JARQ 35(2), 85 – 90 (2001) http://www.jircas.affrc.go.jp

Nucleotide Sequence Variation and Use of Mitochondrial DNA for Phylogenetic Analyses in Anthocorid Bugs (Hemiptera: Anthocoridae)

Masahiko MURAJI[1]*, Kenjiro KAWASAKI[1] and Toru SHIMIZU[2]

[1] Department of Insect Genetics and Breeding, National Institute of Sericultural and Entomological Science (Tsukuba, Ibaraki, 305–8634 Japan)
[2] Department of Environmental Biology, National Institute of Agro-Environmental Science (Tsukuba, Ibaraki, 305–8604 Japan)

Abstract

A total of approximately 1205 bp long nucleotide sequences of portions of the mitochondrial 16S ribosomal DNA (16S rDNA), cytochrome *b* (cyt *b*), and cytochrome oxidase subunit I (COI) genes were determined for 7 species of the family Anthocoridae and one outgroup taxon, *Nezara viridula*. Comparison of the nucleotide and inferred amino acid sequences among anthocorid taxa revealed the biased nature of nucleotide substitutions in the 2 protein-coding genes. In the 16S rDNA, although strong A+T composition bias and A-T transversion bias were evident among anthocorid taxa, uncorrected genetic divergence increased linearly with the increase of the taxonomic levels. Results of preliminary phylogenetic analyses revealed that the nucleotide sequences of the 2 protein-coding genes are useful for the estimation of the relationship among taxa at lower taxonomic levels compared with the 16S rDNA.

Discipline: Biotechnology
Additional key words: PCR, direct sequencing, *Orius*

Introduction

The family Anthocoridae includes a large number of natural enemies predating on various kinds of herbivorous pests such as thrips, aphids, psyllids, scales and mites as well as immature stages of some Lepidopteran and Coleopteran insects. Despite the high potential of this group as biological control agents, ecological information is available only for a small number of species originating from a rather restricted area in the world. Such a situation is partially due to the difficulty in the identification of the species because of their small size and scarcity of clear morphological differences especially among closely related species. Therefore, it is necessary to develop new criteria for anthocorid species identification and for specifying their position in taxonomic grouping.

Recently, a variety of molecular markers has been used to analyze the phylogeny and taxonomy of numerous insect groups. Many researchers have shown that mitochondrial DNA (mtDNA) was a sensitive genetic marker for species relationships and genetic differences among taxa. However, the suitability of the nucleotide sequences for the phylogeny of certain insect groups seems to be different among the genes and regions contained in mtDNA, because the mode and rate of molecular evolution are different among them[6]. These characteristics require a preliminary investigation of nucleotide substitution properties. Such studies would help researchers select appropriate mtDNA fragments suitable for phylogenetic analysis and identification of insect species.

In this context, we have analyzed the nucleotide sequences of the mitochondrial 16S ribosomal RNA (16S rRNA), cytochrome *b* (cyt *b*) and cytochrome oxidase subunit I (COI) genes using 7 anthocorid bugs and one outgroup taxon, *Nezara viridula*. We performed pairwise comparisons of nucleotide and inferred amino acid sequences between taxa to investigate the property of nucleotide substitutions that may affect the phylogenetic use of respective mtDNA fragments. We also performed a preliminary phylogenetic analysis in order to evaluate the usefulness of each mtDNA fragment for the estimation of the anthocorid species relationships.

*Corresponding author: fax +81–298–38–6090, e-mail mmuraji@nises.affrc.go.jp
Received 30 May 2000, accepted 3 October 2000.

Materials and methods

Materials used in this study are listed in Table 1. As a source of template for the PCR, total DNA was extracted from individual adults using GenomicPrep™ Cells and Tissue DNA Isolation Kit (Amersham Pharmacia Biotech). The following primers were used for the amplification: 5'-CCGGTTTGAACTCAGATCATGT and 5'-CGCCTGTTTAACAAAAACAT for the 16S rDNA; 5'-TAGGATATGTTTTACCTTGAGGACA and 5'-CTCCTCCTAATTTATTAGGAATTG for the cyt *b* gene; 5'-AGCAGGAATTTCTTCAATTTT and 5'-CTGTAAATATGTGATGTGCTC for the COI gene. These primers were designed based on the high degree of similarity in aligned homologous sequences of *Apis mellifera* (L06178), *Locusta migratoria* (X80245), *Anopheles quadrimaculatus* (L04272), and *Drosophila yakuba* (X03204). The amplification products correspond to bp 12866 to bp 13417 of the *D. yakuba* sequence for the 16S rDNA, bp 10903 to bp 11388 for the cyt *b* gene, and bp 1923 to bp 2350 for the COI gene.

