Cytochrome Oxidase-1 PCR run on 1.5% agarose gel with bands appearing at approximately 650bp (Fish sample codes: W-BA-F01, W-BA-F02)

The amplified PCR products were purified using enzymatic clean up (ExoSAP-IT) removing unconsumed dNTPs and primers and sequencing reactions were performed in a MJ Research PTC-225 Peltier Thermal Cycler using a ABI PRISM ® BigDyeTM Terminator Cycle Sequencing kits with AmpliTaq®DNA polymerase (FS enzyme) (Applied Biosystems), following the protocols supplied by the manufacturer. Single-pass sequencing was performed on each template using species specific both forward and reverse primers. The fluorescent-labeled fragments were purified from the unincorporated terminators with BigDye®XTerminator™ purification protocol. The samples were re-suspended in distilled water and subjected to electrophoresis in an ABI 3730×l sequencer (Applied Biosystems).

2.5 Data Analysis

The two sequence reads of each sample were processed for trimming followed by assembling via AliView software (Larsson, 2014). The conversion website Endmemo was used to calculate GC- content and the length of the sequences. To confirm the identity of the amplified sequences, I conducted BLAST (Basic Local Alignment Search Tool) searches by inputting the FASTA format of the sequences using the megablast search for highly

similar sequences in the GenBank (https://blast.ncbi.nlm.nih.gov/Blast.cgi).The DNA sequences were submitted to GenBank to obtain accession numbers. A software program MEGA 3.1 (Molecular Evolutionary Genetics Analysis, MEGA Inc., Englewood, NJ) was used to obtain protein sequence. A new multiple sequence alignment program called Clustal Omega (EMBL-EBI, 2019) was used to align the DNA sequences of the samples with the reference sequences of the GenBank.

3. Result

3.1 DNA Barcode

FASTA format of the DNA barcodes of the fish samples from Phewa lake are given below:

>MN087472 *Chagunius chagunio* cytochrome oxidase subunit 1 (CO1) gene, partial cds; mitochondrial

```
AGAATCAGAACAGGTGTTGATATAAAATTGGATCCCCACCCCCTGCCGGGTCGAAGAATGTGGTGTTGAGGTTACG
ATCTGTTAGAAGTATTGTAATTCCTGCGGCTAAAACTGGAAGGGATAAAAGAAGCAGCACGGCAGTTACAAGCACA
GATCACACAAATAAGGGTGTTTGATATTGGGAGATAGCTGGAGGTTTCATATTAATAATTGTGGTGATGAAATTAA
TTGCTCCCAGAATTGATGAAACACCAGCTAAGTGTAAAGAAAAGATGGTTAGGTCGACGGATGCCCCTGCGTGGGC
TAGGTTACCTGCCAGAGGCGGATATACTGTCCATCCTGTTCCGGCTCCGGCTTCAACACCAGAGGAGGCTAAAAGC
AGTAAGAATGAAGGGGGTAATAGTCAAAAGCTTATATTGTTTATTCGTGGGAATGCTATATCGGGGGCTCCAATTA
TGAGGGGTACTAATCAGTTTCCAAAGCCTCCAATAAGAATGGGTATAACTATAAAGAAAATTATTACGAAAGCATG
GGCGGTAACGATTACATTGTAGATTTGATCGTCGCCTAGAAGTGATCCGGGTTGGCTCAGTTCGGCTCGAATGAGG
AGACTTAAAGCAGTTCCTACTATCCCGGCTCAGGCACCAAATACAAGATAAAGAGTGC
```

[GC- content: 44%, DNA sequence length: 666bp]

