Acta Oecologica xxx (2013) 1–8

Contents lists available at ScienceDirect

Acta Oecologica

journal homepage: www.elsevier.com/locate/actoec

Original article

Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*

Yan Chen [a,b], Min Liu [a], Stephen G. Compton [c], Xiao-Yong Chen [a,*]

[a] School of Resources and Environmental Sciences, Tiantong National Station of Forest Ecosystem, Shanghai Key Lab for Urban Ecological Processes and Eco-Restoration, East China Normal University, Dongchuan Road 500, Shanghai 200241, China
[b] Ecological Security and Protection Key laboratory of Sichuan Province, Mianyang Normal University, Mianyang, Sichuan 621000, China
[c] School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom

ARTICLE INFO

Article history:
Received 31 October 2012
Accepted 4 October 2013
Available online xxx

Keywords:
Creeping fig
Ctyb gene
NUMTs
Southeastern China
Wiebesi

ABSTRACT

Nuclear mitochondrial pseudogenes (NUMTs) are nuclear sequences transferred from mitochondrial genomes. Although widespread, their distribution patterns among populations or closely related species are rarely documented. We amplified and sequenced the mitochondrial cytochrome b (*Cytb*) gene to check for NUMTs in three fig wasp species that pollinate *Ficus pumila* (*Wiebesia* sp. 1, 2 and 3) in Southeastern China using direct and cloned sequencing. Unambiguous sequences (332) of 487 bp in length belonging to 33 haplotypes were found by direct sequencing. Their distribution was highly concordant with those of cytochrome c oxidase subunit I (*COI*). Obvious signs of co-amplification of NUMTs were indicated by their uneven distribution. NUMTs were observed in all individuals of 12 populations of *Wiebesia* sp. 3, and 13 individuals of three northern populations of *Wiebesia* sp. 1. Sequencing clones of potential co-amplification products confirmed that they were NUMTs. These NUMTs either clustered as NUMT clades basal to mtDNA *Cytb* clades (basal NUMTs), or together with *Cytb* haplotypes. Basal NUMTs had either stop codons or frame-shifting mutations resulting from deletion of a 106 bp fragment. In addition, no third codon or synonymous substitutions were detected within each NUMT clade. The phylogenetic tree indicated that basal NUMTs had been inserted into nuclei before divergence of the three species. No significant pairwise differences were detected in their ratios of third codon substitutions, suggesting that these NUMTs originated from one transfer event, with duplication in the nuclear genome resulting in the coexistence of the 381 bp copy. No significant substitution differences were detected between *Cytb* haplotypes and NUMTs that clustered with *Cytb* haplotypes. However, these NUMTs coexisted with *Cytb* haplotypes in multiple populations, suggesting that these NUMT haplotypes were recently inserted into the nuclear genome. Both basal and recently inserted NUMTs were rare events, and were absent in most populations of *Wiebesia* sp. 1 and 2. Further studies are needed to distinguish between mechanisms potentially generating this rarity, such as purifying selection, genetic drift or amplification failure.

1. Introduction

Nuclear mitochondrial pseudogenes (NUMTs, pronounced "new-mights") are nuclear sequences that originate from the transfer of mitochondrial genomes (Lopez et al., 1994). They can be co-amplified with, sometimes even in preference to, their paralogous mitochondrial genes when using universal mitochondrial primers (Zhang and Hewitt, 1996). 'Special' NUMTs have been inserted into coding or regulatory regions in some taxa (Ricchetti et al., 2004; Erpenbeck et al., 2011), resulting in potential fitness consequences. For example, in humans, such NUMTs are associated with genetic diseases (Hazkani-Covo et al., 2010; Chatre and Ricchetti, 2011). However, most NUMTs occur in non-coding DNA regions, and larger genomes have more non-coding regions, thereby facilitating NUMTs insertions (Hazkani-Covo et al., 2010). Consequently, total numbers of NUMTs are strongly correlated with genome sizes (Bensasson et al., 2001; Hazkani-Covo et al., 2010), a correlation that even exists among closely related, congeneric species (Rogers and Griffiths-Jones, 2012).

NUMTs have been reported in more than 80 species, covering all well-studied taxa (Bensasson et al., 2001; Hazkani-Covo et al., 2010). The properties and distributions of NUMTs vary among

* Corresponding author. Tel.: +86 21 5434 1125; fax: +86 21 5434 1125.
E-mail address: xychens@hotmail.com (X.-Y. Chen).

1146-609X/$ – see front matter © 2013 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.actao.2013.10.001