Amplification, purification, labeling, and sequencing of mtDNA fragments were carried out according to previously described methods[4]. Phylogenetic analysis was carried out using the PHYLIP (Phylogeny Inference Package) version 3.5c[2]. Sequences obtained in this study have been submitted to the DDBJ / EMBL / GenBank nucleotide sequence databases.

Results and discussion

1) Nucleotide alignment

Approximately 431-, 417-, and 351-base pair long portions of the 16S rDNA, cyt *b*, and COI genes were sequenced using 7 anthocorid species and one additional Hemipteran taxon. From the alignments of these mtDNA regions, 182 (42.2%), 161 (38.6%), and 118 (33.6%) variable sites were detected. Within 139 and 117 amino acid residues of the cyt *b* and COI estimated by translation of the nucleotide sequences, variability was observed at 35 (25.2%) and 12 (10.3%) sites for the respective

Table 1. Species, strains and taxonomic grouping of insects subjected to sequence analysis in this study

Family / Subfamily / Genus / Subgenus / Species	Locality	Code
Anthocoridae		
Anthocorinae		
Orius		
Heterorius		
O. strigicollis	Nagakute, Aichi pref.	A
	Naha, Okinawa pref.	O
O. minutus	Obihiro, Hokkaido	H
	Tsukuba, Ibaraki pref.	I
O. sauteri	Obihiro, Hokkaido	H
	Nangoku, Kouchi pref.	K
O. nagaii	Tsukuba, Ibaraki pref.	I
	Nangoku, Kouchi pref.	K
Paraorius		
O. tantillus	Tamagusuku, Okinawa pref.	
Bilia		
B. japonica	Naha, Okinawa pref.	
Lyctocorinae		
Amphiareus		
A. obscuriceps	Tsukuba, Ibaraki pref.	
Pentatomidae		
Nezara		
N. viridula	Tsukuba, Ibaraki pref.	

Table 2. Average nucleotide content (%) of the mitochondrial DNA fragments of 7 species of Anthocoridae

Genes		A	T	C	G
16S rDNA		45.6	32.8	13.9	7.7
cyt *b*	1st position	32.6	27.4	18.8	21.2
	2nd position	23.7	43.2	22.0	11.1
	3rd position	48.8	36.5	11.4	3.3
	overall	35.1	35.7	17.4	11.9
COI	1st position	27.6	26.0	18.4	28.0
	2nd position	17.5	43.5	21.0	18.1
	3rd position	54.2	31.4	9.4	5.0
	overall	33.1	33.6	16.3	17.0

enzymes.

In the protein-coding genes, nucleotide variability was most pronounced at the third position of codons. In the cyt *b* alignment, 24.8, 11.2, and 64.0% of total variable sites were detected from the first, second, and third positions of codons, respectively. In the case of the COI gene, 20.3, 5.1, and 74.6% of all the variable sites were observed at the respective positions. In contrast, the frequency of replacement substitutions was much higher at the first and second positions of codons (cyt *b*: 65.0 and 88.9%, COI: 34.8 and 100%) than at the third position (cyt *b*: 3.9%, COI: 1.2%).

As in mtDNA of many other insects[3)], a relatively high A+T content was observed in these genes (Table 2). The values were 78.4, 70.8, and 66.7% for the 16S rDNA, cyt *b* and COI genes, respectively. In the case of the protein-coding genes, A+T content bias was obvious at the third position of the codons where 85.3 and 85.6% of all the nucleotides consisted of A or T in the cyt *b* and COI genes, respectively.

2) Nucleotide substitutions

In the mtDNA sequences of several insect species, it was reported that the frequency of base changes was correlated with the nucleotide content of genes and was considerably different among different types of nucleotide substitutions[1)]. In this study, a strong A-T transversion bias was observed in the 16S rDNA (Table 3) in which the nucleotide content was biased for A and T. However, the A-T transversion bias was not apparent in the cyt *b* and COI genes, although these genes also showed relatively high levels of A+T content (Table 2).

