Insect Molecular Biology (2007) **16**(6), 777–784

Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini)

J. Martins Jr*, S. E. Solomon†, A. S. Mikheyev†, U. G. Mueller†, A. Ortiz‡ and M. Bacci Jr*

**Center for the Study of Social Insects, São Paulo State University, Rio Claro, Brazil;* †*Section of Integrative Biology, Patterson Laboratories, University of Texas, Austin, USA; and* ‡*Conservación, Usos y Biodiversidad, Universidad Nacional de Colombia, Medellín, Colombia*

Abstract

Nuclear mitochondrial-like sequences (numts) are copies of mitochondrial DNA that have migrated to the genomic DNA. We present the first characterization of numts in ants, these numts being homologues to a mitochondrial DNA fragment containing loci the 3′ portion of the cytochrome oxidase I gene, an intergenic spacer, the tRNA leucine gene and the 5′ portion of the cytochrome oxidase II gene. All 67 specimens of *Atta cephalotes* (Hymenoptera: Formicidae: Attini) investigated had these homologues, which are within two monophyletic groups that we called numt1 and numt2. Numt1 and numt2 sequences are less variable than mitochondrial sequences and released from the severe purifying selection constraining the evolution of mitochondrial genes. Their formation probably involved bottlenecks related to two distinct transfer events of ancient and fast evolving mitochondrial DNA fragments to comparative slowly evolving nuclear DNA regions.

Keywords: ant, *Atta cephalotes*, Attini, mitochondrial DNA, numt.

Introduction

Mitochondrial DNA has been widely utilized in phylogenetic and population-level studies because it has a great number of copies per cell, a high mutational rate, little or no recombination, haploidy and maternal inheritance (Avise *et al*., 1987; Moritz *et al*., 1987; Avise, 2001). These advantages, however, are limited by heteroplasmy, paternal linkage, non-neutrality, a variable rate of evolution, and the presence of nuclear-like mitochondrial sequences (numts) (Avise, 2001; Ballard & Whitlock, 2004).

Numts are structurally very similar to the functional mitochondrial genes, but may contain unexpected stop codons and lack or contain abnormalities in the regulatory sequences (D'Errico *et al*., 2004). They arise through migration from the mitochondrion to the nucleus either through direct DNA (Lopez *et al*., 1994) or RNA mediated transfer mechanisms (Nugent & Palmer, 1991) involving transposable or viral elements (Farrely & Butow, 1983; Ossario *et al*., 1991; Zullo *et al*., 1991; D'Errico *et al*., 2004).

Since their first discovery by Du Buy & Riley (1967), numts have been found in many groups of organisms including plants, yeast, alveolates, nematodes, insects, and vertebrates such as mammals and birds (reviewed by Bensasson *et al*., 2001). Numts are apparently very common in *Drosophila melanogaster*, in which Harrison *et al.* (2003) identified around 100 numt variants, six of which most likely originated from RNA transfer. However, numts are thought to be virtually absent in fishes (D'Errico *et al*., 2004).

Numts present different degrees of functionality and are located in distinct sites within the nuclear genome, which subject them to different evolutionary pressures. If sequences of numts and mtDNA are used simultaneously in phylogenetic analysis, misleading relationships can result (Sorenson & Quinn, 1998).

In the present investigation, we describe two groups of numts (numt1 and numt2) which are homologous to a mitochondrial DNA segment encompassing cytochrome oxidase I (COI), an intergenic spacer (IGS), tRNA leucine (tRNA-Leu) and cytochrome oxidase II (COII) in the ant *Atta cephalotes*. We also describe protocols and primers to specifically target each of these numts as well as the mitochondrial sequences. The latter primers will be useful to avoid numts interference on reconstructions of phylogenetic and population histories of *A. cephalotes* based on mitochondrial DNA sequences.

Received 21 May 2007; accepted after revision 17 September 2007. Correspondence: Maurício Bacci Jr, Centro de Estudos de Insetos Sociais, Universidade Estadual Paulista, Av. 24A, n. 1515 – Bela Vista, Zip Code: 13506-900 – Rio Claro, SP – Brazil. Tel.: +55 19 3526 4165; fax: +55 19 3534 8523; e-mail: mbacci@rc.unesp.br

Table 1. Primer pairs used to specifically amplify mitochondrial, nuclear mitochondrial-like sequence 1 (numt1) or numt2 loci

Primer pair (5′→3′)	Locus
ANT-F (ATTCATTCTTATCTTGAAATATTATTTC) + ANT-R2 (CAATAGTTTGATTTTCTAG)	Mitochondrial (COI-IGS-tRNA-Leu-COII)
ANT-F + ANT-R3 (CAATGATTATAAATCATGGAAAAAAAATC)	Numt1 (COI-IGS-tRNA-Leu-COII)
ANT-F2 (CTCTTCTATACCCAATTTCTATAACCC) + ANT-R (TTCATAAGTTCAGTATCATTGGTG)	Numt2 (IGS-tRNA-Leu-COII)

