Assaying Mitochondrial COI Sequences and Their Molecular Studies in Hexapoda, PART I: From 2000 to 2009

Wonhoon Lee†, Jongsun Park†, Shin-ichi Akimoto², Sora Kim¹, Yang-Su Kim³, Yeirim Lee¹, Kwang-Ho Kim, Si Hyeock Lee¹, Yong-Hwan Lee¹* and Seunghwan Lee¹**

Crop Protection Division, [Dept. of Agricultural Biology] National Academy of Agricultural Science, RDA, Gyeonggi-do, 441-707, Korea
¹Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea
²Laboratory of Systematic Entomology, Department of Ecology and Systematics, Graduate School of Agriculture, Hokkaido University, Kita-ku, Sapporo, 060-8589, Japan
³Foodborne Disease Prevention & Surveillance Division, Ministry of Food and Drug Safety, Chungcheongbuk-do, 363-700, Korea

ABSTRACT: Since 2000, a large number of molecular studies have been conducted in Hexapoda with generating large amount of mitochondrial sequences. In this study, to review mitochondrial COI sequences and their molecular studies reported in Hexapoda from 2000 to 2009, 488 molecular studies conducted based on 58,323 COI sequences were categorized according to 26 orders and the positions of COI sequences (5’, 3’, and entire regions). The numbers of molecular studies in which the three regions utilized varied largely among the 26 orders; however, seven orders showed preferred positions of COI sequences in the researches: Diptera and Orthoptera revealed the largest number of studies in the 5’ region; while, Coleoptera, Phthiraptera, Odonata, Phasmatodea, and Psocoptera, showed the largest number of studies in the 3’ region. From comparing 84 molecular studies published before 2000, we observed the possibilities that molecular studies in Coleoptera, Diptera, Phthiraptera, and Phasmatodea from 2000 to 2009 had been followed classical studies using the positions of COI sequences well-known until 1999. This study provides useful information to understand the overall trends in COI sequence usages as well as molecular studies conducted from 2000 to 2009 in Hexapoda.

Key words: Cytochrome oxidase subunit I, Systematic, DNA barcode, Hexapoda


검색어: Cytochrome oxidase subunit I, 계통, DNA 바코드, 육각강

*Corresponding author: yonglee@snu.ac.kr
**Co-corresponding author: seung@snu.ac.kr
†These authors contributed equally to this work.
Received October 30 2013; Revised November 20 2013
Accepted November 22 2013
Mitochondrial *cytochrome oxidase subunit I* (COI) gene is ca. 1,500 bp long, and several fragments of this gene have been used in molecular studies such as phylogeny, phylogeography, and/or DNA barcode (Caterino et al., 2000; Hebert et al., 2003). Especially, after the concept of ‘DNA barcode’ was introduced by Hebert et al. (2003), a large number of COI sequences located in 5’ region (COI-F) was accumulated in the GenBank (http://www.ncbi.nlm.nih.gov/genbank/) because of the high success rates of PCR amplification and sequencing (Jinbo et al., 2011). Currently, the number of COI-F sequences and their species can be easily checked in the BOLD system (http://www.barcodinglife.com/) (Ratnasingham and Hebert, 2007), and these sequences can be downloaded in the BOLD system and the GenBank, using key words, ‘taxa’ and/or ‘barcode’. However, in contrast, it is still difficult to estimate the number of other COI sequences (excluding COI-F sequences) and their source species because of the lack of key words which can directly distinguish them and download from the GenBank.

