Distribution of mitochondrial DNA fragments in the nuclear genome of the honeybee

W.X. Du and Y.C. Qin

Department of Entomology, China Agricultural University, Beijing, China

Corresponding author: Y.C. Qin
E-mail: yuchuanqin3@yeah.net

ABSTRACT. Nuclear mitochondrial pseudogenes (numts), which originated from mitochondrial DNA (mtDNA) insertions in the nuclear genome, have been detected in many species. The distribution of numts in the honeybee nuclear genome has not yet been fully reported. By referring to the whole honeybee mtDNA sequence and to the recent version of the honeybee nuclear genome, 236 reference sequences were identified by BLAST, with 90 unmapped. The size of the numts ranged from 219 to 3788 bp, and the homologous identity between numts and their corresponding mtDNA fragments varied from 71 to 93%. Furthermore, identified honeybee numts covered nearly all mitochondrial genes and were distributed over all chromosomes. This study provides useful information for further research related to mitochondrial genes and the evolution of the honeybee.

Key words: Honeybee; Mitochondrial pseudogenes (numts)

INTRODUCTION

The honeybee has been the object of considerable scientific curiosity and for almost a century has served as the classic model organism in diverse research fields, including navigation, face recognition and sensory biology (Dyer et al., 2005; Dacke and Srinivasan, 2007; Somanathan et al., 2009). It is regarded as the premier pollinator of major fruit crops. The honeybee has a long history of association with mankind because of its cavity-nesting lifestyle; it is used for producing honey, bee pollen, wax, and royal jelly (Behura, 2007). Recently, researchers have described mitochondrial DNA in different honeybee species (Crozier and Crozier, 1992; Evans and Lopez, 2002; Tan et al., 2011; Wang et al., 2014; Gibson and Hunt, 2014). Nuclear mitochondrial DNA (numts), or mitochondrial pseudogenes, are DNA sequences that are homogenous to mtDNA fragments, and have been identified in nuclear genomes (Lopez et al., 1994). Since their first discovery (du Buy and Riley, 1967), numts have been identified in numerous species (Bensasson et al., 2001; Leister, 2005; Hassanin et al., 2010; Hazkani-Covo et al., 2010; Nergadze et al., 2010; Song et al., 2013), and an increasing number of researchers are now focusing on the evolution of numts and their possible function and influence in studies involving mitochondrial genes (Ricchetti et al., 2004; Thalmann et al., 2004; Schmitz et al., 2005; Cleaver et al., 2014).

As in the field of genetics, studies of mtDNA polymorphism and genetic variation are increasing. Because of their homology with counterpart mtDNA fragments, numts can significantly influence the outcomes of polymerase chain reaction (PCR) analysis methods, and may even lead to erroneous study conclusions (Wallace et al., 1997; Nergadze et al., 2010). It is therefore important to clarify the distribution of numts in mtDNA. The honeybee genome has been recently published (WGS Project: AADG06) from the approximately 1.8 Gb sequence of a European honeybee, from which whole-genome mapping data facilitated the assembly of super-scaffolds; 7.5-times coverage of the (clonable) 236 megabase (Mb) honeybee genome (The Honeybee Genome Sequencing Consortium, 2006). The publication of the honeybee genome provided an opportunity for analysis of numt distribution in honeybees.

The objective of this study was to detect and clarify the distribution of honeybee numts corresponding to relatively complete mtDNA genes, which could provide important reference data to avoid PCR amplification errors in studies relating to mtDNA, as well as the deepen the understanding of honeybee evolution for further research.

MATERIAL AND METHODS

Basic local alignment search tool (BLASTN) was used to find regions of local similarity between sequences, with the whole honeybee (*Apis cerana*) mitochondrial genome (accession No. GQ162109) submitted as the query sequence in HGSC. The honeybee linear scaffolds databases, released on 14 Jan 2012 (WGS Project: AADG06), were used in the BLAST search, with the maximum expectation value set to e = 10-4 and a max score > 300, in order to recover accession hits of biological significance (Pereira and Baker, 2004; Richly and Leister, 2004).