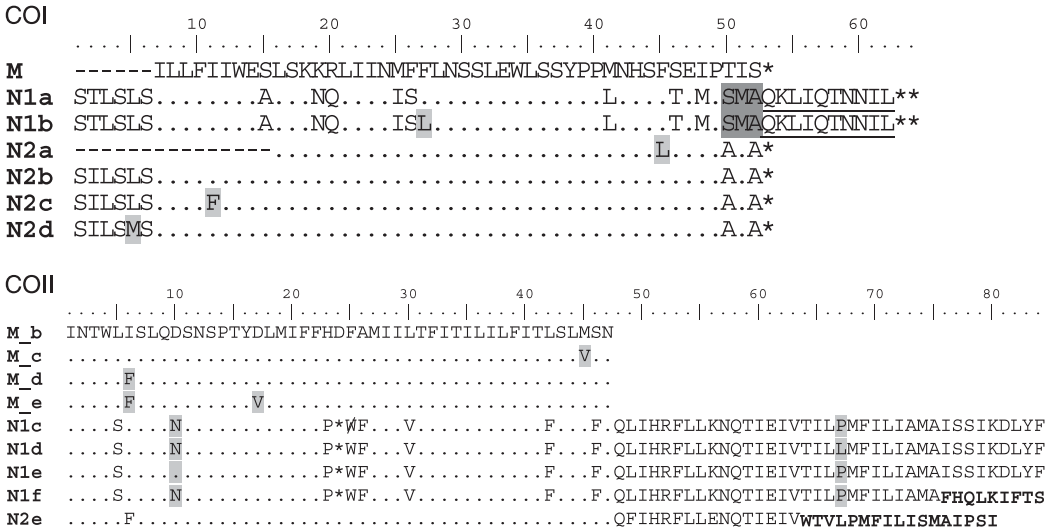

Figure 1. Alignment of the amino acid sequences deduced from cytochrome oxidase I (COI) or COII loci. M, mitochondrial sequences; N1, nuclear mitochondrial-like sequence 1 (numt1); N2, numt2. Dots indicate conserved sites and letters indicate amino acid changes. Differences in amino acid residues within M, N1 or N2 are shaded, insertion of contiguous amino acid residues is underlined, and modifications because of frame shifts are in bold. Stop codons are depicted as (*). Note the 10 amino acid residues insertion seen in N1a and N1b (COI), the internal stop codon seen in N1c, N1d, N1e and N1f (COII), the frame shift seen in N2e (COII) and minor alterations in N2a, N2b, N2c and N2d.

Results

Mitochondrial and numt sequences

Mitochondrial sequences obtained in the present investigation from *A. cephalotes* did not present any of the known diagnostic characteristics of pseudogenes (Pakendorf & Stoneking, 2005), such as premature stop codons, insertions or deletions that caused frame shifts. In addition, all substitutions in COI and most of the substitutions in COII were synonymous in mitochondrial sequences, indicating strong purifying selection acting on these loci. In contrast, numt1 and numt2 sequences presented characteristics of pseudogenes, such as indels and premature stop codons, as well as relaxed selection constraints in COI and COII regions.

Primers ANT-F and ANT-R (Table 1) were designed to specifically target an *Atta sexdens* mitochondrial DNA fragment encompassing the 3′ end of COI, the complete sequences of an IGS and the tRNA-Leu and the 5′ end of COII (Bacci *et al.*, unpubl. data). However, this primer pair resulted in three distinct PCR products from 20 specimens of *A. cephalotes*. These products were cloned and two clones of each specimen were sequenced, revealing one mitochondrial locus (16 haplotypes) and two other loci which we propose as numts: numt1 (11 haplotypes) and numt2 (five haplotypes). These haplotype sequences were used to design specific primers that were combined with ANT-F or ANT-R in order to selectively amplify mitochondrial, numt1 or numt2 segments (Table 1). PCR amplification with the primer pairs depicted in Table 1, followed by direct sequencing of the generated PCR products, showed that mitochondrial and both numt1 and 2 loci were present in every one of the 67 *A. cephalotes* specimens studied.

The analysis of both cloned and directly sequenced fragments from 67 *A. cephalotes* specimens distinguished 30 mitochondrial haplotypes (GenBank accession numbers EUI56242 to EUI56308) that contained 501 to 520 base pairs. Twenty-five of the 48 polymorphic sites including all the indels found in these haplotypes were located in the IGS region. Few substitutions were found in COI or in COII, but no variation was found in the tRNA-Leu gene. The 13 polymorphic sites in COI and seven out of the 10 polymorphic sites in COII corresponded to synonymous mutations, so that all mitochondrial haplotypes coded for a single COI amino acid sequence and four COII amino acid sequences (Fig. 1). The IGS region was represented by 141 base pairs in these mitochondrial sequences.

Table 2. Diversity and selection in *Atta cephalotes* mitochondrial, nuclear mitochondrial-like sequence 1 (numt1) and numt2 loci

Locus	*n*	*h*	π	*dn*	*ds*	ω
Mitochondrial	30	0.975 (0.016)	0.02278 (0.00137)	0.0031	0.0779	0.040
Numt1	11	0.893 (0.111)	0.00314 (0.00072)	0.0041	0.0065	0.631
Numt2	5	0.700 (0.218)	0.00127 (0.00048)	0.0015	0.0008	0.405

Number (*n*) and diversity (*h*) of haplotypes; nucleotide diversity (π); number of replacement substitutions per replacement site (*dn*); number of synonymous substitutions per synonymous site (*ds*); ratio *dn* : *ds* (ω).