More than ten years ago, Caterino et al. (2000) reviewed the trends of insect molecular studies based on ca. 560 molecular phylogenetic studies which had been reported until 1999: most of the molecular studies had been conducted in four orders, Diptera, Lepidoptera, Hymenoptera, and Coleoptera, by mainly using four genes, COI, large subunit ribosomal RNA, 18S ribosomal RNA, and elongation factor-1a. This review provided insights to understand the progress of molecular studies in several hexapod orders and to select suitable molecular markers for various molecular studies. However, after Caterino et al. (2000), any review which reflected recent molecular studies since 2000 in Hexapoda has not been reported. In case of the COI gene, even though a large number of COI sequences have been accumulated in several orders, any reviews about molecular studies using COI sequences have not been conducted, so that there are no information to understand which kinds of molecular studies and which locations and sizes of COI sequences have been preferred in the orders.

**Materials and Methods**

A total of 488 molecular studies were surveyed from the 26 orders based on 58,323 COI sequences downloaded from the GenBank (Table 1). Firstly, to compare tendencies of subjects in molecular studies among the 26 orders, the 488 studies were classified into two categories, systematic and DNA barcode. The systematic category included phylogeny, phylogeography, population genetic, and systematic studies, and the DNA barcode category included DNA barcode, reporting new species, and species identification studies. Secondly, to examine the frequency of the COI sequence positions used in molecular studies according to the 26 orders, the alignments of the 58,323

**Table 1.** List of the numbers of all sequences, species, genus, and family according to 26 hexapod orders about 58,323 COI sequences

<table>
<thead>
<tr>
<th>Order</th>
<th>COI sequence</th>
<th># of taxa</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Species</td>
<td>Genus</td>
</tr>
<tr>
<td>Diptera</td>
<td>12,790</td>
<td>2,470</td>
<td>472</td>
</tr>
<tr>
<td>Coleoptera</td>
<td>12,519</td>
<td>4,608</td>
<td>1,287</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>12,927</td>
<td>3,638</td>
<td>1,146</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>9,265</td>
<td>3,054</td>
<td>827</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>3,618</td>
<td>1,008</td>
<td>421</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>1,454</td>
<td>405</td>
<td>161</td>
</tr>
<tr>
<td>Isoptera</td>
<td>162</td>
<td>102</td>
<td>37</td>
</tr>
<tr>
<td>Phthiraptera</td>
<td>1,343</td>
<td>491</td>
<td>129</td>
</tr>
<tr>
<td>Collembola</td>
<td>600</td>
<td>93</td>
<td>37</td>
</tr>
<tr>
<td>Odonata</td>
<td>701</td>
<td>110</td>
<td>30</td>
</tr>
<tr>
<td>Thysanoptera</td>
<td>539</td>
<td>132</td>
<td>41</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>717</td>
<td>315</td>
<td>140</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>393</td>
<td>132</td>
<td>36</td>
</tr>
<tr>
<td>Phasmatoidea</td>
<td>293</td>
<td>31</td>
<td>10</td>
</tr>
<tr>
<td>Mantodea</td>
<td>154</td>
<td>151</td>
<td>140</td>
</tr>
<tr>
<td>Megaloptera</td>
<td>351</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Blattaria</td>
<td>59</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>Neuroptera</td>
<td>117</td>
<td>69</td>
<td>29</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>30</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Diplura</td>
<td>13</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Psocoptera</td>
<td>83</td>
<td>81</td>
<td>46</td>
</tr>
<tr>
<td>Mantophasmatodea</td>
<td>96</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Zygentoma</td>
<td>20</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Siphonaptera</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mecoptera</td>
<td>51</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>Gryllotalpodea</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Dermaptera</td>
<td>22</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>58,323</strong></td>
<td><strong>17,001</strong></td>
<td><strong>5,047</strong></td>
</tr>
</tbody>
</table>
Table 2. List of 26 complete mitochondrial genomes used for the alignments for COI sequences

