International Journal of Advanced Science and Research
ISSN: 2455-4227
Impact Factor: RJIF 5.12
www.allsciencejournal.com
Volume 3; Issue 1; January 2018; Page No. 60-64

COI vs 16s rRNA sequences: Molecular characterization of *Culex* (Diptera: Culicidae) vector species from India

Gatreddi Srinu
Department of Zoology & Aquaculture, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India

Abstract
Culex is the main source of vector for several diseases in India. In the present study, a comprehensive genetic diversity study of Cytochrome Oxidase I (COI) and 16s rRNA in 2 species of *Culex* mosquitoes (*Culex tritaeniorhynchus* and *Culex quinquefasciatus*) from India was carried out using 257 partial sequences. Genetic divergence analysis based on COI and 16s rRNA sequences revealed a clear gap between intraspecific and interspecific distances of 2 species from India. However, 16sRNA based phylogeny of *Culex* species failed to adequately differentiate species in a whole country context whereas COI sequences were succeed. Average genetic divergence values of COI gene for within species for *Culex quinquefasciatus* and *Culex tritaeniorhynchus* were 0.020% and 0.010% respectively whereas for 16s rRNA gene average genetic divergence within species for *Culex quinquefasciatus* and *Culex tritaeniorhynchus* were 0.001% and 0.005% respectively. Maximum value of A/T content observed in COI sequence (72.0%) followed by 16s rRNA sequences (41.7%) followed by COI partial sequences (34.7%). The GC2 content is dominated followed by GC3 and GC1 in COI sequences indicates the synonymous mutations occur mostly at the GC2 followed by GC3 and GC1. It can be concluded as validation with COI mitochondrial gene was successful and helps in successful estimation of genetic structure of *Culex* species in India.

Keywords: *Culex tritaeniorhynchus*, *Culex quinquefasciatus*, Molecular

Introduction
India is one of the major hotspot for mosquito-vectored diseases. Out of 3500 mosquito species documented worldwide (www.mosquito-taxonomic-inventory.info), nearly 400 mosquitoes have been documented and identified as a carriers for different diseases [1]. Seasonal outbreaks like fatal Acute Encephalitis Syndrome (AES) occur regularly in several parts of India especially since 1955, which mainly transmitted by *Culex* species. Japanese encephalitis disease is the major and consistent outbreak in different parts of the country like Gorakhpur of Utter Pradesh [4], Bankura district of West Bengal, Tamil Nadu [3]. Indian studies have revealed a number of secondary vectors from *Culex* species (*C. pseudovishnui, C. whitmorei, C. gelidus, C. epidesmus*) [2], apart from other phyla organisms.
Morphological identification of mosquitoes is difficult and time consuming [5]. Correct vector identification is important for the development of effectual strategies for the control of transmission vectors [6, 7]. Different genes of mitochondrial origin were proved as efficient for genetic divergence studies in mosquitoes. COI gene based DNA Barcoding gained importance as a taxonomic tool where morphological ambiguities were found and also in case of taxonomy of new species that uses a short genetic marker in an organism's DNA for its identification [8]. It involves the use of the 5' end of the mitochondrial COI gene (~650bp fragment) gene in eukaryotes to diagnose and define the species boundaries [9]. It is preferred over other mitochondrial genes as Indels (Insertions and Deletions) are very rare which create difficultness in sequence alignments [10]. Ribosomal RNA encoded by the mt-genome is commonly used in molecular phylogenetic as well as genetic variation studies because they have conserved sequences and exhibits high level of polymorphism. Moreover, they show ten times faster evolutionary rate than the nuclear genome and it is subjected to maternal and sexual inheritance [14]. The 16s rRNA gene is found in the conserved portion of mtDNA and has been extensively used in phylogenetic as well as evolutionary studies [15, 16].
Many researchers have successfully used DNA Barcoding technique for mosquito's identification [5, 11, 12, 13]. Study of genus *Culex* on DNA based methods have mainly focused on the intra and interspecific identification, phylogenies, and divergence for various species [17, 18]. In the present study, we have used to multi-locus approach to determine genetic divergence using 2 mitochondrial genes i.e. COI and 16s rRNA in *Culex tritaeniorhynchus* and *Culex quinquefasciatus* from India. Finally, we have determined the effectiveness of these 2 genes for genetic divergence and phylogenetic analysis *Culex* species from India.

