Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue

Luis Anderson Ribeiro Leite[1,2]

[1]*Laboratório de Estudos de Lepidoptera Neotropical, Departamento de Zoologia, Universidade Federal do Paraná – UFPR, CP 19020, CEP 81531-980, Curitiba, PR, Brasil*
[2]*Corresponding author: Luis Anderson Ribeiro Leite, e-mail: madison.2@ig.com.br*

Abstract: Molecular tools have been used in taxonomy for the purpose of identification and classification of living organisms. Among these, a short sequence of the mitochondrial DNA, popularly known as DNA barcoding, has become very popular. However, the usefulness and dependability of DNA barcodes have been recently questioned because mitochondrial pseudogenes, non-functional copies of the mitochondrial DNA incorporated into the nuclear genome, have been found in various taxa. When these paralogous sequences are amplified together with the mitochondrial DNA, they may go unnoticed and end up being analyzed as if they were orthologous sequences. In this contribution the different points of view regarding the implications of mitochondrial pseudogenes for entomology are reviewed and discussed. A discussion of the problem from a historical and conceptual perspective is presented as well as a discussion of strategies to keep these nuclear mtDNA copies out of sequence analyzes.
Keywords: COI, molecular, NUMTs.

Resumo: Ferramentas moleculares têm sido utilizadas para os estudos referentes à identificação e classificação dos organismos vivos. Entre estes, uma curta sequência do DNA mitocondrial, popularmente conhecida como DNA barcoding, tornou-se muito popular. No entanto, a utilidade e confiabilidade dos códigos de barras de DNA têm sido recentemente questionadas porque pseudogenes mitocondriais, cópias não-funcionais do DNA mitocondrial incorporados ao genoma nuclear, foram encontrados em vários táxons. Quando estas sequências parálogas são amplificadas juntamente com o DNA mitocondrial, podem passar despercebidas e acabam sendo analisadas como se fossem seqüências ortólogas. Nesta contribuição objetivou-se revisar e discutir os diferentes pontos de vista sobre as implicações de pseudogenes mitocondriais para entomologia. Discutimos também o problema através de uma perspectiva histórica e conceitual, abordando estratégias para eliminar ou evitar a presença dessas cópias nucleares em meio às sequências funcionais de DNA.
Palavras-chave: COI, molecular, NUMTs.

Introduction

The classification and identification of living organisms, conducted by amateurs and professionals alike, has been classically based on the description and analysis of morphological features. While the general interest in documenting species diversity has grown exponentially over the years, the number of taxonomists and other professionals trained in species identification, such as parataxonomists (Jinbo et al. 2011), has steadily declined. Taking this scenario into account, several researchers have attempted to find different ways to accelerate and facilitate the process of species identification making it accessible to non-specialists.

Much of the recent taxonomic research has focused on the use of molecular tools in the classification and identification of living organisms. Among these efforts, the use of a short stretch of the mitochondrial cytochrome c oxidase subunit I, popularly known as DNA barcode, has received much attention (Hebert et al. 2003a, b, 2004, Janzen et al. 2005, Hajibabaei et al. 2006, 2007, Decaëns & Rougerie 2008, Janzen et al. 2009, Strutzenberger et al. 2010). Some authors are so partial to this technique that they have implied, or suggested, that DNA barcoding is superior to the classical, morphologically-based taxonomy, and that it should substitute morphology in species descriptions and identification as well as in studies trying to ascertain the relationships between them (Packer et al. 2009).

Researchers who question the idea that DNA barcoding is a panacea that will solve all taxonomic problems have argued, among other things, that mitochondrial pseudogenes may lead to an overestimation of the actual species diversity, as well as to unreliable or misleading identifications based on barcoding sequences (Song et al. 2008, Buhay 2009, Hlaing et al. 2009, Hazkani-Covo et al. 2010).

In this contribution different views are compared and some of the problems mitochondrial pseudogenes may cause to insect DNA barcoding are discussed.

DNA Barcoding in Entomology

DNA barcoding, a taxonomic method that uses a short, standardized DNA sequence to identify species, has gained increased attention and acceptance from members of the scientific community interested in documenting the Earths' biodiversity (Hebert et al. 2003a, b, Savolainen et al. 2005, Hajibabaei et al. 2007, Borisenko et al. 2009, Ivanova et al. 2009, Janzen et al. 2009). Since its inauguration in 2004, the *Consortium for the Barcode of Life* – CBOL, managed primarily by the *Canadian Centre for DNA Barcoding* at the *Biodiversity Institute of Ontario*, University of Guelph, Ontario, Canada, has gathered partners from all over the world. Their objective is to build, in less than twenty years, a comprehensive database that will include barcode sequences of all extant eukaryotes (Hajibabaei et al. 2005, Ratnasingham & Hebert 2007, Jinbo et al. 2011).

One of the advantages of DNA barcoding with respect to traditional taxonomy is the speed and low costs involved in gathering and analyzing data (Borisenko et al. 2009, Strutzenberger et al. 2010). The creation of the CBOL's online database (*The Barcode of Life Data System* – BOLD: www.barcodinglife.org) has provided an incentive for numerous researchers to join the barcode initiative. The database is easy to access and provides free storage and retrieval of molecular, morphological and geographical data, besides built-in, integrated analysis tools such as tree reconstructions on the basis of genetic similarity (Ratnasingham & Hebert 2007, Frézal & Leblois 2008).

One of the premises on which DNA barcoding relies on is that the genetic variation among species is greater than the variation within species (Hajibabaei et al. 2007). A single, 648bp long sequence, corresponding to the 5' end o of the mitochondrial cytochrome c oxidase subunit I, is used as a standard, universal marker for all living organisms (Hebert et al. 2003a, b, Ratnasingham & Hebert 2007, Strutzenberger et al. 2010).

The choice of a mitochondrial gene as a universal marker was mostly driven by the fact that the mitochondria is maternally inherited, avoiding problems with recombination. Also, the mitochondrial genome has a high mutation rate when compared with the nuclear genome, which results in high degrees of intra-specific polymorphism and divergence, important in evolutionary studies (Williams & Knowlton 2001, Wheat & Watt 2008, Hlaing et al. 2009).