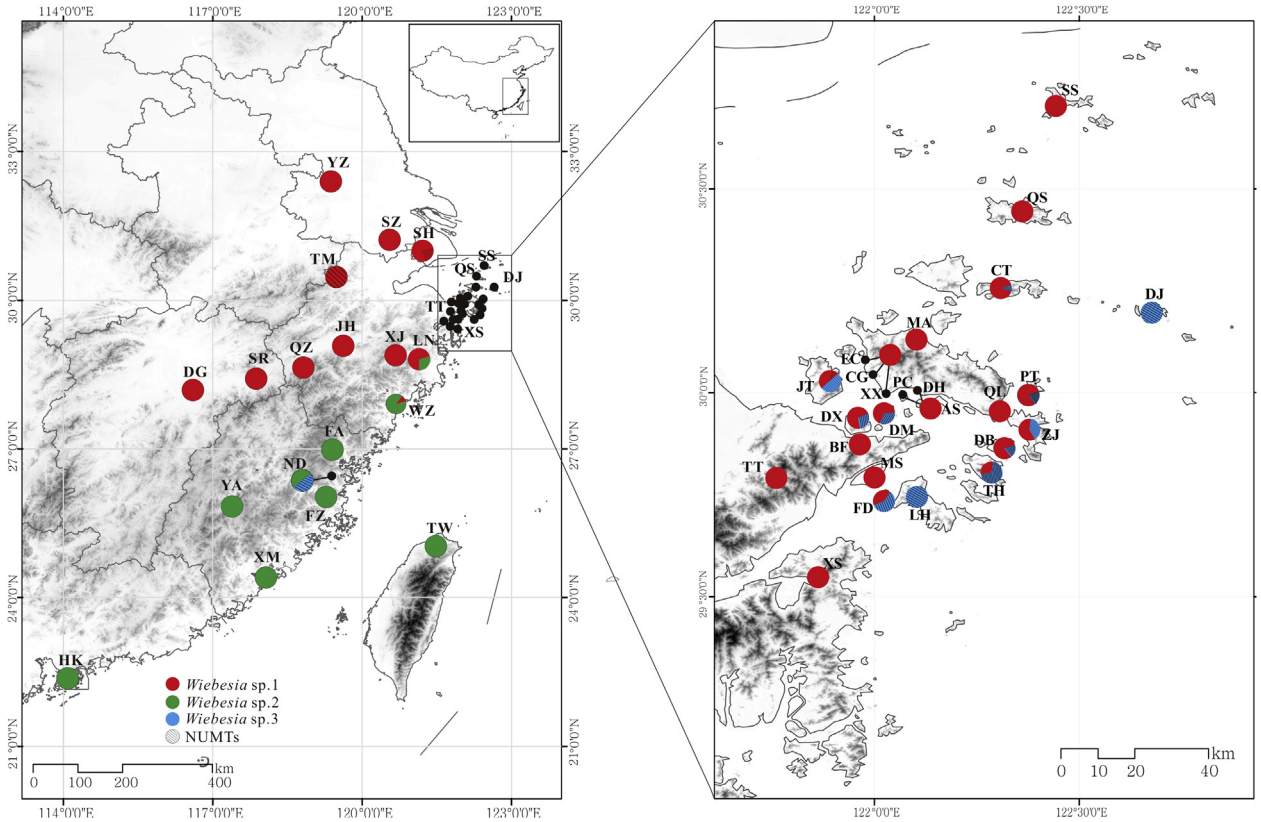

Fig. 1. Distribution of NUMTs in three pollinating fig wasps of *Ficus pumila*. Red, green and blue pies indicate *Wiebesia* sp. 1, 2 and 3, respectively. Patterns in the pies indicate the percentages of NUMTs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

species and populations (Song et al., 2008; Jacques et al., 2010). Song et al. (2008) suggested the transference and insertion of mtDNA happens at the level of individuals, so the prevalence of NUMTs can be both species- and population-specific. The distribution patterns of special NUMTs have been used to infer the phylogenetic relationships of some taxa (Mishmar et al., 2004; Karanth, 2008; Hazkani-Covo, 2009), but very few studies have explored intra-specific variation of NUMTs and the information that their distribution pattern supplies (Jacques et al., 2010). Consequently, patterns of NUMT acquisition and spread are poorly understood.

It has been suggested that NUMTs are dead-on-arrival pseudogenes, which lose their function once they are translocated into nuclear genomes (Bensasson et al., 2001). This implies that their persistence provides no evolutionary cost or benefit (Gherman et al., 2007) and that their acquisition and retention may be driven more by genetic drift than selection (Gherman et al., 2007). This suggestion is consistent with the recorded correlation between the existence of lineage-specific NUMTs and historical bottleneck events in a range of taxa (Gherman et al., 2007; Triant and DeWoody, 2007; Lang et al., 2012).

Based on mitochondrial cytochrome c oxidase subunit I (*COI*) and nuclear internal transcribed spacer (*ITS*) sequences, we have shown that three closely related 'cryptic' species of fig wasps pollinate the figs of *Ficus pumila* in Southeastern China (*Wiebesia* sp. 1, 2 and 3) (Chen et al., 2012). During further studies of mitochondrial cytochrome b (*Cytb*) genes, ambiguous chromatograms with double peaks in some nucleotide sites were often found in some populations when they were sequenced directly, indicating the potential occurrence of NUMTs. In this study, we sampled and sequenced the *Cytb* gene of the three pollinating wasps, in order to answer the following questions. (1) Were the ambiguous chromatograms induced by the co-amplification of the *Cytb* gene and its NUMTs? (2) If NUMTs exist, when did they arise? (3) If they arose prior to the divergence of some or all of the species, do they still exist in all offspring populations?

2. Materials and methods

2.1. Sample collection and DNA preparation

F. pumila L. (Moraceae) is a functionally dioecious evergreen perennial root-climber, with a natural distribution from southern Japan and China to Vietnam (Liu et al., 2013). Two varieties, *F. pumila* var. *pumila* and *F. pumila* var. *awkeotsang*, have been reported. *F. pumila* var. *awkeotsang* is mainly found in Taiwan Island, and shares a pollinator with *F. pumila* var. *pumila* (Wang et al., 2009; Chen et al., 2012). In this study, we focused on the pollinating wasps associated with *F. pumila* var. *pumila*. Between 2005 and 2010, mature male figs of *F. p.* var. *pumila* were collected from 43 populations in Southeastern China (Fig. 1), and were kept in fine-mesh bags to let the adult fig wasps emerge. They were preserved in absolute ethanol and stored in a refrigerator at 4 °C. Total genomic, but not purified mitochondrial genomic DNA was extracted from the whole bodies of single female wasps by a method modified from Sambrook et al. (1989) due to the small body size of fig wasps (less than 2 mm). In total, 409 individuals from 367 *F. pumila* figs were analyzed (more detailed information is available in Chen et al., 2012).

2.2. Direct and cloned sequencing of mitochondrial Cytb genes

A 487-bp fragment of the mitochondrial *Cytb* gene was amplified using the universal primer pair CB-J-10933/CB-N-11367 (CB1/

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001

Table 1
The distributions of sequences with double peaks in direct sequencing chromatograms among populations of pollinating fig wasps from *Ficus pumila* in Southeastern China.