Materials and methods
In total, 257 partial gene sequences obtained for two species with the range of sequences from 3 sequences (16s rRNA sequences of *Culex tritaeniorhynchus*) to 148 sequences (COI sequences of *Culex tritaeniorhynchus*) (Table 1). Number of analysed partial sequences for the gene COI includes 207 sequences (*Culex tritaeniorhynchus*-148 and *Culex quinquefasciatus*-59) whereas for 16s rRNA gene includes 50 sequences (*Culex quinquefasciatus*-47 and *Culex tritaeniorhynchus*-03). All sequences were assembled and

end-trimmed to get homologous region to avoid errors during sequencing and those sequences subjected to aligned using ClustalW analysis tool [19]. Sequences with sufficient length only were considered with the view of bringing uniformity in analysis across all species. To ensure homology in heterogeneous sequences, some bases were trimmed. To bring this homogeneity in some sequences, missing sequence parts were adopted from most conserved regions of the sequences available in NCBI GenBank for the same species. Nucleotide composition (A, T, G, C, GC1, GC2 & GC3) calculated for homologous end-trimmed sequences using MEGA V.7.0 (Molecular Evolutionary Genetic Analysis) [20] software (Arizona). Inter and intra species evolutionary divergences in various hierarchical levels were analysed using Kimura 2 Parameter method [21]. The variation was estimated following the bootstrap method with 5000 bootstrap replicate values. The pair-wise deletion option was selected to treat the gaps or missing data between each compared specimen. Finally, the Maximum Likelihood (ML) tree among species was created to give distance values using bootstrap method in Kimura 2 - parameter mode of analysis those values are in the units of the number of base substitutions per site [22]. To verify the robustness of the nodes of the ML tree, bootstrap analysis was carried out using 5000 pseudo replicates [23]. Both transitions and transversions were cumulated and included as substitutions. Missing bases or gaps were treated by adopting pair wise deletion method employed in MEGA V 7.0. [20].

Results and Discussion

All the retrieved sequences were verified thoroughly and no complexity or ambiguities were observed among them. Out of two analysed sequences, one is (COI) protein coding gene whereas remaining is (16s rRNA) non coding gene. The average length values of the aligned and end-trimmed sequences were presented in table 1.

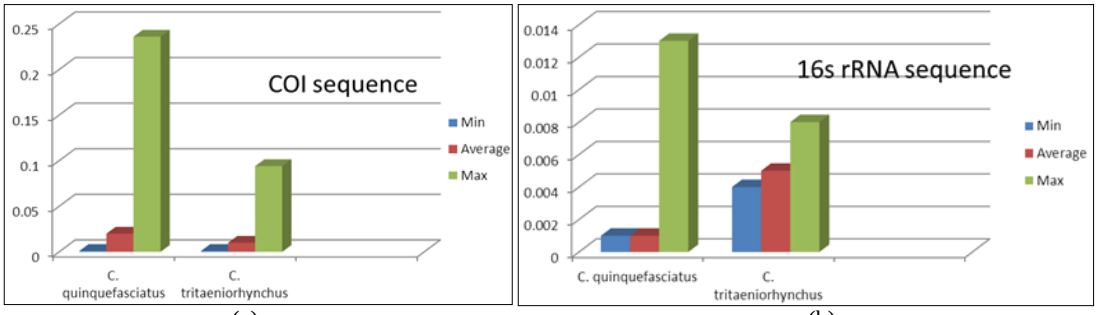
Table 1: List of COI GenBank accession numbers along with their sequences sequence length two mitochondrial genes (N = No. of Sequences)