In the case of the first position in the COI gene, a very strong T-C transition bias was observed, although the nucleotide content at this position was not biased for T and C (Table 2). When triplets estimated from COI sequences were compared, it was revealed that 100% of the T-C transition at this position occurred as silent substitutions between codons (CTN and TTR) within the degenerate codon family encoding leucine. Therefore, it is apparent that nucleotide substitutions at this position were biased by the structure of the COI protein.

Compared with 16S rDNA, the third codon position of the cyt *b* and COI genes showed a relatively lower frequency of A-T transversions, even though there was a very strong A+T content bias (Table 2). When nucleotide

Table 3. Frequency (%) of nucleotide substitutions among 7 species of Anthocoridae

Genes		Transition		Transversion			
		T/C	A/G	A/T	A/C	T/G	G/C
16S rDNA		17.0	12.4	54.4	12.3	3.8	0.0
cyt *b*	1st position	41.2	19.3	22.6	10.0	6.0	1.0
	2nd position	63.6	7.8	7.8	0.0	20.8	0.0
	3rd position	39.6	14.5	31.6	10.1	3.4	0.9
	overall	41.4	15.2	28.0	9.4	5.0	0.9
COI	1st position	92.6	7.4	0.0	0.0	0.0	0.0
	2nd position	22.2	18.5	0.0	55.6	0.0	3.7
	3rd position	29.3	22.3	37.7	8.9	1.7	0.1
	overall	39.6	19.7	30.4	8.7	1.3	0.2

sites of twofold degenerate codon families which allow only A-G or T-C substitutions at the third position were excluded from pairwise comparisons of the COI sequences, the frequency of A-T transversions (58.7%) at the third position became as high as that of the 16S rDNA. Similarly, the cyt *b* gene showed a higher frequency of A-T transversions (49.8%) when silent sites contained in such codon families were excluded. Therefore, the reduced rate of A-T transversions observed at the third position of these genes (Table 3) does not imply the existence of an unbiased mode of nucleotide substitutions, but it indicates that nucleotide substitutions are also biased by the coding function even at the third position of codons.

3) Nucleotide divergence

Fig. 1 shows the percentage of substituted nucleotides estimated by pairwise comparisons between taxa at various taxonomic levels. In the case of the cyt *b* and COI genes, the values were higher than in the 16S rDNA at lower taxonomic levels (Fig. 1A). However, it appeared as if a plateau had been reached at relatively higher taxonomic levels. In the case of the 16S rDNA, the values increased linearly with the increase of the taxonomic levels, indicating that the rate of nucleotide substitutions is constant along taxonomic levels.

In the case of the cyt *b* and COI genes, saturation in the nucleotide divergence became much more obvious when only third positions were considered. The values increased very rapidly and reached a plateau at between-subgenus level in the cyt *b* gene and at within-subgenus level in the COI gene. This phenomenon was reflected in the homoplasy between distantly related taxa. When corresponding COI sequences of several insects in different orders were included in pairwise comparisons, an anthocorid taxon, *Orius nagaii,* showed a higher nucleotide similarity at the third position to *Manduca sexta* (U09843) (73.5%) and *A. mellifera* (68.4%) than to other *Orius* taxa (61.5–67.5%). These results are considered to be due to the fact that frequent multiple substitutions had occurred at this position so that phylogenetically informative nucleotide changes were obscured. Because in the cyt *b* and COI genes, the major part of the nucleotide substitutions (70.6 and 80.9%) occurred at the third position, multiple substitutions at this position are likely to affect the usefulness of nucleotide sequences.

On the other hand, the values for the first and second positions increased very slowly and no plateau was reached among the taxa used in this analysis. However, in the case of the first position in the COI gene, as men-