>MN087473 *Chagunius chagunio* cytochrome oxidase subunit 1 (CO1) gene, partial cds; mitochondrial

```
CGAAGAATCAGAACAGGTGTTGATATAAAATTGGATCCCCACCCCCTGCCGGGTCGAAGAATGTGGTGTTGAGGTT
ACGATCTGTTAGAAGTATTGTAATTCCTGCGGCTAAAACTGGAAGGGATAAAAGAAGCAGCACGGCAGTTACAAGC
ACAGATCACACAAATAAGGGTGTTTGATATTGGGAGATAGCTGGAGGTTTCATATTAATAATTGTGGTGATGAAAT
TAATTGCTCCCAGAATTGATGAAACACCAGCTAAGTGTAAAGAAAAGATGGTTAGGTCGACGGATGCCCCTGCGTG
GGCTAGGTTACCTGCCAGAGGCGGATATACTGTCCATCCTGTTCCGGCTCCGGCTTCAACACCAGAGGAGGCTAAA
AGCAGTAAGAATGAAGGGGGTAATAGTCAAAAGCTTATATTGTTTATTCGTGGGAATGCTATATCGGGGGCTCCAA
TTATGAGGGGTACTAATCAGTTTCCAAAGCCTCCAATAAGAATGGGTATAACTATAAAGAAAATTATTACGAAAGC
ATGGGCGGTAACGATTACATTGTAGATTTGATCGTCGCCTAGAAGTGATCCGGGTTGGCTCAGTTCGGCTCGAATG
AGGAGACTTAAAGCAGTTCCTACTATCCCGGCTCAGGCACCAAATACAAGATAAAGAGTGCT
```

[GC- content: 44%, DNA sequence length: 669bp]

The BLAST result showed significant alignment of the DNA sequence of the fish samples from Phewa lake with the 16 reference sequences of *Chagunius chagunio* from GenBank belonging to accession numbers: AP011373.1, JX066746.1, KF742437.1, KJ476811.1, KJ476810.1, KJ476809.1, KJ476808.1, KU667387.1, MK029815.1, KY853031.1, MG604366.1, KY290058.1, MH545566.1, MG736389.1, MH102304.1 and JN965199.1. (Table 2)

Table 2. BLAST result of the alignment of DNA sequences of the fish samples with the GenBank database

