Extent and divergence of heteroplasy of the DNA barcoding region in *Anapodisma miramae* (Orthoptera: Acrididae)

Ah Rang Kang¹, Min Jee Kim¹, In Ah Park¹, Kee Young Kim², and Iksoo Kim¹

¹College of Agriculture & Life Sciences, Chonnam National University, Gwangju, Korea and ²Department of Agricultural Biology, National Academy of Agricultural Science, Wanju-gun, Korea

**Abstract**

A partial sequence of the mitochondrial cytochrome oxidase subunit I (COI) gene is widely used as a molecular marker for species identification in animals, also termed a DNA barcode. However, the presence of more than one sequence type in a single individual, also known as heteroplasmacy, is one of the shortcomings of barcode identification. In this study, we examined the extent and divergence of COI heteroplasmacy, including nuclear-encoded mitochondrial pseudogenes (NUMTs), at the genomic-DNA level from 13 insect species including orthopteran *Anapodisma miramae*, and a long fragment of mitochondrial DNA and cDNA from *A. miramae* as templates. When multiple numbers of clones originated from genomic DNA were sequenced, heteroplasmacy was prevalent in all species and NUMTs were observed in five species. Long fragment DNA (~13.5 kb) also is a source of heteroplastic amplification, but the divergent haplotypes and NUMTs obtained from genomic DNA were not detected in *A. miramae*. On the other hand, cDNA was relatively heteroplastic-free. Consistently, one dominant haplotype was always obtained from the genomic DNA-origin clones in all species and also from the long fragment- and cDNA-origin clones in the two tested individuals of *A. miramae*. Furthermore, the dominant haplotype was identical in sequence, regardless of the DNA source in *A. miramae*. Thus, one possible solution to avoid the barcoding problem in relationship to heteroplasmacy could be the acquisition of multiple numbers of barcoding sequences to determine a dominant haplotype that can be assigned as barcoding sequence for a given species.

**Introduction**

The cytochrome oxidase subunit I (COI) is a variable gene in the 13 protein-coding genes (PCGs) in animal mtDNA, particularly at the silent site (e.g. Simon et al., 1994; Wan et al., 2013). Due to the availability of highly conserved primer sites, convenient PCR technology, and relative ease to handle a segment of COI gene, the DNA barcoding region, has become an ideal identification marker at the species level in animals (Folmer et al., 1994; Hajibabaei et al., 2006; Hebert et al., 2003; Simmons & Weller, 2001; Simon et al., 1994). In fact, genetic variability among species has been extensively studied in many animal groups, resolving morphological ambiguity (Hajibabaei et al., 2006; Hebert et al., 2003, 2004a,b). Such popular employment of the DNA barcode relies on the fact that mtDNA exists as a single type, termed homoplasmy in an organism (Birky, 2001). However, deviation from such an intrinsic tenet has early been reported in a bivalve species, where mtDNA is inherited in a biparental manner (Zouros et al., 1994).

The heteroplasmacy, which is defined as a mixture of more than one mtDNA genome sequence within a cell or between cells of a single individual, occurs in several ways, such as point mutation, paternal leakage, and recombination (White et al., 2008). Among them, several lines of evidences, including a backcross experiment with *Drosophila*, have shown intraspecific and interspecific paternal leakages (Arunkumar et al., 2006; Ciborowski et al., 2007; Fontaine et al., 2007; Sherengul et al., 2006), even though the mechanisms that prevent paternal leakage exist in specific manner in different organisms (e.g. simple dilution of paternal mtDNA in general, degradation in primates and cow, DNA digestion in fish, autophagy in the nematode, and a post-fertilization mechanism in mouse; Chan & Schon, 2012). Once occurred, heteroplasmacy persists for substantial generations, particularly in insects (e.g. 500 generations) because of a considerably larger population size or fewer cell divisions at oogenesis (Ashley et al., 1989; Rand & Harrison, 1989; Solignac et al., 1984). Recombination between paternal and maternal mtDNA has also been suggested to be one of the causes of heteroplasmacy; however, it occurs at a low frequency, and a clear separation between maternal and paternal mtDNA is difficult (White et al., 2008). Another major source of heteroplasmacy is the nuclear-encoded mitochondrial pseudogenes (NUMTs) that are non-functional copies of mitochondrial sequences in the absence of true heteroplasmacy (White et al., 2008). Several feasible pathways for an mtDNA fragment to transfer to the nuclear genome have been reported (Hazkani-Covo et al., 2010).

**Keywords**

*Anapodisma miramae*, COI, DNA barcoding, heteroplasmacy, mitochondrial DNA, Orthoptera, NUMTs, pseudo-sequence

**History**

Received 10 December 2014
Revised 3 February 2015
Accepted 15 February 2015
Published online 2 April 2015

Correspondence: I. Kim, College of Agriculture & Life Sciences, Chonnam National University, Gwangju 500-757, Republic of Korea. Tel: +82 62 530 5117. Fax: +82 62 530 2079. E-mail: ikkim81@chonnam.ac.kr
In insects, the presence of heteroplasmy has been reported in many groups, such as species of Drosophila, including D. melanogaster in Diptera (de Stordeur, 1997; Kondo et al., 1990; Sherengul et al., 2006; Solignac et al., 1983, 1984), Podisma pedestris in Orthoptera (Bensasson et al., 2001), Thrips tabaci in Thysanoptera (Frey & Frey, 2004), Apis mellifera and A. cerana in Hymenoptera (Meusel & Moritz, 1993; Songram et al., 2006), and Stobion spp. and Magicicada spp. in Hemiptera (Fontaine et al., 2007; Sunnucks & Hales, 1996). These results emphasize that heteroplasmy occurs more frequently than a rare event. Therefore, the remaining question is to uncover the extent and pattern of heteroplasmy and to find a way to select the functional mitochondrial copy that validates DNA barcoding.

In this study, we investigated the prevalence of heteroplasmy of the DNA barcoding region from 13 species. With the consideration of previous reports on the prevalence of heteroplasmy in Orthoptera (Gellissen & Michaelis, 1987; Gellissen et al., 1983; Keller et al., 2006) and our preliminary data, we selected one orthopteran species, Anapodisma miramae Dovnor-Zapol’skii, 1933, to study the extent and divergence of heteroplasmy, including NUMTs. This species is distributed in Korea, China, and Russian Far East and is found frequently in forest and grassland during July–September in Korea (Bae, 1998). This species is characterized by a large head, projected forward; thin antennae extended slightly beyond posterior margin of pronotum, tegmina narrow, lobe-like; and apex rounded (Lee & Lee, 1984). The length of body is 20–28.8 mm and the length of pronotum, tegmina narrow, lobe-like; and apex rounded (Lee & Lee, 1984). Specifically, we sequenced 15–45 clones of the DNA barcoding region amplified from different sources of DNA, such as genomic DNA, long-PCR fragment, and cDNA of A. miramae. These sequences were analyzed for patterns of heteroplasmy and discussed for their possible origin. Finally, we proposed a possible method to avoid ambiguity for species identification using DNA barcoding.

