INTERNATIONAL JOURNAL OF PLANT, ANIMAL AND ENVIRONMENTAL SCIENCES

IJPAES

Volume-3, Issue-2, April-June-2013

ISSN 2231-4490

Coden : IJPAES www.ijpaes.com

Received: 15th Mar-2013 *Revised: 25th Mar-2013* *Accepted: 26th Mar-2013*

Research article

MOLECULAR CHARACTERIZATION OF TWO SPECIES OF *SPILOSTETHUS* (INSECTA: LYAGEIDAE) AND PHYLOGENETIC RECONSTRUCTION USING 16S RIBOSOMAL RNA GENE FRAGMENT

K.P.Sanjayan* and V. Rama

Department of Advanced Zoology and Biotechnology, Guru Nanak College, Chennai 600 042, India

* Corresponding Author: kpsanjayan@yahoo.co.in

ABSTRACT: Molecular characterisation of two species of *Spilostethus* viz. *S.hospes* (Fab.) and *S.pandurus* (Scop.) based on the 16S rRNA sequence was carried out to differentiate the taxa which are otherwise morphologically identical at the nymphal stages. The 16S rRNA gene was amplified by polymerase chain reaction. The primers used for amplification were forward LR-J-13017: 5'-TTA CGC TGT TAT CCT AA-3' and reverse LR-N-13398: 5'-CAC CTG TTT AAC AAA AAC AT-3'. The sequencing was carried out in a ABI PRISM 3730XL Analyzer. The 16S rRNA sequence of *S. hospes* and *S.pandurus* differed in the total number of base pairs as well as in the percentage G+C and A+T content. However, a sequence similarity of 97.86 % was recorded. The nucleotide paired frequencies among the two species showed 274 identical pairs. The ratio of the transitional to transversional pairs was 1.5. The sequence similarity search among other lygaeid bugs was done using BlastN and nucleotide sequences were aligned using the ClustalW option of MEGA5 software. Analysis of the maximum likelihood of 24 different nucleotide substitution patterns, indicated that the GTR+G had the least BIC and AIC scores and therefore considered the best model for the present data set. The phylogenetic tree constructed indicated the two species of *Spilostethus* to share a most recent common ancestor and are clustered at 70%. This clade is a sister group to *Lygaeus equestris* and *Oncopeltus fasciatus*, both being members of subfamily Lygaeinae. Species of the subfamily Rhyparochrominae, *Dieuches* sp and *Laryngodus* sp, have been positioned distantly.

Key words: mt DNA, 16S rRNA, phylogenetic tree, Lygaeidae, *Spilostethus hospes*, *Spilostethus pandurus*, bioinformatics

INTRODUCTION

Insect systematics has in recent years undergone a sea of change with the advent of the use of molecular techniques involving genome organization studies. These studies make use of several molecular features such as DNA content, chromosomal and mitochondrial DNA organization, DNA sequences of ITS (internal transcribed spacer) and IGS (inter genic sequences). The mitochondrial DNA of insects have been used for molecular evolution studies [1-10] because of its high rate of evolution. Animal mitochondrial genomes are small, circular DNA with length ranging from 14,000 bp to 17,000 bp [11,12] and usually encode13 protein subunits of the enzymes of oxidative phosphorylation, the two rRNAs of the mitochondrial ribosome, and the 22 tRNAs necessary for the translation of the proteins encoded by mtDNA. It has become apparent that other mitochondrial gene regions can also provide insights concerning deeper divergences, as shown by a study that employed 16S rDNA sequences to examine the affinities of major gastropod lineages [13]. Therefore, the nucleotide sequence of the small-subunit (12S) and large subunit (16S) ribosomal RNA (srRNA) are proving extremely valuable for probing phylogenetic relationships among distantly related taxa because many regions remain conserved or semiconserved over large periods of time. Woese [14] has suggested that the ribosomal RNAs may be the ultimate molecular clocks. There is a large body of evidence that suggests that the rate of molecular evolution in general is relatively constant with time [15,16]. Because the srRNA molecule is common to virtually all life forms, it becomes possible to test the clock for a single molecule across many phyla.