Population code	Latitude	Species	No. of unambiguous sequences (Haplotypes)	No. of sequences with two bands in agarose gels	No. of sequences with double peaks
YZ	32°24′N	sp. 1	10 (H09)	0	0
SZ	31°18′N	sp. 1	10 (H09)	0	0
SH	31°05′N	sp. 1	5 (H09)	0	2
SS	30°42′N	sp. 1	10 (H01, H04)	0	0
TM	30°31′N	sp. 1	0	0	10
QS	30°16′N	sp. 1	8 (H02, H04, H07)	0	0
CT	30°14′N	sp. 1	9 (H01–H04)	0	0
		sp. 3	0	1	1
DJ	30°14′N	sp. 3	0	10	10
MA	30°08′N	sp. 1	10 (H02, H03)	0	0
FC	30°05′N	sp. 1	8 (H01–H03, H23)	0	0
CG	30°03′N	sp. 1	10 (H01–H03, H05, H08)	0	0
JT	30°03′N	sp. 1	6 (H01–H03)	0	1
		sp. 3	1 (H25)	3	2
DH	30°00′N	sp. 1	10 (H02, H03, H15)	0	0
PT	30°00′N	sp. 1	8 (H03, H04)	0	0
		sp. 3	0	2	2
AS	29°59′N	sp. 1	9 (H02–H04)	0	0
PC	29°59′N	sp. 1	10 (H01–H05)	0	0
XX	29°59′N	sp. 1	10 (H01–H03, H05, H13)	0	0
QL	29°58′N	sp. 1	10 (H01, H04, H07)	0	0
DM	29°57′N	sp. 1	7 (H01–H04)	0	0
		sp. 3	0	2	2
DX	29°56′N	sp. 1	6 (H01, H02, H08)	0	0
		sp. 3	0	2	2
ZJ	29°54′N	sp. 1	7 (H04)	0	0
		sp. 3	2 (H26)	3	1
BF	29°52′N	sp. 1	9 (H01, H02, H08, H12)	0	0
DB	29°52′N	sp. 1	8 (H02–H04)	0	0
		sp. 3	0	2	2
TH	29°48′N	sp. 1	3 (H01, H09)	0	0
		sp. 3	1 (H26)	7	6
MS	29°47′N	sp. 1	10 (H01, H02)	0	0
TT	29°47′N	sp. 1	8 (H01, H02, H18, H23)	0	0
FD	29°44′N	sp. 3	0	10	10
LH	29°44′N	sp. 1	4 (H01, H03, H19)	0	0
		sp. 3	0	7	6
XS	29°29′N	sp. 1	10 (H01, H02)	0	0
JH	29°09′N	sp. 1	10 (H03, H09–H11, H23)	0	0
XJ	28°56′N	sp. 1	10 (H03, H06, H12, H13, H16)	0	0
LN	28°53′N	sp. 1	6 (H06, H12, H14)	0	0
		sp. 2	4 (H21)	0	0
QZ	28°42′N	sp. 1	10 (H09, H27)	0	0
SR	28°30′N	sp. 1	7 (H27, H32)	0	0
DG	28°15′N	sp. 1	10 (H09, H27, H28)	0	0
WZ	27°54′N	sp. 1	1 (H27)	0	0
		sp. 2	7 (H21, H33, H34)	0	0
FA	27°04′N	sp. 2	10 (H21–H23)	0	0
ND	26°26′N	sp. 2	4 (H20–H22)	0	0
		sp. 3	0	5	5
FZ	26°09′N	sp. 2	9 (H21, H22, H29, H30)	0	0
YA	25°54′N	sp. 2	10 (H22, H36)	0	0
TW	25°02′N	sp. 2	0	0	0
XM	24°26′N	sp. 2	9 (H22, H29, H35)	0	0
HK	22°28′N	sp. 2	9 (H22, H29, H31)	0	0

CB2) (Simon et al., 1994). The PCR amplification was carried out in 50 μL volumes on PTC-220 DNA engine DYAD (MJ Research, USA), containing 1 mM $MgCl_2$, 0.25 mM of each dNTP, 0.4 μM of each primer, and 1.25 U PrimerSTAR™ HS Taq DNA polymerase (Takara, Dalian). The thermal program was 94 °C for 3 min; 30 cycles of 94 °C for 30 s, 54 °C for 45 s, 72 °C for 1 min; followed by a final extension of 72 °C for 4 min.

The PCR amplicons were sequenced directly on an ABI PRISM 3730 machine from both directions. The PCR amplicons of some individuals were detected with two bands on agarose gels (about 490 and 380 bp, respectively). The longer band was purified from 1.5% agarose gels using a DP1501 kit (BioTeke, Beijing, China) and used for direct sequencing. Only sequences whose chromatograms from both directions were unambiguous and well-matched were regarded as potential mitochondrial *Cytb* haplotypes and used for further analysis.

Double or triple peaks in certain nucleotide sites were found in chromatograms of some individuals. Their presence in each population was recorded, and seven individuals were selected for cloned sequencing based on their species status, as determined by their mitochondrial *COI* genes (Chen et al., 2012). The amplification protocol was the same as used in direct sequencing. PCR products were purified from 1.5% agarose gels using the DP1501 kit (BioTeke, Beijing, China) and ligated into a pMD 18-T cloning vector (Takara, Dalian, China), then transferred into *Escherichia coli* JM109 (Takara, Dalian, China). Ten to 18 colonies were selected from each individual to be sequenced by an ABI PRISM 3730 machine.

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001

Table 2
The haplotypes of cloned sequences from seven pollinating fig wasp individuals from *F. pumila*. The series L and S were detected only by cloned sequencing, whereas the haplotypes indicated by H were also identified in other individuals by direct sequencing. The copy numbers of each haplotype are presented in parentheses. * indicates haplotypes that have stop codons.

Individuals	Species	Number of sequenced clones	Mitochondrial *Cytb* haplotypes	NUMTs haplotypes		
				487 bp		381 bp
JT05	sp. 1	18	H01 (8)	L01 (2)	L02 (1)	/
TM11	sp. 1	16	H09 (6)	L03 (1)	L04 (4)*	/
SH02	sp. 1	10	H09 (4)	/	L04 (4)*	/
CT10	sp. 3	15	L05 (2)	L08 (2)*	L06 (1)*	S11 (6)
PT06	sp. 3	17	H26 (6)	L07 (1)	L08 (2)*	S11 (4)
TH13	sp. 3	12	H26 (3)	L10 (1)	L09 (1)*	S11 (4)
FD24	sp. 3	12	H26 (6)	L10 (3)	/	S11 (2)

All sequences were aligned using ClustalX 2.0 (Larkin et al., 2007). All the singletons (nucleotide variations unique to one sequence) were regarded as Taq errors (Baldo et al., 2011), and their host sequences were omitted in subsequent analyses.

2.3. Phylogenetic analyses

Neighbour joining trees were initially reconstructed using Mega 4 (Tamura et al., 2007). One of the sequences in each monophyletic clade was used to search for similar sequences in GenBank using a MegaBlast strategy. The top 10 sequences in each run of blast were downloaded. The documented pollinator of *F. pumila*, *Wiebesia pumilae*, has been described as being closely related to some *Blastophaga* species (Machado et al., 2001; Cruaud et al., 2010). *Cytb* sequences of species in the genera *Wiebesia* and *Blastophaga* in GenBank were therefore also downloaded. All the GenBank sequences were aligned with our sequences and used as outgroups in phylogenetic analyses.