S. No.	Species(N)	GenBank accession no.	Length
COI gene			
1	*Culex quinquefasciatus* (59)	KC970292-98, EU259297, FJ536189-94, FJ536181-86, FJ536179, FJ536172-77, FJ536168-70, KU920683-97, KR817730, AB907179, KC250017, FN395204-05, FN395201-02, FM177756, KJ680549, FJ536187-88, FJ536180, FJ536178, FJ536171	491
2	*Culex tritaeniorhynchus* (148)	DQ424952, AY834249, KM350641-69, KM362833-50, KC970299, AY917215, AY917206, HM638227, HM638223-25, HM638220-21, AY729975, KM3552-640	589
16s rRNA gene			
5	*Culex quinquefasciatus* (47)	EU555413-16, EU339128-47, EU711082-101, KX451138, KX524981-82	205
6	*Culex tritaeniorhynchus* (03)	KX085417, AF034469, EF593020	302

Average genetic divergence values of COI gene for within species for *Culex quinquefasciatus* and *Culex tritaeniorhynchus* were 0.020% and 0.010% respectively whereas for 16s rRNA gene average genetic divergence within species for *Culex quinquefasciatus* and *Culex tritaeniorhynchus* were 0.001% and 0.005% respectively. (Table 2). In almost all sequences, the range of genetic distances were comprehensively higher for COI gene followed 16s rRNA partial sequences. (Figure 1). Hence, it is evident that COI gene undergone drastic changes in the process of evolution compared to other genes.

Table 2: Average and range of Kimura 2-parameter distance values within 2 *Culex* species for 2 mitochondrial genes. (AD= Average distance, SE= Standard Error)

S. No.	Species	A.D. ± S.E.	Min	Max
COI gene				
1	*Culex quinquefasciatus*	0.020±0.003	0.00	0.236
2	*Culex tritaeniorhynchus*	0.010±0.002	0.00	0.094
16s rRNA gene				
5	*Culex quinquefasciatus*	0.001±0.000	0.000	0.013
6	*Culex tritaeniorhynchus*	0.005±0.004	0.004	0.008



(a) (b)

Fig 1: The average and range of intra genetic divergences for 2 mt- genes

Pair wise genetic distance matrix of 2 *Culex* species for both mt- genes sequences were represented in Table 3. The highest Kimura's 2-parameter (K-2P) genetic distance was observed between for COI gene was observed between *C. quinquefasciatus* and *C. tritaeniorhynchus* (0.071%) whereas the least value found between *C. quinquefasciatus* and *C. tritaeniorhynchus* (0.000%) for 16s rRNA sequences**.**

Table 3: Inter species pair-wise genetic distances using K-2P model for 2 mt-genes

S. No.	Species	*C. quinquefasciatus*
COI gene		
1	*C. quinquefasciatus*	-
2	*C. tritaeniorhynchus*	0.071±0.015
16s rRNA gene		
3	*C. quinquefasciatus*	-
4	*C. tritaeniorhynchus*	0.000±0.000

The average nucleotide frequencies for two *Culex* species were represented in Table 4. The average nucleotide percentage values for COI gene were A= 29.5%, T= 39.9%, G= 15.25%, C= 15.35%. The average nucleotide percentage values for 16s rRNA sequences were A= 45.3%, T= 38.1%, G= 10.75%, C= 5.81%. A/T content is higher than the G/C content across two mitochondrial genes. Maximum value of A/T content observed in COI sequence (72.0%) followed by 16s rRNA sequences (41.7%) followed by COI partial sequences (34.7%). The comparative analysis for A/T and G/C content values for two mt-genes revealed that COI gene undergone most recent evolution followed by 16s rRNA sequences.

The average GC content in 3 codon positions (GC1, GC2 and GC3) of COI, for two *Culex* species was depicted in figure 2. GC2 content is dominated followed by GC3 and GC1. This indicates that the synonymous mutations occur mostly at the GC2 followed by GC3 and GC1. This non-synonymous mutation effects resulted protein. The highest percentage of GC2 content for COI gene was noticed in *C. tritaeniorhynchus* (46.4%) whereas the lowest value was observed in *C. quinquefasciatus* (42.4%). Average GC content was higher than the COI gene GC content of Arthropod dataset suggested by Keskin & Atar [24].