S. No.	Query cover	E-value	Percent Identity	Scientific Name
1.0	75 % to 100%	0.0	97.29% to 100%	*Chagunius chagunio*

3.2 Multiple Sequence Alignment of DNA Sequences

Multiple Sequence Alignment among 19 homologous DNA sequences of *Chagunius chagunio* was carried out, among which two sequences belonging to MN087473 and MN087472 are from Phewa lake, four sequences belonging to MK572094, MH102304, MK572092 and MK572093 are from Bangladesh and the remaining are from India.

```
MG604366.1     ------------------------------------------------------------ 0

KY853031.1     ---------------------------------------------CCG--------- 3

MK029815.1     --------ATTGGCACTCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 38

KY290058.1     ------------------------------------------------------------ 0

JX066746.1     --AAAGACATTGGCACTCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 44
```

```
KJ476811.1    ----------------TCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 30
KJ476810.1    ----------------TCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 30
MK572094.1    ----------------TCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 30
KU667387.1    ----------------TCTTTATCTTGTATTTGGCGCCTGAGC-----CGG--------- 30
MG736389.1    -----------------------------------------C-----CGG--------- 4
KJ476808.1    ----------------TCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 30
MK572092.1    ----------------TCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 30
MK572093.1    ----------------TCTTTATCTTGTATTTGGTGCCTGAGC-----CGG--------- 30
KF742437.1    ---------------CTTTTATCTTGTATTTGGTGCCTGAGC-----CCG--------- 30
JN965199.1    ---------------------------------------------------------- 0
MH545566.1    ---------------------------------------------------------- 0
MN087472.1    --AGAATCAGAACAGGTGTTGATATAAAATTGGATCCCCACCCCCTGCCGGGTCGAAGAA 58
MN087473.1    GAAGAATCAGAACAGGTGTTGATATAAAATTGGATCCCCACCCCCTGCCGGGTCGAAGAA 60
MH102304.1    --------------------------------------CCCCCTGCCGGGTCGAAAAA 20


MG604366.1    ----------ATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 40
KY853031.1    ---------GATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 44
MK029815.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 79
KY290058.1    -----------TAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 39
JX066746.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 85
KJ476811.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
KJ476810.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
MK572094.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
KU667387.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
MG736389.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 45
KJ476808.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
MK572092.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
MK572093.1    ---------AATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
KF742437.1    ---------GATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 71
JN965199.1    -----------------------------------ATTCGAGCCGAAC---------- 13
MH545566.1    ---------ATAGTAGGAACTGCTTTAAGTCTCCTCATTCGAGCCGAAC---------- 40
MN087472.1    TGTGGTGTTGAGGTTACGATCTGTTAGAAGTATTGTAATTCCTGCGGCTAAAACTGGAAG 118
MN087473.1    TGTGGTGTTGAGGTTACGATCTGTTAGAAGTATTGTAATTCCTGCGGCTAAAACTGGAAG 120
MH102304.1    TGTGGTGTTGAGGTTACAATCTGTTAAAAGTATTGTAATTCCTGCGGCTAAAACTGGAAG 80
                                                     ****  ** *


MG604366.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 100
KY853031.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 104
MK029815.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 139
KY290058.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 99
JX066746.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 145
KJ476811.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
```

```
KJ476810.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
MK572094.1    TAAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
KU667387.1    TAAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
MG736389.1    TAAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 105
KJ476808.1    TAAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
MK572092.1    TAAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
MK572093.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
KF742437.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 131
JN965199.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 73
MH545566.1    TGAGCCAACCCGGATCACTTCTAGGCGACGATCAAATCTACAATGTAATCGTTACCGCCC 100
MN087472.1    GGATAAAAGAAGCAGCACGGCAGTTACAAGCACAGATCACACAAATAAGGGTG------- 171
MN087473.1    GGATAAAAGAAGCAGCACGGCAGTTACAAGCACAGATCACACAAATAAGGGTG------- 173
MH102304.1    GGATAAAAGAAGCAGTACGGCAGTTACAAGCACAGATCACACAAATAAGGGTG------- 133
              *    **    * *  **  *       * *  ** ***    *  ***  **


MG604366.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 159
KY853031.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 163
MK029815.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 198
KY290058.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 158
JX066746.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 204
KJ476811.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
KJ476810.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
MK572094.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
KU667387.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
MG736389.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 164
KJ476808.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
MK572092.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
MK572093.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
KF742437.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 190
JN965199.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 132
MH545566.1    ATGCTTTCGTAATAATTTTCTTTATAGTTATACCCATTCTTATTGGAGGC-TTTGGAAAC 159
MN087472.1    ---------TTTGATATTGGGAGATAGCTGGAGGTT-TCATATTAATAATTGTGGTGATG 221
MN087473.1    ---------TTTGATATTGGGAGATAGCTGGAGGTT-TCATATTAATAATTGTGGTGATG 223
MH102304.1    ---------TTTGATATTGGGAGATAGCTGGAGGTT-TCATATTAATAATTGTGGTGATG 183
              *    *   **     **** *  *     ** ****      * *   *


MG604366.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 219
