EUROPEAN JOURNAL OF ENTOMOLOGY
ISSN (online): 1802-8829
http://www.eje.cz
Eur. J. Entomol. **113**: 482–488, 2016
doi: 10.14411/eje.2016.063
ORIGINAL ARTICLE

The mitochondrial genome of the Mediterranean flour moth, *Ephestia kuehniella* (Lepidoptera: Pyralidae), and identification of invading mitochondrial sequences (numts) in the W chromosome

Katrin Lämmermann[1,3], Heiko Vogel[2] and Walther Traut[3,*]

[1] Universität zu Lübeck, Institut für Neuro- und Bioinformatik, Ratzeburger Allee 160, D-23538 Lübeck, Germany; e-mail: katrin.laemmermann@hotmail.de

[2] Max Planck Institute for Chemical Ecology, Department of Entomology, Hans-Knoell-Strasse 8, D-07745 Jena, Germany; e-mail: hvogel@ice.mpg.de

[3] Universität zu Lübeck, Zentrum für Medizinische Strukturbiologie, Institut für Biologie, Ratzeburger Allee 160, D-23538 Lübeck, Germany; e-mail: traut@bio.uni-luebeck.de

Key words. Lepidoptera, Pyralidae, *Ephestia kuehniella*, mitogenome, Mediterranean flour moth, phylogeny, numts, W chromosome

Abstract. The Mediterranean flour moth, *Ephestia kuehniella* is a widespread pest of stored products and a classical object in experimental biology. In the present study, we determined its complete mitochondrial genome sequence. The genome is circular, consists of 15,327 bp and comprises 13 protein-coding, 2 rRNA- and 22 tRNA-coding genes in an order typical for the Ditrysia clade of the order Lepidoptera. A phylogenetic study of the Lepidoptera based on complete mitochondrial genomes places *E. kuehniella* correctly in the family Pyralidae and supports major lepidopteran taxa as phylogenetic clades. The W chromosome of *E. kuehniella* is an exceptionally rich reservoir of originally mitochondrial sequences (numts). Around 0.7% of the W DNA was found to be of mitochondrial origin, 83% of the mitogenome sequence was represented between 1–11 × in the W chromosome. Phylogenetic analysis further revealed that these numts are an evolutionary recent acquisition of the W chromosome.

ABBREVIATIONS

ATP6 – ATP synthase subunit 6
ATP8 – ATP synthase subunit 8
COX1 – cytochrome c oxidase subunit 1
COX2 – cytochrome c oxidase subunit 2
COX3 – cytochrome c oxidase subunit 3
CYTB – cytochrome b
ME – Minimum Evolution algorithm
ML – Maximum Likelihood algorithm
mitogenome – mitochondrial genome
mtDNA – mitochondrial DNA
NADH1 – NADH dehydrogenase subunit 1
NADH2 – NADH dehydrogenase subunit 2
NADH3 – NADH dehydrogenase subunit 3
NADH4 – NADH dehydrogenase subunit 4
NADH4L – NADH dehydrogenase subunit 4L
NADH5 – NADH dehydrogenase subunit 5
NADH6 – NADH dehydrogenase subunit 6
NJ – Neighbor Joining algorithm
numts – nuclear mitochondrial DNA segments
PCG – protein coding gene

INTRODUCTION

Mitochondrial genomes have some outstanding features compared with nuclear eukaryotic genomes. They are usually circular, non-recombining, present in multiple copies, transmitted from females only, and have a slightly different genetic code. In Metazoa, the mitochondrial genome typically contains 13 well-conserved intron-less protein coding genes (PCGs), 2 rRNA and 22 tRNA genes (Bernt et al., 2013).

Because of the relative ease of obtaining them, mitochondrial DNA sequences have been used extensively in phylogenetic studies. These have mainly been based on PCGs and/or the 16S rRNA gene. The mitochondrial COX1 gene has been sequenced from a great many species in pursuit of the Barcode-Of-Life project and forms the greatest existing global database for a single gene (www.boldsystems.org).

The Mediterranean flour moth, *Ephestia kuehniella* is a pest of stored cereal products, grains and flour. Humans have been largely responsible for spreading this serious

* Corresponding author; e-mail: traut@bio.uni-luebeck.de

economic pest insect around the world via infested goods (Invasive Species Compendium, www.cabi.org/isc). *E. kuheniella* is easy to breed in the laboratory and has been extensively used in developmental studies, and in genetic and cytogenetic research since the 1920s (reviewed by Robinson, 1971). Although the W chromosome and several genomic loci were studied at the molecular level (see e.g. Rohwedel et al., 1993; Sahara et al., 1999; Traut et al., 1999, 2013; Pytelková et al., 2009; Guz et al., 2012; Kobelková et al., 2015), the nuclear genome has not yet been sequenced. In the present study, we describe the complete mitochondrial genome, compare it with that of other lepidopteran species, and investigate the invasion of the W chromosome by mitochondrial sequences called numts (i.e. nuclear mitochondrial DNA segments; Lopez et al., 1994).