The 16S rRNA gene, which encodes the mitochondrial large ribosomal subunit (mt LSU) in animals, has been employed extensively to explore phylogenetic relationships in arthropods at most phylogenetic levels [17], familial level [18] and the genus level and below[19,20]. The wide range in utility of 16S at various taxonomic levels suggests that the differential rates of molecular evolution within 16S, due to varying functional constraints, greatly affect its phylogenetic utility. Here we present the molecular characterization of two species of *Spilostethus*, namely *Siplostethus hospes* and *Spilostethus pandurus* based on sequences of 16S ribosomal RNA by analysing their gene content, base composition, and the nucleotide pair frequencies. Data mining for related information on other species of lygaeids were conducted and a phylogenetic tree was reconstructed after visualizing the best DNA model.

MATERIALS AND METHODS

Genomic DNA isolation

Adults of *S.hospes* and *S.pandurus* were collected from the common host plant *Calotropis gigantea* (L). About 1 gm of body tissue was ground in liquid nitrogen and incubated overnight at 50°C in 10 ml digestion buffer (0.01M NaCl, 0.1 M Tris-HCl, 0.25M EDTA and 0.5% SDS with 100µg/ml proteinase K). The digested samples were extracted twice with an equal volume of Tris-HCl, saturated phenol (pH 8.0) and then centrifuged at 5000g for 15 min to remove protein contaminants and debris. The supernatant was transferred into a fresh tube and treated with RNAase at 37°C for 30 min, followed by chloroform extraction and centrifugation at 5000g for 15 min. Finally the aqueous phase from each tube was transferred separately to clean centrifuge tubes and mixed with 0.1 volume of 3M sodium acetate, pH 5.2. Genomic DNA was precipitated with two volumes of cold ethanol, spooled, washed twice with 70% ethanol, dried and suspended in 10 mM Tris-CL (pH 8.0).

Amplification and sequencing of 16S rRNA

The 415-bp region of 16S rRNA was amplified by polymerase chain reaction. The primers used for amplification were forward forward LR-J-13017: 5'-TTA CGC TGT TAT CCT AA-3' and reverse LR-N-13398: 5'-CAC CTG TTT AAC AAA AAC AT-3' . PCR was carried out in 50 µl using 1 U of *Taq* polymerase (Roche) in a PTC-225 Peltier Thermal Cycler. The PCR conditions were 94°C for 5 min (denaturation), 40°C for 1 min (annealing) and 72°C for 1 min for extension and further 72°C for 10 min for final extension; hold at 4°C and total number of cycles is 40. This amplified product was sequenced using the ABI PRISM 3730XL Analyzer. The 16S rRNA sequence of *S.hospes* and *S.pandurus* were submitted to the Genbank database.

16S rRNA sequence analysis and phylogenetic reconstruction

A Blastn similarity search was conducted using the 16S rRNA sequence of *S. hospes* as the query sequence. The nucleotide collection database was searched with the organism key as Lygaeidae (taxid:7533). The programme was optimized for highly similar sequences (megablast). The search resulted in 22 Blast hits on the query sequence. Of these, top sequences producing significant alignments were selected. These nucleotide sequences were aligned using the ClustalW option present in the MEGA5 software (Tamura *et al*, 2011). After computing the alignment, the data menu was opened and the active data was explored for analysis of various sites such as conserved sites, parsimonious informative sites, variable sites etc., using the highlight section of the sequence data explorer window of the MEGA tool. The statistics of the nucleotide composition was analysed and automatically exported to Microsoft Excel 2007. Further, the aligned sequences were used to find the Best DNA model and to compute the pair-wise distance in order to estimate the evolutionary divergence between the sequences. To construct Phylogenetic trees, the Maximum Likelihood method and the Neighbor-Joining method were employed and the test of phylogeny had 500 bootstrap replication. The General Time Reversible (GRT) model with 5 discrete Gamma categories was opted under the rates and patterns option of the MEGA software.

RESULTS AND DISCUSSION

The 16S rRNA gene sequenced for *S.hospes* and *S.pandurus* was submitted to the Genebank database and listed under accession numbers KC495445 and KC495446 respectively (Table 1). The base statistics for the gene is presented in table 2. There were 291 nucleotides in *S.hospes* and 348 in *S. pandurus*. Both the species had more number of T nucleotide than the other bases. Also, the percentage of A+T was more than G+C. The ratio of AT:CG for *S. hospes* is 3.76 : 1 and for *S.pandurus* is 3.46 : 1.