Table 4: Calculated nucleotide frequecies for coding genes along with their S.E. values

S. No.	Species	A	T	G	C	(GC)1	(GC)2	(GC)3
1	*C. quinquefasciatus*	28.8±0.04	42.0±0.14	14.4±0.09	14.8±0.17	6.89±0.48	42.4±1.15	38.7±0.12
2	*C. tritaeniorhynchus*	30.2±0.06	37.8±0.03	16.1±0.06	15.9±0.10	6.75±0.26	46.4±0.06	42.9±0.08

Table 5: Calculated nucleotide frequecies for non-coding genes along with their S.E. values

S. No.	Species	A	T	G	C
1	*C. quinquefasciatus*	45.9±0.06	38.5±0.03	10.2±0.09	5.35±0.06
2	*C. tritaeniorhynchus*	44.7±0.03	37.7±0.03	11.3±0.23	6.27±0.09

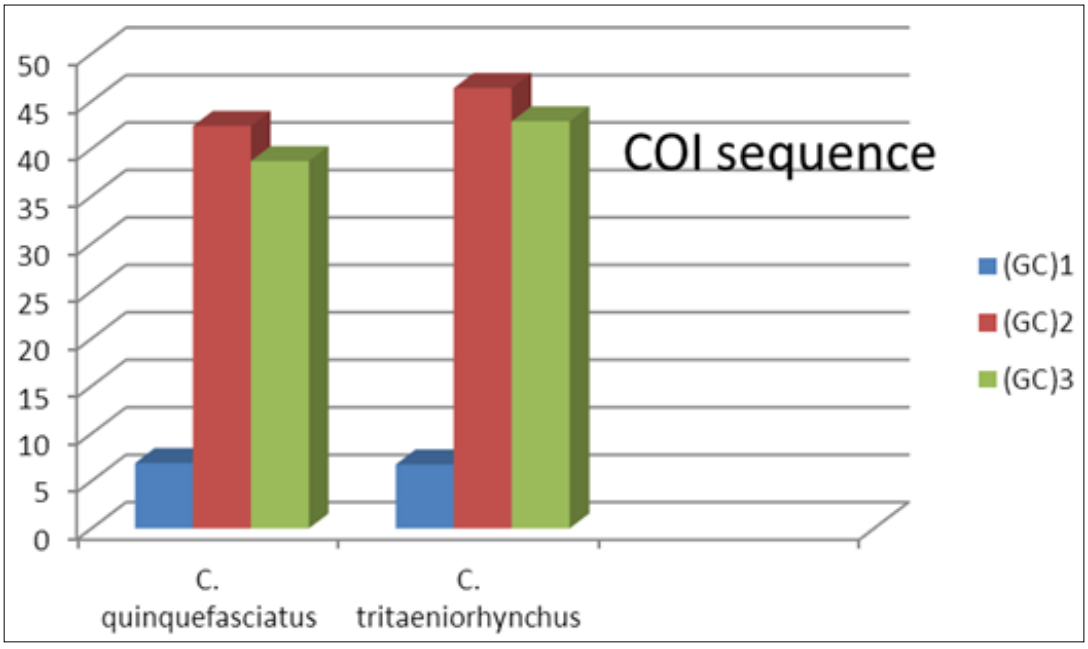
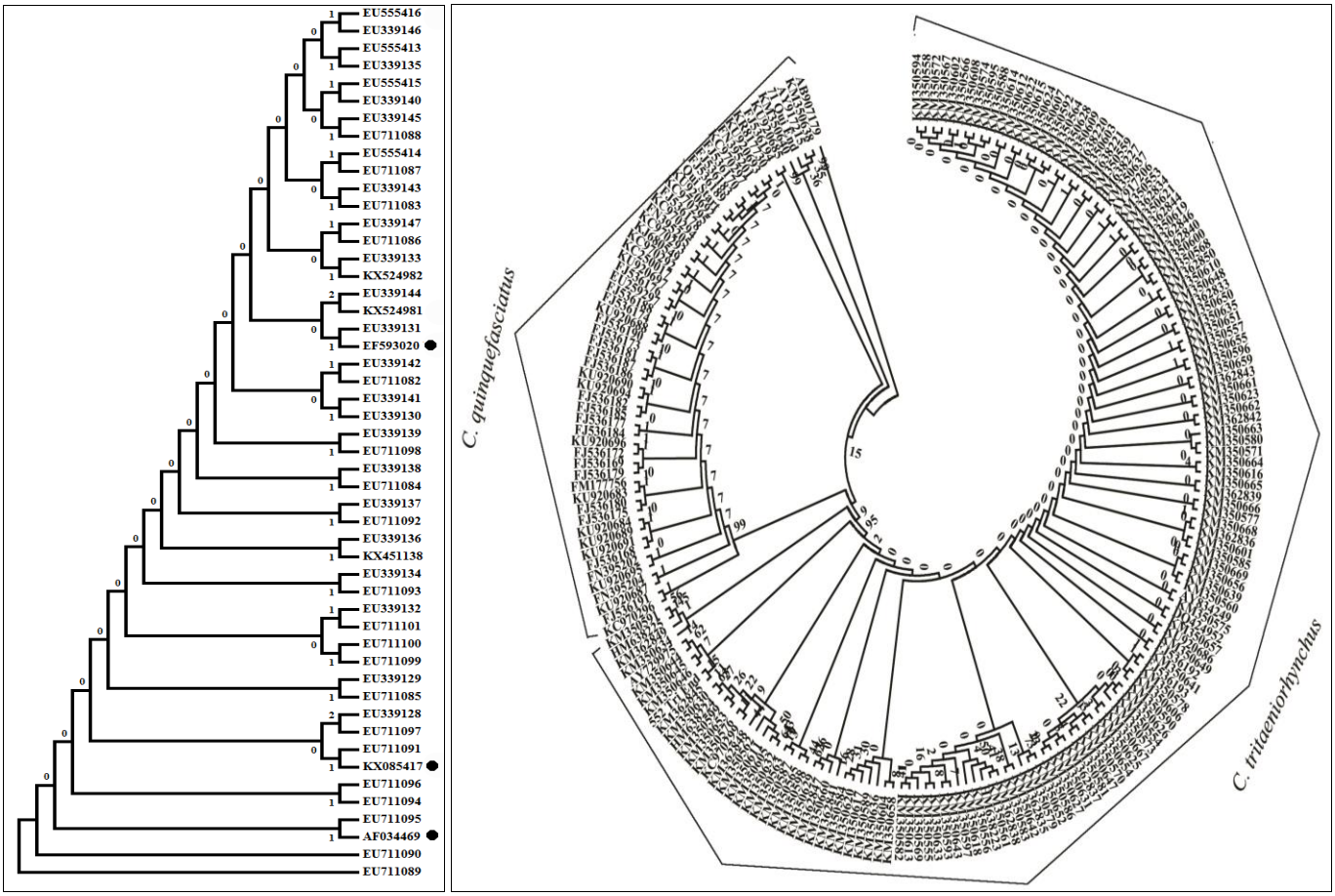