MATERIALS AND METHODS

Animals

For our study, we used the *E. kuehniella* laboratory strain "L", a strain known to have been inbred for many decades, originating from Alfred Kühn's broods of the 1930s. DNA for PCR amplification was extracted from male larvae only, to avoid contamination with known W chromosomal numts (Traut et al., 2013).

DNA sequencing

The mitogenome was assembled from Sanger sequencing of two different sources: (1) DNA enriched for mitochondrial sequences by the alkaline lysis procedure (Tamura & Aotsuka, 1988); and (2) PCR products of primer pairs listed in Table 1. PCRs were performed with Phusion High-fidelity Polymerase (Thermo Fisher Scientific, Dreieich, Germany). The DNA templates for PCR amplifications were isolated using the ZR Tissue & Insect DNA Microprep (Zymo Research, Freiburg, Germany) or by a salting-out protocol (Sunnucks & Hales, 1996).

A small gap around position 6300 of the linearised mitogenome (see acc. no. KF305832) proved difficult to close by PCR. Several primer pairs were tried without success, although the same primers amplified the correct sequences on either side of the gap. This problem was assumed to be caused by a highly stable secondary structure in that region. Eventually, PCRs with prolonged elongation times (3 min for ~500 bp) and using GC Phusion buffer were able to close the gap and thus rendered the products amenable for sequencing. When the mitogenome sequence from position 4000–8000 was scanned with different window sizes for the lowest free energy and local folding potential with RNAslider and mfold (Patzel et al., 1999; Dornseifer & Sczakiel, 2013), a minimum of free energy with about –20 kcal turned up around position 6300 with a potential to form a perfect 40 nt hairpin loop which may have caused the initial PCR problems.

Table 1. Primers used in the reconstruction of the *Ephestia kuehniella* mitogenome.

Name	Position	Sequence
Forward		
mitF2	14217	CAAGCTCTAATTTGATTCTTTC
mitF3	12107	CAGCAAAATCAAATGGTGTC
mitF5	2119	TGGAGGGGGAGATCCTATTC*
mitF10	10606	TAATTGATCTTCCATCCCCC
mitF18	330	TGATTTGGGTGTTGAATTGG
mitF23	6023	AAGCAATTTTGCATTTAATTTC
mitF41	6086	AACCAAAATAGAGGTATATCACTG
s1_3	4586	TATACCTTCCTTTTTTATCCCC
Reverse		
mitR1	458	AATTGAGGCGATAGATTGGG
mitR2	14316	ATGTACATATTGCCCGTCGC
mitR4	3899	GGTATTATTTGAGGGATAAAAG
mitR9	12583	GGGGAATTTTACAACCTTTTTC
mitR17	1913	TTCCAGCTAAATGAAGGGAG
mitR43	6649	AAGAAGTTTTTTAGGGATAATATG
N4552modif	4550	ATGGCCTGCAATTATATTAGC
N6160modif	6125	TCAATTTTATCATTAACAGTGA

* 2 mismatches compared to the genomic sequence (TGGGGGAGGAGATCCTATTC).

Annotation

PCGs and ribosomal RNA genes were predicted with the help of the DOGMA program (Wyman et al., 2004) and by homology with known lepidopteran genes. To predict tRNA genes, ARWEN (Laslett & Canbäck, 2008) was used.

Alignment and phylogenetic analyses

In a preliminary approach with only seven lepidopteran species, we tested whole mitogenomes, concatenated PCG nucleotide, and amino acid sequences as substrates for a phylogenetic study. Phylogenies calculated from the concatenated pre-aligned amino acid sequences of all 13 PCGs conformed best with the lepidopteran phylogeny published by Regier et al. (2013). This procedure was then used in a final series of 49 mitogenomes (47 Lepidoptera and 2 Trichoptera species, Table 2).

Multiple alignments were performed with ClustalW (Larkin et al., 2007) using the standard parameters. Misplacements of amino acids bordering gaps were corrected manually if, and only if, the new alignment was obviously better and no further gaps were created. For phylogenetic analyses, Neighbor Joining (NJ) (Saitou & Nei, 1987), Maximum Likelihood (ML) (Felsenstein, 1973) and Minimum Evolution (ME) (Rhetsky & Nei, 1993) trees were constructed with default parameters using the MEGA6 platform (Tamura et al., 2013). Bootstrapping (1000 repeats) served as a confidence test.