Several contributions have been made to the taxonomy and systematics of insects using DNA barcoding, particularly in the following orders: Hemiptera (Foottit et al. 2009, Lee et al. 2010, Shufran & Puterka 2011), Diptera (Smith et al. 2006, Ekrem et al. 2007, Rivera & Currie 2009), Hymenoptera (Smith et al. 2005, Sheffield et al. 2009, Smith et al. 2009), Coleoptera (Yoshitake et al. 2008, Raupach et al. 2010, Greenstone et al. 2011), and Trichoptera (Salokannel et al. 2010, Geraci et al. 2011, Zhou et al. 2011). Additionally, a considerable number of articles on lepidopteran DNA barcoding have been produced since the beginning of this century (Hebert et al. 2004, Janzen et al. 2005, Hajibabaei et al. 2006, Hulcr et al. 2007, Bravo et al. 2008, Emery et al. 2009, Wilson 2010, Hausmann et al. 2011). The first animals to be used in the DNA barcoding campaign, and to have their sequences incorporated into the CBOL's database, were insects (Lepidoptera), fish and birds (Ratnasingham & Hebert 2007). The *Consortium for the Barcode of Life* – CBOL: (www.barcodeoflife.org) currently has other campaigns that contribute with DNA barcode data for insects such as bees, mosquitoes, fruit flies (Tephritidae: Diptera), Trichoptera and Lepidoptera.

Several characteristics intrinsic to insects, such as their diversity and the economic and epidemiological relevance of some groups, have made them the main target of DNA barcoding studies. The BOLD system currently stores molecular data on approximately one million exemplars (Table 1). This standard database can be used in studies on the taxonomy, phylogeny, ecology, agriculture and conservation of various groups of organisms (Jinbo et al. 2011).

Several contributions focusing on identification using the mtCOI have proved useful in the detection of cryptic insect species. Some of those cryptic species which were initially almost impossible to separate using morphological characters alone, have had their identities corroborated by other characters in their natural history and even characters in their morphology (Hebert et al. 2004, Janzen et al. 2005, Smith et al. 2006, Pfenninger et al. 2007, Decaëns & Rougerie 2008, Vaglia et al. 2008, Wheat & Watt 2008, Dasmahapatra et al. 2010, Hausmann et al. 2011).

Recent studies have suggested that the barcode sequence may be useful when morphological differences are present in the same species, including cases of sexual dimorphism, different castes, or different stages of development (Miller et al. 2005, Geraci et al. 2011, Jinbo et al. 2011). Other applications of DNA barcoding are: identification of host plants by sequencing the stomach contents or plant tissues left on the outside of an insect's body (Jurado-Rivera et al. 2009); identification of the stomach contents of predators in biological control studies (Greenstone et al. 2005, Greenstone 2006); additional data uncovering trophic relationships (Clare et al. 2009, Hrcek et al. 2011); and finally, population genetics, community ecology and biodiversity inventories (Janzen et al. 2005, Hajibabaei et al. 2006, Lukhtanov et al. 2009, Craft et al. 2010).

According to Jinbo et al. (2011), DNA barcoding may be used in the future in official protocols for the identification of insects and

other groups, not as a competitor against traditional taxonomy, but as a strong tool to assist in the discovery and description of new taxa. The projected growth of databases such as the BOLD system, which are capable to integrate morphological, physiological and ecological data, strengthen and give respectability to the method.

Mitochondrial Pseudogenes

Pseudogenes, also known as nuclear mitochondrial DNA (*NUMTs*), are non-functional copies of mitochondrial sequences that have become incorporated into the nuclear genome. The transfer of mitochondrial genes to the nuclear DNA may happen through direct transfer, or may be mediated by RNAs, in which case viral elements are believed to participate (Williams & Knowlton 2001, D'Errico et al. 2004, Frézal & Leblois 2008, Song et al. 2008). According to Strugnell & Lindgren (2007), when transferred to the nucleus, the mitochondrial gene loses its original function, and is free to accumulate mutations, even though the mutation rate of the nuclear DNA is slower than that of the mitochondrial genome.

Pseudogenes have been long known to occur in prokaryotes, where they usually originate when errors in the transcription process

cause a gene to "die". A pseudogene is structurally similar to the stretch of DNA it originates from, but may lack a start codon, have duplicated termination codons, and/or abnormal regulatory sequences on either end. For this reason, unlike functional genes, pseudogenes cannot be translated into functional proteins (D'Errico et al. 2004, Gerstein & Zheng 2006).

Pseudogenes have been detected in several eukaryotes; they vary in number, size and abundance (Bensasson et al. 2001a, Richly & Leister 2004, Timmis et al. 2004, Arthofer et al. 2010). In humans, for instance, *NUMTs* are very common, and five of them are known to cause diseases (Hazkani-Covo et al. 2010).

Besides mitochondrial pseudogenes, two other types of pseudogenes exist: processed and unprocessed. The former are copied from RNA and are not found in the same chromosome they originated from. They lack introns and regulatory sequences. Unprocessed pseudogenes, by contrast, can be found in the same chromosome where they originated, and may have introns and regulatory sequences, just as a functional gene (D'Errico et al. 2004).

NUMTs may originate anywhere in the mitochondrial DNA, and may occur as a unique copy in different parts of the genome. These fragments are usually less than 1kb long, but longer fragments seem to be common in mammals (Bensasson et al. 2001a, Richly & Leister 2004, Arthofer et al. 2010).

According to Sorenson & Quinn (1998), even though *NUMTs* are very similar to their source DNA sequences, they have various degrees of functionality; because they are located in different parts of the cell and away from their origin, they are subject to different evolutionary pressures. This is why misleading conclusions can be reached when pseudogenes are unknowingly included in analyzes of mitochondrial sequences. According to Hazkani-Covo et al. (2010), the generation of *NUMTs* is an important evolutionary process in continuous development.

History of the Relationship Between *NUMTs* and Insects

The first record of a mitochondrial pseudogene in Metazoa was for *Locusta migratoria* (Linnaeus, 1758) (Orthoptera : Acrididae) (Gellissen et al. 1983), when sequences homologous to stretches of the mitochondrial DNA were found in the nuclear genome.