### Materials and methods

#### Samples

Thirteen species in four insect orders (Orthoptera, Heteroptera, Hymenoptera, and Mantodea) were screened (Table 1). Myzus persicae in Heteroptera was obtained from an artificial rearing colony that continued for 4 years, but other species were collected in wild. From the preliminary experiment, we found that Anapodisma miramae is one of the heteroplasmy-rich species in genomic DNA for the DNA barcoding region. Thus, four individuals of A. miramae were collected from at least 60 km distant localities in Korea (named AMYG, AMJJ, AMJS, and AMBB) and sequenced for the double number of clones for each individuals compared with other species using genomic DNA. Among the four individuals, AMJS and AMBB that were newly collected were further used for sequencing using long fragment and cDNA as templates.

#### DNA extraction and primer design

For genomic DNA extraction, a hind leg was used for most species, but for aphid species, the whole body was used after the midgut was removed. DNA extraction was performed using a Wizard Genomic DNA Purification Kit in accordance with the instructions of the manufacturer (Promega, Madison, WI). To amplify and sequence the COI gene corresponding to the DNA barcoding region, a pair of universal primers was adapted from Folmer et al. (1994), but the efficiency for amplification and sequencing were not high enough for some species. Thus, for the two species of Hymenoptera, a set of primers was adapted from Kim et al. (2009). On the contrary, for A. miramae, which requires further casual amplification and sequencing, a pair of primer set was designed from several full-length mitochondrial genomes of Orthoptera (Supplementary Figure 1) as follows: OTLCOF, 5'-TCAACAAACCTAAGACATTGGG-3' and OTHCOR, 5'-ATATGTGAATAAATACCAAATCCTGG-3'. This primer set

### Table 1. Summary of COI sequences obtained from genomic DNA.

| Taxon          | Species                  | Clone | Sea | D   | H   | SD | NUMTs |
|----------------|--------------------------|-------|-----|-----|-----|----|-------|
| Heteroptera    |                          |       |     |     |     |    |       |
| Aphididae      | M. persicae              | 26    | 17  | MP01 (9) | 9 | 0.15–0.61 (4) | 0 |
| U. formosanum  |                          | 17    | 16  | UF01 (11) | 5 | 0.15–0.30 (2) | 1 |
| Orthoptera     |                          |       |     |     |     |    |       |
| Acrididae      | A. miramae (AMYG)        | 35    | 30  | AMYG05 (14) | 8 | 0.14–2.58 (18) | 9 |
|                | A. miramae (AMJJ)        | 39    | 37  | AMJG11 (22) | 9 | 0.14–7.75 (54) | 7 |
|                | A. miramae (AMJS)        | 39    | 39  | AMJS03 (36) | 4 | 0.14–8.03 (56) | 0 |
|                | A. miramae (AMBB)        | 45    | 45  | AMBBG03 (37) | 7 | 0.14–2.15 (15) | 2 |
|                | L. migratoria            | 20    | 16  | LM (12)   | 5 | 0.15–0.30 (2) | 0 |
| Gryllidae      | O. chinensis             | 19    | 15  | OC01 (12) | 3 | 0.15–0.61 (4) | 1 |
|                | T. emma                  | 26    | 12  | LA07 (3)  | 8 | 0.15–5.93 (39) | 2 |
| Gryllotalpidae | G. orientalis            | 21    | 16  | TE01 (8), TE03 (2) | 7 | 0.15–3.34 (22) | 1 |
| Tettigoniidae  | P. ussuriensis           | 21    | 19  | PU01 (13) | 7 | 0.15–3.5 (23) | 0 |
|                | P. nigroantennata        | 17    | 16  | PN02 (9) | 6 | 0.15–0.46 (3) | 2 |
| Hymenoptera    |                          |       |     |     |     |    |       |
| Apididae       | X. appendiculata         | 21    | 20  | XA01 (11) | 10 | 0.15–0.91 (6) | 0 |
| Scoliidae      | C. annulata              | 21    | 19  | CA08 (4)  | 16 | 0.15–0.76 (5) | 0 |
| Mantodea       | S. maculata              | 20    | 14  | SM01 (8) | 7 | 0.15–2.43 (16) | 0 |

Detailed sequence statuses of NUMT are presented in Sup. Table 1. Full species names are as follows: Myzus persicae, Uroleucon formosanum, Anapodisma miramae, Locusta migratoria, Oxya chinensis, Loxoblemmus arietulus, Teleogryllus emma, Gryllotalpa orientalis, Paraltanticus ussuriensis, Phaneroptera nigroantennata, Xylocopa appendiculata, Campsomeris annulata, and Statilia maculata.

|a| Number of clones identifiable as COI sequence. |
|b| Dominant haplotype (number of identical clones). |
|c| H, number of haplotypes not assigned as NUMTs. |
|d| Percent sequence divergence of haplotypes (maximum nucleotide divergence, bp). |
|e| Number of NUMTs. |
amplifies a slightly longer DNA barcoding region, 697 bp, instead of 658 bp.

Considering the reports that the length of NUMTs reaches only to 3.5 kb in orthopteran Locusta migratoria and 7.9 kb in a cat (Gellissen & Michaelis, 1987; Lopez et al., 1994), we expected that amplification of the DNA barcoding region from a template longer than 8 kb would be free from NUMTs, maximizing the chance to amplify a single cytoplasmic copy, if heteroplasmy was not present. Thus, we first amplified and sequenced ~500 bp of the A. miramae CytB gene and ~700 bp of the ND4 gene from genomic DNA. The primers for the short fragments were designed via the alignment of several orthopteran mitochondrial genomes sequenced in their entirety (Fenn et al., 2007; Flook et al., 1995; Kim et al., 2005; Ma et al., 2009; Zhou et al., 2009). Using the sequence information of the short fragments, a pair of species-specific primers was designed to amplify a long fragment of ~13.5 kb in length. The primer sequences for these short and long fragments are listed in Supplementary Table 1.

