HAL
open science

Nuclear mitochondrial DNA: an Achilles' heel of molecular systematics, phylogenetics, and phylogeographic studies of stingless bees

Maykon Cristiano, Tânia Fernandes-Salomão, Karla Yotoko

To cite this version:

Maykon Cristiano, Tânia Fernandes-Salomão, Karla Yotoko. Nuclear mitochondrial DNA: an Achilles' heel of molecular systematics, phylogenetics, and phylogeographic studies of stingless bees. Apidologie, 2012, 43 (5), pp.527-538. 10.1007/s13592-012-0122-4 . hal-01003644

HAL Id: hal-01003644

https://hal.science/hal-01003644v1

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apidologie (2012) 43:527–538
© INRA, DIB and Springer-Verlag, France, 2012
DOI: 10.1007/s13592-012-0122-4

Original paper

Nuclear mitochondrial DNA: an Achilles' heel of molecular systematics, phylogenetics, and phylogeographic studies of stingless bees

Maykon P. Cristiano[1,2], Tânia M. Fernandes-Salomão[1,2], Karla S. C. Yotoko[3,4]

[1]Laboratório de Biologia Molecular, Departamento de Biologia Geral, Universidade Federal de Viçosa—UFV, Avenida P.H. Rolfs, 36570-000 Viçosa, Minas Gerais, Brazil
[2]Programa de Pós-Graduação em Genética e Melhoramento, Universidade Federal de Viçosa, Avenida P.H. Rolfs, 36570-000 Viçosa, Minas Gerais, Brazil
[3]Laboratório de Bioinformática e Evolução, Departamento de Biologia Geral, Universidade Federal de Viçosa, Avenida P.H. Rolfs, 36570-000 Viçosa, Minas Gerais, Brazil
[4]Programa de Pós-Graduação em Entomologia—Universidade Federal de Viçosa, Avenida P.H. Rolfs, 36570-000 Viçosa, Minas Gerais, Brazil

Received 25 September 2011 – Revised 27 December 2011 – Accepted 16 January 2012

Abstract – Mitochondrial-like DNA (numt) has been found in a variety of insect species. In this work, our objective was to create a phylogeographic hypothesis of *Melipona capixaba* based on the complete *CO*I sequence. However, several inconsistencies were found, such as 1–2-bp-long indels and a stop codon within the putative amino acid sequences. This led us to infer that *M. capixaba* has numts. A phylogenetic analysis which included *CO*I sequences of several species of *Melipona* Illiger, 1806 revealed that numts are also common among other species of the genus. Based on our results, we have proposed a checklist to help to identify the most conspicuous numts; however, it does not ensure that all numts will be identified, since not all numts present explicit signals. Therefore, we recommend taking the maximum care in phylogeographic and phylogenetic analysis within *Melipona* as well as Hymenoptera, since several evidences of numts were found within this order.

numts / cytochrome *c* oxidase subunit I / *Melipona*

1. INTRODUCTION

Molecular characters are the first choice of researchers who study the distribution patterns of genetic diversity in a given taxon. In studies of animal species, most researchers prefer to use mitochondrial DNA (mtDNA) because of certain properties of this genome, including maternal transmission, a high substitution rate, and in most cases the absence of intramolecular genetic recombination (Avise 2009; Emerson and Hewitt 2005).

On the other hand, nuclear sequences present some problems, such as several copies (paralogs) and pseudogenes, either of which can compromise the results. However, many studies have shown that the use of mtDNA sequences can also be compromised by the existence of pseudogenes; in this case, non-translated and non-transcribed DNA regions which were transferred from the mitochondrial to nuclear genome (numts), (Lopez et al. 1994; Hazkani-Covo et al. 2010) during the evolutionary process (Gaziev

Electronic supplementary material The online version of this article (doi:10.1007/s13592-012-0122-4) contains supplementary material, which is available to authorized users.

Corresponding author: K.S.C. Yotoko, karla@ufv.br
Manuscript editor: Klaus Hartfelder

and Shaikhaev 2010; Hazkani-Covo et al. 2010). The primers used to amplify mitochondrial targets may also amplify numts (Bensasson et al. 2001), which can lead to erroneous interpretations because different evolutionary constraints act on functional mitochondrial sequences than nonfunctional nuclear sequences (Sorenson and Quinn 1998; Buhay 2009).

Numt sequences were found in diverse groups of organisms including arthropods, fungi, plants, protozoa, and vertebrates (for details see Bensasson et al. 2001; Richly and Leister 2004; Hazkani-Covo et al. 2010). The complete genomes of several insect species have been sequenced, and numt content has been analyzed for some of them, including *Anopheles gambiae* (Richly and Leister 2004), *Aedes aegypti* (Black and Bernhardt 2009; Haling et al. 2009; Behura et al. 2011), *Culex quinquefasciatus* (Behura et al. 2011), *Drosophila melanogaster* (Richly and Leister 2004), *Tribolium castaneum* (Pamilo et al. 2007), *Nasonia vitripennis* (Viljakainen et al. 2010), and *Apis mellifera* (Behura 2007; Pamilo et al. 2007). Among these species, *A. gambiae* appears to have no numts, and *D. melanogaster* contains only a few numt sequences in its entire genome (Richly and Leister 2004), whereas other insects contain hundreds of numt sequences. For example, *A. mellifera* has the largest proportion of numt sequences (Behura 2007; Pamilo et al. 2007) among all metazoa. Numts have also been reported for insect species whose genomes have not been sequenced, such as grasshoppers (Bensasson et al. 2000; Moulton et al. 2010; Berthier et al. 2011) and ants (Martins et al. 2007).