Comparative studies by Habeeb *et al* (2011) also indicated that most insects had higher percentage of T nucleotides than C, A and G. Shouche and Patole [21] used 16S r RNA sequence for molecular taxonomy of mosquitoes and reported the sequence to have 75-78% A+T content.

Table 1: 16S rRNA gene Sequence entry for *S.hospes* and *S.pandurus* in Genbank

```
LOCUS       KC495445     291 bp    DNA     linear   INV 02-MAR-2013
DEFINITION  Spilostethus hospes voucher vr01 16S ribosomal RNA gene,
partial sequence; mitochondrial.
ACCESSION   KC495445
ORIGIN
1 attttaaata gctgcagtat tttgactgta caaaggtagc ataatcattt gtcttttaat
61 tggaggcttg tatgaatgat tggacgagaa ataagctttc tttatttaat aattttgaat
121 ttaatttttt agttaaaaag ctaaaattta tttgtaagac gagaagaccc tatagaattt
181 tattttattt attttattta ttttattttt aaatgttttt aatttaatta aaatttggtt
241 ggggtgactg tgaaaattat ttaactttca ttttttttg ttcataaatt a
```

```
LOCUS       KC495446    348 bp    DNA     linear   INV 02-MAR-2013
DEFINITION  Spilostethus pandurus voucher vr02 16S ribosomal RNA gene,
partial sequence; mitochondrial.
ACCESSION   KC495446
ORIGIN
1 ctgcagtatt ttgactgtac aaaggtagca taatcatttg tcttttaatt gaaggcttgt
61 atgaatgatt ggacgagaaa taatctttct ttatttaatt attttgaatt taatttttta
121 gttaaaaagc taaaatttat ttgtaagacg agaagaccct atagaatttt attttattta
181 ttttatttat tttatttta aatgttttta atttgattaa aatttggttg gggtgactgt
241 gaaaattatt taactttcat ttttttttt catgaattaa tgttttaatg atccggtttt
301 atagattaaa agattaaatt accttaggat aacagcggta agaaaaag
```

Table 2. Base statistics of 16S rRNA gene of *S.hospes* and *S.pandurus*

Species	T(U)	C	A	G	G+C %	A+T %
S.hospes	137	20	93	41	20.96	79.04
S.pandurus	156	26	114	52	22.41	77.59

Sequence similarity

A Blastn similarity search was conducted using the 16S rRNA sequence of *S. hospes* as the query sequence which resulted in 22 Blast hits on the query sequence. Of these, top sequences producing significant alignments were selected (Table 3). The e value is the expect value which refers to the number of BLAST hits that is expected to be seen by chance with the observed score or higher. The values recorded here are all very small indicating that it is not random. This interpretation means that if it is not by chance then the relationship between the two sequences is because of some biologically meaningful relationship. Homology can also be inferred from good alignments that have been observed here. It is interesting to note that the first four hits in the sequence similarity are Lygaeidae bugs belonging to the subfamily Lygaeinae followed closely by yet another Lygaeinae member, *Graptostethus* sp with a 86.64% identity. Members of the Geocorinae (*Germalus* sp), Blissinae (*Ischnodemus* sp), Ischnorhynchinae (*Kleidocerys* sp) and Orsillinae (*Nysius* sp) followed next, with the two species of Rhyparochrominae, *Laryngodus* and *Dieuches* lagging behind.

Table 3: Top Sequences producing significant alignments with 16S r RNA gene of *S. hospes* (GenBank: KC495445.1)

Accession No.	Species	% identity	Alignment length	Mismatch	Gap open	e value	Bit score
KC495446.1	Spilostethus pandurus	97.86	280	5	1	4.00E-141	483
JQ234972.1	Lygaeus equestris	95.56	291	9	1	1.00E-140	481
AY252660.1	Oncopeltus fasciatus	93.81	291	14	4	1.00E-126	435
AY252891.1	Neacoryphus sp	88.01	292	29	6	3.00E-98	340
AY252900.1	Germalus sp	86.39	294	30	10	7.00E-90	313
AY986813.1	Graptostethus sp	86.64	277	35	2	1.00E-87	305
AY252687.1	Ischnodemus sp	85.91	291	27	13	5.00E-86	300
AY252890.1	Kleidocerys sp	85.12	289	37	6	3.00E-83	291
AY252671.1	Nysius sp	84.85	297	32	10	5.00E-82	287
AY252680.1	Laryngodus sp	85.17	290	30	11	2.00E-81	285
AY252891.1	Dieuches sp	84.64	280	31	11	2.00E-76	268