After the first discovery of *NUMTs* in Orthoptera, thirteen years passed until an important contribution involving pseudogenes was published in insect molecular research. In 1996, Sunnucks & Hales reported on numerous transpositions of mitochondrial sequences similar to the cytochrome oxidase I and II in *Sitobion* Mordvilko, 1914 (Hemiptera : Aphididae). According to the authors, the non-mitochondrial copies of at least three species seemed to have originated even before transposition. In the same year, Zhang & Hewitt (1996) detected highly conserved pseudogenes in the nucleus of *Schistocerca gregaria* (Forskal, 1775) (Orthoptera : Acrididae) that had been amplified along with authentic mitochondrial sequences. They observed that pseudogene amplification seemed more common when the source specimen had been preserved dry for longer periods of time, and without drawing further conclusions on that comment, suggested that researchers should seek for *NUMTs* when conducting population biology research using the mitochondrial DNA as a marker in order to avoid potentially misleading evidence.

Five years after the works mentioned above, Bensasson et al. (2001a) found pseudogenes in all 10 species of Acrididae (Orthoptera) studied, distributed in four subfamilies (Podisminae, Calliptaminae, Cyrtacanthacridinae and Gomphocerinae). Until then, grasshoppers were among the groups believed to harbor a great number of pseudogenes. According to Bensasson et al. (2001a), at least in

Table 1. Data obtained from the (*The Barcode of Life Data System* – BOLD: www.barcodinglife.org) relating to the numbers of specimens and species with data in the DNA Barcoding in the database. Date of access: October 26, 2011.

Order	Specimens with DNA Barcodes on BOLD system	Species with DNA Barcodes on BOLD system
Lepidoptera	561.713	64.197
Hymenoptera	134.151	17.099
Diptera	108.679	9.377
Coleoptera	33.344	8.304
Trichoptera	27.731	3.779
Hemiptera	23.285	3.745
Ephemeroptera	9.235	614
Orthoptera	5.218	840
Odonata	4.669	367
Plecoptera	3.863	464
Thysanoptera	1.857	137
Neuroptera	1.714	147
Megaloptera	1.152	112
Isoptera	826	197
Blattaria	684	125
Phthiraptera	624	85
Psocoptera	395	3
Mantodea	374	150
Dermaptera	140	11
Phasmatodea	101	25
Archaeognatha	81	4
Siphonaptera	158	12
Mecoptera	49	27
Embioptera	20	11
Raphidioptera	15	5
Thysanura	14	3
Diplura	10	4
Strepsiptera	9	7
Mantophasmatodea	2	1
Grylloblattodea	1	1
Total	920.114	109.853

Orthoptera, the evolution of *NUMTs* seems to involve two steps. First, horizontal transfer, which is the simple transposition of mitochondrial DNA to the nucleus; second, post-transfer replication in the nucleus, which allows for the continuation of the pseudogene. According to the summary compiled by Bensasson et al. (2001b), pseudogenes have been found in a total of 82 species of eukaryotes, corresponding to approximately 21 species of insects, particularly in Orhtoptera and Hemiptera. The great majority of reports on *NUMTs* in animals have been for vertebrates (Blanchard & Schmidt 1996, Sorenson & Quinn 1998, Bensasson et al. 2001b). Other important publications on insect molecular studies appeared in the beginning of the twenty-first century, reporting on the discovery of mitochondrial pseudogenes. For instance, Harrison et al. (2003) located about 100 pseudogenes in *Drosophila melanogaster* Meigen, 1830 (Diptera: Drosophilidae). Richly & Leister (2004) found pseudogenes in 13 eukaryote species, including *Drosophila melanogaster*, but failed to find any in *Anopheles gambiae* Giles, 1926 (Diptera : Culicidae). The variation in the abundance of *NUMTs* in closely related species in that study, when compared with the variation found for other eukaryotes, was explained as a function of two factors. First, among-species differences in the rate of sequence transfer from the mitochondrial to the nuclear DNA; and second, among-species differences in the rate of loss of *NUMTs* in the nucleus. These conclusions were corroborated by a similar study by D'Errico et al. (2004), who utilized only *D. melanogaster* as a representative of Hexapoda.

In 2006, Brower re-evaluated data on *Astraptes fulgerator* (Walch, 1775) (Lepidoptera: Hesperiidae) and found *NUMTs* among the barcode sequences published by Hebert et al. (2004). Later, Pamilo et al. (2007) searched for pseudogenes in four insect species, *D. melanogaster*; *A. gambiae*; *Apis mellifera* Linnaeus, 1758 (Hymenoptera : Apidae) and *Tribolium castaneum* (Herbst, 1797) (Coleoptera : Tenebrionidae), and suggested that the rate of transfer of mitochondrial genes to the nuclear DNA in *Apis mellifera* and *Tribolium castaneum* is high with respect to the dipterans sampled. After analyzing the number of *NUMTs* (>2000) and the relationship between the number of base-pairs transferred per 1 Kb of nuclear sequence (>1.0) in their samples, they also concluded that *A. mellifera* has the greatest number of *NUMTs* in the animal kingdom. Hlaing et al. (2009) searched for *NUMTs* in the genome of *Aedes aegypti* (Linnaeus, 1762) (Diptera: Culicidae), *D. melanogaster* and *A. gambiae*, concluding that the first species has more *NUMTs* than the other two. They also concluded that many of the *NUMTs* detected had originated more recently and for that reason they were difficult to distinguish from their functional counterparts. The authors suggested that similar cases might pose a great problem for DNA barcoding.

Hazkani-Covo et al. (2010) studied sequences of 85 eukaryotes in search for pseudogenes, which they referred to as "molecular Poltergeists". They found *NUMTs* in 72 species, absent from the study of Richly & Leister (2004). One of their new records was *Bombyx mori* (Linnaeus, 1758) (Lepidoptera: Bombycidae) with 0.0016% of the nuclear genome composed of *NUMTs* (the record for Metazoa, *Apis mellifera*, is 0.081%). Magnacca & Brown (2010) found barcode-like pseudogenes in *Hylaeus* Fabricius, 1793 (Hymenoptera : Colletidae), which were easily distinguished from their functional counterparts in their nucleotide sequences and translated amino-acids.

Implications of *NUMTs* for Entomological Studies using DNA Barcoding

Using the subunit I of the cytochrome c oxidase, Hebert et al. (2004) divided *Astraptes fulgerator* (Lepidoptera : Hesperiidae) into ten different species, most of which (six to seven species) were

corroborated by morphological, ethological and ecological evidence. The remaining (cryptic) species were defined based solely on their barcode sequences. Even though the authors did not rule out the possibility that pseudogenes were a problem, they stressed that only 2.8% of their sequences were likely to have been amplified from *NUMTs*.