RESULTS

There were 236 alignments (complete results are shown in **Table S1**); results of counterpart sequences covered almost the complete mitochondrial genome. Integrations of mitochondrial

fragments occurred in all of the honeybee nuclear genomes (Figure 1), except for 90 unmapped numts. Similarities between numts and corresponding mtDNA fragments varied from 71 to 93%, and the length of numt fragments ranged from 203 to 3788 bp. Further analysis revealed that the numt lengths, and similarities to corresponding mtDNA fragments, ranged from 71 to 93% and 219 to 3788 bp, respectively (Figure 2). The total length of the honeybee numts fragment was 153306 bp, approximately 9.64-times longer than the entire honeybee mitochondrial genome. However, most numts identified corresponded with mitochondrial gene fragments. We identified 101 relatively complete mitochondrial gene regions distributed in 236 numts, in which 13 mitochondria protein-coding genes (ND1, ND2, CoxI, CoxII, CoIII, ATP8, ATP6, ND3, ND4L, ND4, ND5, ND6, and Cytb); all the mt-tRNA genes; and excluding srRNA (small and large subunit ribosomal RNA) were detected.
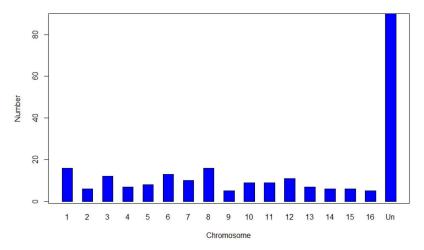

Figure 1. Number of numts in each chromosome. The y-coordinate represents number of numts within chromosomes: the x-coordinate is the chromosome number; Un is numts unmapped to any chromosome.
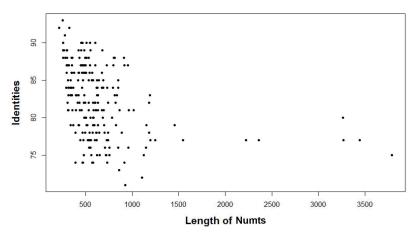

Figure 2. Distributions of the numt lengths and similarities to mtDNA of the honeybee.

DISCUSSION

Although we did not examine the functionality of numts detected here, high similarity between numts and mtDNA caution us that numts may participate in the function and regulation of genes; there is, therefore, a need to study the possible functions of numts. One such study verified that numts may affect DNA barcoding (Kim et al., 2013). In analyzing mtDNA, the mtDNA fragments themselves will greatly predominate detected copy numbers, which is unlikely to influence study outcomes in routine applications. There are few specific situations under which primers would match with higher efficacy to numts than to target mtDNA, resulting in contaminated mtDNA products due to preferential amplification of numts (Willerslev and Cooper, 2005; Liu and Zhao, 2007). The D-loop region is often used for the analysis of evolutionary relationships among different species and indicated possible domestic sites of species (Zhang et al., 1995; Englund et al., 2014; Bloor et al., 2015). The complete numt fragment corresponding to the most variable region of the mitochondrial DNA was not detected in honeybees. Therefore, numts of the D-loop would not influence results relating to this region in the current study.

It is commonly believed that numts were formed by non-homologous recombination (Woischnik and Moraes, 2002). Once integrated in the nuclear genome, numts and their mitochondrial counterparts evolved, showing different patterns (Kazazian, 2004). These different patterns may be useful for analyzing the relationship between numts and their corresponding mitochondrial DNA sequences; for example, the time point of mtDNA insertion in the nuclear genome, and the evolutionary relationship of the same sequence located in mitochondrial and nuclear genomes (Kim et al., 2006). Therefore, numts are an important tool for phylogenetic analysis (Hazkani-Covo et al., 2010), although common mechanisms responsible for mitochondrial insertions into nuclear genomes remain unclear at this time.

CONCLUSION

We identified 236 highly homogenous numt fragments, between the mitochondrial and nuclear genome of the honeybee. These numts included mostly of tRNAs and protein coding genes, except rRNAs in mitochondrial DNA. Therefore, these relatively complete numts could serve as a powerful tool to analysis the relationship of different species of honeybee and reconstruct phylogenetic network by comparing their mtDNA counterparts.

Conflicts of interest

The authors declare no conflict of interest to each other.

ACKNOWLEDGMENTS

Supplementary material

REFERENCES

Behura SK (2007). Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome. *Mol. Biol. Evol*. 24: 1492-1505.
Bensasson D, Zhang D, Hartl DL and Hewitt GM (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends*

Ecol. Evol. 16: 314-321.

Bloor P, Ibáñez C and Viloria-Lagares TA (2015). Mitochondrial DNA analysis reveals hidden genetic diversity in captive populations of the threatened American crocodile (*Crocodylus acutus*) in Colombia. *Ecol. Evol.* 5: 130-40.

Cleaver JE, Brennan-Minnella AM, Swanson RA, Fong KW, et al. (2014). Mitochondrial reactive oxygen species are scavenged by Cockayne syndrome B protein in human fibroblasts without nuclear DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* 111: 13487-13492.

Dacke M and Srinivasan MV (2007). Honeybee navigation: distance estimation in the third dimension. *J. Exp. Biol.* 210: 845-853.

du Buy HG and Riley FL (1967). Hybridization between the nuclear and kinetoplast DNA's of *Leishmania enriettii* and between nuclear and mitochondrial DNAs of mouse liver. *Proc. Proc. Natl. Acad. Sci. U. S. A.* 57: 790-797.