Haplotype diversity (h) and nucleotide diversity (π) found within numt1 or 2 nucleotide sequences were lower than those seen in mitochondrial sequences (Table 2). Cloned and directly sequenced numt1 fragments from the 67 studied *A. cephalotes* specimens revealed 11 haplotypes (594 to 626 base pairs) (GenBank accession numbers EUI56309 to EUI56319 and EUI65374 to EUI65429) which contained 13 polymorphic sites, most of which were in COI (nine sites) and COII (three sites). The IGS contained only one polymorphic site and no variation was found in tRNA-Leu. The numt2 group contained five haplotypes (606 to 651 base pairs) (GenBank accession numbers EUI56320 to EUI56324 and EUI65432 to EUI65493) with four polymorphic sites, all them in the COI region. The IGS region was represented by 136 base pairs in numt1 and by 178 base pairs in numt2.

Contrary to that found in the mitochondrial sequences, in which most of the nucleotide substitutions in the COI and COII coding regions were synonymous, mutations in the COI and COII regions of the numt1 and 2 sequences often resulted in amino acid substitutions or insertions. In addition, these numts presented premature stop codons and frame shifts (Fig. 1).

Numt1 coded for two COI amino acid sequences (N1a and N1b) which differed by a single amino acid and, compared to mitochondrial COI sequence, contained an insertion of 10 amino acid residues followed by two sequential stop codons. Numt1 also coded for four COII amino acid sequences (N1c to N1f), three of which differed by a single amino acid, whereas the fourth was more distinctive as a result of a frame shift in the C terminus. All four COII putative amino acid sequences from the numt1 group (N1c to N1f) had an internal stop codon likely blocking its C terminus transcription. Sequences in the numt2 group coded for four COI amino acid sequences (N2a to N2d) differing from each other by single amino acid substitutions, and for a single COII amino acid sequence (N2e).

Selection strength on COI and COII nucleotide sequences, estimated through the ω calculation (Table 2), showed mitochondrial coding genes submitted to a stronger purifying selection than numt1 or numt2.

The relative rate of evolution of numts and mitochondrial sequences was estimated using *Wasmania auropunctata* sequence (GenBank AF016026) as the outgroup and every *A. cephalotes* mitochondrial or numt sequence as ingroups.

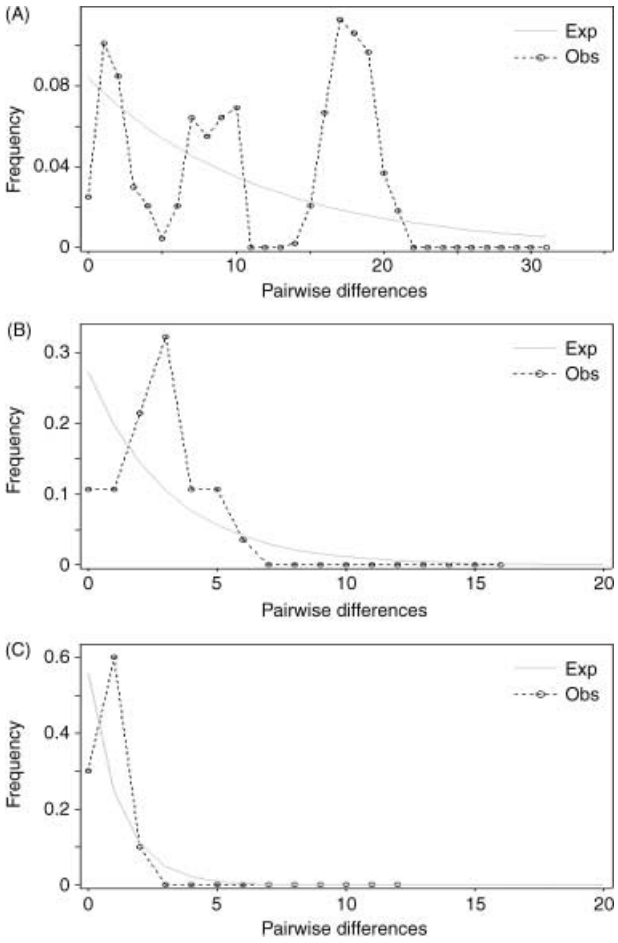

Figure 2. Mismatch distribution of mitochondrial (A), nuclear mitochondrial-like sequence 1 (numt1) (B) or numt2 (C) loci. Exp, expected distribution; Obs, observed distribution.

This test indicated that the numt1 evolution rate was significantly distinct (5% CI) from that of mitochondrial ($\chi^2 = 5.49$) or numt2 ($\chi^2 = 4.00$), as no significant difference was found between numt2 and the mitochondrial groups or within groups.

In addition, population demographic changes analysis (Fig. 2) showed a ragged multimodal mismatch distribution consistent with demographic equilibrium of mitochondrial haplotypes. Conversely, unimodal distributions indicative of population bottleneck or expansion resulted from numt1 or numt2 analysis.