The contribution by Hebert et al. (2004) was criticized by Brower (2006) and Song et al. (2008), who suggested that the number of cryptic species had been overestimated because mitochondrial pseudogenes of *Astraptes fulgerator* had been amplified by the universal primers used by Herbert and collaborators. The critique is based on the fact that when paralogous genes are used in the place of orthologous ones, the assumptions of phylogenetic reconstructions are violated, leading to erroneous reconstructions.

In a study involving DNA barcode sequences as well as pseudogenes of various orthopterans, Song et al. (2008) concluded that the presence of *NUMTs* in their analysis led to an overestimation of the number of species. Even though they expressed some pessimism that *NUMTs* can be completely eliminated, they suggested some strategies to help identify these alien sequences: search for ambiguity among sequences, noise, or double peaks in the electropherogram or chromatogram (Figure 1); sequence translation in search for additional termination codons and the comparison of the amplified sequences with other, published sequences from closely related species.

Hebert et al. (2004) suggested sequencing freshly collected specimens (preserved for less than 10 years) and using reverse transcriptase followed by PCR (RT-PCR) to prevent pseudogene amplification, particularly for taxa known to carry *NUMTs*. Even though they excluded 13 sequences from their analysis because their electropherogram revealed double peaks, they failed to mention other strategies they might have used to look for *NUMTs*, for instance, searching for additional termination codons. Later, Ratnasingham & Hebert (2007) declared that all sequences submitted to the BOLD system are scrutinized with various tools in search for abnormalities typical of pseudogenes, including the search for termination codons and translation into protein for comparison with the cytochrome c oxidase I product.

Zhang & Hewitt (1996) compared the nuclear copies of the mtDNA to mitochondrial heteroplasmy (the presence of more than one type of mtDNA within cells). The latter, a common cause of degenerative diseases, also causes trouble in sequence analysis. The authors mentioned some strategies that can be used to avoid pseudogene contamination in molecular data, such as the use of specific primers, search for well delimited peaks in the chromatograms and termination codons, comparison with other sequences, and re-sampling when radically divergent sequences are found, or contradictory topologies are recovered in an analysis. An additional strategy, mentioned by Calvignac et al. (2011) to avoid *NUMTs* is to evaluate the rate of evolution of potential pseudogenes as they tend to evolve faster with respect to their paralogous counterparts.

Moulton et al. (2010) tested the strategy of using specifically designed primers to amplify the barcode region as a means to avoid co-amplification of *NUMTs* in 11 species of Orthoptera. Their results casted more doubts on the barcode method, because the use of specific primers only eliminated *NUMTs* from sequences of one species, and merely reduced the amount of amplified pseudogenes from the others. Several of the pseudogenes found lacked termination codons, a determining factor that makes their identification difficult. Moulton and collaborators regarded the presence of *NUMTs* as a challenge to insect DNA barcoding, and suggested that a lot more control on sequence quality needs to be exerted, and further studies

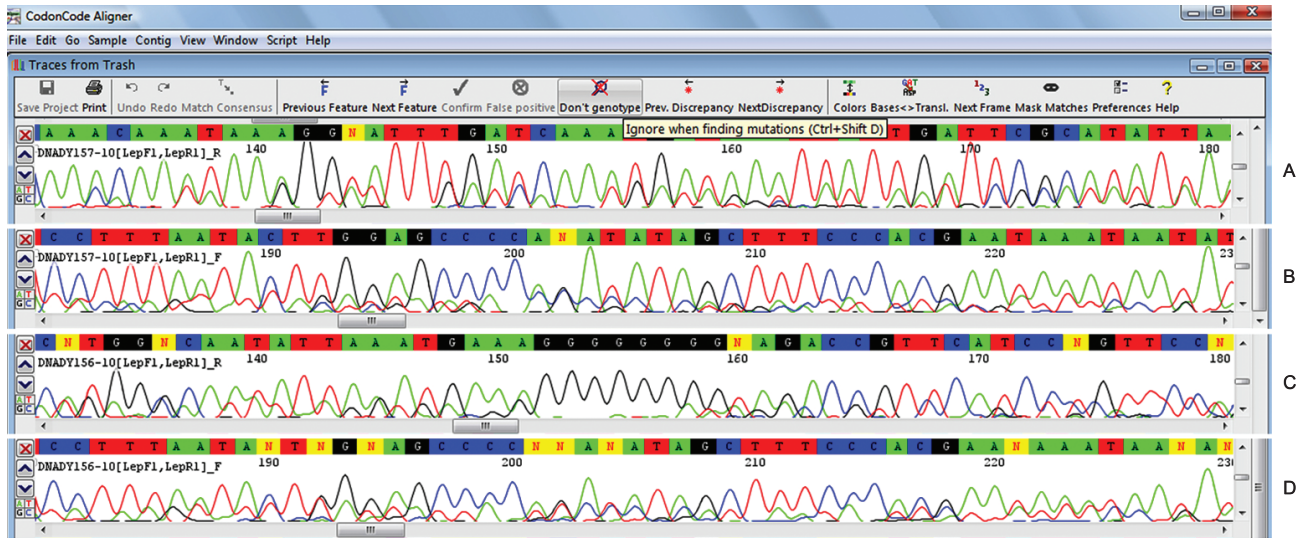Mitochondrial pseudogenes in insect DNA barcoding