Numt searches

For numt searches in the *E. kuehniella* W chromosome we used an existing library of 20,718 contigs with ≥100 bp, assembled from 361,828 Roche/454 and 1,172 Sanger reads of microdissected W chromosome DNA (see Traut et al., 2013). FASTA36 searches were performed with a threshold expectation value of $E = 0.0001$ and the complete *E. kuehniella* mitogenome as a query. To allow for the use of a linear query in the case of a circular genome, we added the missing hits obtained with a query in which the start of the linearised mitogenome was shifted by 300 bp. All alignments of hit sequences were inspected individually. We counted as numts only those sequences that had a minimum of 15 bp flanking non-mitochondrial sequence and, therefore, appeared "chimeric". Completely "mitochondrial" contigs were excluded from this study because they may have been genuine mitochondrial sequences contaminating the W chromosomal library. Further excluded were hits on purely low complexity regions.

BLAST searches for numts in other genomes were performed with the same criteria in the genomes of the *Bombyx mori*, *Danaus plexippus*, *Heliconius melpomene*, *Melitaea cinxia* (EnsembleMetazoa at www.ebi.ac.uk), *Chilo suppressalis* (ento.njau.edu.cn/ChiloDB/), and *Manduca sexta* (agripestbase.org/manduca/).

RESULTS AND DISCUSSION

Sequence and gene order

The *E. kuehniella* mitogenome was assembled from Sanger sequences of mtDNA-enriched samples and PCR products from males. Long sequences of 516–3.282 bp (average 2.239 bp), free of non-mitochondrial segments, with precisely matching overlapping ends, were selected for the assembly to avoid the inclusion of numts. Neverthe-

Table 2. Species included in the phylogenomic study.

Order	Family	Species	Accession no.
Trichoptera	Trichoceridae	*Paracladura trichoptera* (Osten Sacken)	NC_016173.1
	Phryganeidae	*Eubasilissa regina* (McLachlan)	NC_023374.1
Lepidoptera	Bombycidae	*Bombyx mori* (L.)	AB070264
		Rondotia menciana (Moore)	NC_021962.1
	Carposinidae	*Carposina sasakii* Matsumura	NC_023212.1
	Cossidae	*Eogystia hippophaecolus* (Hua et al.)	NC_023936.1
	Crambidae	*Ostrinia nubilalis* (Hübner)	NC_003367.1
		Chilo suppressalis (Walker)	NC_015612.1
	Elachistidae	*Promalactis suzukiella* Matsumura	NC_026697.1
	Erebidae	*Hyphantria cunea* (Drury)	NC_014058.1
		Amata formosae (Butler)	NC_021416.1
	Gelechiidae	*Pectinophora gossypiella* (Saunders)	KM225795.1
	Geometridae	*Phthonandria atrilineata* (Butler)	NC_010522.1
		Biston suppressaria (Guenée)	NC_027111.1
	Hepialidae	*Ahamus yunnanensis* (Yang et al.)	NC_018095.1
		Thitarodes renzhiensis (Yang et al.)	NC_018094.1
	Hesperiidae	*Celaenorrhinus maculosa* (C. & R. Felder)	KF543077.1
		Daimio tethys (Ménétries)	NC_024648.1
	Lasiocampidae	*Dendrolimus punctatus* (Walker)	NC_027156.1
	Lycaenidae	*Coreana raphaelis* (Oberthür)	NC_007976.1
		Lycaena phlaeas (L.)	NC_023087.1
	Lymantriidae	*Lymantria dispar* (L.)	NC_012893.1
		Gynaephora menyuanensis Yan & Chou	NC_020342.1
	Lyonetiidae	*Leucoptera malifoliella* (O. Costa)	NC_018547.1
	Noctuidae	*Sesamia inferens* (Walker)	NC_015835.1
		Spodoptera litura (Fabricius)	KF701043.1
	Nolidae	*Risoba prominens* Moore	NC_026841.1
		Gabala argentata Butler	NC_025842.1
	Notodontidae	*Ochrogaster lunifer* Herrich-Schäffer	NC_011128.1
		Phalera flavescens (Bremer & Gery)	NC_016067.1
	Nymphalidae	*Kallima inachus* (Boisduval)	NC_016196.1
		Fabriciana nerippe (C. & R. Felder)	NC_016419.1
	Papilionidae	*Papilio glaucus* (L.)	NC_027252.1
		Teinopalpus aureus Mell	KP941013.1
	Pieridae	*Artogeia melete* (Ménétries)	EU597124.1
		Pieris rapae (L.)	NC_015895.1
	Plutellidae	*Plutella xylostella* (L.)	NC_025322.1
	Pyralidae	*Ephestia kuehniella* (Zeller)	NC_022476.1
		Plodia interpunctella (Hübner)	NC_021746.1
	Riodinidae	*Apodemia mormo* (C. & R. Felder)	NC_024571.1
		Abisara fylloides (Moore)	NC_021746.1
	Saturniidae	*Antheraea pernyi* (Guérin-Méneville)	NC_004622.2
		Attacus atlas (L.)	NC_021770.1
	Sphingidae	*Manduca sexta* (L.)	NC_010266.1
		Sphinx morio (Rothschild & Jordan)	NC_020780.1
	Stathmopodidae	*Atrijuglans hetaohei* (Yang)	KT581634.1
	Tortricidae	*Adoxophyes honmai* Yasuda	NC_008141.1
		Spilonota lechriaspis Meyrick	NC_014294.1
	Yponomeutidae	*Prays oleae* (Bernard)	NC_025948.1

less, like other mitogenomes, there is always a residual risk of not having assembled an absolutely numt-free sequence.