KY853031.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 223
MK029815.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 258
KY290058.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 218
JX066746.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 264
KJ476811.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
```

```
KJ476810.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
MK572094.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
KU667387.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
MG736389.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 224
KJ476808.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
MK572092.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
MK572093.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
KF742437.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 250
JN965199.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 192
MH545566.1    TGATTAGTACCCCTCATAATTGGAGCCCCCGATATAGCATTCCCACGAATAAACAATATA 219
MN087472.1    AAATTAATTGCTCCCAGAATTGATGAAACACC-------AGCTAAGTGTAAAGAAAAGA 273
MN087473.1    AAATTAATTGCTCCCAGAATTGATGAAACACC-------AGCTAAGTGTAAAGAAAAGA 275
MH102304.1    AAATTAATTGCTCCCAAAATTGATGAAACACC-------AGCTAAGTGTAAAGAAAAGA 235
              ****  *   *  *  **  *****   *     *          *  *  *   ****  **  *  *


MG604366.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 271
KY853031.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 275
MK029815.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 310
KY290058.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 270
JX066746.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 316
KJ476811.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
KJ476810.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
MK572094.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
KU667387.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
MG736389.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 276
KJ476808.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
MK572092.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
MK572093.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
KF742437.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 302
JN965199.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 244
MH545566.1    AGCTTTTGACTATTA--CCCCCTTCATTCTT------ACTGCTTTTAGCCTCCTCTGGTG 271
MN087472.1    TGGTTAGGTCGACGGATGCCCCTGCGTGGGCTAGGTTACCTGCCAGAGGCGGATATACTG 333
MN087473.1    TGGTTAGGTCGACGGATGCCCCTGCGTGGGCTAGGTTACCTGCCAGAGGCGGATATACTG 335
MH102304.1    TGGTTAGGTCGACGGATGCCCCTGCGTGGGCTAGGTTACCTGCCAGAGGCGGATATACTG 295
              *  **   *  *  *      *****  *  *          **       **  *    *  *   **


MG604366.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 331
KY853031.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 335
MK029815.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 370
KY290058.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 330
JX066746.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 376
KJ476811.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
KJ476810.1    TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
```

```
MK572094.1     TTGAAGCCGGAGCTGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
KU667387.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
MG736389.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 336
KJ476808.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
MK572092.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
MK572093.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTTTGGCAGGTAACCTAGCCC 362
KF742437.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 362
JN965199.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 304
MH545566.1     TTGAAGCCGGAGCCGGAACAGGATGGACAGTATATCCGCCTCTGGCAGGTAACCTAGCCC 331
MN087472.1     TCCATCCTGTTCCGGCTCCGGCTTCAACACCAGAGGAGGCTAAAAGCAGTA------AGA 387
MN087473.1     TCCATCCTGTTCCGGCTCCGGCTTCAACACCAGAGGAGGCTAAAAGCAGTA------AGA 389
MH102304.1     TCCATCCTGTTCCGGCTCCGGCTTCAACACCAGAGGAGGCTAAAAGCAGTA------AAA 349
               *   *  * *    * *    * *  *  *** * *   * **      ***


MG604366.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 383
KY853031.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 387
MK029815.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 422
KY290058.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 382
JX066746.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 428
KJ476811.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
KJ476810.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
MK572094.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
KU667387.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
MG736389.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 388
KJ476808.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
MK572092.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
MK572093.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
KF742437.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 414
JN965199.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 356
MH545566.1     ACGCAGGGGCATCCGTCGACCTAACCATCTTTTCTTTACACTTAGCTGG--TGT------ 383
MN087472.1     ATGAAGG--GGGTAATAGTCAAAAGCTTATATTGTTTATTCGTGGGAATGCTATATCGGG 445
MN087473.1     ATGAAGG--GGGTAATAGTCAAAAGCTTATATTGTTTATTCGTGGGAATGCTATATCGGG 447
MH102304.1     ATGAAGG--GGGTAATAGTCAAAAGCTTATATTGTTTATTCGTGGGAATGCTATATCGGG 407
               * * ***       * *  ** * *  ** **** *  *      * *


MG604366.1     TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 443
KY853031.1     TTCATCAATTTTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 447
MK029815.1     TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 482
KY290058.1     TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 442
JX066746.1     TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 488
KJ476811.1     TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
KJ476810.1     TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
MK572094.1     TTCATCAATTCTGGGGGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
```

```
KU667387.1   TTCATCAATTCTGGGGGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
MG736389.1   TTCATCAATTCTGGGGGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 448