Polymerase chain reaction (PCR)

PCR for the DNA barcoding region from genomic DNA and a 13.5-kb long fragment was conducted using the HotStart Highfidelity Polymerase Kit (Qiagen, Hilden, Germany). Each reaction contained 1 μL of DNA diluted 1:10, 5 μL of 5× Q-solution, 0.5 μL of MgSO₄, 0.5 μL of polymerase, 1 μL of 10 pmol forward and reverse primers and was filled up to 25 μL with nuclease free water. The reaction was performed in a Biometra thermal cycler (model T-gradient Thermoblock, Hamburg, Germany) using a 5-min initial denaturation at 95 °C, followed by 40 cycles of a 15-s denaturation at 95 °C, 1-min annealing at 50–55 °C, and 1-min extension at 72 °C, and finalized with 10 min at 72 °C. For the amplification of the long fragment, different concentrations of genomic DNA (1/10, 1/50, 1/100, 1/1000, and 1/2000 from the extracted DNA) were mixed with Long-Accurate (LA) Taq™ polymerase (Takara Biomedical, Tokyo, Japan), and PCR was conducted under the following conditions: initial denaturation for 1 min at 96 °C, followed by 30 cycles of 10 s at 98 °C, and 15 min at 55 °C, and a subsequent 10-min final extension at 72 °C. To verify successful PCR amplification, electrophoresis was conducted using 0.5× TAE buffer on a 1% agarose gel, and successful amplicons were purified with QIAquick PCR Purification Kit reagents (QIAGEN, Hilden, Germany) for subsequent experiments.

Total RNA extraction and cDNA synthesis

To obtain functional copies of the DNA barcoding sequences, two A. miramae individuals (AMBB and AMJS) were subjected to PCR amplification using the cDNA synthesized from total isolated RNA. Total RNA was first isolated from one rear leg of each individual using the SV Total RNA Isolation System (Promega, Madison, WI). RNA purity and concentration were measured using a ND-1000 Spectrophotometer (NanoDrop, Wilmington, DE). The RNA was then reverse-transcribed using AccuPower™ RT Premix (Bioneer, Daejeon, Korea) and oligo dT primers. cDNA synthesis was performed for 60 min at 42 °C, and reverse transcriptase was inactivated for 5 min at 94 °C. The resulting cDNA was used as a template for the amplification of the DNA barcoding region using the same primers, polymerase, and PCR conditions as above.

Cloning and sequencing

To test the extent of heteroplasy, the genomic DNA-origin COI amplicons were cloned into the pGEM-T Easy vector (Promega, Madison, WI) and transformed into XL1-Blue competent cells (Stratagene, La Jolla, CA). Approximately 15–26 colonies per species for the 12 species and 35–45 colonies for A. miramae were selected, cultured for 16 h (Model IB-22, Jelo Tech, Seoul, Korea), and isolated for plasmid DNA using the Wizard Plus SV Miniprep DNA Purification System (Promega, Madison, WI). For the long fragment-origin COI amplicons, 20 colonies from both A. miramae AMJS and AMBB were randomly selected, cultured, and isolated for plasmid DNA. For the eDNA-origin COI amplicons, 15 colonies from both A. miramae AMJS and AMBB were isolated for plasmid DNA. DNA sequencing was performed using the ABI PRISM® BigDye® Terminator v3.1 Cycle Sequencing Kit and the ABI PRISM™ 3100 Genetic Analyzer (PE Applied Biosystems, Foster City, CA). All fragments were sequenced from both strands.

Restriction enzyme digestion analysis

To complement the sequence data, a few COI amplicons obtained from genomic DNA were digested with a restriction enzyme, and their digests were separated on agarose gel to observe banding pattern. Using Webcutter 2.0 (Heiman, 1997), HindIII gene, which recognizes the ’AAGCTT’ sequence, was selected after a preliminary experiment. The digestion reaction contained 2 μL of PCR product, 2 μL of 10× restriction enzyme buffer, 1 μL of HindIII (Beamsbio, Korea), and 15 μL of double distilled water to make up a 20-μL reaction. The reaction was incubated for 12 h at 37 °C. Heteroplasmic amplicons were expected to be digested into 278-bp- and 468-bp-long fragments, including the primer sequences (255-bp- and 442-bp-long fragments, excluding the primer sequences), instead of a 746-bp-long non-digested amplicon (697 bp, excluding primer sequence).

Polymerase error test

To estimate the error ratio of the polymerase (HotStar Hifidelity polymerase Kit; QIAGEN, Hilden, Germany), a COI amplicon that has already been sequenced was re-amplified and cloned. Sixteen of the selected 20 clones were successfully sequenced. The intrinsic mutational ratio of the polymerase is known to be 2.3 × 10⁻⁶ base pairs.

Sequence analysis

Sequences of each clone were finalized by checking nucleotide peaks of chromatograms and subsequent sequence alignment of forward and reverse strands using the Clustal W program (Thompson et al., 1994). Individual sequences of clones were then compared using Blast Search at NCBI (http://www.ncbi.nlm.nih.gov/blast) and the BOLD System (Ratnasingham & Hebert, 2007) for their appropriateness as respective species. Nucleotide sequences were translated into amino acid sequences to determine the presence of stop codons, unusually inserted or deleted codon, and frameshifts, which can confidentially assign NUMTs on the basis of the invertebrate mtDNA genetic code using EMBOSS Transeq (http://www.ebi.ac.uk/emboss/transeq). When a homologous sequence from two clones differed by one or more nucleotides, the sequences were considered different haplotypes. Sequence divergence between haplotypes (and NUMTs) was calculated for uncorrected pairwise distance using PAUP ver. 4.0b10 (Swoford, 2002). All haplotypes and NUMTs obtained in this study are provided with GenBank accession numbers in Supplementary Table 2.

Phylogenic analysis

For each A. miramae individual, phylogenetic analysis was performed using all available haplotypes and NUMTs to understand the divergence and relationships between sequences in
each individual. The maximum-likelihood (ML) method was performed using PHYML (Guindon et al., 2005) with the GTR (Lanave et al., 1984) + G model selected as the best-fit substitution model on the basis of the Akaike information criterion score (Akaike, 1974) using Modeltest ver. 3.7 (Posada & Crandall, 1998). The number of substitution rate categories was specified as four, and the proportion of invariable sites and the number of substitution site were sets as the values obtained from the model test with the starting tree set as a BIONJ distance-based tree. The confidence values were evaluated via the bootstrap test with 1000 iterations. Three orthopteran species, Acrida cinerea and A. willemsei belonging to the subfamily Acridinae in Acrididae (Fenn et al., 2008; Liu & Huang, 2010) and Ognevia longipennis belonging to the subfamily Catantopinae in Acrididae, were obtained from GenBank for outgroups.