In this paper, our initial goal was to use an mtDNA gene, cytochrome oxidase subunit I (*CO*I), to study the genetic variability in *Melipona capixaba* Moure and Camargo (1994) (Hymenoptera: Apidae: Meliponini) using a phylogeographic approach. This species, endemic to the Atlantic Rainforest, is restricted to high-altitude areas in Espírito Santo State, Brazil. However, we found several inconsistencies in our results, which led us to consider the existence of numts in *M. capixaba*. To our knowledge, there are no descriptions of numts for Meliponini species. Thus, our new goals were (1) to look for analytical evidence that these inconsistencies are due to the amplification of nuclear copies of *CO*I paralogues (numts) in *Melipona* species, and (2) to detect how common these *numts* are within *Melipona* and their presumed effects on phylogenetic studies.

2. MATERIALS AND METHODS

We sampled 74 *M. capixaba* individuals from seven localities of Espírito Santo State, Brazil. We extracted their DNA and sequenced their complete mitochondrial *CO*I, which is the most commonly used marker for performing phylogeographic studies of insects (Tanaka et al. 2001; Bell et al. 2007; Lohman et al. 2008; Solomon et al. 2008; Batalha-Filho et al. 2010; Berger et al. 2010; Resende et al. 2010). In order to generate two overlapping fragments (initial and final portions) and to ensure the proper sequencing of the whole *CO*I sequence, we used two primer pairs (Batalha-Filho et al. 2010), Table S1, based on the complete mitochondrial genome of *Melipona bicolor* (NC_004529.1), positions 1,571–3,130 (Figure 1).

However, in 72 of 74 samples, the resulting fragments (initial and final) did not truly overlap, because within the putative overlap region (positions 2,178–2,455, Figure 1) several nucleotide substitutions prevented the formation of contigs between the sequences of an individual. In addition, when we tried to translate these fragments, we found a stop codon within the putative amino acid sequences of the final fragment (except for sequence JN315123). Also, most of these sequences contain 1- or 2-bp indels. We therefore analyzed the initial and final portion of *CO*I separately, and our preliminary data analyses (not shown) yielded completely different phylogenetic results. Since we could not identify any evidence of numts in the initial portion of the *CO*I (positions 1,388–2,455), we chose to report only our results regarding the final portion of the *CO*I (primers COX2–COX4, positions 2,178–3,130—Figure 1).

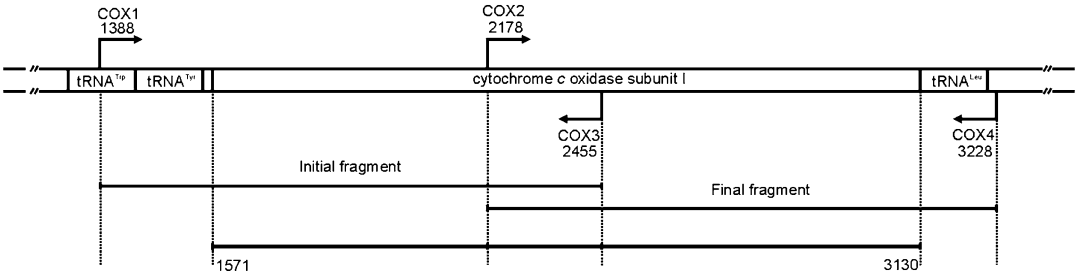

Figure 1. Representation of the amplified fragment and the position of the primers used regarding the entire mitochondrial DNA sequence of *M. bicolor*.

2.1. DNA preparation, amplification, purification, cloning, and sequencing

Total DNA was isolated from each adult using a modified phenol–chloroform protocol (Fernandes-Salomão et al. 2005). A 1,050-bp fragment of the final portion of the COI gene plus the tRNA-Leu gene was amplified using the standard PCR technique. The amplification conditions consisted of a 5 min denaturation at 94°C, followed by 35 cycles of 94°C for 1 min, 52°C for 1 min and 20 s and 64°C for 2 min, with a final extension at 64°C for 10 min. PCR products were cleaned out of non-specific fragments, primers and enzymes using the PureLink™ Quick Gel Extraction Kit (Invitrogen™, Carlsbad, CA, USA). The amplicons were linked into the pGEM-T Vector (Promega™, Madison, WI, USA), and transformed into *Escherichia coli* cells DH5α (Inoue et al. 1990) by heat shock. The recombinant plasmids were recovered with the Miniprep kit (Invitrogen™, Carlsbad, CA, USA) and the target fragment was sequenced using the DYEnamic ET Dye Terminator Cycle Sequencing kit (GE Healthcare, Freiburg, Germany) with the pUC/M13 (Forward and Reverse) primers. Sequencing was performed using an automated sequencing machine (MegaBACE™500, GE healthcare).

2.2. Sequence evaluation and alignment

The resulting chromatograms were evaluated and edited with the program Consed (Gordon et al. 1998). The edited sequences were translated to their putative amino acid sequences and aligned using ClustalW (Thompson et al. 1994). The alignment was corrected manually using the program Mega 5.0 (Tamura et al. 2011), and returned to the nucleotide sequences, which were used in further phylogenetic analyses.