<table>
<thead>
<tr>
<th>Order</th>
<th>Species</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coleoptera</td>
<td>Tribolium castaneum</td>
<td>NC_003081</td>
</tr>
<tr>
<td>Diptera</td>
<td>Drosophila yakuba</td>
<td>NC_001322</td>
</tr>
<tr>
<td>Hymenoptera</td>
<td>Apis mellifera</td>
<td>NC_001566</td>
</tr>
<tr>
<td>Lepidoptera</td>
<td>Bombyx mori</td>
<td>AB070264</td>
</tr>
<tr>
<td>Megaloptera</td>
<td>Corydalus cornutus</td>
<td>FJ171323</td>
</tr>
<tr>
<td>Neuroptera</td>
<td>Ascaloptyn appendiculatus</td>
<td>FJ171324</td>
</tr>
<tr>
<td>Archaeognatha</td>
<td>Nesomachilis australica</td>
<td>NC_006895</td>
</tr>
<tr>
<td>Zygoptera</td>
<td>Thermobia domestica</td>
<td>NC_006080</td>
</tr>
<tr>
<td>Collembola</td>
<td>Cryptopygus antarcticus</td>
<td>NC_010533</td>
</tr>
<tr>
<td>Diptera</td>
<td>Campodea fragilis</td>
<td>NC_008233</td>
</tr>
<tr>
<td>Isoptera</td>
<td>Reticulitermes flavipes</td>
<td>NC_009498</td>
</tr>
<tr>
<td>Orthoptera</td>
<td>Acrida wilemenset</td>
<td>NC_011303</td>
</tr>
<tr>
<td>Hemiptera</td>
<td>Bemisia tabaci</td>
<td>NC_006279</td>
</tr>
<tr>
<td>Phthiraptera</td>
<td>Bothriometopus macronemis</td>
<td>NC_009983</td>
</tr>
<tr>
<td>Pscopteria</td>
<td>Lepidoponoid sp.</td>
<td>NC_004816</td>
</tr>
<tr>
<td>Thysanura</td>
<td>Atelura formicaria</td>
<td>NC_011197</td>
</tr>
<tr>
<td>Thysanoptera</td>
<td>Thrips imaginis</td>
<td>NC_004371</td>
</tr>
<tr>
<td>Phasmatoidea</td>
<td>Timema californicum</td>
<td>DQ241799</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>Pteronarcy princeps</td>
<td>NC_006133</td>
</tr>
<tr>
<td>Mantodea</td>
<td>Tamolanica tamolana</td>
<td>NC_007702</td>
</tr>
<tr>
<td>Mantophasmatoidea</td>
<td>Sclerophasma paresisene</td>
<td>NC_007701</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>Parafromurus youi</td>
<td>NC_011359</td>
</tr>
<tr>
<td>Odonata</td>
<td>Orthetrum triangulare melania</td>
<td>AB126005</td>
</tr>
<tr>
<td>Grylloblattodea</td>
<td>Grylloblatta sculleni</td>
<td>DQ241796</td>
</tr>
<tr>
<td>Raphidioptera</td>
<td>Mongoloraphidia harmandi</td>
<td>FJ859902</td>
</tr>
<tr>
<td>Blattaria</td>
<td>Periplaneta fuliginosa</td>
<td>NC_006076</td>
</tr>
</tbody>
</table>

COI sequences were done using CLUSTALW 1.83 (Thompson et al., 1997) under default parameters and a BLAST search (Altschul, 1990) in the insect mitochondrion genome database (IMGD; http://www.imgd.org) (Lee et al., 2009). The complete COI sequences (ca. 1,500 bp) were used as the standard sequences in the alignments (Table 2). Based on the alignments, the focal sequences of the 488 studies were distinguished into three groups: i) molecular studies using COI sequences located in the 5' region (750 bp), ii) using COI sequences located in the entire region (range 1 bp - 1,500 bp), and iii) using COI sequences located in the 3' region (750 bp). Additionally, 84 molecular studies which had been published until 1999 (Caterino et al., 2000) were classified into the three groups according to the 26 orders.

