International Journal of Entomology Research
ISSN: 2455-4758
Impact Factor: RJIF 5.24
www.entomologyjournals.com
Volume 2; Issue 5; September 2017; Page No. 79-82

Phylogenetic characterization of some species of family Calliphoridae using *COII* gene marker

[*1] Naina Khullar, [2] Devinder Singh, [3] Chandan Kumar Jha
[1, 2] Department of Zoology and Environmental Sciences, Punjabi University, Patiala, Punjab, India
[3] Department of Human Genetics, Punjabi University, Patiala, Punjab, India

Abstract
Cytochrome oxidase II (*COII*) gene is the second most widely used gene in phylogenetic analysis of different metazoans after the COI gene. This gene shows more percentage of conserved regions in comparison to COI gene. In the analysis four different forensically important flies were considered. The phylogenetic relationships were deciphered using four different approaches: UPGMA, Neighbour joining, Maximum Liklihood and Minimum Evolution.

Keywords: mtDNA, COII gene, interspecific, calliphorids

Introduction
COII gene is phylogenetically informative mitochondrial gene. It is a protein coding gene, so alignment is easier due to the presence of amino acids. Predetermined primers are known for it. Highly conserved and variable regions are closely associated in the *COI* gene, so perfect for tracing phylogeny. However in comparison to *COI* gene, it has lesser percentage of variable region. But still it has proved to be a good marker for tracing phylogeny of Calliphorids. In the present study three different species belonging to three different genera were chosen namely *Calliphora vicina*, *Lucilia porphyria, Chrysomya ruffifacies.*

Methodology
Different populations of the three species were chosen for the process. DNA was extracted using Qiagen DNA purification kit. Polymerase Chain Reaction was performed using specific *COII* gene primer pair. The amplicons were then sequenced and sequences were deposited in the Gen Bank. Phylogenetic analysis was carried out using MEGA 5 to interpret the interspecific relationship. The reaction protocol and the primer used are depicted in Table 1.

Table 1: PCR program for (C2-J-3400 and C2-N-3661) *COII* primer set.

Stage	Temperature °C	Time	Cycles
Initial Denaturation	98	2 minutes	Hold
Denaturation	98	30 seconds	38 cycles
Annealing	52.6	40 seconds	
Elongation	72	1 minute	
Final Elongation	72	7 minutes	Hold
Store	4	Infinity	Hold

Observations and Results
UPGMA: Phylogenetic tree was constructed using UPGMA method. The optimal tree with the sum of branch length equal to 0.23337565 is shown in Figure1. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. A good posterior probability (above 90%) was observed. The evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. The analysis involved 6 nucleotide sequences. There were a total of 297 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

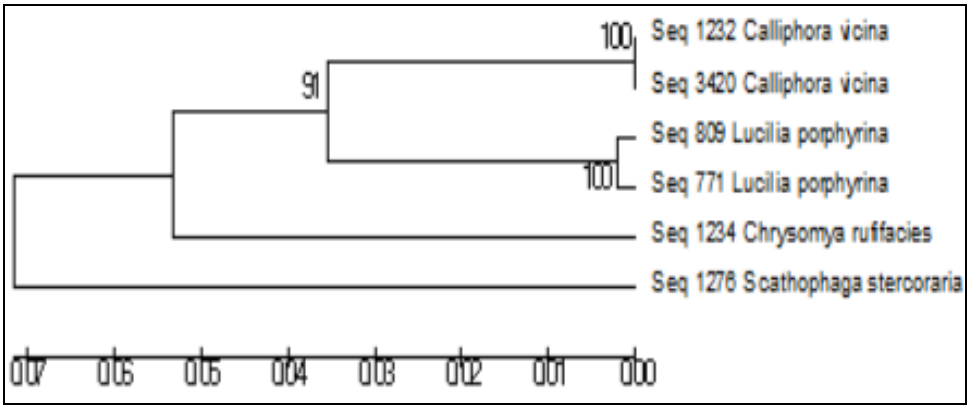

Fig 1: Phylogenetic tree (based on UPGMA method) depicting the evolutionary relationships among different taxa.

Maximum likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood (-598.2095) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 6 nucleotide sequences. There were a total of 247 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. Again, the confidence (posterior probability) based on bootstrapping was above 80, which strongly supported the analysis.