The mitochondrial genome of *E. kuehniella* was circular with a size of 15,327 bp (acc. no. KF305832). Like most metazoans, the mitogenome codes for 13 PCGs, 2 rRNAs, and 22 tRNAs (Table 3). Their order is shown in Fig. 1. An overlap of 7 bp coding sequence was found at the transition from the ATP8 to the ATP6 gene.

A phylogenetically informative tRNA gene rearrangement has been described by Cao et al. (2012) and Timmermans et al. (2014). In insects, the inferred ancestral order of tRNA genes between the control region and the NADH2 gene is tRNA-Ile / tRNA-Gln / tRNA-Met. Non-ditrysian moths from the superfamilies Hepialoidea, Adeloidea, and Nepticuloidea share this order, while in Ditrysia, Palaephatidae, and Tischeriidae, the order is tRNA-Met / tRNA-Ile / tRNA-Gln, thereby supporting the assumed monophyletic origin of this group. In *E. kuehniella*, the order of tRNA genes between the control region and the NADH2 gene, corresponding to nucleotide positions 1–203, is identical to that of other ditrysians.

Phylogenomics

After preliminary tests with a small set of mitogenomes (see Material and methods), we used concatenated prealigned amino acid sequences of all 13 PCGs for a straightforward approach of phylogenetic analysis. Only species with completely known mitogenomes were included. To avoid oversampling of some groups, we selected at most two species from all families of which such data were available (see Table 2). Two species of Trichoptera, the sister group of Lepidoptera, served as outgroups.

The Neighbor Joining (NJ) tree constructed on this dataset is shown in Fig. 2. The tree topology is rather robust. Maximum Likelihood (ML) and Minimum Evolution (ME) trees display identical topologies (except for one node with < 50% bootstrap support, not shown). *E. kuehniella* is grouped with the Indian meal moth, *Plodia inter-*

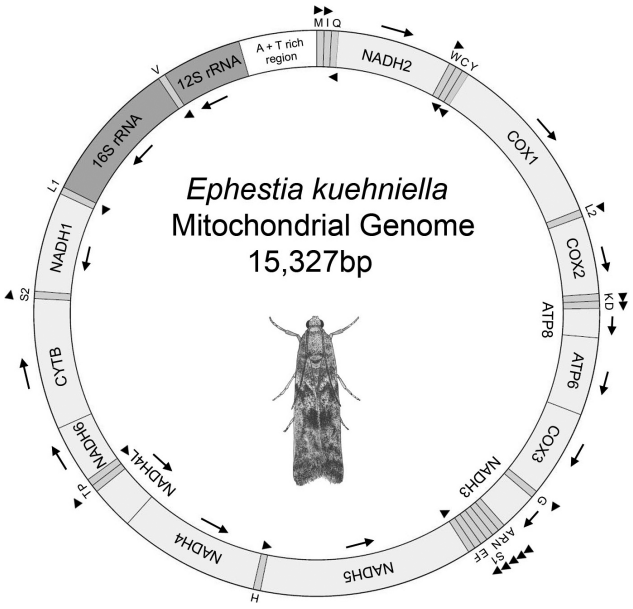

Fig. 1. The mitogenome of *Ephestia kuehniella*. Arrows indicate sense direction of genes. Names of tRNA genes abbreviated by one letter amino acid code.

punctella, another pyralid species and its closest relative with a sequenced mitogenome (Liu et al., 2016).

Members of the same family are correctly joined, and superfamilies, with a few exceptions, are represented as monophyletic groups. A conspicuous exception are the "butterflies", Papilionoidea plus Hesperioidea. As already noted by Kristensen & Skalski (1999) in their earlier phylogenetic analysis of the Lepidoptera, based mainly on comparative morphology, Papilionoidea are paraphyletic. This superfamily forms a single clade only together with Hesperioidea. Bombycoidea are monophyletic only when Lasiocampidae (i.e. *Dendrolimus punctatus)* are omitted. Similarly, Yponomeutoidea appear as monophyletic only without Lyonetiidae (i.e. *Leucoptera malifoliella)*. As to Pyraloidea, Pyralidae (i.e. *E. kuehniella* and *Plodia interpunctella)* and Crambidae (i.e. *Chilo suppressalis* and *Ostrinia nubilalis*), just fail to form a common dichotomous branching node, the bootstrap value is 45%.