Fig 2: Variation in GC content (%) of COI sequences for *Culex* species

The Maximim-Likelihood trees for two mt genes were derived by using all 257 sequences of 2 species for 2 mt-partial sequences with MEGA 6.0 are shown in figure 3. Both the M-L trees resulted from COI sequences and 16s rRNA sequences were showed contrasting results. 16s rRNA sequences were failed to form clades with same species where as COI sequences formed the clades with species belongs to same species with significant boot strap values.



(a) (b)

Fig 3: Traditional Kimura 2- Parameter distance Neighbour Joining tree constructed from partial mitochondrial gene sequences (a) 16s rRNA (b) COI.

Conclusion

It is concluded that the standard barcode region which is based on COI partial sequence contains sufficient genetic divergence levels with different hierarchical levels compared to 16s rRNA sequence Therefore, it is confirmed that validation with COI mitochondrial gene was successful and helps in successful estimation of genetic structure of *Culex* species in India. This study also provided the efficacy of COI gene in delineating the members of the evolutionary closely related organisms particularly in *Culex* species.

References

1. http://ccs.in/400-mosquito-species-india-all-carriers-diseases, accessed on 20th Jan, 2018.
2. Kanojia PC, Shetty PS, Geevarghese G. A long-term study on vector abundance & seasonal prevalence in relation to the occurrence of Japanese encephalitis in Gorakhpur district, Uttar Pradesh. Indian Journal of Medical Research. 2003; 117:104.
3. Namachivayam V, Umayal K, editors. Proceedings of the National Conference on Japanese Encephalitis. New Delhi: Indian Council of Medical Research, 1982, 30(3).
4. NVBDC P. Directorate General of Health services Ministry of Health and Family Welfare. New Delhi cited, 2009. Available from: http://nvbdcp.gov.in/je-cd.html
5. Ashfaq M, Hebert PDN, Mirza JH, Khan AM, Zafar Y, Mirza MS. Analyzing mosquito (Diptera: Culicidae) diversity in Pakistan by DNA barcoding. PLoS One. 2014; 9(5):e97268.
6. Otranto D, Capelli G, Genchi C. Changing distribution patterns of canine vector borne diseases in Italy: leishmaniosis vs. dirofilariosis. Parasites & Vectors. 2009; 2(1):S2.
7. Wilke ABB, Vidal PO, Suesdek L, Marrelli MT. Population genetics of neotropical Culex quinquefasciatus (Diptera: Culicidae). Parasites & vectors. 2014; 7(1):468.
8. Hebert PD, Cywinska A, Ball SL. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B: Biological Sciences. 2003; 270(1512):313-321.
9. Luo A, Zhang A, Ho SY, Xu W, Zhang Y, Shi W, *et al*. Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. Bmc genomics. 2011; 12(1):84.
10. Doyle JJ, Gaut BS. Evolution of genes and taxa: a primer. In Plant Molecular Evolution Springer Netherlands, 2000, 1-23.
11. Hemmerter S, Šlapeta J, Beebe NW. Resolving genetic diversity in Australasian Culex mosquitoes: incongruence between the mitochondrial cytochrome c oxidase I and nuclear acetylcholine esterase 2. Molecular phylogenetics and evolution. 2009; 50(2):317-325.

12. Kumar NP, Rajavel AR, Natarajan R, Jambulingam P. DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). Journal of medical entomology. 2007; 44(1):1-7.
13. Moore M, Sylla M, Goss L, Burugu MW, Sang R, Kamau LW, *et al*. Dual African origins of global Aedes aegypti sl populations revealed by mitochondrial DNA. PLoS neglected tropical diseases. 2013; 7(4):e2175.
14. Awadalla P, Eyre-Walker A, Smith JM. Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science. 1999; 286(5449):2524-2525.
15. Clary DO, Wolstenholme DR. The mitochondrial DNA molecule ofDrosophila yakuba: nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution. 1985; 22(3):252-271.
16. Beard CB, Hamm D, Collins FH. The mitochondrial genome of the mosquito Anopheles gambiae: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect molecular biology. 1993; 2(2):103-124.
17. Laurito M, de Oliveira TM, Almiron WR, Sallum MAM. COI barcode versus morphological identification of Culex (Culex) (Diptera: Culicidae) species: a case study using samples from Argentina and Brazil. Memórias do Instituto Oswaldo Cruz. 2013; 108:110-122.
18. Araki AS, Maia DA, Gil-Santana HR, de Mello CF, Martins ADJ, Alencar J. Variation in Mitochondrial Cytochrome c Oxidase I DNA Can Successfully Identify Culex (Melanoconion) pedroi (Diptera: Culicidae) and Culex (Melanoconion) ribeirensis (Diptera: Culicidae). Journal of medical entomology. 2016; 54(2):485-488.
19. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic acids research. 1997; 25(24):4876-4882.
20. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution. 2016; 33(7):1870-1874.
21. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980; 16:111-120.
22. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987; 4:406-425.
23. Felsinstin J. PHYLIP (Phylogeny Inference Package) version 3.50 Distributed by the author. Department of Genetics, University of Washington, Seattle, USA, 1993.
24. Keskin E, Atar HH. DNA barcoding commercially important aquatic invertebrates of Turkey. Mitochondrial DNA. 2013; 24(4):440-450.