Results

Restriction enzyme digestion of A. miramae

Digestion of genomic DNA-origin PCR products of the four A. miramae individuals all revealed additional restriction fragments when digested with HindIII gene. For example, the A. miramae AMYG revealed two digested fragments with sizes of ~442 bp and ~255 bp, along with an undigested 697-bp fragment, indicating that more than two products were amplified in the PCR process, indicating the heteroplasmic nature of the DNA barcoding region (Supplementary Figure 2).

Genomic DNA-origin PCR products of 13 species

Genomic DNA-origin PCR products mostly provided reasonably proper COI sequences through cloning, but some were unsuccessful because these sequences were unusually short (e.g. ≤400 bp), lacked typical primer sites, contained a large intermittent deletion, or a mixture of these aberrant sequences, even though double peaks in the chromatograms were minimum. These sequences were excluded in the subsequent analysis, treating them as failed clones for sequencing, due to alignment and clear identification of the DNA barcoding region impossible. Eventually, successfully sequenced clones ranged from 46.2% (Loxoblemmus arietulus in Orthoptera) to 100% (Gryllotalpa orientalis and A. miramae AMBB in Orthoptera) in 13 species (Table 1). Even excluding those aberrant sequences, five of the 13 species including three individuals of A. miramae amplified NUMTs. Characteristics of NUMTs are described in Supplementary Table 3.

Along with NUMTs, the genomic DNA-origin PCR products provided three (Oxya chinensis in Orthoptera) to 16 haplotypes (Camposomeris annulata in Hymenoptera) that are translatable in the 13 species, including four A. miramae individuals (Table 1). Most haplotypes in each species showed moderate to low sequence divergence (~1–2%), along with immediately close haplotypes to each other (<1%), but L. arietulus, which provided seven haplotypes, revealed overall higher divergence among the seven haplotypes, with an average divergence of 2.51% (Supplementary Table 4). Furthermore, at least one highly divergent haplotype that was distinguished from others was amplified in several species (Supplementary Table 4). For example, Teleogryllus emma provided one haplotype (TE09) that was divergent from the remaining six haplotypes, providing a sequence divergence range from 2.28% (15 bp) to 3.34% (22 bp), whereas the six haplotypes ranged only from 0.15% (1 bp) to 1.52% (10 bp). For Paratlanticus ussuriensis, one haplotype (PU08) was divergent from the remaining six haplotypes, ranging from 2.89% (19 bp) to 3.50% (23 bp), whereas the other six ranged only from 0.15% (1 bp) to 1.22% (8 bp) among the six.

The maximum sequence divergence of A. miramae was variable among individuals. For AMYG, AMJS, and AMBB, only one sequence (AMYGG65, AMJSG65, and AMBB68, respectively) diverged from the other sequences, ranging from 2.15 to 2.58% (15–18 bp), 7.89 to 8.03% (55 and 56 bp), and 1.87 to 2.15% (13–15 bp), respectively (Supplementary Table 5). On the contrary, AMJJ revealed that two of the nine haplotypes were divergent from the remaining seven, with a sequence divergence range from 7.46% (52 bp) to 7.75% (54 bp), and even these seven were divergent from each other, 0.14% (1 bp) to 7.57 (54 bp). Summarized, it is obvious that all species tested possessed a certain level of heteroplasmy, along with obvious NUMTs and the degree of sequence divergence among haplotypes were variable among species.

Long fragment-origin PCR products of A. miramae

A long mitochondrial fragment spanning from Cytb to ND4, including a COI gene, was successfully amplified from A. miramae AMJS and AMBB. The genomic DNA as a template for long-fragment PCR was diluted into 1:10–1:2000, and the most diluted DNA was utilized as a template for subsequent COI amplification to minimize unwanted PCR products (Supplementary Figure 3). Length approximation using several orthopteran mitochondrial genomes confirmed the expected product length as ~13.5 kb (Fenn et al., 2008; Flook et al., 1995; Liu & Huang, 2010; Ma et al., 2009; Sun et al., 2010; Zhao et al., 2010).

COI sequences from long fragment-origin PCR products provided all readable and translatable sequences each from 20 clones of AMJS and AMBB (Table 2). Unlike genomic DNA-origin products, no sequence was peculiar enough to identify as non-DNA barcoding sequence, and no NUMTs were observed. Nevertheless, it was surprising to have a high number of haplotypes from both individuals (17 from AMJS and 14 haplotypes from AMBB) (Table 2). Pairwise comparisons between haplotypes have shown that the 17 haplotypes of AMJS diverged from 0.14 to 0.72% (5 bp), and the 14 haplotypes of AMBB diverged from 0.14 to 1.72% (12 bp) (Supplementary Table 6). The haplotypes of AMJS were closely related, with a

Table 2. Summary of COI sequences obtained long PCR product as template.

| Individuals | Clone | Sea | Da | Ha | SDa | NUMTe |
|-------------|-------|-----|----|----|-----|-------|
| A. miramae (AMJS) | 20 | 20 | AMJSLO5 (3) | 17 | 0.14–0.72% (5) | 0 |
| A. miramae (AMBB) | 20 | 20 | AMBBLO2 (7) | 14 | 0.14–1.72% (12) | 0 |

aNumber of clones identifiable as COI sequence.

bDominant haplotype (number of identical clones).
cH, number of haplotypes not assigned as NUMTs.
dPercent sequence divergence of haplotypes (maximum nucleotide divergence, bp).
eNumber of NUMTs.
gradual divergence from each other, whereas those of AMBB revealed four distinctively divergent haplotypes (AMBBL04, AMBBL05, ABMML06, and AMBBL14) that showed sequence divergence from 0.57% (4 bp) to 1.44% (10 bp) to the remaining closely related haplotypes. Even these four haplotypes were divergent from 1.15% (8 bp) to 1.72% (12 bp) from each other.

cDNA-origin PCR products of *A. miramae*

COI sequences of cDNA-origin PCR products from AMJS and AMBB provided each 15 readable sequences (Table 3). All sequences were properly translated with the expected size, without NUMTs. Three haplotypes were obtained from AMJS, indicating that heteroplasmy did exist, unless extra copies were derived from experimental errors (i.e., polymerase error), whereas a single haplotype was obtained from AMBB (AMBBC01). The sequence divergence of the AMJS haplotypes ranged from 0.14 to 0.29% (2 bp) (Supplementary Table 7), indicating very low sequence divergence compared with that obtained from genomic DNA-origin haplotypes (8.03%) and long fragment-origin haplotypes (0.72%) (Tables 1 and 2).