Major taxonomic groups like Ditrysia, Obtectomera, and Macroheterocera are well supported as monophyletic in our mitogenome tree. The internode branch lengths of some other early ditrysian lineages are rather short and collapse when a bootstrap cut-off value of 50% is applied as in Fig. 2. The difficulty to resolve the phylogeny of early ditrysian lineages has been observed in previous studies of lepidopteran phylogeny, both when using molecular as well as comparative morphologic approaches. This difficulty is supposed to reflect a period of radiation at the basis of ditrysian evolution, presumably correlated with the radiation of flowering plants (Mutanen et al., 2010).

Table 3. Genes of the *Ephestia kuehniella* mitogenome.

Gene	Product	Position	Size in bp	Start codon	Stop codon
trnM	tRNA-Met	1–68	68		
trnI	tRNA-Ile	70–133	63		
trnQ	tRNA-Gln	complement(135–203)	69		
nad2	NADH2	246–1259	1014	ATC	TAA
trnW	tRNA-Trp	1257–1329	73		
trnC	tRNA-Cys	complement(1320–1388)	69		
trnY	tRNA-Tyr	complement(1391–1455)	65		
cox1	COX1	1460–2995	1535	CGA?	TAA
trnL2	tRNA-Leu	2990–3059	70		
cox2	COX2	3059–3740	682	ATT	T*
trnK	tRNA-Lys	3740–3812	73		
trnD	tRNA-Asp	3811–3883	73		
atp8	ATP8	3883–4044	162	ATC	TAA
atp6	ATP6	4038–4718	681	ATG	TAA
cox3	COX3	4718–5506	489	ATG	TAA
trnG	tRNA-Gly	5508–5575	68		
nad3	NADH3	5572–5928	357	ATA	TAA
trnA	tRNA-Ala	5943–6010	68		
trnR	tRNA-Arg	6015–6081	67		
trnN	tRNA-Asn	6078–6143	66		
trnS1	tRNA-Ser	6142–6219	78		
trnE	tRNA-Glu	6220–6297	78		
trnF	tRNA-Phe	complement(6326–6393)	68		
nad5	NADH5	complement(6393–8109)	1717	ATT	T*
trnH	tRNA-His	complement(8128–8196)	69		
nad4	NADH4	complement(8197–9535)	1339	ATG	T*
nad5	NADH4L	complement(9538–9825)	288	ATG	TAG
trnT	tRNA-Thr	9828–9892	65		
trnP	tRNA-Pro	complement(9892–9959)	68		
nad6	NADH6	9961–10488	528	ATT	TAA
cytb	CYTB	10542–11690	1149	ATG	TAA
trnS2	tRNA-Ser	11706–11773	68		
nad1	NADH1	complement(11793-12728)	936	ATA	TAA
trnL	tRNA-Leu	complement(12731–12800)	70		
rrnL	16S rRNA	complement(12824–14151)	1328		
trnV	tRNA-Val	complement(14163–14231)	69		
rrnS	12S rRNA	complement(14233–15005)	773		

* Stop codon is completed by the addition of a poly(A) tail.