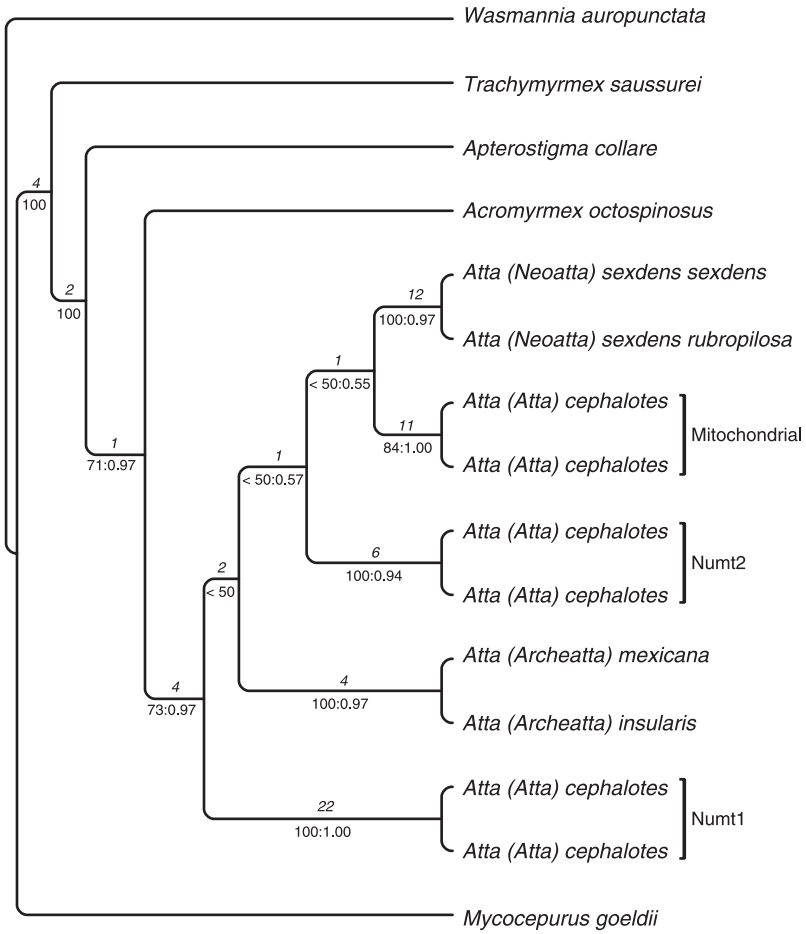

Figure 3. The single most parsimonious tree obtained from 126 parsimony-informative characters from 12 cytochrome oxidase I (COI), tRNA leucine (tRNA-Leu) and COII sequences. Length = 397; consistency index = 0.627; retention index = 0.570. Numbers above nodes represent decay index values and those below nodes represent bootstrap values : posterior probability values. Mitochondrial sequences are represented by haplotypes 2 and 16 (GenBank accession numbers EUI56246 and EUI56276), Numt2 sequences are represented by haplotypes 1 and 3 (GenBank EUI56309 and EUI56311) and Numt1 sequences are represented by haplotypes 1 and 4 (GenBank EUI56320 and EUI56323). Identical topology was recovered in the single most likely tree (–ln L = 1941.0549) and in the 50% majority rule consensus tree recovered from the Bayesian analysis, both generated from the same data set.

Phylogenetic relationships

The phylogenetic positions of numt1, numt2 and mitochondrial groups from *A. cephalotes* were inferred from a matrix containing two representative sequences in each of these groups as well as other mitochondrial sequences from two ant species within each of the currently accepted *Atta* divisions, ie the subgenera *Archeatta* (GenBank EUI65430 and EUI65431), *Atta* (EUI56246 and EUI56276) and *Neoatta* (AF016018 and AF016019) (Gonçalves, 1986). This matrix also contained homologous mitochondrial sequences, which were retrieved from GenBank and corresponded to *Trachymyrmex* cf. *saussurei* (AF016024), *Acromyrmex octospinosus* (AF016014), the lower Attini species, *Mycocepurus goeldii* (AF016017) and *Apterostigma collare* (AF016013), as well as *Wasmannia auropunctata* (AF016026), this last used as the outgroup. The IGS region was highly variable with probably many homoplasies. This region was not included in the phylogenetic analysis.

The ILD test resulted in a *P* value of 0.9, indicating that COI, tRNA-Leu and COII sequences could be concatenated and analysed in a single matrix. The phylogenetic analysis through parsimony resulted in a single most parsimonious tree (Fig. 3), which showed numt1 and numt2 with high support in a clade with the mitochondrial sequences of *Atta* species. High support was also found in each of the clades within the *Atta* group (Archeatta, Atta, Neoatta, numt1 and numt2 clades). However, low bootstrap and decay index values indicated poor support for phylogenetic relationships within the *Atta* group. There was, for example, poor support for a basal numt1 and a more derived numt2. An identical topology to that shown in Fig. 3 was found in the single most likely tree resulting from the maximum likelihood analysis and in the 50% majority rule tree generated from the Bayesian analysis.

Discussion

Numt1 and numt2 of *A. cephalotes* are the first nuclear mitochondrial-like sequences characterized in ants. Thus, our present findings increase the list of animal groups in which numts exist. Amplification of *A. cephalotes* numts resulted from the lack of specificity of primers ANT-F and ANT-R to the mitochondrial sequences of *A. cephalotes*,

although this primer pair usually generates a single mitochondrial PCR product from other ant species within the *Atta* genus (Bacci *et al*., unpubl. data).

When the *A. cephalotes* specimens were amplified with primers that were designed specifically for numt1 or numt2 loci, we observed that each of the 67 analysed specimens contained these loci. In addition, the use of the primer pair ANT-F and ANT-R2 specifically amplified mitochondrial loci in *A. cephalotes*, eliminating the interference from numt1 and numt2. As they frequently occur in a great number of copies, numts are often mistakenly amplified instead of their mitochondrial paralogues, this fact impairing the separation of mitochondrial DNA from numts in small insects, from which enriched mitochondrial DNA preparations must be obtained in order to preferentially amplify mitochondrial genes (Bensasson *et al*., 2000). Thus, the specific amplification of *A. cephalotes* mitochondrial genes described here will facilitate the targeting of mitochondrial genes in further studies on the phylogeny of this ant species.