2.3. Phylogenetic data analyses

To determine how common numts are within *Melipona*, we searched GenBank for the final *CO*I portion in other species of this genus. Table I shows the species names and accession numbers in GenBank. Two sequences, from *Melipona rufiventris* Lepeletier, 1836 (EU163147) and *Melipona fuscopilosa* Moure and Kerr, 1950 (EU163136) were identified as *CO*I-like sequences. These species belong to the subgenus *Michmelia* Moure, 1975 (see Camargo and Pedro 2008, for review), which is the same subgenus as *M. capixaba*. We therefore decided to include sequences of six individuals of *Melipona scutellaris* Latreille, 1811, another species of this same subgenus available in our laboratory. Phylogenetic relationships among the final portion of the *CO*I were estimated using Bayesian Inference (Yang and Rannala 1997) with MrBayes 3.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). The program MrModelTest (Nylander 2004) was used to identify the best substitution model for the fragment. Fifteen million generations of Monte Carlo Markov chains with a sampling frequency of 1,000 were run in a Bayesian approach to generate a majority rule consensus tree. The frequencies of the observed bipartitions (i.e., posterior probabilities) that were higher than 50% were shown next to internal branches of the tree.

Table I. *Melipona* sequences used in this study and their accession numbers in GenBank.

Melipona spp.		GenBank accession number
Eomelipona		
	M. amazonica	EU163166
	M. asilvai	EU163157
	M. bicolor (genoma)	AF466146
	M. bicolor bicolor	EU163158
	M. illustris	EU163145
	M. marginata	EU163153
	M. micheneri	EU163139
	M. ogilviei	EU163140
Melikerria		
	M. beecheii	EU163126
	M. compressipes compressipes	EU163137
	M. compressipes interrupta	EU163143
	M. compressipes interrupta	EU163128
	M. grandis	EU163169
	M. grandis	EU163130
	M. grandis	EU163117
	M. quinquefasciata	EU163155
	M. triplaridis	EU163154
Melipona		
	M. favosa	EU163127
	M. mandacaia	EU163156
	M. quadrifasciata anthidioides	EU163131
Michmelia		
	M. aff. costaricensis	EU163134
	M. aff. crinita	EU163133
	M. aff. crinita	EU163164
	M. capixaba	JN315063 and JN315064
	M. captiosa	EU163142
	M. costaricensis	EU163129
	M. costaricensis	EU163121
	M. fuliginosa	EU163141
	M. fuliginosa (*fallax*)	EU163119
	M. fulva	EU163125
	M. illota	EU163167
	M. lateralis	EU163144
	M. melanopleura	EU163113
	M. melanoventer	EU163135
	M. nebulosa	EU163146
	M. panamica	EU163096
	M. rufiventris	EU163132

Table I (continued)

Melipona spp.		GenBank accession number
	M. rufiventris flavolineata	EU163149
	M. rufiventris rufiventris	EU163151
	M. scutellaris	EU163152
	M. scutellaris	From JN315137 to JN315139
	M. seminigra atrofulva	EU163138
	M. solani	EU163160
(Putative Numts)		
Michmelia		
	M. capixaba	From JN315065 to JN315136
	M. fuscopilosa	EU163136
	M. rufiventris flavolineata	EU163147
	M. scutellaris	From JN315140 to JN315142
(Outgroups)		
Cephalotrigona sp.		EU163161
Friesella schrottkyi		EU163103
Frieseomelitta silvestrii		EU163104
Geotrigona kraussi		EU163112
Lestrimelitta danuncia		EU163111
Meliwillea bivea		EU163114
Plebeia franklii		EU163098
Scaptotrigona polysticta		EU163101
Tetragonisca angustula		EU163097
Trigona amalthea		AF214669
Trigona sp.		EU163105

2.4. Sequence comparisons

2.4.1. Putative codon position

Given that coding sequences evolve under different selective constraints than non-coding sequences (e.g., numts), we expect that mitochondrial coding sequences present different substitution rates among the three codon positions, which did not happen with non-coding sequences. Thus, we classified our sequences into mitochondrial sequences (overall *Melipona* sequences) and putative numts found in *M. capixaba* and *M. scutellaris*, put all these sequences in their putative reading frames and calculated the overall *p*-distances, as well as the *p*-distances regarding the first, second and third putative codon positions using the program MEGA 5.0 (Tamura et al. 2011).

2.4.2. Specific sites comparison

Based on the alignment of all sequences (mitochondrial and nuclear) and the phylogenetic hypothesis derived from this alignment, we have identified some sites that could highlight differences between mts (mitochondrial sequences) and numts. For this, we compared the two mts of *M. capixaba* (JN315063 and JN315064) with the 72 putative numts (sequences with premature stop codons within their putative aminoacid sequences and indels of 1 and 2 bp, JN315065–JN315136). To include a given site in this comparison, this site (1) must be reciprocal identical in *M. capixaba* mts and different from most putative numts (more than 60 of the 72 numts); and (2) must be identical to most other *Melipona* mitochondrial sequences. We used the same procedure for the three

mts (JN315137–JN315139) and putative numts (JN315140–JN315142) of *M. scutellaris*, as well as the three mts of *M. rufiventris* (EU163132, EU163149, and EU163151) and the unique *CO*I-like sequence of this species (EU163147), all of which are available in GenBank. After choosing the nucleotide sites to be compared, we also included the *CO*I-like sequence of *M. fuscopilosa* (EU163136), the *Melipona nebulosa* Camargo, 1988 (EU163146) sequence (which grouped with the clade containing the numt sequences—see Figure 2) and the *M. scutellaris* sequence obtained in GenBank (EU163152, which did not cluster with other *M. scutellaris* mitochondrial sequences, Figure 2).