Results and Discussion

The 26 orders were largely divided into the three groups based on the number of molecular studies (Fig. 1). The 1st group is four ‘major’ orders, Coleoptera, Diptera, Hymenoptera, and Lepidoptera, with more than 50 studies, and the 2nd group is two orders, Hemiptera and Orthoptera, with more than 10 studies. On the other hand, the 3rd group contained the rest 20 orders each with less than 10 studies.

We classified the 488 studies into two categories, systematic and DNA barcode (Fig. 1). The systematic category included a total of 430 studies which had been reported in the 26 orders.

Fig. 1. The number of molecular studies belonging to two categories, systematic and DNA barcode, according to the 26 orders.
Fig. 2. Simple arrangements of the 488 molecular studies according to the 26 orders based on the alignments of COI sequences. The blue, green and pink bars indicated the molecular studies belonging to three regions, 5′, 3′, and entire, of COI gene, respectively. The length of each bar indicates the length of COI sequences and the thickness of the bar reflects the number of molecular studies. The number near the bar indicated references (Appendix S1).

Coleoptera showed the highest number of studies (121 studies; 28.13%), followed by Hymenoptera, Diptera, Lepidoptera, and Hemiptera, which contained between 36 and 83 studies (a total of 254 studies; 59.06%). In contrast, the rest 21 orders had less than 13 studies for a total of 55 studies (12.79%). The DNA barcode category included a total of 58 studies which had been conducted in ten orders, Blattaria, Coleoptera, Diptera, Ephemeroptera, Hemiptera, Hymenoptera, Lepidoptera, Orthoptera, Phthiraptera, and Thysanoptera. Among the 58 studies, 17 studies (29.31%), 13 studies (22.41%), and 11 studies (18.97%) were conducted in Diptera, Coleoptera, and Lepidoptera, respectively; whereas less than 5 studies were conducted in the rest seven orders (in total, 14 studies (22.41%). Among the 58 studies, 55 studies analyzed COI sequences of low taxonomic
level groups such as genus, family, and/or superfamily within one order. On the other hand, three studies (5.17%) analyzed COI sequences from several species belonging to more than two orders. For example, Hebert et al. (2003) and Armstrong and Ball (2005) analyzed a large number of species belonging to seven orders (Coleoptera, Hymenoptera, Orthoptera, Blattaria, Diptera, Ephemeroptera, and Plecoptera) and two orders (Diptera and Lepidoptera), respectively. In contrast, Rowley et al. (2007) analyzed five species belonging to three orders, Diptera, Coleoptera, and Hymenoptera.

Fig. 1 showed that the systematic and the DNA barcode studies showed various tendencies among the 26 orders. For example, Diptera (3rd rank in systematic studies) revealed the highest number of DNA barcode studies; whereas Coleoptera
(2nd rank in DNA barcode studies) showed the highest number of systematic studies. However, in common, the number of DNA barcode studies was very lower than that of the systematic studies in all orders. This result suggests that the systematic studies had been mainly conducted in the 26 orders from 2000 to 2009; while the DNA barcodes studies were in still initiative stages in the 26 orders until 2009.