Compared to mitochondrial sequences, nucleotide diversity (π) was seven times lower in numt1 and 18 times lower in numt2 (Table 2). This relatively low diversity of numts is the result of opposite forces modulating their evolution. For example, numt1 and 2 COI and COII are submitted to a low constraint, as deduced from the ω values shown in Table 2. These values indicate that the proportion of replacement substitutions is 25 times lower than the proportion of synonymous substitutions in mitochondrial sequences, as the corresponding value dropped to 2.5 in numt2 and to 1.6 in numt1. This indicates that numt1 and 2 have escaped the purifying selection that operates on COI and COII mitochondrial regions. This escape would act to favour diversity within numt groups.

However, a population bottleneck seems involved in numts formation, which is consistent with the unimodal mismatch distribution resulting from numts analysis, differently from the ragged multimodal distribution of mitochondrial sequences analysis (Fig. 2), which indicates population equilibrium. This bottleneck would have acted to decrease diversity within the numt groups.

Another important force constraining the diversity within numts may be the low evolution rate. A low evolution rate of nuclear pseudogenes compared to their mitochondrial homologues has been proposed for primates (Brown *et al*., 1982; Zischler *et al*., 1995) and, in insects, nuclear genes generally evolve slower than mitochondrial ones (Chung-Pin & Danforth, 2004). Thus, it is conceivable that numt1 and 2 evolve slower than their mitochondrial homologues. At least for numt1, which seems to have originated before numt2 (see below), the Tajima's test was able to capture a distinct evolution rate compared to that seen in mitochondrial sequences. The low evolution rate of numt1 (and numt2) is also suggested by the high similarity (98–99.7 and 99.9%, respectively) of their haplotypes within their own groups compared to mitochondrial sequences, which were 93–99% similar to each other. It is also suggested by the analysis of the IGS region, which is a hotspot for mutations in mitochondrial sequences, as it contained 25 out of the 48 mutations found in these sequences. However, the IGS region is apparently frozen in numts, in which this region contained only one out of the 17 mutations found. The remaining 16 mutations of numt1 and 2 were concentrated in the COI and COII regions. Thus, the transfer of mitochondrial sequences to the nucleus, which seems to have been the origin of numt1 and 2, resulted in the shift of the mutation hotspot from the IGS to the COI and COII regions.

In addition to the low evolution rate, it was also possible to propose an ancient divergence of numt1 and 2 from the mitochondrial sequences because there was low similarity between numt1 (69–74%) or numt2 (75–78%) and these mitochondrial sequences. Ancient divergence and low evolution rates are characteristics of nuclear mitochondrial-like sequences, which have been described as molecular fossils (Zischler *et al*., 1995), resulting from gene transfer from the relatively fast evolving mitochondrial genome to the nuclear genome (Pakendorf & Stoneking, 2005).

Numts 1 and 2 are monophyletic with mitochondrial sequences representative of species of the three main subdivisions of *Atta* (Fig. 2). Numt1 accumulated more differences (26–31%) in mitochondrial sequences than did numt2 (22–25%) and had 11 haplotypes while numt2 had five. In addition, numt1 is less sensitive to selective constraints than numt2 (Table 2). Taken together, these results suggest that numt1 was transferred to the nucleus before numt2, although they may also indicate that numt1 evolves faster than numt2. An earlier numt1 radiation basal to the *Atta* genus was also suggested by the tree topology (Fig. 3), albeit with low bootstrap, decay index and posterior probability supports, most likely because of the small number of nucleotides sampled from these numts. Therefore, further investigation is needed to determine the order of radiation and the phylogenetic placement of numt1 and numt2 with respect to the *Atta* clade.

Experimental procedures

Ant specimens and DNA extraction

Specimens of *A. cephalotes* Linnaeus (1758), *A. sexdens sexdens* Linnaeus (1758), *Atta sexdens rubropilosa* Forel (1908), *Atta insularis* Guérin, (1844) and *Atta mexicana* (Smith) 1858 (Hymenoptera: Formicidae: Attini) were collected in different sites in Brazil, Colombia, Cuba, the USA and Venezuela (Table 3). For genomic DNA extraction, individual specimens were crushed in 1.5 ml microtubes immersed in liquid nitrogen and then suspended in iced 500 µl TNES (Tris-NaCl-EDTA-SDS) solution (250 mM Tris Base (Promega, Madison, WI), pH 7.5, 2 M NaCl (Merck, Darmstadt, Germany), 100 mM EDTA (Ethylenediamine tetraacetic acid) (Sigma-Aldrich Corp., St. Louis,

Table 3. Specimens codes, species names, collection sites and coordinates associated with each of the *Atta* species studied in the present investigation.