Dyer AG, Neumeyer C and Chittka L (2005). Honeybee (*Apis mellifera*) vision can discriminate between and recognise images of human faces. *J. Exp. Biol.* 208: 4709-4714.

Englund T, Strömstedt L and Johansson AM (2014). Relatedness and diversity of nine Swedish local chicken breeds as indicated by the mtDNA D-loop. *Hereditas* 151: 229-33.

Evans JD and Lopez DL (2002). Complete mitochondrial DNA sequence of the important honeybee pest, *Varroa destructor* (*Acari: Varroidae*). *Exp. Appl. Acarol.* 27: 69-78.

Hassanin A, Bonillo C, Nguyen BX and Cruaud C (2010). Comparisons between mitochondrial genomes of domestic goat (*Capra hircus*) reveal the presence of numts and multiple sequencing errors. *Mitochondrial DNA* 21: 68-76.

Hazkani-Covo E, Zeller RM and Martin W (2010). Molecular poltergeists: mitochondrial DNA copies (NUMTs) in sequenced nuclear genomes. *PloS Genet.* 6: e1000834.

Gibson JD and Hunt GJ (2014). The complete mitochondrial genome of the invasive Africanized Honeybee, *Apis mellifera scutellata* (*Insecta: Hymenoptera: Apidae*). *Mitochondrial DNA* DOI: 10.3109/19401736.2014.905858.

Kazazian HH Jr (2004). Mobile elements: drivers of genome evolution. *Science* 303: 1626-32.

Kim JH, Antunes A, Luo SJ, Menninger J, et al. (2006). Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. *Gene* 366: 292-302.

Kim SJ, Lee KY and Ju SJ (2013). Nuclear mitochondrial pseudogenes in *Austinograea alayseae* hydrothermal vent crabs (*Crustacea: Bythograeidae*): effects on DNA barcoding. *Mol. Ecol. Resour.* 13: 781-787.

Leister D (2005). Origin, evolution and genetics effects of nuclear insertion of organelle DNA. *Trends Genet.* 21: 655-63.

Lopez JV, Yuhki N, Masuda R, Modi W, et al. (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.

Nergadze SG, Lupotto M, Pellanda P, Santagostino M, et al. (2010). Mitochondrial DNA insertions in the nuclear horse genome. *Anim. Genet.* 41 (Suppl 2): 176-185.

Pereira SL and Baker AJ (2004). Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. *BMC Evol. Biol.* 25: 4-17.

Ricchetti M, Tekaia F and Dujon B (2004). Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol.* 2: E273.

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

Schmitz J, Piskurek O and Zischler H (2005). Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates. *J. Mol. Evol.* 61: 1-11.

Somanathan H, Warrant EJ, Borges RM, Wallèn R, et al. (2009). Resolution and sensitivity of the eyes of the Asian honeybees *Apis florea*, *Apis cerana* and *Apis dorsata*. *J. Exp. Biol.* 212: 2448-2453.

Song S, Jiang F, Yuan J, Guo W, et al. (2013). Exceptionally high cumulative percentage of NUMTs originating from linear mitochondrial DNA molecules in the *Hydra magnipapillata* genome. *BMC Genomics*. 14: 447.

Tan HW, Liu GH, Dong X, Lin RQ, et al. (2011). The complete mitochondrial genome of the Asiatic cavity-nesting honeybee *Apis cerana* (*Hymenoptera: Apidae*). *PLoS One* 6: e23008.

Thalmann O, Hebler J, Poinar HN, Paabo S, et al. (2004). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Mol. Ecol.* 13: 321-335.

The Honeybee Genome Sequencing Consortium (2006). Insights into social insects from the genome of the honeybee *Apis mellifera. Nature* 443: 931-949.

Wallace DC, Stugard C, Murdock D, Schurr T, et al. (1997). Ancient mtDNA sequences in the human nuclear genome: A potential source of errors in identifying pathogenic mutations. *Proc. Natl. Acad. Sci. U. S. A.* 94: 14900-14905.

Wang D, Liu L, Zhu X, Wu W, et al. (2014). Hesperidin alleviates cognitive impairment, mitochondrial dysfunction and oxidative stress in a mouse model of Alzheimer's disease. *Cell. Mol. Neurobiol.* 34: 1209-1221.

Willerslev E and Cooper A (2005). Ancient DNA. *Proc. Biol. Sci.* 272: 3-16.

Woischnik M and Moraes CT (2002). Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* 12: 885-893.

Zhang DX, Szymura JM and Hewitt GM (1995). Evolution and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* 40: 382-91.