3. RESULTS

3.1. Sequence alignment

The complete alignment of 74 sequences of *M. capixaba*, six sequences of *M. scutellaris*, and 43 sequences of other *Melipona* species, with 11 outgroup species, consists of 688 base pairs, 364 of which are variable. From these 124 sequences, 79 (72 *M. capixaba*, three *M. scutellaris*, one *M. rufiventris*, and one *M. fuscopilosa*) presented at least one evidence of numts.

Our alignment revealed a transversion (C–A) at position 2,766 (based on the complete

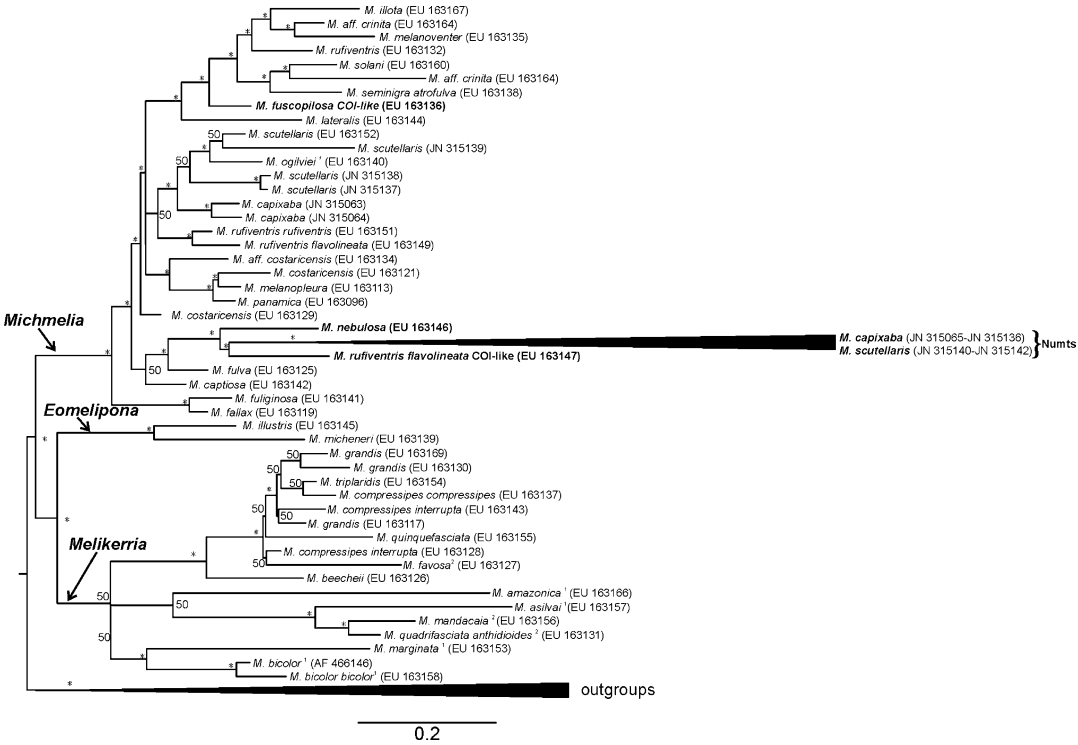

Figure 2. Bayesian phylogenetic consensus tree of 688 bp of the *CO*I and *CO*I-like sequences of the genus *Melipona*, summarized in Table I. The values alongside ancestral nodes correspond to the posterior probabilities (majority rule consensus values) of each node. Asterisks mean P.P. = 100%. Putative numt sequence names are written in bold. Black triangles represent monophyletic groups. The upper triangle contains 72 sequences of *M. capixaba* and three of *M. scutellaris*. The lower triangle contains 11 outgroups sequences. The complete tree is shown in Supplemental Material Fig. S1. The subgenus containing the majority of species in each group is identified in the interior node corresponding to their most recent common ancestor in the tree. The exceptions are represented by superscript numbers: subgenus *Eomelipona* (1) and subgenus *Melipona* (2).

mitochondrial genome of *M. bicolor* Lepeletier, 1836) in several *M. rufiventris* (EU163147), *M. scutellaris*, and *M. capixaba* sequences. This transversion induced a stop codon in the *CO*I-like sequence of each of these species, except for specimen JN315123 of *M. capixaba*.

3.2. Phylogenetic analysis

Figure 2 shows the Bayesian consensus tree. All sequences of the genus *Melipona* were grouped monophyletically. However, considering its division into subgenera, several species were misplaced in our phylogenetic tree. For example, the members of the subgenus *Eomelipona* Moure, 1992 were scattered in the subgenera *Michmelia* (*Melipona ogilviei* Schwarz, 1932) and *Melikerria* Moure, 1992 (*Melipona favosa* Fabricius, 1798, *Melipona amazonica* Schulz, 1905, *Melipona marginata* Lepeletier, 1836 and *M. bicolor*). In addition, the three species of the subgenus *Melipona* appeared within *Melikerria*, in two different groups: *Melipona mandacaia* Smith, 1863 and *Melipona quadrifasciata* Lepeletier, 1836 grouped with *Melipona asilvai* Moure, 1971 (*Eomelipona*) (P.P.=1.0), while *M. favosa* grouped with nine *Melikerria* species (P.P.=1.0).