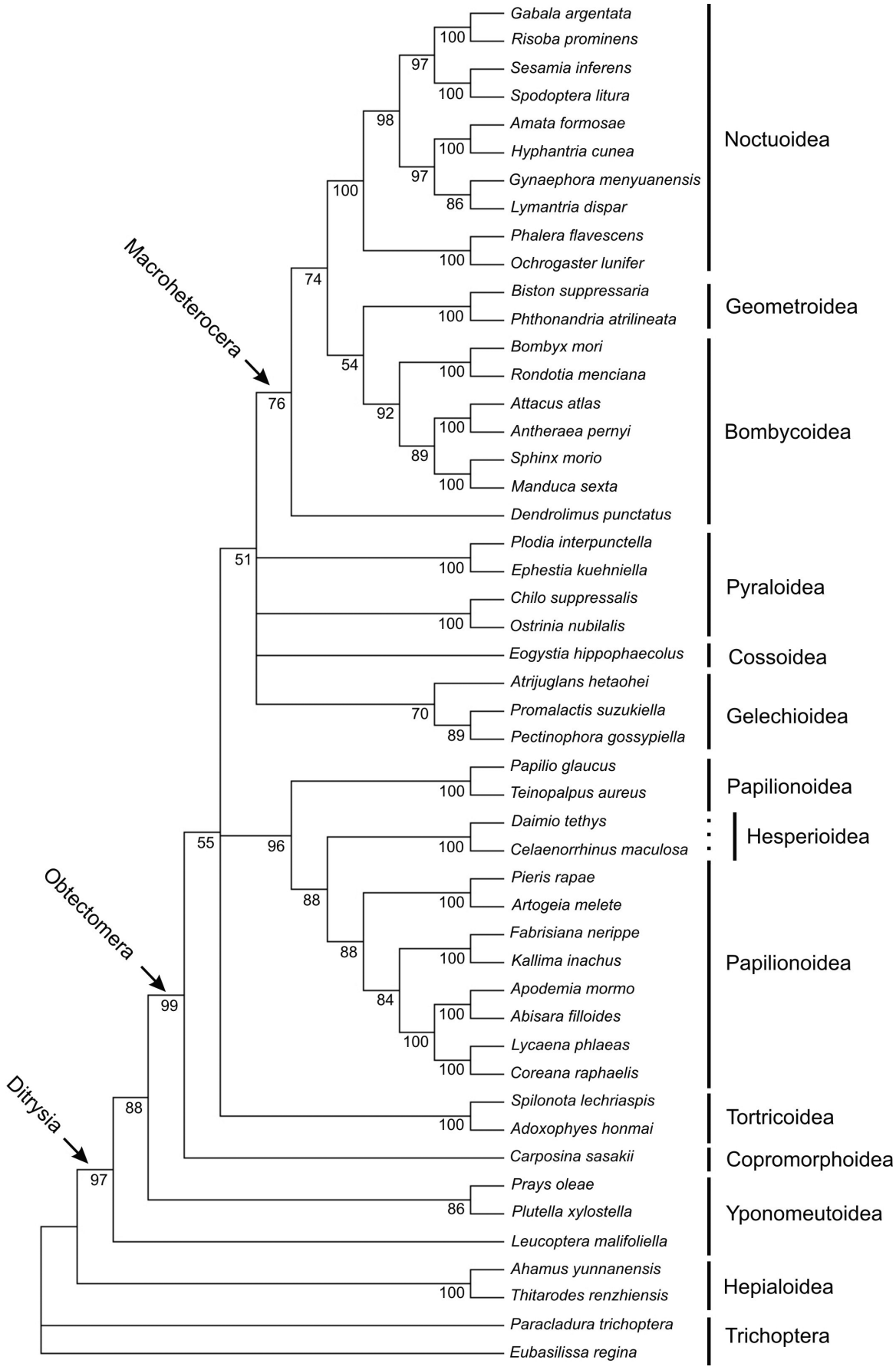

Fig. 2. Phylogenetic tree of 47 selected lepidopteran species, inferred from the mitochondrial PCGs using the Neighbor-Joining method. Nodes were collapsed at the 50% bootstrap cutoff value. Two trichopteran species served as outgroups. Numbers denote bootstrap values.

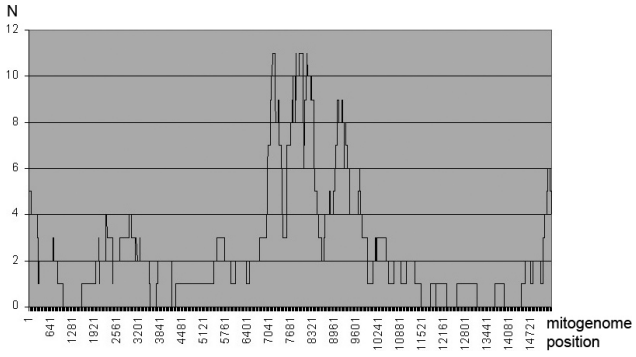

Fig. 3. Source (position in mitogenome) and frequency (N) of numt sequence in the W chromosome of *Ephestia kuehniella*.

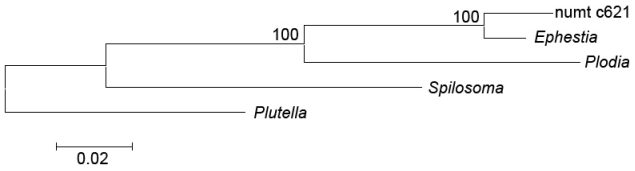

Fig. 4. ME tree of numt c621 (numt in contig c621) and homologous mitochondrial segments from *Ephestia kuehniella* (Pyralidae), *Plodia interpunctella* (Pyralidae), *Spilonota lechriaspis* (Tortricidae), and *Plutella xylostella* (Plutellidae, outgroup).

Numts in the W chromosome

Two libraries, Roche/454 and Sanger sequences, of W chromosomal DNA from *E. kuehniella* had been established in a previous investigation and shown to contain >1% reads with DNA of mitochondrial origin (Traut et al., 2013). The libraries were prepared from microdissected W chromatin. Here, we use the contig library assembled from the combined Roche/454 and Sanger sequences for a detailed analysis.

Using the *Ephestia* mitogenome as a query, FASTA searches exposed 235 contigs as partly or fully consisting of mitochondrion-derived sequence. We could not exclude the possibility that genuine mitochondrial sequences were contaminants of the library. Therefore, we considered as numts only those that had at least 15 bp flanking non-mitochondrial sequence besides mitochondrion-derived sequence. That left 146 contigs with numt sequences.

When the W-chromosomal numts were mapped against the mitogenome, 83.3% of the mitogenome were also found to be present in the W chromosome. Each position of the mitogenome was represented 0–11 × in W chromosomal DNA (Fig. 3). There were 7.28 bp numts per kb mitogenome in the contig library.