Figure 1. Modified chromatogram from (**CodonCode Aligner v.3.0.1 copyright© 2002-2009**) related to the COI project with the genus *Dynamine* Hübner, [1819] (Lepidoptera : Nymphalidae). A-D. Sequences of *Dynamine myrrhina* (Doubleday, 1849). A, C. 3'--- 5'. B, D. 5'--- 3'.
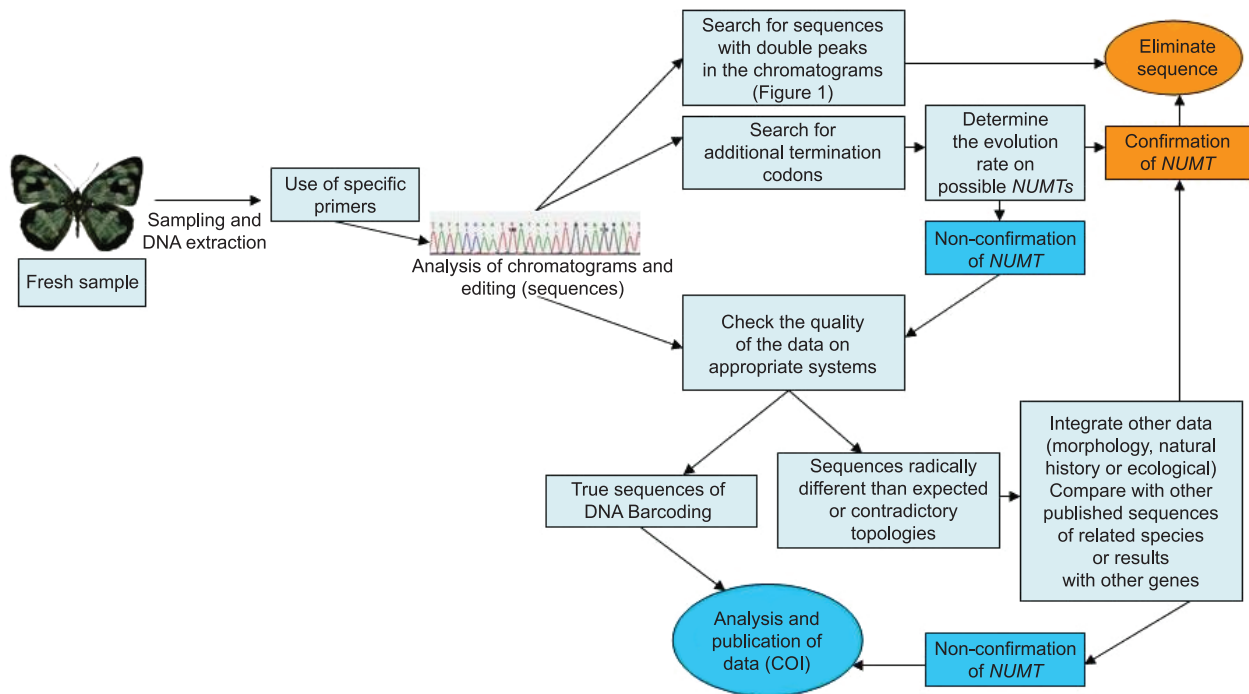

Figure 2. Suggested steps for the future studies with DNA Barcoding in insects seeking the elimination or reduction of the presence of NUMTs in sequences.

on the ubiquity of *NUMTs* need to be conducted, if the COI is to be used as a universal marker.

Conclusions

Based on all previously published data and discussions about mitochondrial pseudogenes in DNA barcoding, we conclude that, if ignored, *NUMTs* pose a major problem for taxonomic and phylogenetic studies based exclusively on barcode sequences. Increased control on submission sequences, amplification from fresh material, use of specific primers, careful analysis of chromatograms, and comparison with other sequences should be mandatory to reduce the risk of contamination with *NUMTs*. In other words, the barcode protocol needs to be adjusted to accommodate for the new information regarding the prevalence of pseudogenes.

Different steps must be prioritized in future studies using DNA barcoding (Figure 2), as previously suggested by others (Zhang & Hewitt 1996, Song et al. 2008, Calvignac et al. 2011), and methodologies must be specified in such a manner as to allow other researchers to make inferences on the reliability of each dataset. Sequences that resemble a pseudogene should be removed as early as possible, beginning with chromatogram analysis in search for suspicious peaks.

Since the barcode sequence was proposed as a universal marker by Hebert et al. (2003a, b), the number of entomological studies using it have grown exponentially. However, despite the problems discussed in the present study, the majority of them fail to mention the possibility of contamination with pseudogenes in their data, or have neglected to use methodologies aimed to mitigate the problem (Janzen et al. 2005, Hajibabaei et al. 2006, Craft et al. 2010, Dasmahapatra et al. 2010). There are two possible explanations for this behavior: ignorance regarding the prevalence of pseudogenes and/or hurry to publish. Fewer studies using COI in insect taxonomy, however, are more reliable despite the fact that they fail to mention *NUMTs*, because they also use morphological and/or ecological information (Burns et al. 2007, Decaëns & Rougerie 2008) as corroborating evidence.

The DNA Barcoding revolution has introduced a strong tool to aid in the taxonomy and phylogenetic systematics, being particularly useful in pairing individuals of different sexes and uncovering cryptic species, which are very important to understand our biodiversity. However, it should not be treated as a substitute for any technique, nor should it be used, along with their programs and system (BOLD), as the only source of evidence in the place of morphological, ecological and natural history evidence, as it has been the case in some studies. The COI is simply a tool that provides additional evidence, and therefore should be treated as such.

More studies should be conducted in order to understand the prevalence of mitochondrial pseudogenes in the various insect orders. Existing data to date are not very informative and only report on the presence and quantity of *NUMTs* in some species within a few orders. The pseudogenes are definitely important contaminants in molecular studies using DNA barcoding, and should be searched for, analyzed, and disposed of when detected.

References

ARTHOFER, W., AVTZIS, D.N., RIEGLER, M. & STAUFFER, C. 2010. Mitochondrial phylogenies in the light of pseudogenes and *Wolbachia*: re-assessment of a bark beetle dataset. Zookeys 56:269-280. http://dx.doi.org/10.3897/zookeys.56.531

BENSASSON, D., ZHANG, X., HARTL, D.L. & HEWITT, G.M. 2011a. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16:314-321. PMid:11369110.

BENSASSON, D., ZHANG, X. & HEWITT, G.M. 2011b. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Mol. Biol. Evol. 17:406-415.

BLANCHARD, J.L. & SCHMIDT, G.W. 1996. Mitochondrial DNA Migration Events in Yeast and Humans: Integration by a Common End-joining Mechanism and Alternative Perspectives on Nucleotide Substitution Pattern. J. Mol. Evol. 13:537-548. PMid:8754225.