KJ476808.1   TTCATCAATTCTGGGGGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
MK572092.1   TTCATCAATTCTGGGGGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
MK572093.1   TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
KF742437.1   TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 474
JN965199.1   TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 416
MH545566.1   TTCATCAATTCTGGGAGCAATTAATTTCATCACCACAATTATTAATATGAAACCTCCAGC 443
MN087472.1   GGCTCCAATTATGAGGGGTACTAATCAGTTTCCAAAGCCTCCAATAAGAATGGGTATAAC 505
MN087473.1   GGCTCCAATTATGAGGGGTACTAATCAGTTTCCAAAGCCTCCAATAAGAATGGGTATAAC 507
MH102304.1   GGCTCCAATTATGAGGGGTACTAATCAGTTTCCAAAGCCTCCAATAAAAATGGGTATAAC 467
              *   ***** ** * *  * ****    *  * *    *   *  *   *    *  * *


MG604366.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 487
KY853031.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 491
MK029815.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 526
KY290058.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 486
JX066746.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 532
KJ476811.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
KJ476810.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
MK572094.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
KU667387.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
MG736389.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 492
KJ476808.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
MK572092.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
MK572093.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTATGATCTGTGCTTG 518
KF742437.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 518
JN965199.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 460
MH545566.1   TATCTCCCAATATCAAAC---------------ACCCTTATTTGTGTGATCTGTGCTTG 487
MN087472.1   TATAAAGAAAATTATTACGAAAGCATGGGCGGTAACGATTACATTGTAGATTTGATCGTC 565
MN087473.1   TATAAAGAAAATTATTACGAAAGCATGGGCGGTAACGATTACATTGTAGATTTGATCGTC 567
MH102304.1   TATAAAAAAAATTATTACAAAAGCATGGGCGGTAACAATTACATTGTAAATTTGATCGTC 527
              ***     **  *   **              **  *** *     ** **  * *


MG604366.1   TAACTGCCGTACTGCTTCTTTTGTCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 547
KY853031.1   TAACTGCCGTACTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 551
MK029815.1   TAACTGCCGTACTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 586
KY290058.1   TAACTGCCGTACTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 546
JX066746.1   TAACTGCCGTACTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 592
KJ476811.1   TAACTGCCGTACTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
KJ476810.1   TAACTGCCGTACTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
MK572094.1   TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
KU667387.1   TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
```

```
MG736389.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 552
KJ476808.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
MK572092.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
MK572093.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
KF742437.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 578
JN965199.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCC------------------ 501
MH545566.1    TAACTGCCGTGCTGCTTCTTTTATCCCTTCCAGTTTTAGCCGCAGGAATTACAATACTTC 547
MN087472.1    GCCTAGAAGTGATCCGGGTTGGCTCAGTTC----------GGCTCGAATGAGGAGACTTA 615
MN087473.1    GCCTAGAAGTGATCCGGGTTGGCTCAGTTC----------GGCTCGAATGAGGAGACTTA 617
MH102304.1    CCCTAAAAGTGATCCGGGTTGGCTCAGTTC----------GGCTCAAATGAGGAGACTTA 577
                   **   *  *    **    **   ***


MG604366.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 607
KY853031.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 611
MK029815.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCA---- 642
KY290058.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 606
JX066746.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 652
KJ476811.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
KJ476810.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
MK572094.1    TAACAGATCGTAATCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
KU667387.1    TAACAGATCGTAATCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
MG736389.1    TAACAGATCGTAATCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 612
KJ476808.1    TAACAGATCGTAATCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
MK572092.1    TAACAGATCGTAATCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
MK572093.1    TAACAGATCGTAATCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
KF742437.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATTT 638
JN965199.1    ------------------------------------------------------------ 501
MH545566.1    TAACAGATCGTAACCTCAACACCACATTCTTCGACCCGGCAGGGGGTGGGGATCCAATT- 606
MN087472.1    AAGCAGTTCCTACTATCC-----------------CGGCTCAGGCACCAAATACAAGAT 657
MN087473.1    AAGCAGTTCCTACTATCC-----------------CGGCTCAGGCACCAAATACAAGAT 659
MH102304.1    AAGCAGTTCCTACTATTC-----------------CGGCTCAGGCACCAAATACAAAAT 619


MG604366.1    TATATCAGCATCTGTTTTGATTCTTTGGCCACCCTGAAGTC    648
KY853031.1    TATATCAACATCTGTTTTGATTCTTTGGCCACCA-------    645
MK029815.1    -----------------------------------------    642
KY290058.1    TATATCAACATCTGTTT------------------------    623
JX066746.1    TATATCAACATCTGTTTTGATTCTTT---------------    678
KJ476811.1    TATATCAACATCTGTTT------------------------    655
KJ476810.1    TATATCAACATCTGTTT------------------------    655
MK572094.1    TATATCAACACCTGTTT------------------------    655
KU667387.1    TATATCAACACCTGTTT------------------------    655
MG736389.1    TA---------------------------------------    614
```

```
KJ476808.1      TATATCAACACCTGTTT----------------------    655

MK572092.1      TATATCAACACCTGTTT----------------------    655

MK572093.1      TATATCAACACCTGTTT----------------------    655

KF742437.1      TATATCAACACCTG-------------------------    652

JN965199.1      --------------------------------------    501

MH545566.1      --------------------------------------    606

MN087472.1      AAAGAGTGC-----------------------------    666

MN087473.1      AAAGAGTGC-----------------------------    668

MH102304.1      AAAAAGTGCCAATGTCTTTGTGGTTGGTTGACTG------    653
```