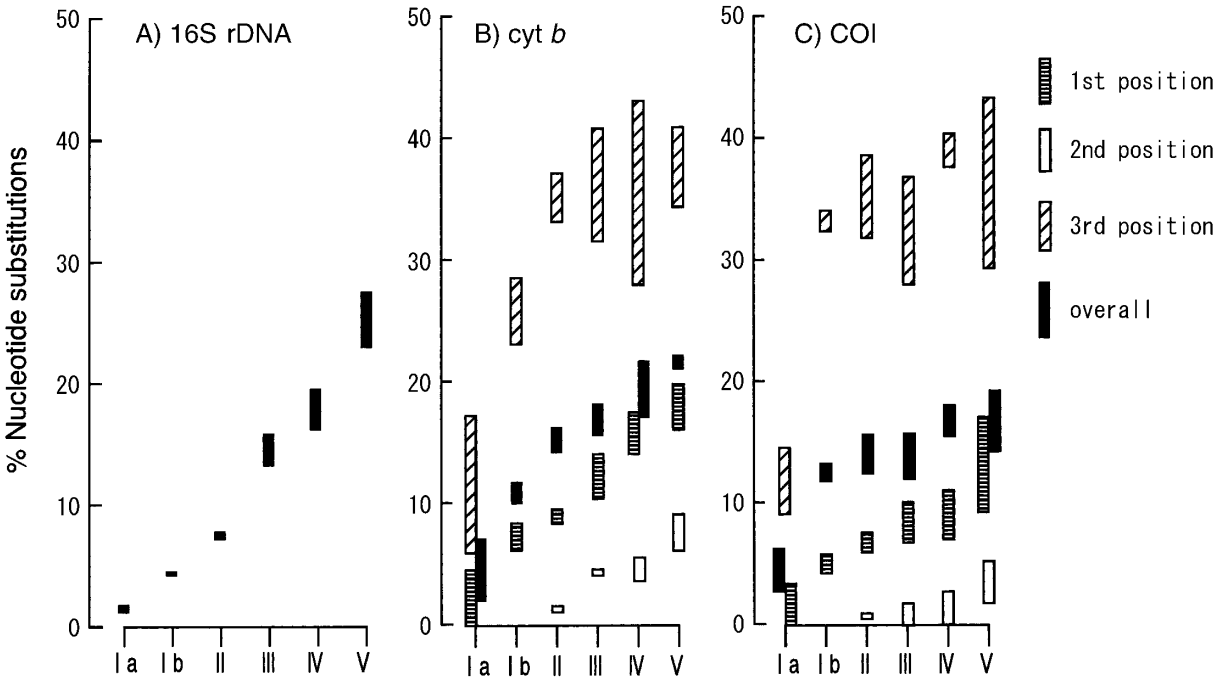

Fig. 1. Percentage of nucleotide substitutions in the mtDNA fragments

Vertical lines show the range of nucleotide divergence between taxa at a given taxonomic level. Pairwise comparisons were made between species for 6 different taxonomic levels: I, within subgenus [Ia, between pair of 3 *Heterorius* taxa (*O. strigicollis*, *O. minutus*, and *O. sauteri*); Ib, between *O. nagaii* and 3 other *Heterorius* taxa]; II, between subgenera; III, between genera; IV, between subfamilies; V, between families.

tioned previously, most of the nucleotide changes occurred as silent substitutions within the degenerate codon family encoding leucine. Although substitutions of this kind are highly accountable for the nucleotide variability in data sets including the first position (88.9%) and the first and second positions (76.2%), the number of leucine-coding sites was restricted (16.4% of total amino acids). Therefore, nucleotide variability in these data sets may be also saturated when distantly related taxa are included in the analysis.

4) Phylogenetic analysis

In order to evaluate the usefulness of the mtDNA sequences, phylogenetic methods generally employed, i.e. the maximum parsimony analysis, UPGMA analysis based on Jukes-Cantor distances, and neighbor-joining analysis based on Kimura's 2-parameter distances, were applied to the data sets obtained in this study. In the case of the 16S rDNA, because of the low genetic divergences, none of these methods could reveal the relationship among closely related species; *O. sauteri*, *O. strigicollis*, and *O. minutus* (Fig. 2A). However, in the case of the relationships among other species, phylogenetic trees agreed well with each other and with the conventional classification of the anthocorid genera based on morphological characters[7] (Table 1). Therefore, the 16S rDNA sequence may be useful for the analysis of rather distant relationships among taxa. It is interesting to note that although the 16S rDNA sequence of anthocorid bugs apparently did not support the assumption for the Jukes-Cantor equation[6], this method is still useful for the analysis of distant relationships. This may be correlated with the Nei's arguments[5] that simple methods enable to estimate the correct phylogenetic relationship in distant trees very often when the rate of nucleotide substitutions is more or less constant and the number of nucleotides examined is not very large. Such a situation was apparently the case for the 16S rDNA of the taxa used in this study (Fig. 1).