Our numt sequences of *M. capixaba* and *M. scutellaris*, as well as the *CO*I-like sequences of *M. fuscopilosa* and *M. rufiventris*, were placed within the *Michmelia* clade in two separate groups. The first group included the putative numts of *M. capixaba*, *M. scutellaris*, the *CO*I-like sequence of *M. rufiventris* and the unique sequence of *M. nebulosa* with 100% Bayesian posterior probability. The *CO*I-like sequence of *M. fuscopilosa* (another presumed numt), was placed in the second group, with putative mitochondrial sequences (P.P.=100%).

3.3. Sequence comparisons

3.3.1. Comparison of specific sites

Using the criteria established in Section 2.4.2 of Section 2, we compared a total of 19 alignment sites (Table II), 11 of which are on the first or second putative codon positions. Most of these 19 sites show no variation among mitochondrial fragments of the 41 *Melipona* species included in this study, except for sites 2,359 (third codon position), 2,444 (first), and 2,563 (third), each of which contained one exception; site 2,479 (third) with three exceptions; and site 2,512 (third) with five exceptions.

Among the numt fragments, *M. capixaba* and *M. scutellaris* share 19 sites, 13 of which were also shared by *M. nebulosa* and seven of the 19 were shared by *M. rufiventris*. *M. fuscopilosa*, on the other hand, shares all the 19 sites with most the mt sequences of other *Melipona* species.

3.3.2. Putative codon position

The *p*-distance regarding the codon position among species (Figure 3) revealed that the *Melipona* putative mitochondrial sequences presented the expected pattern of most substitutions found in the third position, followed by the first position and then for the second position. However, the putative numt sequences of *M. capixaba* and *M. scutellaris* revealed no significant differences among the codon positions.

4. DISCUSSION

The initial goal of this work, to study the genetic variability in *M. capixaba* using a complete *CO*I sequence, failed due to several inconsistencies in our results, which led us to hypothesize the existence of numts within *M. capixaba*. Thus, we searched for some analytical evidences of numts in *M. capixaba* and used a phylogenetic approach to infer how common numts are within *Melipona*, and their potential effects on phylogenetic studies. Our results add to the list of Hymenoptera species with mitochondrial-like sequences in their nuclear genomes, which already includes honeybees (Behura 2007; Pamilo et al. 2007) and ants (Martins et al. 2007; Beckenbach 2009).

The first evidence of numts was the lack of overlapping in the final and initial portions of

Table II. Alignment of the 19 sites which differentiate some putative numts from mitochondrial sequences of *M. capixaba* (72 numts and 2 mts), *M. scutellaris* (3 numts and 3 mts), and/or *M. rufiventris* (1 numt and 1 mt). These sites were also evaluated for putative mts of 38 *Melipona* species, as well as putative numts of *M. nebulosa* and *M. fuscopilosa*. The GenBank accession numbers of each sequence are shown in Table I. Capital letters represent the nucleotides found in all or most specimens in a particular site, while lowercase letters represent nucleotides found in one or very few individuals.

		Site Position[1]																		
	N	2229 (2)	2267 (1)	2340 (2)	2359 (3)*	2376 (2)	2393 (1)	2436 (2)	2444 (1)	2455 (3)	2479 (3)	2512 (3)	2542 (3)	2563 (3)	2647 (3)	2670 (2)	2706 (2)	2707 (3)*	2766 (2)**	2781 (2)
Mit. fragments																				
Melipona spp.	38	C	T	T	A/c	T	A	A	A/t	A	T/c	A/t	A	T/c	A	A	T	A	C	A
M. capixaba	02	C	T	T	A	T	A	A	A	A	T	A	A	T	A	A	T	A	C	A
M. scutellaris	03	C	T	T	A	T	A	A	A	A	T	A	A	T	A	A	T	A	C	A
M. scutellaris (EU163152)	01	C	T	T	A	T	A	A	A	A	T	A	A	T	A	A	T	A	C	A
M. rufiventris (EU163149)	01	C	T	T	A	T	A	A	A	A	T	A	A	T	A	A	T	A	C	A
Numt fragments																				
M. capixaba	72	A/c	G/t	C/t	C/a	T	A/t	T/a	G/a/t	G/a	C/t	A	T	C/t/a	A/t	A	T	A/g	A/c	A
M. scutellaris	03	A	G	C	C	T	A	T	G	G	C	A	T	C	A	A	T	A	A	A
M. rufiventris (EU163147)	01	C	G	T	C	G	G	A	A	G	C	T	T	C	T	G	C	G	A	G
M. nebulosa (EU163146)	01	C	T	T	C	T	A	A	G	G	C	A	T	C	A	A	T	A	C	T
M. fuscopilosa (EU163136)	01	C	T	T	A	T	A	A	A	A	T	A	A	T	A	A	T	A	C	A

* Changes the amino acid
** Inserts a stop codon
[1] Site position, based on *M. bicolor* genome complete mitochondrial

*CO*I in 72 of our 74 sequences, which was confirmed by the different phylogenies based on these separated portions (data not shown), which showed two distinct fragments rather than complementary fragments. Berthier et al. (2011) reviewed a list of procedures to avoid the amplification of numts, and one of their recommendations is to use the long-PCR technique (see Bensasson et al. 2001), which could prevent numt amplification but also prevent numt detection.