For a comparison, we determined the numt density in the genome libraries of other lepidopteran species, which are publicly available. The fraction of numt sequences in these genomes was 2–3 orders of magnitude lower than present in the *E. kuehniella* W chromosome (Table 4). However, there was no opportunity to study numt density in their W chromosomes because three of the libraries were assembled from male genomes, i.e. they consisted of autosomal and Z chromosomal sequences only. Three other libraries included the W chromosome, but since W-chromosomal contigs or scaffolds have not been earlier identified in such libraries, we could not determine the numt density in the W chromosomes therein.

Recorded numt densities in other nuclear genomes are also mostly in the order of 0.001–0.01 bp/kbp but vary from 0 in the mosquito, *Anopheles gambiae* to 1.7 bp/kbp in the model plant species, *Arabidopsis thaliana* (Richly & Leister, 2004; Pamilo et al., 2007). Thus the numt density of the *Ephestia* W chromosome is more than 4× that of the highest density recorded to date in a nuclear genome.

We attribute the high proportion of numt sequences in the W chromosome to its non-recombining nature. W chromosomes cannot eliminate invading sequence as long as these sequences are not decreasing the organism's overall fitness. Hence, numts accumulate very much like transposon sequences do in the W chromosome. Besides non-recombination, there is probably no selection acting on mitochondrial sequences, once they are inserted in the W chromosome. Hence, mutability is supposed to be high. We indeed found deletions, insertions, inversions and base substitutions in originally protein-coding and other numt sequences. Nevertheless, identity with the genuine mitochondrial sequence was still very high (94–100%). This indicates that the detected numts were rather recent insertions in the W chromosome.

To test this assumption, we constructed phylogenetic trees from the longest numts and their corresponding mitogenomic segments of *E. kuehniella* (Pyralidae), *Plodia interpunctella* (Pyralidae), *Spilonota lechriaspis* (Tortricidae), and *Plutella xylostella* (Plutellidae). *Plodia* is the closest relative of *Ephestia* with a sequenced mitogenome, whereas *Plutella* serves as an outgroup. In all 14 such trees, the numt sequence was on the same sub-terminal node as the *Ephestia* mitochondrial sequence, as shown in Fig. 4. We conclude from this that these sequences were inserted into the W after divergence of the *Plodia* lineage from that of *Ephestia*. A search with the mitogenome of *Plodia* in the W chromosme contig library did not reveal any numt with equal distance to *Ephestia* and *Plodia*. This indicates that none of the numts had been inserted in the W before divergence of the *Ephestia* and *Plodia* lineages. We had expected older numts besides more recently inserted numts. It is conceivable, however, that older numts have escaped

Table 4. Numt density in lepidopteran genomes.

	Chromosomes	Database	Size, bp	Hits	Numts, bp	Fraction
Bombyx mori	A, Z	EnsemblMetazoa	431,707,935	46	13,555	3.10×10^{-5}
Chilo suppressalis	A, Z, W	ChiloDB	269,658,870	81	16,311	4.40×10^{-5}
Danaus plexippus	A, Z, W	EnsemblMetazoa	272,853,388	17	25,677	9.40×10^{-5}
Heliconius melpomene	A, Z	EnsemblMetazoa	269,658,870	81	7,814	2.90×10^{-5}
Melitaea cinxia	A, Z, W	EnsemblMetazoa	360,985,414	40	6,072	1.70×10^{-5}
Manduca sexta	A, Z	ManducaBase	419,441,879	13	1,217	2.90×10^{-6}
Ephestia kuehniella	W	Contig library	8,921,293	142	64,978	7.28×10^{-3}

detection due to major DNA changes since their insertion and only those from a rather recent burst of numt insertions had retained sufficient similarity.

ACKNOWLEDGEMENTS. We thank G. Sczakiel and M. Schütt (Universität zu Lübeck) for calculating the minimum energy profile and interpretation of potential secondary structures; M. Bernt (Universität Leipzig, Germany), D. Cherkasov (Genovoxx, Marburg, Germany), R. Duden (Universität zu Lübeck), and B. Misof (Universität Bonn, Germany) for their valuable suggestions concerning the writing of the manuscript; C. Reuter for technical assistance; and H.D. Loxdale for his editorial suggestions which have also helped to improve the manuscript.

REFERENCES

Bernt M., Braband A., Schierwater B. & Stadler P.F. 2013: Genetic aspects of mitochondrial genome evolution. — *Mol. Phylogenet. Evol.* **69**: 328–338.

Cao Y.-Q., Ma C., Chen J.-Y. & Yang D.-R. 2012: The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: the ancestral gene arrangement in Lepidoptera. — *BMC Genomics* **13**: 276.

Dornseifer S. & Sczakiel G. 2013: Computational identification of biologically functional non-hairpin GC-helices in human Argonaute mRNA. — *BMC Bioinform.* **14**: 122, 8 pp.