Bayesian trees were constructed to assess phylogenetic relationships, based on the sequences obtained. BEAST v.1.6.1 (Drummond and Rambaut, 2007; Heled and Drummond, 2010) was run to explore the appropriate models. Bayes Factors (BF, estimated as the harmonic mean of the posterior likelihoods of each model) (Suchard et al., 2001), were calculated by TRACER 1.5 (Rambaut and Drummond, 2007), and used as criteria based on the lnBF table of Kass and Raftery (1995). The combination of each substitution, clock and population models was tried step by step. lnBFs indicated that a combination of Hasegawa–Kishino–Yano (HKY), strict clock and expansion models was the best. Two independent runs of 50 million iterations were then performed using the models determined with BEAST, sampling the genealogies and model parameters every 1000 iterations. The phylogenetic tree was summarized by TreeAnnotator v.1.6.1, and then viewed using FigTree v.1.3.1 (Rambaut, 2006). Phylogenetic networks of all our samples and the haplotype frequencies were constructed with TCS 1.2.1 (Clement et al., 2000).

2.4. Within-clade and pairwise substitution patterns

When mitochondrial genes are integrated into non-coding regions of nuclear genomes they will lose their function and become dead-on-arrival pseudogenes. As they are free from selection, their substitution rates are expected to be different in mitochondrial and nuclear genomes and the ratios of the third codon substitutions to the first and second codon substitutions should also change. We translated the sequences into amino acid sequences using the invertebrate mitochondrial genetic code to check for stop codons and indels (Song et al., 2008). The first, second and third codon substitutions within clades were calculated using Mega 4. The percentages of the mutations in each codon position were then compared among clades. The pairwise differences in third codon substitution ratios were tested using post hoc multiple comparisons with Bonferroni correction using R 2.15.0 (R Development Core Team, 2012). Ratios of transitions to transversions (T_S/T_V) and the substitutions in each codon position were also calculated with Mega 4.0, as were the percentages of non-synonymous mutations (d*S*) and synonymous mutations (d*N*). The Kimura 2-parameter (K2P) distances between each species were also calculated using Mega.

3. Results

3.1. Direct and cloned sequence distributions

Two bands were observed in all PCR amplicons of individuals of *Wiebesia* sp. 3 using agarose gels (one about 490 bp, another about 380 bp), whereas only one band (about 490 bp) was present in amplicons of *Wiebesia* sp. 1 and 2 (Table 1; Supplementary Material Fig. S1). The 332 directly sequenced sequences were unambiguous and well-matched in two directions. Among them, 33 haplotypes of 487 bp were obtained (named as H haplotypes, in subsequent analyses; GenBank accession numbers: KF657678–KF657710). Most of them (26 out of 33 H haplotypes) were represented by multiple individuals and populations (Table 1). H01 was the most common haplotype in *Wiebesia* sp. 1, shared by 51 individuals from 16 populations, whereas H22 was the most common haplotype in *Wiebesia* sp. 2, displayed by 23 individuals from 6 populations. Only two H haplotypes (H25 and H26) were detected in individuals of *Wiebesia* sp. 3.

Sequence ambiguities were detected in 13 individuals from three populations (SH, TM, JT) of *Wiebesia* sp. 1 and 49 individuals from 12 populations of *Wiebesia* sp. 3 (Table 1). One individual from each of the three populations of *Wiebesia* sp. 1 (SH02, TM11, JT05) and four individuals (PT06, CT10, FD24, TH13) of *Wiebesia* sp. 3 were cloned and sequenced. A fragment of 487 bp in length was obtained from all individuals, and an additional 381 bp fragment was found in four individuals of *Wiebesia* sp. 3, showing a deletion of 106 bp. After excluding haplotypes with singletons, a total of 11 haplotypes were obtained (10 haplotypes of 487 bp in length named as L haplotypes, and one 381 bp haplotype named as S haplotype in the following descriptions; GenBank accession numbers: KF657711–KF657721). Among them, five L haplotypes and the single S haplotype were detected in more than one cloned sequence from the same (L01, L05) or different (L04, L08, L10, S11) individuals (Table 2). All individuals except CT10 had a haplotype which had been detected in multiple populations by direct sequencing (H haplotype), however, three L haplotypes were detected in CT10 (Table 2, Fig. 2). Five out of the seven individuals of cloned sequencing were found to have another two copies of L haplotypes, while another two individuals (SH02, FD24) had only one L haplotype. All four individuals of *Wiebesia* sp. 3 shared one S haplotype (Table 2).

3.2. Phylogenetic relationship of the haplotypes

The Bayesian tree clustered all haplotypes of the pollinating fig wasps of *F. pumila* (33 H haplotypes obtained from direct and cloned sequences, 10 L and 1 S haplotypes only obtained from cloned sequences and a GenBank sequence (GQ368014) of *Cytb* gene of *Wiebesia pumilae*) into 6 clades, separately from 35 GenBank sequences of pollinating wasps of other *Ficus* species (Fig. 2).

Within the sequences reared from pollinators of *F. pumila*, all cloned sequences with stop codons or deletions formed two clades