Specimen	Species	Site	Latitude (°)	Longitude (°)
694	*Atta cephalotes*	Santafé de Antioquia, Colombia	6.53583	−75.88472
689	*A. cephalotes*	Santafé de Antioquia,	6.53583	−75.88500
693	*A. cephalotes*	Santafé de Antioquia, Colombia	6.53528	−75.88388
688	*A. cephalotes*	Santafé de Antioquia,	6.53472	−75.88444
690	*A. cephalotes*	Santafé de Antioquia, Colombia	6.53333	−75.88528
691	*A. cephalotes*	Santafé de Antioquia, Colombia	6.53333	−75.88333
692	*A cephalotes*	Santafé de Antioquia,	6.53333	−75.88333
659	*A. cephalotes*	San Roque, Colombia	6.48500	−75.01805
651	*A. cephalotes*	San Roque, Colombia	6.46917	−75.03472
652	*A. cephalotes*	San Roque, Colombia	6.46917	−75.03472
655	*A. cephalotes*	San Roque, Colombia	6.46917	−75.03555
656	*A. cephalotes*	San Roque, Colombia	6.46917	−75.03361
657	*A. cephalotes*	San Roque, Colombia	6.46917	−75.03361
650	*A. cephalotes*	San Roque, Colombia	6.46833	−75.03555
699	*A. cephalotes*	Venecia, Colombia	5.96722	−75.71777
626	*A. cephalotes*	Sonsón, Colombia	5.70250	−75.40083
625	*A. cephalotes*	Sonsón, Colombia	5.70083	−75.40138
695	*A. cephalotes*	Andes, Colombia	5.66722	−75.85083
697	*A. cephalotes*	Andes, Colombia	5.66722	−75.85027
488	*A. cephalotes*	Andes, Colombia	5.66694	−75.88555
SES040305-01	*A. cephalotes*	Kukenan Camp, Venezuela	5.10819	−60.82000
SES040208-06	*A. cephalotes*	Macapá, AP, Brazil	0.62059	−51.69100
SES040208-08	*A. cephalotes*	Macapá, AP, Brazil	0.61917	−51.70161
SES040208-10	*A. cephalotes*	Macapá, AP, Brazil	0.60082	−51.75435
SES040208-09	*A. cephalotes*	Macapá, AP, Brazil	0.60029	−51.75318
SES040204-03	*A. cephalotes*	Belém, PA, Brazil	−1.68755	−48.54977
SES040205-02	*A. cephalotes*	Belém, PA, Brazil	−1.68755	−48.54977
SES040214-03	*A. cephalotes*	Alenquer, PA, Brazil	−1.90357	−54.64113
SES040214-11	*A. cephalotes*	Alenquer, PA, Brazil	−1.91676	−54.62662
SES040215-03	*A. cephalotes*	Alenquer, PA, Brazil	−1.92631	−54.63676
SES040212-01	*A. cephalotes*	Santarém, PA, Brazil	−2.55767	−54.72733
SES040220-01	*A. cephalotes*	Carreiro da Várzea, AM, Brazil	−3.65538	−60.26173
SES040120-01	*A. cephalotes*	Frei Caneca, PE, Brazil	−8.72038	−35.84425
SES040120-04	*A. cephalotes*	Frei Caneca, PE, Brazil	−8.72038	−35.84425
SES040121-01	*A. cephalotes*	Frei Caneca, PE, Brazil	−8.72038	−35.84425
SES040121-03	*A. cephalotes*	Frei Caneca, PE, Brazil	−8.72038	−35.84425
SES040121-04	*A. cephalotes*	Frei Caneca, PE, Brazil	−8.72038	−35.84425
E958	*A. cephalotes*	Alta Floresta, MT, Brazil	−9.55650	−55.99700
SES040130-02	*A. cephalotes*	Alta Floresta, MT, Brazil	−9.59214	−56.02422
SES040129-03	*A. cephalotes*	Alta Floresta, MT, Brazil	−9.75642	−55.86272
SES040129-06	*A. cephalotes*	Alta Floresta, MT, Brazil	−9.75642	−55.86272
SES040131-12	*A. cephalotes*	Alta Floresta, MT, Brazil	−9.86249	−56.07603
SES040131-13	*A. cephalotes*	Alta Floresta, MT, Brazil	−9.86249	−56.07603
SES040131-17	*A. cephalotes**	Alta Floresta, MT, Brazil	−9.86249	−56.07603
E947	*A. cephalotes*	Alta Floresta, MT, Brazil	−10.03715	−55.43497
SES040131-03	*A. cephalotes*	Alta Floresta, MT, Brazil	−10.05337	−55.43224
SES040131-06	*A. cephalotes*	Alta Floresta, MT, Brazil	−10.05337	−55.43224
SES040131-04	*A. cephalotes*	Alta Floresta, MT, Brazil	−10.05337	−55.43224
SES040124-08	*A. cephalotes*	Ipiaú, BA, Brazil	−14.09619	−39.78103
SES040125-09	*A. cephalotes*	Farm, BA, Brazil	−14.20028	−39.81581
SES040125-10	*A. cephalotes*	Farm, BA, Brazil	−14.20028	−39.81586
E909	*A. cephalotes*	Aurelino Leal, BA, Brazil	−14.20177	−39.82375
E907	*A. cephalotes*	Aurelino Leal, BA, Brazil	−14.41333	−39.33083
SES040125-01	*A. cephalotes*	Fazenda de Cascata, BA, Brazil	−14.41356	−39.33089
SES040125-02	*A. cephalotes*	Fazenda de Cascata, BA, Brazil	−14.41356	−39.33089
SES040125-03	*A. cephalotes*	Fazenda de Cascata, BA, Brazil	−14.41356	−39.33089
ASM040123-01	*A. cephalotes*	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040123-01	*A. cephalotes*	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040123-02	*A. cephalotes*	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040123-03	*A. cephalotes*	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040123-06	*A. cephalotes*	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040123-07	*A. cephalotes**	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040123-08	*A. cephalotes*	CEPLAC, BA, Brazil	−14.75536	−39.23256
SES040124-01	*A. cephalotes*	Ubaitaba, BA, Brazil	−14.83975	−39.02689
SES040124-02	*A. cephalotes*	Ubaitaba, BA, Brazil	−14.83975	−39.02689
SES040124-03	*A. cephalotes*	Ubaitaba, BA, Brazil	−14.83975	−39.02689
E899	*A. cephalotes*	Ilhéus, BA, Brazil	−14.83975	−39.02688
SES040205-05	*A. sexdens sexdens*	Belém, PA, Brazil	1.56850	−48.52650
E305	*A. sexdens rubropilosa**	Rio Claro, SP, Brazil	−22.53640	−47.56580
E399	*A. insularis**	Pinar Del Rio, Cuba	22.23540	−83.65840
E334	*A. mexicana**	Austin, TX, USA	30.35420	−97.85460

*Specimens from which sequences were used for building the tree shown in Fig. 3.