Nucletotide substitution among the sequences

In the study of molecular evolution, it is important to know the number of nucleotide substitutions per site (d) between DNA sequences. Two important factors that are considered in the estimation of d are the inequality of the rates of transitional and transversional nucleotide substitution (transition-transversion bias) and the deviation of the G+C content from 0.5 (G+C-content bias). Transitions refer to the substitution of a purine (A or G) by another purine or the substitution of a pyrimidine (T or C) by another pyrimidine; transversions are the substitutions of a purine by a pyrimidine or a pyrimidine by a purine. With this understanding the nucleotide substitution among the sequences of the two species of *Spilostethus* was analysed.

The nucleotide pair frequencies (table 4) computed for among the two species showed 274 identical pairs among a total of 279 pairs of which there were 132 TT pairs. The ratio of Transitional Pairs versus Transversional pairs was 1.50. Habeeb *et al* [22] observed an average ratio to be 0.467 for the 16S rRNA of insects they had studied. When two DNA sequences are derived from a common ancestral sequence, the descendant sequences gradually diverge by nucleotide substitution. A simple measure of the extent of sequence divergence is the proportion of nucleotide sites at which the two sequences are different. This is estimated as the p-distance for nucleotide sequences. It is useful to know the frequencies of different nucleotide pairs between the two sequences. Since there are four nucleotides, there are 16 different types of nucleotide pairs. There are four pairs of identitical nucleotides (AA,TT, CC,GG represented as O), four transition-type pairs (AG,GA,TC,CT represented as P) and remaining 8 transversion-type pairs (represented as Q). The p distance for nucleotide sequence, given by the relationship p=P + Q. was calculated to be 5 (ie. 3+2). If nucleotide substitution occurs at random, Q is expected to be about two times higher than P when p is small which was not the case in the present investigation. In general, transition usually occur more frequently than transversions. Therefore P may be greater than Q. When the extent of divergence is low, the ratio (R) of transitions to transversions can be estimated from the observed values of P and Q. R is usually 0.2-2 in many nuclear genes, but in mitochondrial DNA it can be as high a 15 (Vigilant *et al* 1991). In the present study the value of R was 1.5. The analysis of the p value indicate that no synonymous substitution occurs in the first three codons (p for Ist codon could not be calculated, 2nd codon is 0.02 and 3rd codon is 0.03).

Molecular Evolutionary Genetic Analysis

The nucleotide sequences of the 16srRNA were aligned using the ClustalW option present in the MEGA5 software [23]. All phylogenetic methods make assumptions about the process of DNA substitution [24]. An assumption common to phylogenetic methods is a bifurcating tree to describe the phylogeny of species [25]. All methods of phylogenetic inference depend on their underlying models.

To have confidence in inferences it is necessary to have confidence in the models [26]. Because of this all methods based on explicit models of evolution should explore which is the model that fits the data best. Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. Table 5 provides details of the Maximum Likelihood of 24 different nucleotide substitution models. The GTR+G had the least BIC scores and therefore considered the best model for the present data set. GTR (General Time Reversible) + G (Gamma distribution) model is characterized by unequal base frequencies and transition rate is not equal to transversion rate. Another way of selecting the most appropriate model for a data set is to use the Akaike information criterion (AIC) [27], which can be thought of as the amount of information lost when a particular model is used to approximate reality.

Table 4 Nucleotide pair frequencies between *S.hospes* and *S.pandurus* sequences : Undirectional (10 pairs)

	ii	si	sv	R	TT	TC	TA	TG	CC	CA	CG	AA	AG	GG	Total
Avg	274.0	3.0	2.0	1.5	132.0	0.0	1.0	1.0	20.0	0.0	0.0	85.0	3.0	37.0	279.0
1st	92.0	0.0	0.0	nc	36.0	0.0	0.0	0.0	9.0	0.0	0.0	37.0	0.0	10.0	92.0
2nd	92.0	1.0	1.0	1.0	50.0	0.0	0.0	1.0	7.0	0.0	0.0	22.0	1.0	13.0	94.0
3rd	90.0	2.0	1.0	2.0	46.0	0.0	1.0	0.0	4.0	0.0	0.0	26.0	2.0	14.0	93.0

ii = Identical Pairs; si = Transitionsal Pairs; sv = Transversional Pairs; R = si/sv
1st, 2nd, 3rd Codon position
TC AG -Transition; TA TG CA CG - Transversion, TT,CC,AA,GG- Identical pairs