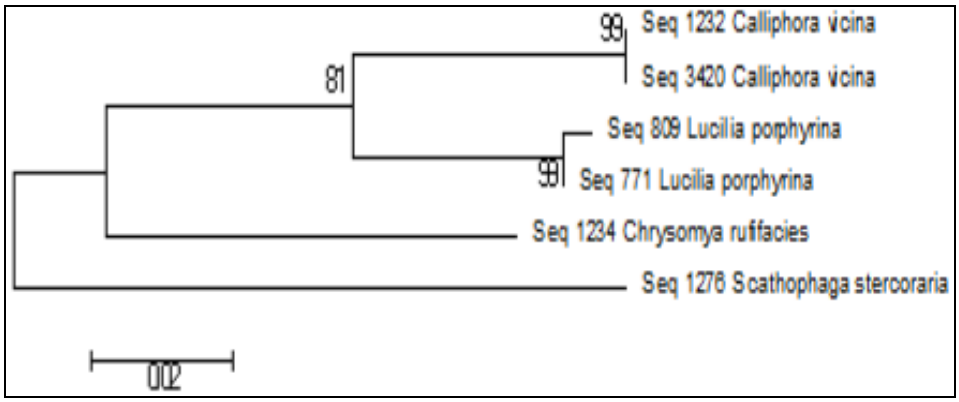

Fig 2: Phylogenetic tree (based on Maximum Likelihood method) depicting the evolutionary relationships among different taxa.

Minimum Evolution Method

The evolutionary history was inferred using the Minimum Evolution method. The optimal tree with the sum of branch length equal to 0.23678698 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighborjoining algorithm was used to generate the initial tree. The analysis involved 6 nucleotide sequences. There were a total of 297 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. Again, the posterior probability was observed to be above 85.

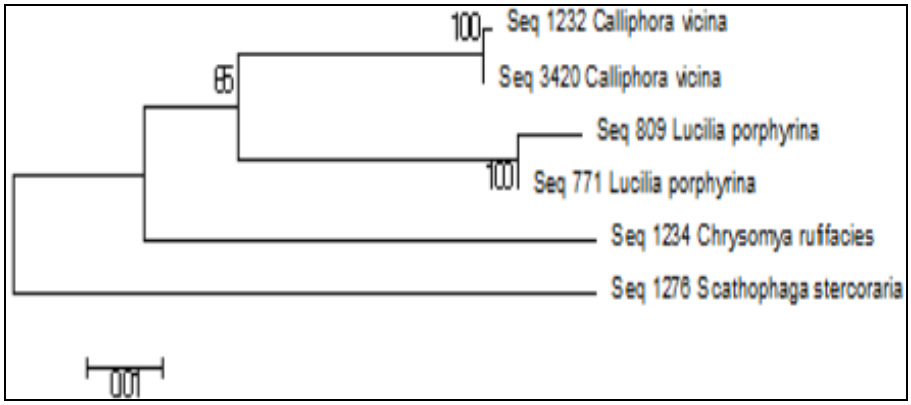

Fig 3: Phylogenetic tree (based on Minimum Evolution method) depicting the evolutionary relationships among different taxa.

Neighbour joining method

The evolutionary history was also inferred using the Neighbour joining method. The optimal tree with the sum of branch length equal to 0.23678698 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura-Nei method and are in the units of the number of base substitutions per site. The analysis involved 6 nucleotide sequences. There were a total of 297 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

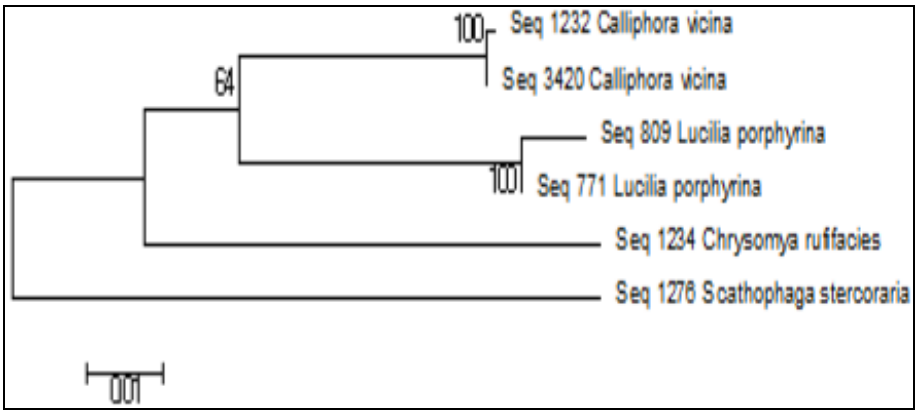

Fig 4: Phylogenetic tree (based on Neighbour joining method) depicting the evolutionary relationships among different taxa.