MO), 2% SDS (Sodium dodecyl sulfate) (Promega, Madison, WI). Then, 5 μl of 20 mg/l proteinase K (USB Corporation, Cleveland, OH, cat # 76230Y) was added to the microtubes and the mixture was incubated for 3 h at 55 °C, followed by addition of 5 μl of 4 mg/ml RNAse A (Calbiochem, San Diego, CA, cat # 556746) and 30 min incubation at 37 °C. Proteins were then precipitated by adding 200 μl of 5 M NaCl, the supernatant was collected, DNA was precipitated with 100% isopropanol, washed with 70% ethanol, centrifuged, dried and dissolved in 35 μl pH 8.0 TE buffer.

DNA amplification and products purification

PCR was performed with Pure Taq Ready-To-Go PCR Beads (GE Healthcare UK Ltd., Little Chalfont, Bucks, U.K., cat # 27-9557-01), 6 pm primers (Table 1), ~100 ng DNA template and ultra-pure water added to a final volume of 25 μl. PCR included a 5 min 94 °C incubation followed by 30 cycles of 1 min at 94 °C, 1 min at 47 °C and 2 min at 68 °C and ended with a 15 min 68 °C extension. For numt2 amplification, primer annealing was for 30 s at 50 °C. PCR products were purified utilizing the GFX PCR DNA and Gel Band Purification kit (GE Healthcare cat # 27-9602-01).

Cloning and sequencing

In *A. cephalotes*, PCR products generated with nonspecific ANT-F and ANT-R primers (Table 1) were ligated into the pGEM-T Vector (Promega, Madison, WI, USA), transformed into *Escherichia coli* cells DH10α by heat shock (Sambrook, 2001), purified by a miniprep procedure (Sambrook, 2001) and sequenced with the T7 and SP6 vector primers. PCR products generated with the specific primers ANT-R2, ANT-F2 and ANT-R3 were directly sequenced. In the other *Atta* species (*A. sexdens sexdens*, *A. sexdens rubropilosa*, *A. mexicana* and *A. insularis*) ANT-F and ANT-R primers specifically amplified the mitochondrial DNA and the amplicons were directly sequenced. Sequencing of PCR products was carried out with the same primers as for their amplification. The sequencing reaction contained 800 ng recombinant plasmids or 100 ng PCR templates, 6 pm primers, 2 μl BigDye (Applied Biosystems, Foster City, CA, USA, cat # 4336919), 2.5 μl buffer (200 mM Tris.HCl, 5 mM $MgCl_2$) and 1 μl ultra-pure water. These reactions included a 95 min 5 °C-incubation, followed by 30 cycles of 1 min at 95 °C, 1 min at 50 °C and 4 min at 72 °C. The products of the sequencing reaction were purified and then resolved in an ABI377 automated sequence machine (Applied Biosystems).

Sequence analysis

Sequences were inspected to determine the number and position of synonymous and nonsynonymous substitutions, insertions, deletions and frame shifts using the programs DnaSP 4.0 (Rozas *et al*., 2003) and MEGA3 (Kumar *et al*., 2004). Sequences generated in this study, as well as others which were retrieved from GenBank, were edited with the program BioEdit 7.0.5 (Hall, 1999) and aligned with ClustalW (Thompson *et al*., 1994) using default parameters. Genetic diversity within mitochondrial, numt1 or numt2 sequences was estimated using haplotype diversity (*h*) and nucleotide diversity (π) as implemented in DnaSP 4.0 (Rozas *et al*., 2003). The number of replacement substitutions per replacement site (*dn*), the number of synonymous substitutions per synonymous sites (*ds*), the ratio (ω) *dn* : *ds* and the Tajima's relative rate test (Tajima, 1993) were calculated with the program MEGA3 (Kumar *et al*., 2004). The incongruence-length difference test (ILD) (Farris *et al*, 1995) was implemented by paup* 4.0b10 (Swofford, 2003), in order to test whether COI, tRNA-Leu and COII could be concatenated in a single matrix for analysis. Phylogenetic analyses were carried out by maximum parsimony using a heuristic search with 1000 replicates, tree-bisection-reconnection (TBR) branch swapping, random sequence addition, and the *collapse* and *multrees* options implemented. Bootstrap (Felsenstein, 1985) with 1000 pseudoreplicates, each of them with 100 heuristic searches, was performed with the same settings as for tree search. Decay Index (Bremer, 1994) were calculated with the program prap 1.0 (Müller, 2004), which creates a constraint file containing instructions to be executed by paup*. Maximum likelihood analysis was also applied to the dataset. The Transitional Model (TIM) of substitution and gamma distribution was significantly more likely than the next best model, as determined by the hierarchical likelihood ratio test implemented by Modeltest 3.7 (Posada & Crandall, 1998). The maximum likelihood analysis was then run using paup* with a starting tree obtained by stepwise addition of taxa, heuristic search with 1000 replicates, TBR branch swapping, random sequence addition, *multrees* options implemented. Bayesian analysis was carried out with the program MrBayes 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003), considering the commands for the TIM+G model which were generated by ModelTest. Two separated runs were conducted, each of which starting from different randomly chosen trees and with four independent chains. Each chain was run for 1 000 000 update cycles and the chain states were sampled every 100th cycle. The first 2000 samples were discarded as the 'burn-in' of the Markov chain. Using paup*, the 50% majority rule consensus tree was built from 8001 trees sampled to calculate the posterior probability of the tree nodes. The program Tree Graph 1.0 (Müller & Müller, 2004) was used to draw the tree.