Table 5. Maximum Likelihood of 24 different nucleotide substitution models

Model	BIC	AICc	InL	Invariant	Gamma	R	*f*A	*f* T	*f*C	*f* G
GTR+G	2809.5	2628.2	-1283.8	n/a	0.3546	0.529	0.310	0.457	0.079	0.152
GTR+G+I	2816.9	2629.6	-1283.5	0.3266	0.8183	0.534	0.310	0.457	0.079	0.152
HKY+G	2818.3	2661	-1304.3	n/a	0.3323	0.616	0.310	0.457	0.079	0.152
TN93+G	2820.8	2657.5	-1301.5	n/a	0.3286	0.650	0.310	0.457	0.079	0.152
TN92+G	2820.9	2675.7	-1313.7	n/a	0.3479	0,565	0.384	0.384	0.116	0.116
HKY+G+I	2825.7	2662.4	-1304	0.3502	0.8273	0.626	0.310	0.457	0.079	0.152
TN93+G+I	2828.8	2659.5	-1301.5	0	0.3286	0.650	0.310	0.457	0.079	0.152
TN92+G+I	2828.9	2677.7	-1313.7	0	0.3479	0.565	0.384	0.384	0.116	0.116
HKY+I	2829.7	2672.5	-1310	0.5347	n/a	0.572	0.310	0.457	0.079	0.152
TN93+I	2831.5	2668.2	-1306.9	0.5321	n/a	0.597	0.310	0.457	0.079	0.152
GTR+I	2837.2	2655.9	-1297.6	0.5280	n/a	0.332	0.310	0.457	0.079	0.152
GTR	2938.1	2762.7	-1352.1	n/a	n/a	0.294	0.310	0.457	0.079	0.152
T92	2947.6	2808.4	-1381	n/a	n/a	0.441	0.384	0.384	0.116	0.116
HKY	2951.1	2799.9	-1374.7	n/a	n/a	0.445	0.310	0.457	0.079	0.152
TN93	2955.2	2798	-1372.8	n/a	n/a	0.448	0.310	0.457	0.079	0.152
TN92+I	2955.6	2810.4	-1381	0.00001	n/a	0.441	0.384	0.384	0.116	0.116
JC+G	3023.9	2890.8	-1423.2	n/a	0.4316	0.5	0.25	0.25	0.25	0.25
JC+I	3031	2897.9	-1426.8	0.5173	n/a	0.5	0.25	0.25	0.25	0.25
JC+G+I	3032	2892.8	-1423.2	0	0.4316	0.5	0.25	0.25	0.25	0.25
KR+G	3043.8	2904.7	-1429.2	n/a	0.4270	0.751	0.25	0.25	0.25	0.25
KR+G+I	3051.9	2906.7	-1474.4	0	0.4270	0.751	0.25	0.25	0.25	0.25
JC	3118.2	2991.2	-1478.4	n/a	n/a	0.5	0.25	0.25	0.25	0.25
K2	3134.2	3001.1	-1478.4	n/a	n/a	0.674	0.25	0.25	0.25	0.25
K2+I	3142.3	3003.2	-1478.4	0.00001	n/a	0.674	0.25	0.25	0.25	0.25

Abbreviations:
BIC scores = Bayesian Information Criterion.
AICc value = Akaike Information Criterion, corrected,
(*lnL*) = Maximum Likelihood value
(*R*) = estimated values of transition/transversion bias
(*f*) = nucleotide frequencies
GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

The AIC implements best-fit model selection by calculating the likelihood of proposed models, and imposing a penalty based on the number of model parameters. Parameter-rich models incur a larger penalty than more simple models so that fitting an excessively complex model is not likely. The best fitting model is the one with the smallest AIC value. The GTR+G had the smallest AIC value (Table 5).