BORISENKO, A.V., SONES, J.E. & HEBERT, P.D.N. 2009. The front-end logistics of DNA barcoding: challenges and prospects. Mol. Ecol. Resour. 9:27-34. http://dx.doi.org/10.1111/j.1755-0998.2009.02629.x

BRAVO, J.P., SILVA, J.L.C., MUNHOZ, R.E.F. & FERNANDEZ, M.A. 2008. DNA barcode information for the sugar cane moth borer *Diatraea saccharalis*. Genet. Mol. Res. 7:741-748. PMid:18767242.

BROWER, A.V.Z. 2006. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). Syst. Biodivers. 4:127-132. http://dx.doi.org/10.1017/S147720000500191X

BUHAY, J.E. 2009. 'COI-like' sequences are becoming problematic in molecular systematic and DNA barcoding studies. J. Crustacean Biol. 29:96-110. http://dx.doi.org/10.1651/08-3020.1

BURNS, J.M., JANZEN, D.H., HAJIBABAEI, M., HALLWACHS, W. & HEBERT, P.D.N. 2007. DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. J. Lep. Soc. 61:138-153.

CALVIGNAC, S., KONECNY, L., MALARD, F. & DOUADY, C.J. 2011. Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). Mitochondrion. 11:246-254. PMid:21047564.

CLARE, E.L., FRASER, E.E., BRAID, H.E., FENTON, M.B. & HEBERT, P.D.N. 2009. Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. Mol. Ecol. 18:2532-2542. PMid:19457192.

CRAFT, K.J., PAULS, S.U., DARROW, K., MILLER, S.E., HEBERT, P.D.N., HELGEN, L. E., NOVOTNY, V. & WEIBLEN, G.D. 2010. Population genetics of ecological communities with DNA barcodes: an example from New Guinea Lepidoptera. P. Natl. Acad. Sci. 11:5041-5046. http://dx.doi.org/10.1073/pnas.0913084107

DASMAHAPATRA, K.K., ELIAS, M., HILL, R.I., HOFFMAN, J.I. & MALLET, J. 2010. Mitochondrial DNA barcoding detects some species that are real, and some that are not. Mol**.** Ecol**.** Resour. 10:264-273. http://dx.doi.org/10.1111/j.1755-0998.2009.02763.x

D'ERRICO, I., GADALETA, G. & SACCONE, C. 2004. Pseudogenes in metazoa: Origin and features. Brief. Funct. Gen. Prot. 3:157-167. PMid:15355597.

DECAËNS, T. & ROUGERIE, R. 2008. Descriptions of two new species of Hemileucinae (Lepidoptera: Saturniidae) from the region of Muzo in Colombia - evidence from morphology and DNA barcodes. Zootaxa. 1944:34-52.

EKREM, T., WILLASSEN, E. & STUR, E. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol. Phylogenet. Evol. 43:530-542. http://dx.doi.org/10.1016/j.ympev.2006.11.021

EMERY, V.J., LANDRY, J.F. & ECKERT, C.G. 2009. Combining DNA barcoding and morphological analysis to identify specialist floral parasites (Lepidoptera:Coleophoridae: Momphinae: Mompha). Mol. Ecol. Resour. 9:217-223. http://dx.doi.org/10.1111/j.1755-0998.2009.02647.x

FOOTTIT, R.G., MAW, H.E.L., HAVILL, N.P., AHERN, R.G. & MONTGOMERY, E. 2009. DNA barcodes to identify species and explore diversity in the Adelgidae (Insecta: Hemiptera: Aphidoidea). Mol. Ecol. Resour. 9:188-195. PMid:21564978.

FRÉZAL, L. & LEBLOIS, R. 2009. Four years of DNA barcoding: Current advances and prospects. Infect. Genet. Evol. 8:727-736. PMid:18573351.

GERACI, C.J., MOHAMMED, A.A. & ZHOU, X. 2011. DNA barcoding facilitates description of unknown faunas: a case study on Trichoptera in the headwaters of the Tigris River, Iraq. J. N. Am. Benthol. Soc. 30:163-173. http://dx.doi.org/10.1899/10-011.1

GELLISEN, G., BRADFIELD, J.Y., WHITE, B.N. & WYATT, G.R. 1983. Mitochondrial DNA sequences in the nuclear genome of a locust. Nature. 301:631-634. http://dx.doi.org/10.1038/301631a0

GERSTEIN, M. & ZHENG, D. 2006. The real life of pseudogenes. Sci. Am. 295:48-55. PMid:16866288.

GREENSTONE, M.H., ROWLEY, D.L., HEIMBACH, U., LUNDGREN, G., PFANNENSTIEL, R.S. & REHNER, S.A. 2005. Barcoding generalist predators by polymerase chain reaction: carabids and spiders. Mol. Ecol. 14:3247-3266. PMid:16101789.

GREENSTONE, M.H. 2006. Molecular methods for assessing insect parasitism. B. Entomol. Res. 96:1-13. http://dx.doi.org/10.1079/BER2005402

GREENSTONE, M.H., VANDENBERG, N.J. & HU, J.H. 2011. Barcode haplotype variation in north American agroecosystem lady beetles (Coleoptera: Coccinellidae). Mol. Ecol. Resour. 11:629-637. http://dx.doi.org/10.1111/j.1755-0998.2011.03007.x

HAJIBABAEI, M., DEWAARD, J.R., IVANOVA, N.V., RATNASINGHAM, S., DOOH, R.T., KIRK, S.L., MACKIE, P.M. & HEBERT, P.D.N. 2005. Critical factors for assembling a high volume of DNA barcodes. Philos T. R. Soc. B. 360:1959-1967. PMid:16214753.

HAJIBABAEI, M., JANZEN, D.H., BURNS, J.M., HALLWACHS, W. & HEBERT, P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. P. Natl. Acad. Sci. USA. 103:968-971. http://dx.doi.org/10.1073/pnas.0510466103

HAJIBABAEI, M., SINGER, G.A.C., HEBERT, P.D.N. & HICKEY, D.A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23:167-172. PMid:17316886.

HARRISON, P.M., MILBURN, D., ZHANG, Z., BERTONE, P. & GERSTEIN, M. 2003. Identification of pseudogenes in the *Drosophila melanogaster* genome. Nucleic Acids Res. 31:1033-1037. PMid:12560500.