Nucleotide diversity (π) was calculated through the DnaSP program using the Jukes & Cantor model. Selection constraints acting on COI and COII sequences in mitochondrial or numt sequences were inferred through the ratio (ω) of the number of replacement substitutions per replacement site (*dn*) to the number of synonymous substitutions per synonymous sites (*ds*).

Signatures of population demographic changes were examined by comparing mismatch distributions in each of the mitochondrial, numt1 and numt2 markers with those distributions expected in constant populations using DnaSP (Rozas *et al*., 2003).

Acknowledgements

We acknowledge the financial support from CAPES, CNPq and FAPESP. Scott E. Solomon was supported by a Doctoral Dissertation Improvement Grant from the NSF (DEB0407772) and by a Dorothea Bennett Memorial Graduate Fellowship from UT-Austin. Joaquim Martins Jr was supported by CAPES. We also thank Dr Roger Butlin (University of Leeds, UK) for his valuable comments on sequence analysis.

References

Avise, J.C. (2001) *Phylogeography: the History and Formation of Species*. Harvard University Press, London.

Avise, J.C., Arnold, J., Ball, R.M., Birmingham, E., Lamb, T., Neigel, J.E. *et al*. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* **18**: 489–522.

Ballard, J.W.O. and Whitlock, M. (2004) The incomplete natural history of mitochondria. *Mol Ecol* **13**: 729–744.

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

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

Bremer, K. (1994) Branch support and tree stability. *Cladistics* **10**: 295–304

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982) Mitocondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* **18**: 225–239.

Chung Pin, L. and Danforth, B.N. (2004) How do insect nuclear and mitochondrial gene substitution pattern differ? Insights from Bayesian analyses of combined datasets. *Mol Phylogenet Evol* **30**: 686–702.

D'Errico, I., Gadaleta, G. and Saccone, C. (2004) Pseudogenes in metazoa: origin and features. *Brief Funct Gen Prot* **3**: 157–167.

Du Buy, H.G. and Riley, F.L. (1967) Hybridization between the nuclear and kinetoplast DNA's of *Leishmania enriettii* and between nuclear and mitochondrial DNA's of mouse liver. *Proc Natl Acad Sci USA* **57**: 790–797.

Farrely, F. and Butow, R.A. (1983) Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **301**: 296–301.

Farris, J.S., Källersjö, M., Kluge, A.G. and Bult, C. (1995) Testing significance of incongruence. *Cladistics* **10**: 315–319.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Gonçalves, C.R. (1986) Filogenia do gênero *Atta* Fabricius (Hymenoptera, Formicidae). Anais do VI Congresso Brasileiro de Zoologia. *Plublicações Avulsas do Museu Nacional, Rio de Janeiro* **65**: 13–17.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Series* **41**: 95–98.

Harrison, P.M., Milburn, D., Zhang, Z., Bertone, P. and Gerstein, M. (2003) Identification of pseudogenes in the *Drosophila melanogaster* genome. *Nucleic Acids Res* **31**: 1033–1037.

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

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinfo* **5**: 150–163.

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

Moritz, C., Dowling, T.E. and Brown, W.M. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* **18**: 269–292.

Müller, J. and Müller, K. (2004) TreeGraph: automated drawing of complex tree figures using an extensible tree description format. *Mol Ecol Notes* **4**: 786–788.

Müller, K. (2004) PRAP – Computation of Bremer support of large data sets. *Mol Phylogenet Evol* **31**: 780–782.

Nugent, J.M. and Palmer, J.D. (1991) RNA-mediated transfer of the gene COXII from the mitochondrion to the nucleus during flowering plant evolution. *Cell* **66**: 473–481.

Ossario, P.N., Sibley, D.L. and Boothroyd, J.C. (1991) Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. *J Mol Biol* **22**: 525–536.

Pakendorf, B. and Stoneking, M. (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* **6**: 165–183.

Posada, D. and Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes version 3.0: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Sambrook, J. and Russel, D.W. (2001) *Molecular cloning*: a laboratory manual. 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sorenson, M.D. and Quinn, T.W. (1998) Numts: a challenge for avian systematics and population biology. *Auk* **115**: 214–221.

Swofford, D.L. (2003) *PAUP* : Phylogenetic Analysis Using Parsimony*, version 4.0b10. Smithsonian Institution, Washington DC.

Tajima, F. (1993) Simple methods for testing molecular clock hypothesis. *Genetics* **135**: 599–607.

Thompson, J.D., Higgings, D.G. and Gibbson, T.J. (1994) CusltalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gaps penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Zischler, H., Geisert, H., von Haesele, R.A. and Pääbo, S. (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**: 489–92.

Zullo, S., Leang, C.S., Slighton, J.L., Hadler, H.I. and Eisenstadt, J.M. (1991) Mitochondrial D-loop sequence are integrated in the rat nuclear genome. *J Mol Biol* **221**: 1223–1235