In the present study, maximum parsimony, neighbor-joining and UPGMA analyses of the cyt *b* sequences did not enable to determine the phylogenetic position of *B.*

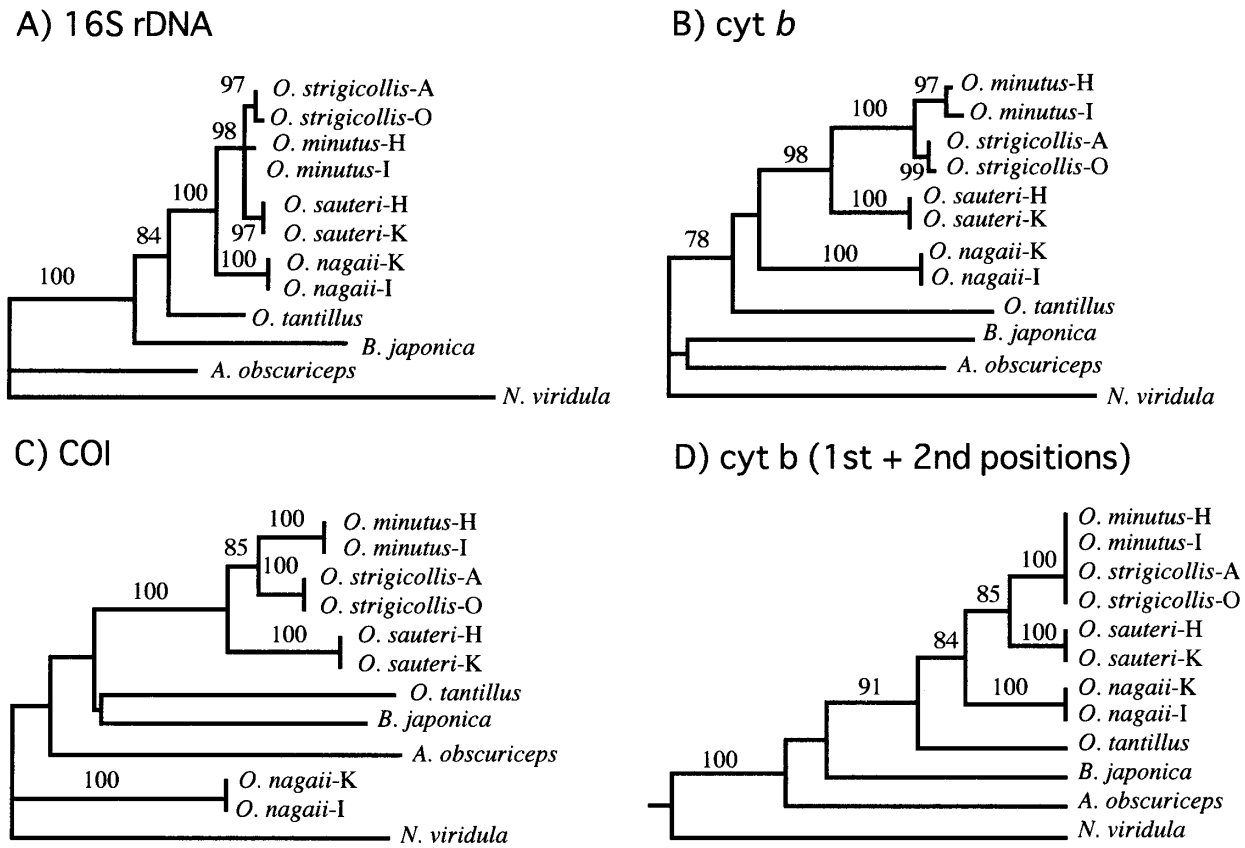

Fig. 2. Phylogenetic trees generated by neighbor-joining analysis based on Kimura's 2-parameter distances (A, B, and C) and UPGMA method based on Jukes-Cantor distances (D)

For the UPGMA analysis of the cyt *b* gene, the data set generated, excluding the third codon position, was used. Bootstrap confidence limits are shown above the branches of clades supported in more than 60% of 500 replicates.

japonica and *A. obscuriceps* precisely. However, the results from these methods agreed well with each other in terms of the positioning of taxa within the genus *Orius* (Fig. 2B). Relationship estimated by these methods also agreed with the morphological observations. Therefore, it is suggested that the use of the cyt *b* sequence is suitable for the estimation of close relationships. This assumption may be supported by the fact that base substitutions in this gene were saturated at higher taxonomic levels and the rate of base changes was rather constant among the taxa within the genus *Orius* (Fig. 1A). In the case of the COI gene, none of the phylogenetic analyses could reveal the relationship among distantly related taxa, *O. tantillus*, *B. japonica*, and *A. obscuriceps* (Fig. 2C). Phylogenetic trees generated by these methods were in agreement with morphological observations only for the relationships among the most closely related species, *O. minutus*, *O. strigicollis*, and *O. sauteri*.