HAUSMANN, A., HASZPRUNAR, G. & HEBERT, P.D.N. 2011. DNA Barcoding the Geometrid Fauna of Bavaria (Lepidoptera): Successes, Surprises, and Questions. PLoS ONE. 6:e17134. http://dx.doi.org/10.1371/journal.pone.0017134

HAZKANI-COVO, E., ZELLER, R. M. & MARTIN, W. 2010. Molecular poltergeists: Mitochondrial DNA copies in sequenced nuclear genomes. PLoS Genet. 6:1-11. http://dx.doi.org/10.1371/journal.pgen.1000834

HEBERT, P.D.N., CYWINSKA, A., BALL, S.L. & DEWAARD, J.R. 2003a. Biological identifications through DNA barcodes. P. Roy. Soc. Lond. B. 270:313-321. http://dx.doi.org/10.1098/rspb.2002.2218

HEBERT, P.D.N., RATNASINGHAM, S. & DEWAARD, J.R. 2003b. Barcoding animal life:cytochrome c oxidase subunit 1 divergences among closely related species. P. Roy. Soc. Lond. B. 270:S96-S99. http://dx.doi.org/10.1098/rsbl.2003.0025

HEBERT, P.D.N., PENTON, E.H., BURNS, J.M., JANZEN, D.H. & HALLWACHS, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. P. Natl. Acad. Sci. USA. 101:14812-14817. http://dx.doi.org/10.1073/pnas.0406166101

HLAING, T., TUN-LIN, W., SOMBOON, P., SOCHEAT, D., SETHA, T., MIN, S., CHANG, M.S. & WALTON, C. 2009. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. BMC Genet. 10:11. http://dx.doi.org/10.1186/1471-2156-10-11

HRCEK, J., MILLER, S.E., QUICKE, D.L.J. & SMITH, M.A. 2011. Molecular detection of trophic links in a complex insect host-parasitoid food web. Mol. Ecol. Resour. 11(5):786-94. http://dx.doi.org/10.1111/j.1755-0998.2011.03016.x

HULCR, J., MILLER, S.E., SETLIFF, G.P., DARROW, K., MUELLER, M.D., HEBERT, P.D.N. & WEIBLEN, G.D. 2007. DNA barcoding confirms polyphagy in a generalist moth, *Homona mermerodes* (Lepidoptera: Tortricidae). Mol. Ecol. Notes 7:549-57. http://dx.doi.org/10.1111/j.1471-8286.2007.01786.x

IVANOVA, N.V., BORISENKO, A.V. & HEBERT, P.D.N. 2009. Express barcodes: racing from specimen to identification. Mol. Ecol. Resour. 9:35-41. http://dx.doi.org/10.1111/j.1755-0998.2009.02630.x

JANZEN, D.H., HAJIBABAEI, M., BURNS, J.M., HALLWACHS, W., REMIGIO, E. & HEBERT, P.D.N. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philo. T. Roy. Soc. B. 360:1835-45. http://dx.doi.org/10.1098/rstb.2005.1715

JANZEN, D.H., HALLWACHS, W., BLANDIN, P., BURNS, J.M., CADIOU, J-M., CHACON, I., DAPKEY, T., DEANS, A.R., EPSTEIN, M.E., ESPINOZA, B., FRANCLEMONT, J.G., HABER, W.A., HAJIBABAEI, M., HALL, J.P.W., HEBERT, P.D.N., GAULD, I.D., HARVEY, D.J., HAUSMANN, A., KITCHING, I.J., LAFONTAINE, D., LANDRY, J-F., LEMAIRE, C., MILLER, J.Y., MILLER, J.S., MILLER, L., MILLER, S.E., MONTERO, J., MUNROE, E., GREEN, S.R., RATNASINGHAM, S., RAWLINS, J.E., ROBBINS, R.K., RODRIGUEZ, J.J., ROUGERIE, R., SHARKEY, M.J., SMITH, M.A., SOLIS, M.A., SULLIVAN, J.B., THIAUCOURT, P., WAHL, D.B., WELLER, S.J., WHITFIELD, J.B., WILLMOTT, K.R., WOOD, D.M., WOODLEY, N.E. & WILSON, J.J. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Mol. Ecol. Resour. 9:1-26. http://dx.doi.org/10.1111/j.1755-0998.2009.02628.x

JINBO, U., KATO, T. & ITO, M. 2011. Current progress in DNA barcoding and future implications for entomology. Entomol. Sci. 14(2):107-24. http://dx.doi.org/10.1111/j.1479-8298.2011.00449.x

JURADO-RIVERA, J.A., VOGLER, A.P., REID, C.A.M., PETITPIERRE, E. & GÓMEZ-ZURITA, J. 2009. DNA barcoding insect-host plant associations. P. Roy. Soc. B-Biol. Sci. 276:639-648. http://dx.doi.org/10.1098/rspb.2008.1264

LEE, W., KIM, H., LIM, J., CHOI, H-R., KIM, Y., KIM, Y-S., JI, J-Y., FOOTTIT, R.G. & LEE, S. 2010. Barcoding aphids (Hemiptera:Aphididae) of the Korean Peninsula: updating the global data set. Mol. Ecol. Resour. 11(1):32-7. http://dx.doi.org/10.1111/j.1755-0998.2010.02877.x

LUKHTANOV, V.A., SOURAKOV, A., ZAKHAROV, E.V. & HEBERT, P.D.N. 2009. DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. Mol. Ecol. Resour. 9:1302-1310. PMid:21564901.

MAGNACCA, K.N. & BROWN, M.J.F. 2010. Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus* (*Nesoprosopis*) bees (Hymenoptera: Colletidae). BMC Evol. Biol. 10:174. http://dx.doi.org/10.1186/1471-2148-10-174

MILLER, K.B., ALARIE, Y., WOLFE, G.W. & WHITING, M.F. 2005. Association of insect life stages using DNA sequences: the larvae of *Philodytes umbrinus* (Motschulsky) (Coleoptera: Dytiscidae). Syst. Entomol. 30:499-509. http://dx.doi.org/10.1111/j.1365-3113.2005.00320.x

MOULTON, M.J., SONG, H. & WHITING, M.F. 2010. Assessing the effects of primer specificity on eliminating numt coamplification inDNAbarcoding: a case study from Orthoptera (Arthropoda: Insecta). Mol. Ecol. Resour. 10:615-627. PMid:21565066.

PACKER, L., GIBBS, J., SHEFFIELD, C. & HANNER, R. 2009. DNA barcoding and the mediocrity of morphology. Mol. Ecol. Resour. 9:42-50. PMid:21564963.