The second evidence of numt fragment amplification came from the translation of the putative amino acid sequences, which revealed premature putative stop codons, as well as 1–2 bp indels in most sequences of the final portion of the *CO*I (Figure 1). Third, we found no significant substitution rates differences among the "codon positions" of the putative numt sequences, which would be expected within coding sequences. Based on our results, we have summarized a checklist to detect the most conspicuous numt sequences in Figure 4.

The accidental amplification of the *CO*I-like numt sequence in most sequences of *M. capixaba* and *M. scutellaris* with primers COX2–COX4 (Figure 1) may be due to preferential amplification of these sequences. If their genomes show abundant mitochondrial-like sequences, numts sequences will be amplified in higher proportion than mitochondrial sequences in the PCR reaction (Bensasson et al. 2001; Buhay 2009). In many organisms that have numts in their nuclear genomes, mitochondrial-like sequences represent more than 50% of the PCR amplificons (Bensasson et al. 2001). Moreover, primers tend to become more specific to numt sequences than mitochondrial sequences due to the mitochondrial genome's higher rate of evolution compared to nuclear sequences (Bensasson et al. 2001; Beckenbach 2009).

To investigate the incidence of numts within *Melipona*, we performed a phylogenetic analysis which included our complete set of the final portion of the *CO*I sequences, in addition to these sequences of all *Melipona* deposited in GenBank. We found that most conspicuous numt sequences (*M. capixaba*, *M. scutellaris*, and *M. rufiventris*) group with *M. nebulosa* (EU163146) with high statistical support

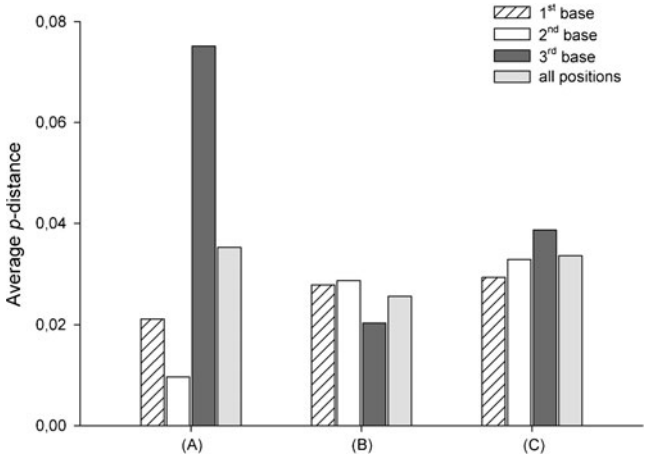

Figure 3. First, second, third, and overall *p*-distances calculated using alignments including (*A*) mitochondrial sequences of *Melipona* species; (*B*) *M. capixaba* putative numt sequences, and (*C*) *M. scutellaris* putative numt sequences.

(Figure 2). We inferred that the sequences of at least three of the four species descended from a unique event of mitochondria to nucleus sequence migration, prior to their diversification. In the 19 nucleotide sites selected to highlight the differences between putative numts and mts, we found that *M. nebulosa* shares more sites (13) with numts of *M. capixaba* and *M. scutellaris* than *M. rufiventris*, which shares only seven sites with these two species. This suggests that the sequence of *M. nebulosa* is also a numt, regardless of the absence of other numt evidences, which would require further sequencing for confirmation.

On the other hand, the *M. fuscopilosa CO*I-like sequence shares no substitutions with any other putative numt sequence (Table II). Indeed, it shares the 19 loci with mts of most *Melipona* species. This could be explained by a two-part amplification of the *CO*I, but contrary to what we found in *M. capixaba* and *M. scutellaris* numt sequences, the initial portion (which contains a stop codon) would have been amplified from the nucleus and the final portion from the mitochondria. However, in addition to the *M. fuscopilosa CO*I-like sequence, this cluster contains a single sequence of *M. rufiventris* (EU163132) and two sequences of *Melipona aff. crinita* Moure and Kerr, 1950 (EU163133 and EU163164) in separated positions. This suggests that at least one of the two sequences of *M. aff. crinita* and *M. rufiventris* sequences should be numts, even though they show no other numt evidences. These potential numts were inferred by the presence of more than one sequence of each species within the phylogeny. Therefore, our data show that the use of a single sequence per species may not be sufficient to obtain good phylogenetic hypotheses.

Moreover, the misplacement of several species among the subgenera suggests that regardless the lack of conspicuous numts evidence in most of them (Figure 4), it is possible that there are more numts within *Melipona* than we were able to detect, since not all numts present explicit signals such as indels or premature stop codons (Berthier et al. 2011). However, these misplacements may also be due to classification problems within the genus *Melipona*, and this possibility needs to be evaluated in future works of the taxonomy of this genus.

Therefore, we suggest that in their future works, the researchers working with phylogenetic and phylogeographic hypothesis of species within *Melipona*, and possibly other Hymenoptera, take the possibility of numts when interpreting and publishing their results. This is because the amplification of an accidental numt may violate the basic assumption of phylogenetic and phylogeographical analysis, which must be based on the comparison of orthologous sequences.