Maximum Likelihood Estimate of Gamma Parameter for Site Rates

It is well known that different amino acid residues of a protein may have different functional constraints such that the substitution rate varies among the sites. The gamma distribution has been widely used for modeling the rate variation among sites. According to the gamma distribution, the substitution rate often varies from site to site within a sequence. The shape of this distribution is determined by gamma parameter, also known as the shape parameter. In the present analysis, the estimated value of the shape parameter for the discrete Gamma Distribution is 0.3546. Substitution pattern and rates were estimated under the General Time Reversible model (+G). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+*G*]). Mean evolutionary rates in these categories were 0.01, 0.08, 0.33, 0.98, 3.61 substitutions per site. The nucleotide frequencies are A = 31.07%, T/U = 45.72%, C = 7.96%, and G = 15.25%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -1283.780. The analysis involved 12 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 265 positions in the final dataset. Evolutionary analyses were conducted in MEGA.

Distance matrix

The number of base substitutions per site from between sequences are shown in table 6. Analyses were conducted using the Maximum Composite Likelihood model. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 12 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 265 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. The evolutionary divergence between *S.hospes* and *S.pandurus* was at a distance of only 0.11. The maximum distance of 0.327 was recorded between *Laryngodus* sp and *Nysius* sp.

Table 6 Estimates of evolutionary Divergence between species.

No	Species	1	2	3	4	5	6	7	8	9	10	11
1	Spilostethus hospes	-										
2	Spilostethus pandurus	0.011	-									
3	Lygaeus equestris	0.032	0.036	-								
4	Graptostethus sp	0.153	0.153	0.147	-							
5	Neacoryphus sp	0.149	0.154	0.155	0.154	-						
6	Ischnodemus sp	0.205	0.212	0.199	0.234	0.225	-					
7	Nysius sp	0.160	0.177	0.183	0.238	0.195	0.268	-				
8	Germalus sp	0.177	0.173	0.197	0.202	0.173	0.180	0.216	-			
9	Dieuches sp	0.229	0.229	0.245	0.293	0.250	0.305	0.274	0.287	-		
10	Kleidocerys sp	0.195	0.190	0.196	0.279	0.228	0.245	0.179	0.233	0.209	-	
11	Oncopeltus fasciatus	0.062	0.062	0.76	0.171	0.153	0.219	0.189	0.184	0.288	0.237	-
12	Laryngodus sp	0.241	0.236	0.256	0.262	0.293	0.248	0.327	0.279	0.182	0.197	0.273

Evolutionary tree involving 16S rRNA sequence

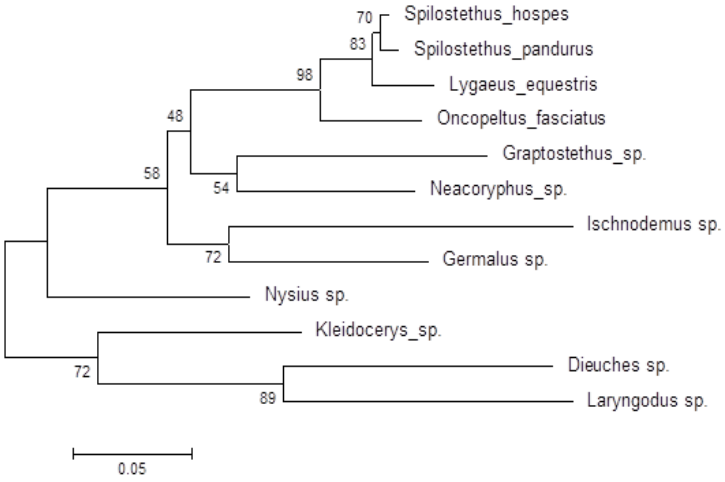
Fig 1 provides details of the molecular phylogenetic analysis based on two phylogenetic estimators, namely a) Maximum Likelihood (ML) and b) Neighbor-Joining method. The Neighbor Joining algorithm is different from ML because it uses the model to calculate pairwise genetic distances between sequences, and reconstructs a topology based on those distances.

Maximum likelihood method use the sequence data directly to reconstruct a tree, thereby utilizing information in specific nucleotide differences instead of summarizing changes with a genetic distance. Due to these differences, ML offers noteworthy statistical properties in comparison with genetic distance-based methods, but is much more computationally intensive [28-30]. In both these phylogenetic estimators, the evolutionary history was inferred based on the General Time Reversible model. In both these methods the tree branching pattern was identical. In Fig 1A, the tree with the highest log likelihood (-1283.8212) 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. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.3721)).