Figure 2. Multiple sequence alignment of 19 homologous DNA sequences based on Clustal Omega. The stars show number of matches among the sequences

The percentage of similarity among the aligned sequences is calculated based on the following equation.

$$Percentage\ of\ similarity = \frac{Number\ of\ matches}{Length\ of\ the\ shortest\ DNA\ sequence} \times 100$$
$$Percentage\ of\ similarity = \frac{197}{501} \times 100$$
$$= 39.3\%$$

(1)
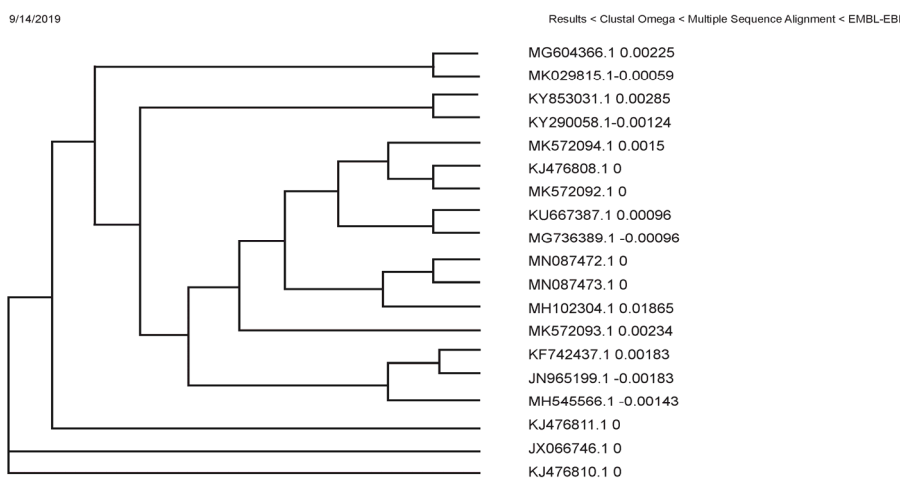
Figure 3. Multiple sequence alignment based Neighbour-Joining Tree without distance corrections

4. Discussion

The DNA sequence from the two samples of fish from Phewa lake scored the highest alignment scores against 16 reference sequences of *Chagunius chagunio* in BLAST result in terms of Query Cover (75% to 100%) and Percentage Identity (97.29% to 100%).The Expect Value (E-value) remained 0.0. This confirmed the identification of the samples as *Chagunius chagunio*. The two DNA sequences with the same percentage of GC-content and similar BLAST results provide strong evidence that the two samples of fish belong to the same species.

Sequence alignment or sequence comparison is a way of arranging biological sequences of DNA, RNA or proteins in order to distinguish regions of similarity which help us to determine how closely or distantly the organisms are related. Multiple sequence alignment is a sequence alignment of three or more biological sequences to find out their homology so that phylogenetic analysis can be obtained. In evolutionary biology, homology refers to any similarity between characters that is due to their shared ancestry. The homology among proteins and DNA is often concluded on the basis of sequence similarity.

Sequence similarity and sequence identity are synonymous for nucleotide sequences. An identity of 39.3% in the present multiple sequence alignment of nucleotide sequences is highly desirable and suggests similarity of function and structure among the aligned sequences. According to NEB (2019), an identity of 25% or higher suggests the potential for similarity of function or structure of the aligned DNA sequences.

The most widely used approach to multiple sequence alignment is progressive technique that builds a hierarchical or tree diagram showing relationship between the sequences based on neighbor- joining method or UPGMA. In the multiple sequence alignment, if the species are closely related or while comparing individuals of the same species, it is better to use the DNA sequences. This is because the protein sequences are too similar and only a few results would be obtained. Based on the Neighbour-Joining tree, MN087473 and MN087472 which includes the DNA sequences of *Chagunius chagunio* from Phewa lake are monophyletic and they show very close relationship with MH102304 that includes *Chagunius chagunio* from Bangladesh.

5. Conclusion

The main purpose of the present study is to give identification to *Chagunius chagunio* of Phewa lake by barcoding its mitochondrial DNA and depositing the barcode sequence in the GenBank. There is no availability of DNA sequence of *Chagunius chagunio* in the GenBank database from Nepal as the molecular study of fish of Nepal is still at preliminary stage. This is the first deposition of DNA sequence of Nepalese *Chagunius chagunio* in the GenBank database which can act as a reference sequence for future studies.

The multiple sequence alignment of the DNA sequences show that the *Chagunius chagunio* of Phewa lake has close relationship with *Chagunius chagunio* from Bangladesh.