Based on our results, we suggest an initial check for the most conspicuous numts (Figure 4),

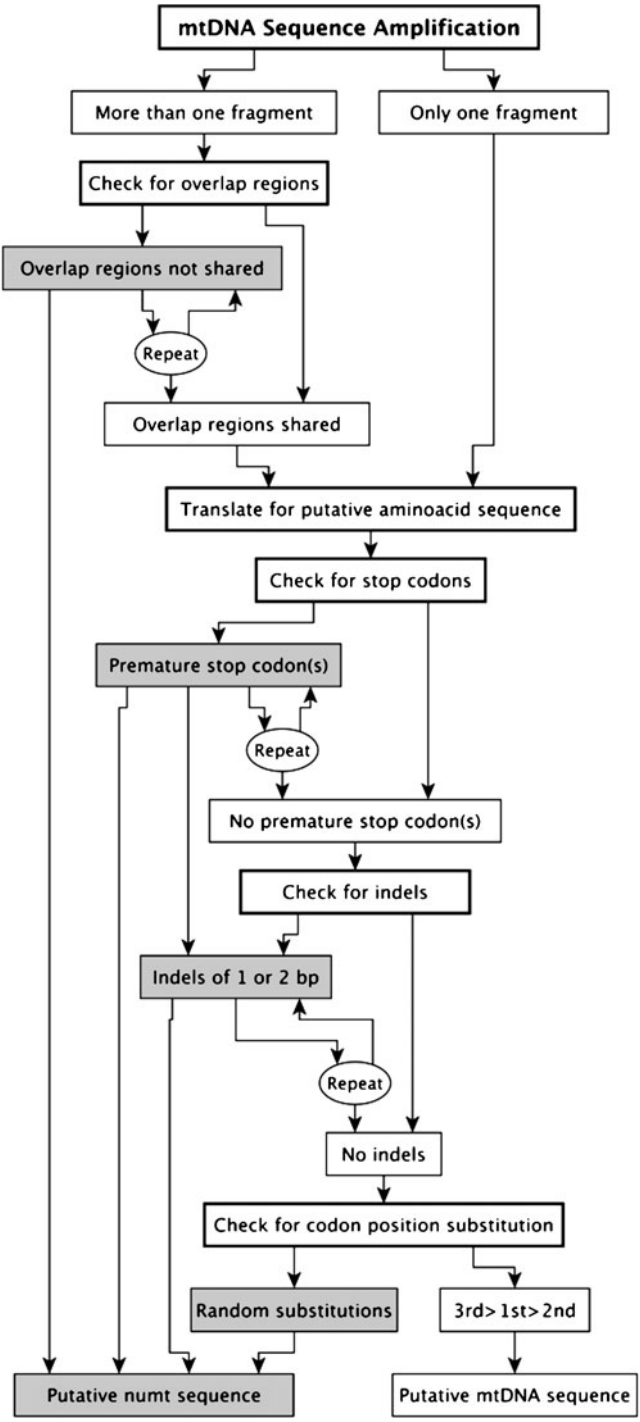

Figure 4. Flow chart showing a checklist to detect the most conspicuous numts. Gray boxes represent diagnostics for numts. Highlighted boxes represent analytical procedures.

which unfortunately does not ensure that all numts will be identified. A complementary strategy to avoid numts is the inclusion of at least two different sequences per species in

phylogenetic studies, preferably generated with different primer pairs, as well as multiple clones from each individual in phylogeographic studies to minimize the consequences of numts. In studies involving molecular identification via mitochondrial barcoding, the consequences could be even worse (Song et al. 2008; Casiraghi et al. 2010; Moulton et al. 2010), and extra care must be taken to properly identify species.

ACKNOWLEDGMENTS

This research was part of the M.Sc. dissertation of MPC who was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, Process numbers: CBB-1331/06) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). We are grateful to Dr. Lucio Antonio de Oliveira Campos for his assistance and encouragement, to the M. Sc. Helder Resende and Íris Stanciola for collecting samples, and to the bee keepers that kindly provided the specimens analyzed.

L'ADN mitochondrial dans le noyau : un talon d'Achille dans la systématique moléculaire, la phylogénie et les études de phylogéographie des abeilles sans aiguillon.

séquences d'ADN mitochondrial insérées / sous-unité I de la cytochrome C oxydase / *Melipona*

Nukleäre mitochondriale DNA: eine Achillesferse für molekulare Systematik, Phylogenie und phylogeographische Studien an Stachellosen Bienen.

Numts / Cytochrome c Oxidase Untereinheit I / *Melipona*

REFERENCES

Avise, J.C. (2009) Phylogeography: retrospect and prospect. J. Biogeogr. **36**, 3–15

Batalha-Filho, B., Waldschmidt, A.N., Campos, L.A.O., Tavares, M.G., Fernandes-Salomão, T.M. (2010) Phylogeography and historical demography of the neotropical stingless bee *Melipona quadrifasciata* (Hymenoptera, Apidae): incongruence between morphology and mitochondrial DNA. Apidologie **41**, 534–547

Beckenbach, A.T. (2009) Numts and mitochondrial pseudogenes. Myrmecol. News **12**, 217–218

Behura, S.K. (2007) Analysis of nuclear copies of mitochondrial sequences in Honeybee (*Apis mellifera*) genome. Mol. Biol. Evol. **24**, 1492–1505

Behura, S.K., Lobo, N.F., Haas, B., deBruyn, B., Lovin, D.D., Shumway, M.F., Puiu, D., Romero-Severson, J., Nene, V., Severson, D.W. (2011) Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect Biochem. Mol. Biol. **41**, 770–777

Bell, K.L., Moritz, C., Moussalli, A., Yeates, D.K. (2007) Comparative phylogeography and speciation of dung beetles from the Australian Wet Tropics rainforest. Mol. Ecol. **16**, 4984–4998