In Fig 1B, the evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.95412430 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches.

The results of the phylogenetic relationships among these taxa indicate a topology consistent with the generally accepted relationship. The two species of *Spilostethus* share a most recent common ancestor and are clustered at 70% (ML method) and 87% (Neighbor-Joining method). This clade is a sister group to *Lygaeus equestris* and *Oncopeltus fasciatus*, both being members of Lygaeinae. Species of the subfamily Rhyparochrominae, *Dieuches* sp and *Laryngodus* sp have been positioned distantly.

A



B

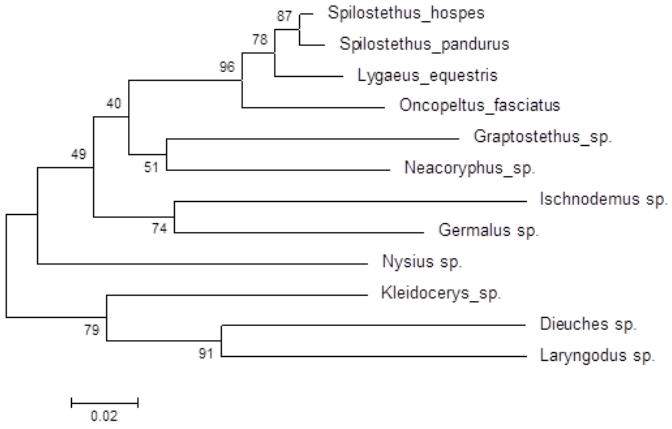


Figure 1. Molecular Phylogenetic anaylsis A) by Maximum Likelihood method based on GTR model B) Neighbor-Joining method with 500 bootstrap test

CONCLUSION

The two species of *Spilostethus* could be distinguished based on their 16S rRNA sequence pattern. The 16S rRNA sequence of *S. hospes* and *S.pandurus* differed in the total number of base pairs as well as in the percentage G+C and A+T content. However, a sequence similarity of 97.86 % was recorded. The nucleotide paired frequencies among the two species showed 274 identical pairs. The ratio of the transitional to transversional pairs was 1.5. Analysis of the maximum likelihood of 24 different nucleotide substitution pattern to reconstruct the phylogenetic tree, indicated that the GTR+G had the least BIC and AIC scores and therefore can be considered the best model for the present data set. The phylogenetic tree constructed indicate the two species of *Spilostethus* to share a most recent common ancestor and are clustered at 70-87%. This clade is a sister group to *Lygaeus equestris* and *Oncopeltus fasciatus*, both being members of the subfamily Lygaeinae. Species of the subfamily Rhyparochrominae, *Dieuches* sp and *Laryngodus* sp have been positioned distantly. The tree developed here is in consistence with the mode of evolution of feeding in lygaeid insects, the Rhyparochrominae being more primitive are litter dwellers feeding on seeds in the habitat, while the more evolved Lygaeinae are arboreal insects.

REFERENCES

[1] Kambhampati, S. 1995. A phylogeny of cockroaches and related insects based on DNA sequence of mitochindrial ribosomal RNA genes. Proc. Natl. Acad. Sci. USA 92: 2017–2020

[2] Tang J, Pruess K, Cupp E. W. and Unnasch T. R.1996. Molecular phylogeny and typing of blackflies (Diptera) that serve as vectors of human or bovine onchocerciasis; Med. Vet. Entomol. 10: 228–234

[3] Tang J and Unnasch T R 1997 Heteroduplex analysis in medical entomology: A rapid and sensitive sequence based tool for population and phylogenetic studies; Parasitology Today. 13: 271–274

[4] Misof B, Anderson C. L. and Hadrys H. 2000. A phylogeny of the damselfly genus calopteryx (Odonata) using mitochondrial 16S rDNA markers; Mol. Phylogenet. Evol. 15: 5–14

[5] Hall, M. J., Edge W, Testa J. M., Adams Z. J. and Ready P.D. 2001. Old world screwworm fly, Chrysomya bezziana, as two geographical races. Med. Vet. Enomol., 15, 393-402.