A dominant haplotype from different DNA sources of *A. miramae*

Genomic DNA-origin haplotypes of all species, including four *A. miramae* individuals, revealed one dominant haplotype that often accounted for the majority of sequences, with a frequency from 25.0% (*L. arietulus*) to 92.3% (*A. miramae* AMJS), whereas nearly all other haplotypes were singleton (Table 1). Furthermore, two *A. miramae* individuals (AMJS and AMBB) also provided a single dominant haplotype throughout the tested DNA sources, such as genomic DNA, long fragment, and cDNA (Table 4). The frequency of the dominant haplotype of *A. miramae* AMJS was 92.3% from genomic DNA (AMJSG03), 15% from long fragment (AMJSL05), and 86.7% from cDNA (AMJSC03). For AMBB, the dominant haplotype accounted for 82.2% (AMBBG03), 35% (AMBBL02), and 100% (AMBBC01) from the genomic DNA, long fragment, and cDNA, respectively. Importantly, the dominant haplotype obtained each from different DNA sources was identical in sequence in each individual (Supplementary Table 8). On the contrary, other haplotypes obtained from different DNA sources were all different in each *A. miramae* individual (Supplementary Table 8).

Polymerase error test for *A. miramae*

To test whether the heteroplasmy obtained in this study was artificially generated by an unexpected error ratio of the *Taq* polymerase, a plasmid already sequenced was randomly chosen, re-amplified, isolated for plasmid DNA, and successfully sequenced for 16 clones. All sequences were translatable, without indels and NUMT. Fifteen of the 16 clones were identical to the original sequence, but one clone differed by one nucleotide, consequently accounting for an error ratio of 8.97/10^16 base pairs [1 bp/(697 bp × 16 clones = 11,152 bp) or 0.063 bp per reaction]. This estimate is ~39-fold higher than the error ratio known for the HotStar Highfidelity polymerase we used (2.3 × 10^-6 base pairs, Qiagen, Zürich, Switzerland). Most likely, several factors might have caused the detected error ratio, although we did not check other factors (e.g. number of clones tested, number of cycles, etc.). Nevertheless, the number of haplotypes obtained with all sources of PCR products remained as they were when the detected error ratio was subtracted from the pairwise distance values. For example, when the three haplotypes obtained from AMJS cDNA as template were subtracted with the background mutational ratio (0.06 bp/687 bp), one nucleotide variation became 0.94 nucleotide per 698-bp long haplotype, resulting in change in sequence divergence among the three haplotypes from 1–2 bp (0.14–0.29%) to 0.94–1.94 bp (0.13–0.27%), but the number of haplotypes obtained remained as they were (Table 5).

### Table 3. Summary of COI sequences obtained from cDNA as a template.

| Individuals | Clone | Se | D'H | H | SD | NUMT |
|-------------|-------|----|-----|---|----|------|
| *A. miramae* (AMJS) | 15 | 15 | AMJSC03 (13) | 3 | 0.14–0.29% (2) | 0 |
| *A. miramae* (AMBB) | 15 | 15 | AMBBC02 (15) | 1 | – | 0 |

Table 3. Summary of COI sequences obtained from cDNA as a template.

### Table 4. Summary of heteroplasmy diversity of *Anapodisma miramae* COI region.

| Individual | Genomic DNA | Long fragment | cDNA |
|------------|-------------|---------------|------|
| AMYG       | 30/8/AMYGG05/14/46.7 | – | – |
| AMJJ       | 37/9/AMJJG11/22/59.5 | – | – |
| AMJS       | 39/4/AMJSG03/36/92.3 | –20/17/AMJSL05/3/15 | 15/3/AMJSC03/13/86.7 |
| AMBB       | 45/7/AMBBG03/37/82.2 | 20/14/AMBBL02/7/35 | 15/1/AMBBC01/15/100 |

Table 4. Summary of heteroplasmy diversity of *Anapodisma miramae* COI region.
Table 5. Pairwise sequence divergence among COI haplotypes obtained from A. miramae AMJS cDNA as a template.

|   | 1     | 2     | 3     |
|---|-------|-------|-------|
| 1 | AMJSC03 | –     | 0.13/0.14 | 0.14/0.13 |
| 2 | AMJSC17 | 0.94/1 | –     | 0.27/0.29 |
| 3 | AMJSC22 | 0.94/1 | 1.94/2 | –     |

Values above the diagonal are percent difference and those below the diagonal are nucleotide difference. First values are unrooted pairwise difference without error ratio subtraction, whereas second values are those subtracted polymerase error ratio.

Phylogenetic analyses of each A. miramae individual

For each A. miramae individual, phylogenetic analysis with the haplotypes and NUMTs from all DNA sources was performed to detect any discernible pattern present (Figure 1). Regardless of DNA sources, there was a tendency to form a monophyletic group only among haplotypes excluding NUMTs, but AMJ was exceptional. Seven of the nine haplotypes of AMJJ formed a strong monophyletic group (99.1%), but two haplotypes (AMJJG28 and AMJJG37) were interspersed in the tree clustering together with the NUMTs, indicating that NUMTs are not distinguishable from haplotypes (Figure 1B). Not robust but consistently in all A. miramae individuals, the dominant haplotype was always placed at the derived position in the tree (AMYG05 for AMYG, AMJJG11 for AMJ, AMJSG03 = AMJSL05 = AMJSC03 for AMJS, and AMBBG03 = AMBBL02 = AMBBC02 for AMBB) and grouped together with the majority of haplotypes, indicating the relative recency of dominant haplotype and its allied haplotypes, compared with NUMTs (Figure 1). The hierarchical branching pattern of haplotype groups was evident in the tree, even though all haplotypes are within-individual origin, indicating different timings of origin and antiquity of some haplotypes (e.g. AMYGG65 for AMYG, AMJJG41 AMJJG28 AMJJG37 for AMJJ, AMJSG65 for AMJS, and AMBBG68 for AMBB).

In contrast to the analysis based on genomic DNA-origin haplotypes, the long fragment-origin haplotypes have shown a conscious topology, forming each A. miramae AMJS and AMBB haplotypes strong monophyly without hierarchical branching pattern (98% and 100%, respectively; Supplementary Figure 4). Nevertheless, no further inference was possible due to unresolved relationships between most haplotypes in both individuals. The absence of an independent group and hierarchical branching pattern unlike the topology obtained from genomic DNA-origin haplotypes may indicate that the long fragment serve as a source of homogeneous haplotypes, absent of NUMTs.