Bensasson, D., Zhang, D.-X., Hewitt, G.M. (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. Mol. Biol. Evol. **17**, 406–415

Bensasson, D., Zhang, D., Hartl, D.L., Hewitt, G.M. (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. **16**, 314–321

Berger, D., Chobanov, D.P., Mayer, F. (2010) Interglacial refugia and range shifts of the alpine grasshopper *Stenobothrus cotticus* (Orthoptera: Acrididae: Gomphocerinae). Org. Divers. Evol. **10**, 123–133

Berthier, K., Chapuis, M.-P., Moosavi, S.M., Tohidi-Esfahani, D., Sword, G.A. (2011) Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. Syst. Entomol. **36**, 285–299

Black IV, W.C., Bernhardt, S.A. (2009) Abundant nuclear copies of mitochondrial origin (numts) in the *Aedes aegypti* genome. Insect Mol. Biol. **18**, 705–713

Buhay, J.E. (2009) "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. J. Crust. Biol. **29**, 96–110

Camargo J.M.F., Pedro, S.R.M. (2008) Meliponini Lepeletier, 1836. In Moure, J.S., Urban, D. and Melo, G.A.R. (Orgs). Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region. Sociedade Brasileira de Entomologia, Curitiba, pp. 272–578 (online version. Available at http://www.moure.cria.org.br/catalogue).

Casiraghi, M., Labra, M., Ferri, E., Galimberti, A., De Mattia, F. (2010) DNA barcoding: a six-question tour to improve users' awareness about the method. Brief. Bioinform. **11**, 440–453

Emerson, B.C., Hewitt, G.M. (2005) Phylogeography. Curr. Biol. **15**, R367–R371

Fernandes-Salomão, T.M., Rocha, R.B., Campos, L.A. O., Araújo, E.F. (2005) The first internal transcribed spacer (ITS-1) of *Melipona* species (Hymenoptera, Apidae, Meliponini): characterization and phylogenetic analysis. Insect. Soc. **52**, 11–18

Gaziev, A.I., Shaikhaev, G.O. (2010) Nuclear mitochondrial pseudogenes. Mol. Biol. **44**, 358–368

Gordon, D., Abajian, C., Green, P. (1998) Consed: a graphical tool for sequence finishing. Genome Res. **8**, 195–202

Haling, T., Tunn-Lin, W., Somboon, P., Scheat, 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

Hazkani-Covo, E., Zeller, R.M., Martin, W. (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. **6**, e1000834

Huelsenbeck, J.P., Ronquist, F. (2001) MrBayes: Bayesian inference of phylogenetic tree. Bioinformatics **17**, 754–755

Inoue, H., Nojima, H., Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene **96**, 23–28

Lohman, D.J., Peggie, D., Pierce, N.E., Meier, R. (2008) Phylogeography and genetic diversity of a widespread Old World butterfly, *Lampides boeticus* (Lepidoptera: Lycaenidae). BMC Evol. Biol. **8**, e301

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W., O'Brien, S.J. (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. **39**, 174–190

Martins Jr., J., Solomon, S.E., Mikheyev, A.S., Mueller, U.G., Ortiz, A., Bacci Jr., M. (2007) Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini). Insect Mol. Biol. **16**, 777–784

Moulton, M.J., Song, H., Whiting, M.F. (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). Mol. Ecol. Resour. **10**, 615–627

Nylander, J.A.A. (2004) MrModeltest v2. Program distributed by the author. Uppsala University: Evolutionary Biology Center.

Pamilo, P., Viljakainen, L., Vihavainen, A. (2007) Exceptionally high density of NUMTs in the honeybee genome. Mol. Biol. Evol. **24**, 1340–1346

Resende, H.C., Yotoko, K.S.C., Delabie, J.H.C., Costa, M.A., Campiolo, S., Tavares, M.G., Campos, L.A. O., Fernandes-Salomão, T.M. (2010) Pliocene and Pleistocene events shaping the genetic diversity within the central corridor of the Brazilian Atlantic Forest. Biol. J. Linn. Soc. **101**, 949–960

Richly, E., Leister, D. (2004) NUMTs in sequenced eukaryotic genomes. Mol. Biol. Evol. **21**, 1081–1084

Ronquist, F., Huelsenbeck, J.P. (2003) MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**, 1572–1574

Solomon, S.E., Bacci Jr., M., Martins Jr., J., Vinha, G.G., Mueller, U.G. (2008) Paleodistributions and comparative molecular phylogeography of leafcutter ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. PLoS One **3**, e2738

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. Proc. Natl. Acad. Sci. USA **105**, 13486–13491

Sorenson, M.D., Quinn, T.W. (1998) Numts: a challenge for avian systematics and population biology. The Auk **155**, 214–221

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**, 2731–2739

Tanaka, H., Roubik, D.W., Kato, M., Liew, F., Gunsalam, G. (2001) Phylogenetic position of *Apis nuluensis* of northern Borneo and phylogeography of *A. cerana* as inferred from mitochondrial DNA sequences. Insect. Soc. **48**, 44–51

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. **22**, 4673–4680

Viljakainen, L., Oliveira, D.C., Werren, J.H., Behura, S. K. (2010) Transfers of mitochondrial DNA to the nuclear genome in the wasp *Nasonia vitripennis*. Insect Mol. Biol. **19**, 27–35

Yang, Z., Rannala, B. (1997) Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. **14**, 717–724