[6] Zehner R, Amendt J, Schütt S, Sauer J, Krettek R and Povolný D. 2004. Genetic identification of forensically important flesh flies (Diptera: sarcophagidae). Int. J. Legal Med., 118, 245-247.

[7] Cummings, M A, and Krafsur, E.S. 2005. Spatial diversity in Mitochondrial Cytochrome c oxidase in houseflies. Med.Vet. Entomol., 19, 53-59.

[8] Segura M. D., Callejas C, Fernández M.P. and Ochando M.D. 2006. New contributions towards the understanding of Phylogenetic relationships among economically important fruit flies (Diptera: Tephritidae). Bull. Entomol. Res., 96: 279-288.

[9] Angilla Aline F, Luis HS Gil, Luis HP da Silva, Paulo EM Ribolla. 2007. Population structure of the malaria vector Anopheles darlingi in Rendonia, Brazilia Amazon, based on mitochondrial DNA. Mem. Inst. Oswaldo Cruz, 102, 953-958.

[10] Habeeb, S.K.M. and K.P. Sanjayan. 2011. Sequencing and phylogenetic analysis of the mitochondrial cytochrome C oxidase subunit 1 of Oxycarenus laetus (Hemiptera: Lygaeidae). International J of Plant, Animal and environmental Science, 1(3): 85-92.

[11] Boore, J L. 1999 Animal Mitochondrial genomes. Nucleic Acids. Res., 27: 1767- 1780

[12] Cameron S. L., Lambkin C. L., Barker S. C. and Whiting M. F. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst. Entomol., 32, 40–59.

[13] Thollesson, M. 1999. Phylogenetic analysis of Euthyneura (Gastropoda) by means of the 16S rRNA gene: use of a fast gene for higher-level phylogenies. Proc. R. Soc. London B 266:75–83.

[14] Woese, C R. 1987. Bacterial evolution. Microbial Rev., SI: 221-271.

[15] Wilson, A.C., Carlson, S.S. and White, T. J. 1977. Biochemical evolution. Annu. Rev. Biochem. 46:473- 639.

[16] Ayala, F. J. 1986. On the virtues and pitfalls of the molecular evolutionary clock. J. Hered., 77: 226-235.

[17] Flook, P.K. and C.H.F. Rowell. 1995. The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). Molecular Phylogenetics and Evolution, 8:177–192.

[18] Black, W.C. and J. Piesman, 1994. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rRNA sequences. Proceedings of the National Academy of Sciences, USA, 91:10034–10038.
[19] DeSalle, R. 1992.The phylogenetic relationships of flies in the family Drosophiladae deduced from mtDNA sequences.Mol. Phylogenetic Evol.1 31–40
[20] Bond, J.E., M.C. Hedin, M.A. Ramirez and B.D. Opell. 2001. Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider Aptostichus simus. Molecular Ecology, 10: 899–910.
[21] Shouche Y. S. and M.S. Patole. 2000. Sequence analysis of mitochondrial 16S ribosomal RNA gene fragment from seven mosquito species. J.Biosci. 25(4) 361-366
[22] Habeeb, S.K.M., Sirish Kumar, Sanjayan, K.P., Anuradha, V., and Praveena, A. 2011.Comparative evolutionary analysis of insects using mitochondrial genes. International Journal of Current Research, 3(11) 66-73.
[23] Tamura K, Daniel Peterson, Nicholas Peterson, Glen Stecher, Masatoshi Nei, and Sudhir Kumar 2011 MEGA5. Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum, Parsimony Methods. Mol. Biol.Evol. 28(10): 2731-2739.
[24] Felsenstein J 1988. Phylogenies from molecular sequences: Inference and reliability. Annual Review of Genetics. 22:521-565.
[25] Huelsenbeck, J. P. and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Review of Ecology and systematics 28: 437-466
[26] Goldman, N.1993. Statistical tests of models of DNA substitution. Journal of Molecular Evolution, 36:182-198
[27] Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans Autom Contr, 19:716–23.
[28] Kuhner, M K and Felsenstein, J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol Biol Evol 11:459–68.
[29] Huelsenbeck, J. P. 1995 The robustness of two phylogenetic methods: four-taxon simulations reveal a slight superiority of the maximum likelihood over neighbor joining. Mol Biol Evol 12:843–9.
[30] Yang Z, 1994 Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Syst Biol. 43:329–42.