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001

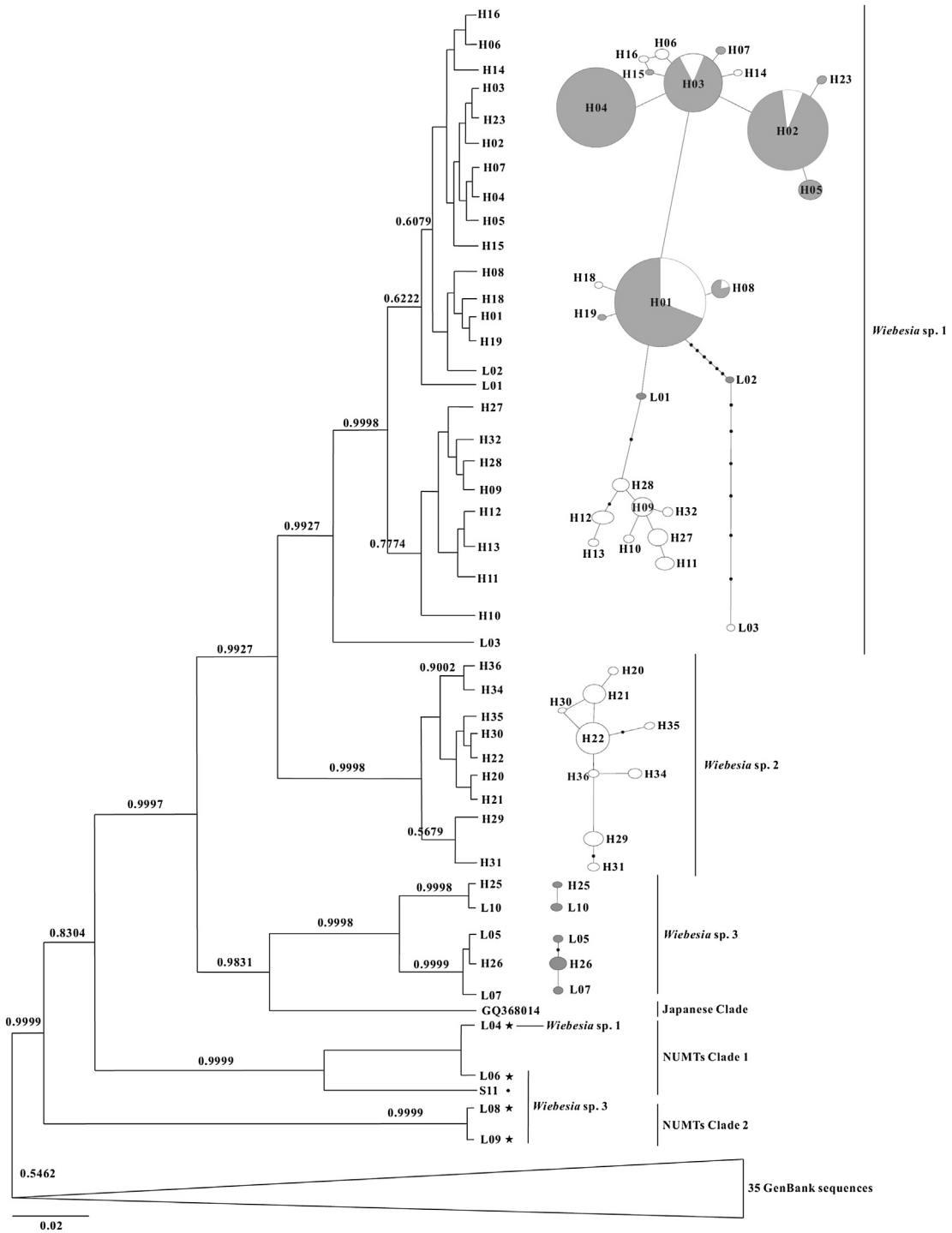

Fig. 2. The phylogenetic tree and networks of sequences for the pollinating fig wasps from *F. pumila*. Left: the Bayesian tree of the haplotypes of the mitochondrial *Cytb* gene and its NUMTs. Thirty-six *Cytb* sequences from GenBank are also included. Posterior probability node supports of the main branches are shown above them. Haplotype series of H were exposed by direct and cloned sequencing, while those of L and S were only identified from cloned sequences. Stars and dots highlight the haplotypes with stop codons and deletions, respectively. The corresponding mitochondrial *COI* clades of each individual as described in Chen et al. (2012) are indicated for *Wiebesia* sp. 1, 2 or 3. Right: the networks of each *Cytb* clade. The areas of each circle indicate the haplotype frequencies, highlighting frequencies of mainland and island populations with white and grey respectively. Black dots indicate missing haplotypes.

(NUMT clades 1, 2) basal to other haplotypes, indicating they were obtained before the divergence of three pollinator species. All H haplotypes and six L haplotyes clustered into four monophyletic ingroup clades with high posterior probability node supports (>0.98). The topology, individual attributions (Fig. 2), and pairwise genetic distances within and between ingroup clades (Table 3) were highly concordant with those of the mitochondrial *COI* genes of the three pollinating wasps. The sequence of *Wiebesia pumilae* (GQ368014) downloaded from GenBank clustered initially with *Wiebesia* sp. 3. However, its genetic distances to *Wiebesia* sp. 1, 2

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001

Table 3
K2P distances within and between four clades of the pollinator fig wasps associated with *F. pumila*.

	Wiebesia sp. 1	*Wiebesia* sp. 2	*Wiebesia* sp. 3	Japanese clade
Wiebesia sp. 1	0.011			
Wiebesia sp. 2	0.073	0.007		
Wiebesia sp. 3	0.114	0.120	0.029	
Japanese clade	0.120	0.114	0.118	0.008

and 3 were equivalent to those between these three species, suggesting that this single sequence (from a specimen collected from Naha City, Okinawa, Japan) may represent an additional fourth clade of pollinators associated with *F. pumila*. The ingroup clades will be referred to as the *Wiebesia* sp. 1, 2, 3 and Japanese clades, accordingly (Fig. 2).

The networks of ingroup clades showed similar phylogenies to those revealed by *COI* genes (Chen et al., 2012). The network of *Wiebesia* sp. 1 was typically star-like. Almost all haplotypes were separated from the most common four haplotypes (H01–H04) by only a few mutations. Island haplotypes were especially close, as all departed from the four dominant haplotypes by only one mutation. However, two haplotypes revealed by cloned sequencing (L02, L03) were separated from the dominant haplotypes by many more mutations. There were no obvious dominant haplotypes in *Wiebesia* sp. 2 and 3, although the network of haplotypes of *Wiebesia* sp. 2 showed loops. The haplotypes of *Wiebesia* sp. 3 could not all be connected together under a 95% statistical parsimony criterion, which may indicate that there have been too many lost haplotypes to allow their connection.

NUMT clades contained individuals of both *Wiebesia* sp. 1 and 3. Only the cloned sequences of JT05 were clustered with *Cytb* haplotypes of *Wiebesia* sp. 1 (Fig. 2).

3.3. Substitution patterns of each clade

Five basal haplotypes (L04, L06, L08, L09, S11) showed signs of pseudogenization. Four of these L haplotypes were found to include stop codons when translated with the invertebrate mitochondrial genetic code, with two in the 73rd amino position and two in the 139th amino position. The S haplotype showed a 106 bp deletion, and thus a frame-shifting mutation. All *Cytb* haplotypes of the *Wiebesia* spp. had neither in-frame stop codons nor frame-shifting mutations (Table 2).