The 488 studies were rearranged based on the positions of COI sequences. Fig. 2 shows that molecular studies have utilized various fragments of COI sequences according to the 26 orders. For example, 77 different regions of COI sequences were identified in 134 coleopteran studies, and 46 different regions had been used in 73 lepidopteran molecular studies. In addition, 13 different regions were detected in 14 orthopteran
molecular studies. To confirm whether each order has preferred positions of COI sequences in their molecular studies, we distinguished the 488 studies into three groups based on arrangement results (Fig. 3). In total, the 5', 3', and entire regions covered 166, 124, and 198 studies, respectively. Among the 26 orders, Diptera showed the highest number of studies, 50 (30.12% out of the 166 studies) in the 5' region, and Coleoptera showed the highest numbers of studies, 74 (37.37% out of the 198 studies) and 37 (29.84% out of the 124 studies), in the 3' and entire region, respectively. Before 2000, COI sequences had been used in 84 molecular studies of eleven orders, Lepidoptera, Diptera, Dermaptera, Hemiptera, Phasmatodea, Phthiraptera, Coleoptera, Hymenoptera, Orthoptera, Thysanoptera, and Odonata (Caterino et al., 2000). Among the 84 studies, 74 studies have been conducted in the four ‘major’ orders; while the rest studies were reported from the rest seven orders, Dermaptera, Hemiptera, Phasmatodea, Phthiraptera, Orthoptera, Thysanoptera, and Odonata. In this study, we confirmed that 445 molecular studies had been reported in the same eleven orders from 2000 to 2009, and 43 studies had been newly reported in 15 orders, Trichoptera, Collembola, Zygentoma, Neuroptera, Mecoptera, Diplura, Blattaria, Psocoptera, Mantophasmatodea, Ephemeroptera, Isoptera, Plecoptera, Megaloptera, Grylloblattodea, and Mantodea. It shows that a large number of molecular studies using COI sequences have been actively done over the 26 orders since 2000.

Among the 26 orders, seven orders, Coleoptera, Diptera, Orthoptera, Phthiraptera, Odonata, Phasmatodea, and Psocoptera, demonstrated the preferred positions of COI sequences in their molecular studies (Fig. 3). Two orders, Diptera and Orthoptera, showed the largest number of studies in the 5' region (50 out of 91 studies in Diptera; 9 out of 14 in Orthoptera). On the other hand, five orders, Coleoptera, Phthiraptera, Odonata, Phasmatodea, and Psocoptera, showed the largest number of studies in the 3'
region (74 out of 134 in Coleoptera; 6 out of 9 in Phthiraptera; 4 out of 5 in Odonata; 6 out of 7 in Phasmatodea; 3 out of 4 in Psocoptera). In Fig. 3, we confirmed that Coleoptera showed the largest number of studies in the 3’ region, and Diptera exhibited the largest number of studies in the 5’ region before 2000. In addition, in Phthiraptera and Phasmatodea, only one molecular study was conducted in the 3’ region, respectively. These results indicate the possibilities that molecular studies in Coleoptera, Diptera, Phthiraptera, and Phasmatodea from 2000 to 2009 had been followed classical studies using the positions of COI sequences well-known until 1999.

Since 2000, because of developing molecular techniques and computer sciences, a large amount of molecular studies have been conducted in Hexapoda, together with an elucidation of various mitochondrial and nuclear genes sequences. As a result, the taxa analyzed in molecular studies have been rapidly enlarged in hexapod orders, and the number of sequences has been explosively increased. However, this large scale of molecular studies and sequences have made difficult to understand the trends in molecular studies in Hexapoda. This study provides useful information to know the progress of COI sequences and their molecular studies in the 26 orders from 2000 to 2009; however, it is still not enough to understand overall tendencies of all mitochondrial and nuclear genes as well as all molecular studies in 33 hexapod orders. Much reviews for several genes and studies should be conducted in near future.

**Acknowledgement**

This research was supported by Grants-in-Aid for Scientific Research (No. P.11079 for Postdoctoral Fellowship for Foreign Researchers) from the Japan Society for the Promotion of Science, a grant from the Agricultural R&D Promotion Center and the project on survey and excavation of Korean indigenous species of the National Institute of Biological Resources (NIBR) under the Ministry of Environment, Korea, the National Research Foundation of Korea grants (2009-0063340 and 2009-0080161) and grants from the Biogreen 21 (20080401-034-044-009-01-00), the TDPAF (309015-04-SB020), and the Crop Functional Genomics Center (2009K001198). This study was also carried out with the support of “Cooperative Research Program for Agricultural Science & Technology Development” and “Evaluation of impact on the essential problem according to the new scenario of climate change (PJ009394)” Project of National Academy of Agricultural Science, Rural Development Administration, Republic of Korea.

**Literature Cited**