The DNA barcoding is expensive and takes nearly three months for the completion of the process including submission to the GenBank. However, it is a useful tool to quickly and accurately identify species and has the potential to prompt the discovery of new species.

Acknowledgement

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press, Cambridge, Massachusetts.

Avise, J. C., & Walker, D. (1999). Species realities and numbers in sexual vertebrates: Perspectives from an asexually transmitted genome. *Proc. Natl Acad. Sci., 96*. 992-995. USA.

Birky, C. W. (2001). The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annu Rev Genet, 35*, 125-48.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Res, 27*, 1767-1780.

Cowan, R. S., Chase, M. W., Kress, W. J., & Savolainen, V. (2006). 300,000 species to identify: problems, progress, and prospects in NDA barcoding of land plants. *Taxon, 55*, 611-616.

EMBI-EBI. (2019). *European Molecular Biology Laboratory-European Bioinformatics Institute*.

Giri, K. (2013). *Fish Distribution Pattern and Status of Fisher Community in Phewa lake*. LAP LAMBERT Academic Publishing.

Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A., (2007). DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics, 23*(4), 167-172.

Hebert, P. D. N., Ratnasingam, S., & Dewar, J. R. (2003). Barcoding animal life: Cytochrome c oxidases subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B (Suppl.), 270*, s96-s99.

Hedrick, P. W. (2001). Conservation genetics: Where are we now? *Trends in Ecology and Evolution, 16*(11), 629-636.

Hlaing, T., Tun-Lin, W., Somboon, P., Socheat, D., Setha, T., Min, S., Chang, M. S., & Walton, C. (2009). Mitochondrial pseudogenes in the nuclear genome of Aedesaegypti mosquitoes: Implications for past and future population genetic studies. *BMC Genet,* 10-11.

Hu, M., Chilton, N. B., & Gasser, R. B. (2004). The mitochondria genomics of parasitic nematodes of socioeconomic importance :recent progress, and implications for population genetics and systematics. *Adv Parasitol, 56*, 133-212

Hubert et al. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *PloS One, 3*, 24-90.

Ingman, M., Kaessmann, H., Paabo, S., & Gyllensten, V. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature, 408*, 708-713.

Ivanova, N., Zemlak, T., Hanner, R., & Hebert, P. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes, 7*(4), 544-548.

Khan, H. A., Arifia, A., Homaidan, A., & AL Farhan, A. H. (2008). Application of 16s rRNA, cytochrome b and control, region dequences for understanding the phylogenetic relationships in Oryx species. *Gent. Mol. Res., 7*, 1392-1397.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution, 35*, 1547-1549.

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics, 30*(22), 3276-3278.

NEB. (2019). *The Beginners Guide to DNA sequences Alignment*. New England Biolabs 240 County Road.

Olivo, D. D., Van de Walle, M. J., Laipis, P. J., & Hauswirth, W. W. (1983). Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature, 306*, 400-402.

Pokharel, S. (2003). *Lessons from Nepal on Developing a Strategic Plan for the Integrated Lake Basin Management: Conservation of Phewa Lake of Pokhara, Nepal (PDF)*. International Lake Environment Committee: World Lake Database.

Rai, A. K. (2000). Evaluation of natural food for planktivorous fish in Lakes Phewa, Begnas, and Rupa in Pokhara Valley, Nepal (PDF). *Limnology, 1*, 81-89.

Shrestha, P. (2003). Conservation and management of Phewa Lake ecosystem, Nepal (PDF). *Aquatic Ecosystem Health and Management Society* (pp. 1-4).

Shrestha, T. K. (2019). *Ichthyology of Nepal*. B.J.Shrestha Publisher, Kathmandu, Nepal.

Ward, R. D., Zemlac, T. C., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B, 360*, 1847-1857.

Wheat, C. W., & Watt, W. B. (2008). A mitochondrial –DNA-based phylogeny for some evolutionary – genetic model specis of Colias butterflies (Lepidoptera, Pieridae). *Mol. Phylogenet Evol, 47*, 893-902.

Williams, S. T., & Knowlton, N. (2001). Mitochondrial pseudogenes are pervasive and often indidious in the snapping shrimp Genus Alpheus. *Mol Biol Evol, 18*, 1484-93.

Zhang, D. X., & Hewitt, G. M. (1997). Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. *Insect Molecular Biology, 6*, 143-150.