Discussion

Extent of heteroplasmy

The PCR products of the COI region from genomic DNA have obviously provided multiple sequences from all species, including A. miramae individuals. Evidence of enzyme digestion pattern (Supplementary Figure 2) and obviously higher divergence than the known polymerase error ratio (Table 5) along with the existence of clonal variation in all species demonstrate that COI sequences are not present as homoplasmy in a diverse species, if not at all, in insects, and this finding is consistent with previous studies that investigated heteroplasmy in insects (Bensasson et al., 2001; Fontaine et al., 2007; Sherengul et al., 2006; Songram et al., 2006; Sunnucks & Hales, 1996). These multiple sequences were either translatable haplotypes with a typical length in nucleotide and amino acid sequence or NUMTs that could be classified into several categories (Supplementary Table 3). Furthermore, the sequence divergence among haplotypes was substantial within an individual in some species. For example, Campsomeris annulata in Hymenoptera has shown 76.2% haplotype diversity (16 haplotypes in 21 clones), and two individuals of A. miramae (AMJJ and AMJS) have shown each 7.75% and 8.03% of the maximum sequence divergence in nine and four haplotypes (Table 1). Similar to our findings, 64 clones sequenced from a total of four individuals of T. tabaci in Thysanoptera have shown 22 different haplotypes (Frey & Frey, 2004). Magnacca & Brown (2010) examined four species of Hawaiian bees and found that all four species have heteroplasmy, showing that the uncorrected sequence divergence of the dominant haplotype to others ranged from 1.22 to 4.44%. With the inclusion of NUMTs, the maximum sequence divergence was well over 10% in many of our studied species, with varying degrees of divergence between the haplotypes and NUMTs (data not shown). Similarly, 10 grasshopper species, using enriched mtDNA, have shown that multiple ND5-like sequences were either minimally (e.g. <1%), moderately (e.g. 1–2%), or largely (e.g. >10%) divergent, suggesting differences in the ability of different genomes to gain or lose NUMTs (Bensasson et al., 2001). Collectively, the general fact that mtDNA, including the DNA barcoding region, exists as homoplasy in insects does not hold true anymore, and the sequence divergence among haplotypes and with the inclusion of NUMTs are moderate to high, requiring caution for species identification in insects.

Possible sources of heteroplasmy

From a general perspective, the presence of heteroplasmy can be explained mainly in three aspects. First, the NUMTs that have been transferred to nuclear DNA are one major source. At present, the only way we could obviously distinguish them from the functional mitochondrial copy is the detection of (1) in-frame stop codons that indicate an obvious non-functional copy, (2) the frameshift mutation that resulted in an abnormal amino acid alignment, and (3) insertion and deletion of unique triple indels, resulting in a slightly longer or shorter amino acid length. Our current data show abundant examples of the first case, and the other cases are present but rarely (Supplementary Table 3). Other grasshopper data have shown the presence of many paralogs that cannot readily be categorized as NUMTs because they lack in-frame stop codons and differ from the orthologous mtDNA sequences by one or two nucleotides (Song et al., 2008). This may suggest that several haplotypes assigned as haplotypes in our study, in fact, might also be NUMTs but are indistinguishable from heteroplasmic copies. Second, the individuals utilized for this study possess heteroplasmy through somatic mutations acquired during life. It has been shown that mutations in mtDNA accumulate with age in a variety of post-mitotic mammal tissues (Wallace, 2010). The main cause for the accumulation of the mutation is the oxidation of DNA bases by free radicals, generated especially during cellular energy production by the mitochondria (Wallace, 2005). Without data for an age-specific mutational ratio, further reasonable inference may not be possible. In contrast, largely divergent haplotypes found in some species, including the four individuals of A. miramae, cast a doubt on life span mutations. Furthermore, the hierarchical branching pattern and clustering of heteroplasmic copies and NUMTs together in the phylogeny of A. miramae individuals may suggest the antiquity of some haplotypes, instead of life span mutations (Figure 1). On the contrary, the two extra haplotypes found in the cDNA of A. miramae AMJS can further casually be assigned as heteroplasmic copies that may have been acquired during life, considering the low sequence divergence to the
dominant haplotype obtained from the cDNA. Consistent to this rationale, several genomic DNA-derived haplotypes that are close to the dominant haplotypes (e.g. 1–2 nucleotides) might also be assigned as heteroplasmic copies, which possibly could have been originated by the same mechanism of cDNA-derived heteroplasmic copies. The third source could be paternal leakage. Previous studies have shown that the leakage of mitochondrial genome to offspring in several organisms, such as scorpions,}

Figure 1. Maximum-likelihood phylograms of haplotypes and NUMTs obtained from the genomic DNA as template. (A) A. miramae AMYG, (B) A. miramae AMJJ, (C) A. miramae AMJS, and (D) A. miramae AMBB. The log-likelihood value of the best tree is $-3335.453064$, $-3136.925600$, $-2045.613611$, and $-2329.585355$, respectively, and the numbers at each node specify the bootstrap percentage of 1000 pseudoreplicates. The scale bar indicates the number of substitutions per site. The underlined haplotype indicates the dominant haplotype, and NUMTs are marked with an N after sequence names. Three orthopteran species, Acrida cinerea, A. willemsei, and Ognevia longipennis, were used as the outgroups.
lizards, and drosophila (Gantenbein et al., 2005; Sherengul et al., 2006; Ujvari et al., 2007). However, the frequency of paternal leakage is difficult to determine, mostly due to genetic similarity between paternal and maternal haplotypes in populations. If leakage occurs between distantly related groups, then differences in mtDNA haplotypes should be obvious and the detection of paternal mtDNA straightforward. In this regard, some heteroplasmic copies that have rather larger sequence divergence are plausible to assign as paternal leakage. However, current casual PCR and sequencing technologies have a limited power for the detection of paternal leakage, requiring further specific detection methods (Milligan, 1992).

The incorporation of the observed polymerase error ratio in the pairwise comparison did not lessen the extent of heteroplasy (Table 5). Nevertheless, we cannot rule out the possibility of error-generating methods, including Taq polymerase during PCR because no Taq polymerase works totally error-free, and we did not test other factors critically. In particular, several haplotypes close to the long fragment-origin dominant haplotype of A. miramae AMBB and AMIS (Supplementary Table 6) are plausible to some degree because the procedure to amplify COI region after the amplification of long mitochondrial fragment from genomic DNA may double the chance of generating the dominant haplotype-related haplotypes. In fact, we obtained a total of each nine and six haplotypes that are immediately close to the dominant haplotypes (e.g. one bp difference) out of each 20 clones from the long fragments of AMIS and AMBB (Supplementary Table 6). The frequency of immediately close haplotypes in long fragment-derived haplotypes has higher frequency compared with genomic DNA-derived ones (two out of 39 in AMJS and four out of 45 clones in AMBB) (Supplementary Table 8). Furthermore, the calculation of codon position-based substitution between sites shows that each codon position has an equal frequency of substitution both in AMJS (each three in a total of nine) and AMBB (each two in a total of six) (data not shown). This frequency is highly deviated from general mutational bias in mitochondrial genomes, including Orthoptera, because the third codon position has a higher mutational pressure with a degenerative nature (Perna & Kocher, 1995).