The within-clade substitution patterns were also different between basal NUMT and *Cytb* clades. The *Cytb* clades showed a strong bias towards third codon position substitutions and synonymous substitutions, but no third codon and synonymous substitutions were detected within the two NUMT clades (Table 4). When differences in the ratios of substitutions in third codon positions to those in first and second codon positions were tested pairwise, all haplotypes of NUMT clades showed significant differences in the ratio of third codon substitution with most ingroup haplotypes, but no significant differences between themselves. However, one *Cytb* haplotype (H07) also showed significant differences with four of the H haplotypes and haplotype L02 (Table S1).

The other six L haplotypes (L01–L03, L05, L07, L10), which clustered within *Cytb* clades, showed no significant differences in substitution characteristics with H haplotypes. However, differences in distribution patterns were detected between them and the H haplotypes (Table 2).

4. Discussion

4.1. Co-amplification of the mitochondrial Cytb gene and its NUMTs

NUMTs can be co-amplified, or even amplified in preference to their corresponding mtDNA genes (Bensasson et al., 2001; Song et al., 2008). Evidence for co-amplifications may be seen in gel electrophoresis and sequence chromatograms, with ghost bands, double peaks and high noise in chromatograms (Song et al., 2008). Such signs were obvious in the PCR and sequence procedures of the *Cytb* genes in populations of the pollinating fig wasps associated with *F. pumila*; two agarose gel bands (Fig. S1) and double chromatogram peaks were detected in the PCR products of some individuals. NUMTs were verified in their cloned sequences by the phylogenetic structure and substitution patterns.

H haplotypes were detected in all individuals whose chromatograms were unambiguous by direct sequencing, and in almost all cloned individuals except CT10. Most of them (26 out of 33) were found in multiple populations or multiple individuals of one population (Tables 1 and 2). The substitutions showed a strong bias towards third codon location and transitions among H haplotypes (Table 4), and no significant differences in the ratios of third codon substitutions were detected between most haplotypes when pairwise differences were tested. The distribution pattern and substitution patterns of H haplotypes suggested they were mitochondrial *Cytb* genes. However, one H haplotype (H7) showed a significant difference in third codon substitution ratio with some *Cytb* haplotypes (Table S1), failing one of the criteria for NUMTs transfer events (Bensasson et al., 2000; Baldo et al., 2011). Whether H07 represented a NUMT needs further studies.

Five cloned sequences clustered into two basal NUMT clades. All of them had in-frame stop codon or frame-shifting mutations (Table 2), suggesting that they were non-functional genes. No synonymous and third codon substitutions were detected between haplotypes within each NUMT clade (Table 4), and all sequences of NUMT clades had significant differences in substitution ratios of third codons when compared to most *Cytb* haplotypes. All these characteristics have been recognized as evidence for the presence of NUMTs (Mundy et al., 2000; Song et al., 2008; Baldo et al., 2011). Based on their basal position compared with *Cytb* sequences, we

Table 4
The substitution patterns of each *Wiebesia* sp. clade, showing average pairwise differences within clades. A GenBank Cytb sequence of *Wiebesia pumila* (GQ368014) is included, but the short cloned sequence S11 was excluded due to its frame-shifting mutation. 1st, 2nd and 3rd indicate the average substitution numbers in the first, second and third codon positions, respectively. T_S/T_V is the ratio of transition to transversion substitutions. d*S*/d*N* indicates the ratio of synonymous mutations to non-synonymous ones.

Clade	1st%	2nd%	3rd%	T_S	T_V	T_S/T_V	d*S*	*dN*	d*S*/d*N*	Number of haplotypeswith stop codons
Wiebesia sp. 1 ($N=25$)	15.33	0	84.67	4.89	0.54	9.06	5.05	0.23	21.66	0
Wiebesia sp. 2 ($N=9$)	15.67	7.84	76.49	2.39	0.44	5.43	2.39	0.44	5.38	0
Wiebesia sp. 3 ($N=5$)	8.45	0	91.55	12.40	1.80	6.89	14.20	0.00	/	0
Basal clade 1 (L04, L06)	0	100	0	1.00	0.00	/	0.00	1.00	0.00	2
Basal clade 2 (L08, L09)	66.67	33.33	0	1.00	2.00	0.50	0.00	3.00	0.00	2

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001

may conclude that these sequences of NUMT clades were inserted into ancestral nuclear genomes before the divergence of the three Chinese fig wasp species about 6 mya (Chen et al., 2012). The Japanese clade may represent a further species, based on the deep divergence from the other three pollinators, but more information is required to confirm this.

The third codon substitution ratios are expected to be significantly different between NUMT sequences if they were integrated into nuclear genomes at different times, because the mutation rates are different in nuclear and mitochondrial genomes and NUMTs are free from selection (Bensasson et al., 2000; Baldo et al., 2011). However, in our case, none of the five sequences of NUMT clades showed significant differences in their ratios of third codon substitutions with the others when pairwise sequences were compared, though the NUMTs were clustered as two clades. This suggests that these five basal NUMTs were obtained from one transfer event that took place before divergence of the cryptic pollinators of *F. pumila*. These sequences were then duplicated in the nuclear genomes, resulting in coexistence of two types of NUMTs (L and S haplotypes) in each individual of *Wiebesia* sp. 3.

We found no significant difference in substitution patterns between the L haplotypes clustered in *Cytb* clades and H haplotypes (Table 4). It was hard to tell whether these L haplotypes were mitochondrial or nuclear genes on the basis of the substitution patterns and phylogenetic relations. However, mitochondrial heteroplasmy could be rejected because of the unambiguous amplification of *COI* genes. Only one *COI* haplotype was detected in each individual, and neither ghost bands nor ambiguous chromatograms were observed. All *COI* haplotypes were clustered into three monophyletic clades with high posterior probabilities (over 0.99) (Chen et al., 2012). All these suggested there was only one mitochondrial genome in the pollinating wasps.