Because in the case of the protein-coding genes, the A+T content bias and saturation in nucleotide divergence were most obvious at the third position of the codons, we performed a phylogenetic analysis using the data sets generated by excluding the third codon position. Because of the low genetic variability, these data sets could not discriminate between *O. minutus* and *O. strigicollis*. However, the UPGMA method generated topologies that were in agreement with that estimated from the 16S rDNA and conventional classification (Fig. 2D). Other methods did not improve the resolution of the phylogenetic estimate. Subsequently, we performed the maximum parsimony analysis using data sets generated by excluding variable polymorphic sites that contain transitional substitutions from the cyt *b* and COI alignments. However, these data sets did not produce realistic topologies more than that including all the codon-positions.

In this study, we found that the mtDNA sequences of anthocorid bugs were biased in both nucleotide composition and base substitutions, which may restrict their phylogenetic utility. For example, saturation in the nucleotide divergence is likely to restrict the range of taxonomic levels to be analyzed. Homoplasy due to multiple substitutions will lead to an underestimation of the genetic distances between taxa. For a phylogenetic estimation using biased sequences, it may be appropriate to use some sophisticated algorithms to compensate for such factors[5, 6]. However, relatively simple methods can also be used to evaluate the general suitability of the DNA sequences.

Preliminary analyses using conventional phylogenetic methods showed that taxonomic levels at which phylogenetic analysis can be applied vary with the genes. Compared with the 16S rDNA, the use of nucleotide sequences of the protein-coding genes was more suitable for phylogenetic analyses at rather lower taxonomic levels. Therefore, in order to estimate the phylogeny of a wide variety of anthocorid taxa, it may be appropriate to combine the results estimated separately for distantly and closely related taxa using different kinds of mtDNA sequences.

In addition to providing information for phylogenetic analysis, the mitochondrial DNA sequences can also be used to develop diagnostic markers to identify closely related species that are morphologically similar. Because *O. minutus* and *O. strigicollis* are closely related morphologically and they frequently coexist on the same plant, it had been difficult to evaluate each species separately as a biological control agent against agricultural pests in the field. In the present study, *O. minutus* and *O. strigicollis* differed in 3, 7, and 7 nucleotide sites, at which no intraspecific variations were observed among strains, in the 16S rDNA, cyt *b* and COI sequences, respectively. These differences lead to differences in the recognition sites of several restriction enzymes and they can be used to identify the 2 species in PCR-RFLP analysis. Because in the anthocorid species, both the adult and nymphal stages are predators, mitochondrial DNA markers are particularly useful for analyzing the ecology of these species in the case of immature stages which are not morphologically distinct.

References

1) DeSalle, R. et al. (1987): Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. Mol. Evol.,* **26**, 157–164.
2) Felsenstein, J. (1993): Phylogeny inference package (PHYLIP), version 3.5c. Department of Genetics, University of Washington, Seattle.
3) Jermin, L. S. & Crozier, R. H. (1994): The cytochrome *b* region in the mitochondrial DNA of the ant *Tetraponera rufoniger.* Sequence divergence in Hymenoptera may be associated with nucleotide content. *J. Mol. Evol.,* **38**, 282–294.
4) Muraji, M. et al. (2000): Nucleotide sequence variation and phylogenetic utility of the mitochondrial COI fragment in anthocorid bugs (Hemiptera: Anthocoridae). *Appl. Entomol. Zool.,* **35**, 301–307.
5) Nei, M. (1996) Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.,* **30**, 371–403.
6) Simon, C. et al. (1994): Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.,* **87**, 651–701.
7) Yasunaga, T. (1997): The flower bug genus *Orius* Wolff (Heteroptera: Anthocoridae) from Japan and Taiwan, Part 2. *Appl. Entomol. Zool.,* **32**, 379–386.