Implications for DNA barcoding
DNA barcoding is generally considered a proper molecular marker for species identification as evidenced in numerous cases (Hebert et al., 2003; Lawrence et al., 2009). In spite of several positive factors, including the nature of molecule (e.g. multiple copies in a cell), technical easiness, and a certain level of genetic distance, several studies questioned the reliability in insects (e.g. Meier et al., 2006). They include the quality of DNA, particularly for museum specimens, necessitating an alternative method (Hajibabaei et al., 2006), larger intraspecific variation than interspecific difference (Meier et al., 2006; Meyer & Paulay, 2005), analytical method (Munch et al., 2008; Virgilio et al., 2010), and NUMTs (Bensasson et al., 2001; Rubinoff et al., 2006; Song et al., 2008). Low-frequency and low sequence variation in heteroplasmic and careful selection of a genuine copy may not disrupt a proper interpretation in the large population-based molecular ecological and multi-gene-based phylogenetic results. However, dependence on limited number of sequences from a limited number of individuals, where DNA barcoding for species delimitation is often practiced, may introduce serious ambiguity if not properly handled. The genomic DNA, which is the major source of PCR amplification, may also be problematic because it is more prone to amplify heteroplasmic copies, including large divergent NUMTs, compared with other sources of DNA, such as long fragment and cDNA. Between genomic DNA and long fragment DNA, both provide multiple copies, but the latter provided overall similar haplotypes without divergent haplotypes and NUMTs (Supplementary Table 8). A phylogenetic approach may be helpful for some cases because haplotypes tended to position in the most derived group, whereas all NUMTs were branched from earlier nodes in A. miramae individuals (Figure 1). This may explain that the pseudo-sequence types are “old” transposed to nuclear DNA and evolved from ancestral mtDNA.

Practically, obvious pseudo-sequence types can easily be excluded for the species identification purpose, but discrimination of orthologous copies from paralogous copies might be a difficult task, possible solution to overcome such an obstacle includes the acquisition of multiple numbers of barcoding sequences, possibly using different PCR products of an individual to determine a dominant haplotype. The functional mtDNA, which is present abundantly in the mitochondria in eukaryotic cells (Michaels et al., 1982), will predominantly be amplified by targeted primers (Roehrdanz, 1993). On the other hand, accumulated substitutions and indels in the process of time in the primer sites may lower the stochastic chance of amplification of the NUMTs that have identical sequences, and the lower copy numbers of paralogous sequences compared with the functional copy may serve as limited templates for the amplification of identical heteroplasmic sequences in the cytoplasm. Also, separate transposition events may further hamper an amplification of identical NUMTs the NUMTs branched hierarchically in the tree indicate that they may have evolved from ancestral mtDNA in separate transposition events (Figure 1). On the other hand, a consistent amplification of a single dominant haplotype throughout different DNA sources in both the A. miramae AMJS and AMBB may suggests that the most dominant haplotype, regardless of DNA source, is a true functional cytoplasmic copy (Table 4). Previously, an analysis of 64 clones from four individual thrips from a laboratory culture and subsequent confirmation from cDNA of reverse-transcribed mRNA evidenced that the most common haplotype is of mitochondrial origin (Frey & Frey, 2004). Also, Sunnucks & Hales (1996) found each one dominant COI-COII haplotype derived from genomic DNA of three apid species and confirmed the dominant haplotypes from purified mtDNA. Therefore, multiple number of PCR and sequencing of an individual might practically be an immediate choice, if NUMTs or heteroplasmic copies are encountered. Nevertheless, if unclear peaks in the chromatograms are encountered, cloning-based multiple numbers of sequences from genomic DNA might be required to determine a conclusive representative haplotype for species identification.

Declaration of interest
The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this manuscript. This work was supported by research grants from National Institute of Biological Resources – “The Genetic Evaluation of Important Biological Resources” and “Origin of Biological Diversity of Korea: Molecular Phylogenetic Analyses of Major Korean Taxa” awarded to Iksoo Kim.

References
Akaike H. (1974). A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–23.
Arunkumar KP, Metta M, Nagaraju J. (2006). Molecular phylogeny of silkmoths reveals the origin of domesticated silkmoth, Bombyx mori from Chinese Bombyx mandarina and paternal inheritance of Antheraea proylei mitochondrial DNA. Mol Phylogenet Evol 40: 419–27.
Ashley MV, Laipis PJ, Hauswirth WW. (1989). Rapid segregation of heteroplasmic bovine mitochondria. Nucleic Acids Res 17:7325–31.
Bae YJ. (1998). Insects’ life in Korea. I. Apterygota, exopterygota (in part), and aquatic insects. Seoul, Korea: Korean Entomological Institute, Korea University.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM. (2001). Mitochondrial pseudogenes: Evolution’s misplaced witnesses. Trends Ecol Evol 16: 314–21.

Birky CW. (2001). The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. Annu Rev Genet 35: 125–48.

Chan DC, Schon EA. (2012). Eliminating mitochondrial DNA from sperm. Dev Cell 22: 469–70.

Ciborowski KL, Consuegra S, de Lainiz CG, Beaumont MA, Wang J, Jordan WC. (2007). Rare and fleeting: An example of interspecific recombination in animal mitochondrial DNA. Biol Lett 3:554–7.

d’Oster E. (1997). Nonrandom partition of mitochondria in heteroplasmic Drosophila. Heredity 79:615–23.

Fenn JD, Cameron SL, Whiting MF. (2007). The complete mitochondrial genome sequence of the Mormon cricket (Anabrus simplex: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol Biol 16:239–52.

Fenn JD, Song H, Cameron SL, Whiting MF. (2008). A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial data. Mol Phylogenet Evol 49:59–68.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–9.

Fontaine KM, Cooley JR, Simon C. (2007). Evidence for paternal leakage in hybrid peridical cicadas (Hemiptera: Magicicada spp.). PLoS One 2:e892.

Flook PK, Rowell CHF, Gellissen G. (1995). The sequence, organization, and evolution of the Locusta migratoria mitochondrial genome. J Mol Evol 41:928–41.

Frey JE, Frey B. (2004). Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of Thrips tabaci (Thysanoptera: Thripidae): Mitochondrial heteroplasy or nuclear integration? Hereditas 140:92–8.