Discussion

The phylogenetic analysis was conducted using different substitutional models. All these methods provided similar relationship among the taxa under study. This thus, verified and established that the analysis was correct. *Scathophaga* acted as the out group and clearly got separated as a separate branch in the phylogenetic tree. Also, *Calliphor*a and *Lucilia* showed greater similarity to one another than to *Chrysomya*. This is in concurrence with the morphological classification where *Chrysomya* is considered different from Lucilia and Calliphora on the basis of presence of hair on the stem vein (which is absent in both *Calliphora* and *Lucilia*. Thus, *COII* gene has been observed as a valuable marker for phylogenetic analysis, as its analysis indicates an apomorphic trait. The number of base substitutions per site from between sequences is shown in Table 1. Analyses were conducted using the Tamura-Nei model. The analysis involved 6 nucleotide sequences. There were a total of 297 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. The overall proportion of conserved sites was 245/297 and that of variable sites was only 52/297. This clearly indicates that COII gene is highly conserved. Out of the 52 variable sites, 26 sites were found to be parsimoniously informative. In Table 29, each entry shows the probability of substitution (r) from one base (row) to another base (column). For simplicity, the sum of r values was made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in italics. The nucleotide frequencies are 33.43% (A), 35.90% (T/U), 15.88% (C), and 14.78% (G). The transition/transversion rate ratios are k1 = 22.701 (purines) and k2 = 16.395 (pyrimidines). The overall transition/transversion bias was R = 8.304, where R = [A*G*k1 + T*C*k2]/ [(A+G)*(T+C)]. The analysis involved 6 nucleotide sequences. There were a total of 297 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

Table 1: Estimates of Evolutionary Divergence between Sequences

Species					
Seq_1232_Calliphora_vicina					
Seq_1234_Chrysomya_rufifacies	0.10				
Seq_1276_Scathophaga_stercoraria	0.14	0.15			
Seq_809_Lucilia_porphyrina	0.07	0.12	0.15		
Seq_3420_Calliphora_vicina	0.00	0.10	0.14	0.07	
Seq_771_Lucilia_porphyrina	0.07	0.11	0.13	0.00	0.07

Table 2: Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution

	A	T	C	G
A	-	*1.67*	*0.69*	**16.77**
T	*1.55*	-	**11.27**	*0.74*
C	*1.55*	**27.37**	-	*0.74*
G	**35.29**	*1.67*	*0.69*	-

Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution: Each entry shows the probability of substitution (r) from one base (row) to another base (column). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in italics. The nucleotide frequencies are 33.43% (A), 35.90% (T/U), 15.88% (C), and 14.78% (G). The transition/transversion rate ratios are k1 = 22.701 (purines) and k2 = 16.395 (pyrimidines). The overall transition/transversion bias is R = 8.304, where R = [A*G*k1 + T*C*k2]/[(A+G)*(T+C)]. The analysis involved 6 nucleotide sequences. There were a total of 297 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

Table 3: Maximum Likelihood Estimate of Substitution Matrix

	A	T/U	C	G
A	-	*7.40*	*3.01*	**4.33**
T/U	*6.96*	-	**13.05**	*3.21*
C	*6.96*	**32.06**	-	*3.21*
G	**9.39**	*7.40*	*3.01*	-

Maximum Likelihood Estimate of Substitution Matrix

Each entry represents the probability of substitution (r) from one base (row) to another base (column). Substitution pattern and rates were estimated under the Tamura-Nei (1993) [2] model. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in italics. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 100. The nucleotide frequencies are A = 33.81%, T/U = 35.96%, C = 14.64%, and G = 15.59%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -598.209. The analysis involved 6 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 247 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

Maximum Likelihood Estimate of Transition/Transversion Bias

The estimated Transition/Transversion bias (R) is 1.16. Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model. The nucleotide frequencies are A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -632.535. The analysis involved 6 nucleotide sequences. There were a total of 247 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. Average frequency of each nucleotide calculated for the targeted COII gene fragment.

Disclosure

The authors are not conversant of any memberships, financial holdings, affiliations that could raise a conflict of interest.

References

1. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences J Mol. Evol. 1980; 16:111-120.
2. Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. Mol. Biol. Evol. 1993; 10(3):5-12-526.