Felsenstein J. 1973: Maximum-likelihood and minimum steps methods for estimating evolutionary trees from data on discrete characters. — *Syst. Zool.* **22**: 240–249.

Guz N., Kilincer N. & Aksoy S. 2012: Molecular characterization of *Ephestia kuehniella* (Lepidoptera: Pyralidae) transferrin and its response to parasitoid *Venturia canescens* (Hymenoptera: Ichneumonidae Gravenhorst). — *Insect Mol. Biol*. **21**: 139–147.

Kobelková A., Závodská R., Šauman I., Bazalová O. & Doležel D. 2015: Expression of clock genes period and timeless in the central nervous system of the Mediterranean flour moth, *Ephestia kuehniella*. — *J. Biol. Rhythms* **30**: 104–116.

Kristensen N.P. & Skalski A.W. 1999: Phylogeny and palaeontology. In Kristensen N.P. (ed.): *Handbook of Zoology, Lepidoptera, Moths and Butterflies. Vol. 1: Evolution, Systematics, and Biogeography.* Walter de Gruyter, Berlin, New York, pp 7–25.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R. et al. 2007: Clustal W and Clustal X version 2.0. — *Bioinformatics* **23**: 2947–2948.

Laslett D. & Canbäck B. 2008: ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. — *Bioinformatics* **24**: 172–175.

Liu Q.-N., Chai X.-Y., Bian D.-D., Zhou C.-L. & Tang B.-P. 2016: The complete mitochondrial genome of *Plodia interpunctella* (Lepidoptera: Pyralidae) and comparison with other *Pyraloidea* insects. — *Genome* **59**: 37–49.

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

Mutanen M., Wahlberg N. & Kaila L. 2010: Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. — *Proc. R. Soc. (B)* **277**: 2839–2848.

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

Patzel V., Steidl U., Kronenwett R., Haas R. & Sczakiel G. 1999: A theoretical approach to select effective antisense oligodeoxiribonucleotides at high statistical probability. — *Nucl. Acids Res.* **27**: 4328–4334.

Pytelková J., Hubert J., Lepšík M., Šobotník J., Šindelka R., Křížková I., Horn M. & Mareš M. 2009: Digestive alpha-amylases of the flour moth *Ephestia kuehniella* – adaptation to alkaline environment and plant inhibitors. — *FEBS J.* **276**: 3531–3546.

Regier J.C., Mitter C., Zwick A., Bazinet A.L., Cummings M.P., Kawahara A.Y., Sohn J.-C., Zwickl D.J., Cho S., Davis D.R. et al. 2013: A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). — *PLoS ONE* **8**: e58568.

Rhetsky A. & Nei M. 1993: Theoretical foundation of the minimum-evolution method of phylogenetic inference. — *Mol. Biol. Evol.* **10**: 1073–1095.

Richly E. & Leister D. 2004: NUMTs in sequenced eukaryotic genomes. — *Mol. Biol. Evol*. **21**: 1081–1084.

Robinson R. 1971: *Lepidoptera Genetics*. Pergamon Press, Oxford, 687 pp.

Rohwedel J., Weichenhan D., Meier C. & Traut W. 1993: Different modes of hypervariability in $(GATA)_n$ simple sequence repeat loci. — *Insect Mol. Biol.* **2**: 49–58.

Sahara K., Marec F. & Traut W. 1999: TTAGG telomeric repeats in chromosomes of some insects and other arthropods. — *Chromosome Res* **7**: 449–460.

Saitou N. & Nei M. 1987: The neighbor-joining method: a new method for reconstructing phylogenetic trees. — *Mol. Biol. Evol.* **4**: 406–425.

Sunnucks P. & Hales D.F. 1996: Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). — *Mol. Biol. Evol.* **13**: 510–524.

Tamura K. & Aotsuka T. 1988: Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. — *Biochem. Genet.* **26**: 815–819.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. 2013: MEGA6: Molecular evolutionary genetics analysis version 6.0. — *Mol. Biol. Evol.* **30**: 2725–2729.

Timmermans M.J.T.N., Lees D.C. & Simonsen T.J. 2014: Towards a mitogenomic phylogeny of Lepidoptera. — *Mol. Phylogenet. Evol.* **79**: 169–178.

Traut W., Sahara K., Otto T.D. & Marec F. 1999: Molecular differentiation of sex chromosomes probed by comparative genomic hybridization. — *Chromosoma* **108**: 173–180.

Traut W., Vogel H., Glöckner G., Hartmann E. & Heckel D.G. 2013: High-throughput sequencing of a single chromosome: a moth W chromosome. — *Chromosome Res.* **21**: 491–505.

Wyman S.K., Jansen R.K. & Boore J.L. 2004: Automatic annotation of organellar genomes with DOGMA. — *Bioinformatics* **20**: 3252–3255.

Received May 10, 2016; revised and accepted July 18, 2016
Published online September 15, 2016