There are usually hundreds of mitochondria in each cell, so the copies of mitochondrial genes are typically more abundant than those of nuclear genes (Calvignac et al., 2011). According to their distribution patterns in each cloned individual, the five L haplotypes (L01, L02, L03, L07, L10) that clustered in *Cytb* clades were demonstrated to be NUMTs (Table 2). Those NUMTs showed no significant differences in substitution patterns with the H haplotypes, which may indicate that they have been transferred into nuclear genomes recently and have not had time to accumulate sufficient mutations (Bensasson et al., 2001). Recent studies have demonstrated that some 'cryptic' NUMTs can differ by only 1–3 bp with their paralogous mtDNA genes (Bertheau et al., 2011).

L05 was detected only in individual CT10, in which no H haplotype was found. In this individual, cloned sequencing also revealed two L and one S haplotypes (Table 2), which were in accord with the NUMT copies of other individuals of *Wiebesia* sp. 3, suggesting that L05 was a mitochondrial haplotype.

4.2. Distributions of NUMTs within species and populations

Although the basal NUMTs were integrated into the fig wasps' nuclear genomes before they had diverged, they are not distributed evenly within the three wasp species (Table 1). They were ubiquitous in *Wiebesia* sp. 3 populations, but rare, or even absent, in most populations of *Wiebesia* sp. 1 and 2. Smaller genome sizes and fewer non-coding regions are frequently used to explain variation in rates of NUMT deletions (Bensasson et al., 2001; Hazkani-Covo et al., 2010). However, they cannot explain the between-population differences within *Wiebesia* sp. 1, given that only three out of 34 populations had NUMTs, because it is difficult to imagine that their genome sizes and non-coding gene region ratios have been sufficiently different for them to account for the distribution patterns of NUMTs among the three *Wiebesia* species.

Other factors, such as selection, genetic drift, and different amplifications may also contribute to the distribution patterns of NUMTs in closely related species or populations. Potentially beneficial (Hazkani-Covo et al., 2010; Muradian et al., 2010) and deleterious (Hazkani-Covo et al., 2010; Chatre and Ricchetti, 2011) evolutionary impacts have been suggested as a consequence of NUMTs insertion, so natural selection would result in the persistence or loss of such NUMTs. However, neutral NUMTs have also been proposed (Bensasson et al., 2001; Hazkani-Covo et al., 2010), and the loss of such NUMTs could then be driven by stochastic events. Acquisitions of lineage-specific NUMTs have been found to be associated with bottleneck events over evolutionary time scales in some taxa (Triant and DeWoody, 2007; Lang et al., 2012), suggesting that genetic drift rather than selection has influenced their retentions (Gherman et al., 2007).

The three Chinese pollinators of *F. pumila* have undergone different historical demographic dynamics, as revealed by mitochondrial *COI*. *Wiebesia* sp. 3 has suffered from severe bottleneck events while *Wiebesia* sp. 1 and 2 have retained relatively stable population sizes since the last glacial maximum (Chen et al., 2012). The rarity of *Cytb* NUMTs in most populations of *Wiebesia* sp. 1 and 2 suggests that the NUMTs may be disadvantageous, and have been removed by purifying selection. Genetic drift in small population can fix even deleterious mutations (Travis et al., 2007; Burton and Travis, 2008), and thus the *Cytb* NUMTs could be persisting in *Wiebesia* sp. 3 and a few populations of *Wiebesia* sp. 1 near its expansion front (populations JT, TM, SH; their precise locations were given in Chen et al., 2012).

Amplification failure resulting from mutations in primer binding sites may have contributed to our failure to detect NUMTs in most populations of *Wiebesia* sp. 1 and 2. NUMTs are prone to mutations, and several point mutations or an indel of several bases or longer in the primer regions can lead to amplification failure. Given that several factors can influence the distribution patterns of NUMTs, revealing the mechanisms of *Cytb* NUMTs in the three *Wiebesia* species will need further studies.

Based on our results here and in a previous study (Chen et al., 2012), NUMTs were detected in *Cytb* but not in *COI* in the pollinating fig wasps of *F. pumila*. This does not mean that *Cytb* is more prone to be transferred to the nucleus than *COI*. The presence of NUMTs in target markers is both species-, and population-specific (Song et al., 2008). So far, no transfer of hotspot regions of mitochondrial genes has been verified (Bensasson et al., 2001), but recent genomic studies have revealed that NUMTs can cover most mitochondrial genomic regions (Hazkani-Covo et al., 2010; Erpenbeck et al., 2011).

4.3. Implications for phylogeographic/phylogenetic studies and DNA barcoding

Mitochondrial genes have been widely used in phylogeographic and phylogenetic studies and DNA barcoding. One fundamental assumption underlying these studies is the orthology of the sequences adopted. However, more and more NUMTs (nuclear paralogous sequences of mitochondrial genes) are being documented in a variety of taxa, complicating the inferences made by such studies, especially where they are based on just a short region of mitochondrial genes, such as with DNA barcoding. As Song et al. (2008), pointed out, the presence of NUMTs makes DNA barcoding overestimate species richness. Although ancient NUMTs, usually containing stop codons or indels of long fragments, can be relatively easy to identify and remove from further analyses of DNA barcoding, more recently divergent NUMTs without stop codons are difficult to identify. Therefore, additional sequences other than the target mtDNA gene should be analyzed,

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001

especially when a large number of NUMTs are suspected in a gene (Song et al., 2008).

Recently divergent NUMTs usually have one or a few mutations that make them different from the target mtDNA gene. If they are not excluded in the barcode, their occurrences may increase apparent intra-specific variation, thus reducing if not fully eliminating the gap between intra- and inter-specific genetic distances. Even if these NUMTs do not confuse species status, their occurrences will overestimate haplotype richness, potentially biasing inferences from phylogeographic studies. Therefore, it is necessary to use multiple markers, in combination with nuclear markers, when mitochondrial genes are used in phylogeographic studies.

5. Conclusions

The transference of mitochondrial genes into nuclei is an ongoing process and NUMTs have been found to be common in many taxa. Co-amplification of NUMTs and mitochondrial *Cytb* gene sequences in the three pollinating fig wasps of *F. pumila* in Southeastern China were concordant. Both ancient and recent NUMTs were detected in at least two of the three species. *Cytb* NUMTs were not evenly distributed in the three species. They were prevalent in *Wiebesia* sp. 3, while rare in most populations of *Wiebesia* sp. 1 and 2. Further studies are needed to assess the potential mechanisms (such as selection, drift and amplification failure) that have resulted in this distribution pattern.