Gantenbein B, Fet V, Gantenbein-Ritter IA, Balloux F. (2005). Evidence for recombination in scorpion mitochondrial DNA (Scorpionida: Buthidae). Proc R Soc Lond B Biol Sci 272:697–704.

Gellissen G, Bradfield JY, White BN, Wyatt GR. (1983). Mitochondrial DNA sequences in the nuclear genome of a locust. Nature 301: 631–4.

Gellissen G, Michaelis G. (1987). Gene transfer: Mitochondria to nucleus. Ann N Y Acad Sci 503:391–401.

Guindon S, Lethiec F, Duroux P, Gascoel O. (2005). PHYML Online: A webserver for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:557–9.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. (2006). DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA 103:968–71.

Haknani-Covo E, Zeller RM, Martin W. (2010). Molecular poltergeists: Evolution’s misplaced witnesses. Trends Ecol Evol 16:101–7.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. (2004b). Mitochondrial DNA sequence variation of the bumblebee, G. mellifera L. (Hymenoptera: Apidae). J Asia Pac Entomol 12:133–9.

Kondo R, Satta Y, Matsuura ET, Ishiwa H, Takahata N, Chigusa SI. (1990). Incomplete paternal transmission of mitochondrial DNA in Drosophila. Genetics 126:657–63.

Meusel MS, Moritz RFA. (1993). Transfer of paternal mitochondrial DNA during fertilization of honeybee (Apis mellifera L.) eggs.Curr Genet 24:539–43.

Lanave C, Preparata G, Saccone C, Serio G. (1984). A new method for calculating evolutionary substitution rates. J Mol Evol 20:836–93.

Lawrence JA, Millar CD, Imber MJ, Crockett DE, Robins JH, Scofield RP, Taylor GA, et al. (2009). Molecular evidence for the identity of the M. domestica petrel. Mol Ecol 18:3022–34.

Lee H-S, Lee CE. (1984). Taxonomic revision of the Catantopinae from Korea (Orthoptera: Acrididae) I. Podismini. Korean J Entomol 14: 63–80.

Liu N, Huang Y. (2010). Complete mitochondrial genome sequence of Acrida cinerea (Acrididae: Orthoptera) and comparative analysis of mitochondrial genomes in Orthoptera. Comp Funct Genomics 2010;2010:319486.

Lopez JV, Yukhi N, Masuda R, Modii W, O’Brien SJ. (1994). Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 39:174–90.

Ma C, Liu C, Yang P, Kang L. (2009). The complete mitochondrial genomes of two band-winged grasshoppers, Gastrimargus marmoratus and Oedaleus asiaticus. BMC Genomics 10:156.

Magnacca KN, Brown MJF. (2010). Tissue segregation of mitochondrial haptotypes in heteroplasmic Hawaiian bees: Implication for DNA barcoding. Mol Ecol Resour 10:60–8.

Meier R, Kwong S, Vaidya G, Ng PKL. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Syst Biol 55:715–28.

Meyer CP, Paulay G. (2005). DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol 3:2229–38.

Michaels GS, Hauswirth WW, Laipis PJ. (1982). Mitochondrial DNA copy number in bovine oocytes and somatic cells. Dev Biol 94:246–51.

Milligan BG. (1992). Is organelle DNA strictly maternally inherited? Power analysis of a binomial distribution. Am J Bot 79:1325–8.

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R. (2008). Statistical assignment of DNA sequences using Bayesian phylogenetics. Syst Biol 57:295–70.

Perna NT, Kocher TD. (1995). Patterns of nucleotide composition at fourth degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–8.

Posada D, Crandall KA. (1998). MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817–18.

Rand DM, Harrison RG. (1989). Molecular population genetics of mtDNA size variation in crickets. Genetics 121:551–69.

Ratnasingham S, Hebert PDN. (2007). BOLD: The barcode of life data system. Mol Ecol Notes 7:355–64.

Reeck RL. (1993). An improved primer for PCR amplification of mitochondrial DNA in a variety of insect species. Insect Mol Biol 2:89–91.

Rubinoff D, Cameron S, Will K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for “barcoding” identification. J Hered 97:581–94.

Sherengul W, Kondo R, Matsuura ET. (2006). Analysis of paternal transmission of mitochondrial DNA in Drosophila. Genes Genet Syst 81:399–404.

Simmons RB, Weller SJ. (2001). Utility and evolution of cytochrome b in insects. Mol Phylogenet Evol 10:60–8.

Songram O, Sittipraneed S, Klinbunga S. (2006). Mitochondrial DNA diversity and genetic differentiation of the honeybee (Apis cerana) in Thailand. Biochem Genet 44:256–69.
Sun H, Zheng Z, Huang Y. (2010). Sequence and phylogenetic analysis of complete mitochondrial DNA genomes of two grasshopper species *Gomphocerus rufus* (Linnaeus, 1758) and *Primnoa arctica* (Zhang and Jin, 1985) (Orthoptera: Acridoidea). Mitochondrial DNA 21:115–31.

Sunnucks P, Hales DF. (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase 1-11 in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol Biol Evol 13:510–24.

Swofford DL. (2002). PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sunderland, MA: Sinauer Associates.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–80.

Ujvari B, Dowton M, Madsen T. (2007). Mitochondrial DNA recombination in a free-ranging Australian lizard. Biol Lett 3: 189–92.

Virgilio M, Backeljau T, Nevado B, Meyer MD. (2010). Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics 11:206.

Wallace DC. (2005). A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. Annu Rev Genet 39:359–407.

Wallace DC. (2010). Mitochondrial DNA mutations in disease and aging. Environ Mol Mutagen 51:440–50.

Wan X, Kim MJ, Kim I. (2013). Description of new mitochondrial genomes (*Spodoptera litura*, Noctuoidea and *Cnaphalocrocis medinalis*, Pyraloidea) and phylogenetic reconstruction of Lepidoptera with the comment on optimization schemes. Mol Biol Rep 40:6333–49.

White DJ, Wolfe JN, Pierson M, Gemmell NJ. (2008). Revealing the hidden complexities of mtDNA inheritance. Mol Ecol 17:4925–42.

Zhou ZJ, Huang Y, Shi FM, Ye HY. (2009). The complete mitochondrial genome of *Deracantha onos* (Orthoptera: Bradyporidae). Mol Biol Rep 36:7–12.

Zhao L, Zheng ZM, Huang Y, Sun HM. (2010). A comparative analysis of mitochondrial genomes in Orthoptera (Arthropoda: Insecta) and genome descriptions of three grasshopper species. Zool Sci 27:662–72.

Zouros E, Ball AO, Saavedra C, Freeman KR. (1994). A new type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc Natl Acad Sci USA 91:7463–7.