PAMILO, P., VILJAKAINEN, L. & VIHAVAINEN, A. 2007. Exceptionally high density of NUMTs in the honeybee genome. Mol. Biol. Evol. 24:1340-1346. http://dx.doi.org/10.1093/molbev/msm055

PFENNINGER, M., NOWAK, C., KLEY, C., STEINKE, D. & STREIT, B. 2007. Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. Mol. Ecol. 16(9):1957-68. http://dx.doi.org/10.1111/j.1365-294X.2006.03136.x

RATNASINGHAM, S. & HEBERT, P.D.N. 2007. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol. Ecol. Notes. 7:355-64. http://dx.doi.org/10.1111/j.1471-8286.2007.01678.x

RAUPACH, M.J., ASTRIN, J.J., HANNING, K., PETERS, M.K., STOECKLE, M.Y. & WÄGELE, J-W. 2010. Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. Front. Zool. 7:26. http://dx.doi.org/10.1186/1742-9994-7-26

RICHLY, E. & LEISTER, D. 2004. NUMTs in Sequenced Eukaryotic Genomes. Mol Biol. Evol. 21:1081-1084. http://dx.doi.org/10.1093/molbev/msh110

RIVERA, J. & CURRIE, D.C. 2009. Identification of Nearctic black flies using DNA Barcodes (Diptera: Simuliidae). Mol. Ecol. Resour. 9:224-236. PMid:21564982.

SALOKANNEL, J., RANTALA, M.J. & WHALBERG, N. 2010. DNA-barcoding clarifies species definitions of Finnish *Apatania* (Trichoptera: Apataniidae). Entomol. Fennica 21:1-11.

SAVOLAINEN, V., COWAN, R.S., VOGLER, A.P., RODERICK, G.K. & LANE, R. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos. T. Roy. Soc. B. 360:1805-1811. http://dx.doi.org/10.1098/rstb.2005.1730

SHEFFIELD, K.A., HEBERT, P.D.N., KEVAN, P.G. & PACKER, L. 2009. DNA Barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Mol. Ecol. Resour. 9:196-207. http://dx.doi.org/10.1111/j.1755-0998.2009.02645.x

SHUFRAN, K.A. & PUTERKA, G.J. 2011. DNA Barcoding to Identify All Life Stages of Holocyclic Cereal Aphids (Hemiptera: Aphididae) on Wheat and Other Poaceae. Ann. Entomol. Soc. Am. 104:39-42. http://dx.doi.org/10.1603/AN10129

SMITH, M.A., FISHER, B.L. & HEBERT, P.D.N. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos. T. Roy. Soc. B. 360:1825-1834. http://dx.doi.org/10.1098/rstb.2005.1714

SMITH, M.A., WOODLEY, N.E., JANZEN, D.H., HALLWACHS, W. & HEBERT, P.D.N. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). P. Natl. Acad. Sci. USA. 103:3657-3662. http://dx.doi.org/10.1073/pnas.0511318103

SMITH, M.A, FERNANDEZ-TRIANA, J., ROUGHLEY, R. & HEBERT, P.D.N. 2009. DNA barcode accumulation curves for understudied taxa and areas. Mol. Ecol. Resour. 9:208-216.

SONG, H., BUHAY, J.E., WHITING, M.F. & CRANDALL, K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. P. Natl. Acad. Sci. USA. 105:13468-13491. PMid:21564980.

SORENSON, M.D. & QUINN, T.W. 1998. Numts: a challenge for avian systematics and population biology. The Auk. 115:214-221.

STRUGNELL, J.M. & LINDGREN, A.R. 2007. A barcode of life database for the Cephalopoda? Considerations and concerns. Rev. Fish Biol. Fisher. 17:337-344. http://dx.doi.org/10.1007/s11160-007-9043-0

STRUTZENBERGER, P., BREHM, G. & FIEDLER, K. 2010. DNA barcoding-based species delimitation increases species count of *Eois* (Geometridae) moths in a well-studied tropical mountain forest by up to 50%. Insect Sci. 18:349-362. http://dx.doi.org/10.1111/j.1744-7917.2010.01366.x

SUNNUCKS, P., HALES, D.F. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol. Biol. Evol. 13:510-523.

TIMMIS, J.N., AYLIFFE, M.A., HUANG, C.Y. & MARTIN, W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5:123-135. PMid:14735123.

VAGLIA, T., HAXAIRE, J., KITCHING, I.J., MEUSNIER, I. & ROUGERIE, R. 2008. Morphology and DNA barcoding reveal three cryptic species within the *Xylophanes neoptolemus* and *loelia* species-groups (Lepidoptera: Sphingidae). Zootaxa. 1923:18-36.

WHEAT, C.W. & WATT, W.B. 2008. A mitochondrial-DNA-based phylogeny for some evolutionary-genetic model species of Colias butterflies (Lepidoptera, Pieridae). Mol. Phylogenet. Evol. 47:893-902. PMid:18442929.

WILLIAMS, S.T. & KNOWLTON, N. 2001. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp Genus *Alpheus*. Mol. Biol. Evol. 18:1484-1493.

WILSON, J.J. 2010. Assessing the Value of DNA Barcodes and Other Priority Gene Regions for Molecular Phylogenetics of Lepidoptera. PloS ONE 5:e10525. http://dx.doi.org/10.1371/journal.pone.0010525

YOSHITAKE, H., KATO, T., JINBO, U. & ITO, M. 2008. A new *Wagnerinus* (Coleoptera: Curculionidae) from northern Japan: description including a DNA barcode. Zootaxa. 1740:15-27.

ZHANG, D.X. & HEWITT, G.M. 1996. Nuclear integrations: Challenges for mitochondrial DNA markers. Trends Ecol. Evol. 11:247-251. http://dx.doi.org/10.1016/0169-5347(96)10031-8

ZHOU, X., ROBINSON, J.L., GERACI, C.J., PARKER, C.R., FLINT, J.R.O.S., ETNIER, D.A., RUITER, D., DEWALT, R.E., JACOBUS, L.M., HEBERT, P.D.N. 2011. Accelerated construction of a regional DNA-barcode reference library: caddisflies (Trichoptera) in the great smoky mountains national park. J. N. Am. Benthol. Soc. 30:131-162.