Acknowledgements

We thank Shou-Hiesh Li, Yong Chen, Wen-Shan Wu, Bin Ai, Miao-Miao Shi, Da-Wei Gao, Dong-Wei Shen, Rong Wang, Xiao-Yan Wang and Jian-Hui Li for invaluable assistance in sample collection and Lu-Shui Zhang for help in polishing the figures. This study was financed by the National Natural Science Foundation of China (31270416, 31270387), the Doctoral Program of the Ministry of Education of China (20110076110013), Key program of the Education Department of Sichuan Province (11ZA162) and Graduate School of East China Normal University (CX2011003 and XRZZ2011023).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.actao.2013.10.001.

References

Baldo, L., de Queiroz, A., Hedin, M., Hayashi, C.Y., Gatesy, J., 2011. Nuclear-mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping bristletail *Mesomachilis*. Mol. Biol. Evol. 28, 195–210.

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

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

Bertheau, C., Schuler, H., Krumbock, S., Arthofer, W., Stauffer, C., 2011. Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs. Mol. Ecol. Resour. 11, 1056–1059.

Burton, O.J., Travis, J.M.J., 2008. Landscape structure and boundary effects determine the fate of mutations occurring during range expansions. Heredity 101, 329–340.

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.

Chatre, L., Ricchetti, M., 2011. Nuclear mitochondrial DNA activates replication in *Saccharomyces cerevisiae*. PLoS One 6, e17235.

Chen, Y., Compton, S.G., Liu, M., Chen, X.Y., 2012. Fig trees at the northern limit of their range: the distributions of cryptic pollinators indicate multiple glacial refugia. Mol. Ecol. 21, 1687–1701.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659.

Cruaud, A., Jabbour-Zahab, R., Genson, G., Cruaud, C., Couloux, A., Kjellberg, F., Van Noort, S., Rasplus, J.Y., 2010. Laying the foundations for a new classification of Agaonidae Hymenoptera: Chalcidoidea., a multilocus phylogenetic approach. Cladistics 26, 359–387.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Erpenbeck, D., Voigt, O., Adamski, M., Woodcroft, B.J., Hooper, G.W., Degnan, B.M., 2011. NUMTs in the sponge genome reveal conserved transposition mechanisms in metazoans. Mol. Biol. Evol. 28, 1–5.

Gherman, A., Chen, P.E., Teslovich, T.M., Stankiewicz, P., Withers, M., Kashuk, C.S., Chakravarti, A., Lupski, J.R., Cutler, D.J., Katsanis, N., 2007. Population bottlenecks as a potential major shaping force of human genome architecture. PLoS Genet. 3, 1223–1231.

Hazkani-Covo, E., 2009. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol. Biol. Evol. 26, 2175–2179.

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

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Jacques, N., Sacerdot, C., Derkaoui, M., Dujon, B., Ozier-Kalogeropoulos, O., Casaregola, S., 2010. Population polymorphism of nuclear mitochondrial DNA insertions reveals widespread diploidy associated with loss of heterozygosity in *Debaryomyces hansenii*. Eukaryot. Cell 9, 449–459.

Karanth, K.P., 2008. Primate NUMTs and reticulate evolution of capped and golden leaf monkeys Primates: Colobinae. J. Biosci. 33, 761–770.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Lang, M., Sazzini, M., Calabrese, F.M., Simone, D., Boattini, A., Romeo, G., Luiselli, D., Attimonelli, M., Gasparre, G., 2012. Polymorphic NUMTs trace human population relationships. Hum. Genet. 131, 757–771.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Liu, M., Zhang, J., Chen, Y., Compton, S.G., Chen, X.Y., 2013. Contrasting genetic responses to population fragmentation in a coevolving fig and fig wasp across a mainland-island archipelago. Mol. Ecol. 22, 4384–4396.

Lopez, J., Yuhki, N., Masuda, R., Modi, W., O'Brien, S., 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.

Machado, C.A., Jousselin, E., Kjellberg, F., Compton, S.G., Herre, E.A., 2001. Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps. Proc. Roy. Soc. B Biol. Sci. 268, 685–694.

Mishmar, D., Ruiz-pesini, E., Brandon, M., Wallace, D.C., 2004. Mitochondrial DNA-like sequences in the nucleus NUMTs.: insights into our African origins and the mechanism of foreign DNA integration. Hum. Mutat. 23, 125–133.

Mundy, N.I., Pissinatti, A., Woodruff, D.S., 2000. Multiple nuclear insertions of mitochondrial Cytochrome b sequences in Callitrichine primates. Mol. Biol. Evol. 17, 1075–1080.

Muradian, K.K., Lehmann, G., Fraifeld, V.E., 2010. NUMT ("new mighty") hypothesis of longevity. Rejuv. Res. 13, 152–155.

R Development Core Team, 2012. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rambaut, A., 2006. FigTree v1.3.1. Available from: http://tree.bio.ed.ac.uk/software/figtree.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from: http://beast.bio.ed.ac.uk/Tracer.

Ricchetti, M., Tekaia, F., Dujon, B., 2004. Continued colonization of the human genome by mitochondrial DNA. PLoS Biol. 2, 1313–1324.

Rogers, H.H., Griffiths-Jones, S., 2012. Mitochondrial pseudogenes in the nuclear genomes of Drosophila. PLoS One 7, e32593.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

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.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time markov chain evolutionary models. Mol. Biol. Evol. 18, 1001–1013.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis MEGA software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Travis, J.M.J., Münkemüller, T., Burton, O.J., Best, A., Dytham, C., Johst, K., 2007. Deleterious mutations can surf to high densities on the wave front of an expanding population. Mol. Biol. Evol. 24, 2334–2343.

Triant, D.A., DeWoody, J.A., 2007. Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. Gene 401, 61–70.

Wang, R., Ai, B., Gao, B.Q., Yu, S., Li, Y.Y., Chen, X.-Y., 2009. Spatial genetic structure and restricted gene flow in a functionally dioecious fig, *Ficus pumila* L. var. *pumila* (Moraceae). Popul. Ecol. 51, 307–315.

Zhang, D.X., Hewitt, G.M., 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol. Evol. 11, 247–251.

Please cite this article in press as: Chen, Y., et al., Distribution of nuclear mitochondrial pseudogenes in three pollinator fig wasps associated with *Ficus pumila*, Acta Oecologica (2013), http://dx.doi.org/10.1016/j.actao.2013.10.001