Accepted on 4 July 2008
J Zool Syst Evol Res doi: 10.1111/j.1439-0469.2008.00492.x

[1]*Laboratoire de Biologie des Ligneux et des Grandes Cultures, UPRES-EA-1207, Université d'Orléans, BP6759, Rue de Chartres, Orléans cedex;* [2]*INRA, Centre d'Orléans, Unité de Zoologie Forestière, BP20619 Ardon, Olivet cedex, France*

Incorporation of mitochondrial fragments in the nuclear genome (*Numts*) of the longhorned beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae)

F. A. Koutroumpa[1], F. Lieutier[1] and G. Roux-Morabito[2]

Abstract

A mitochondrial DNA (mtDNA) study, based on 43 European populations (33 of them sampled in France) of *Monochamus galloprovincialis*, vector of the pinewood nematode, and 14 populations of its sister species *Monochamus sutor* was realized. Sequencing of 792 bp of the cytochrome oxidase I (COI) and 521 bp of the COII genes revealed numerous ambiguities on multiple nucleotide sites for half of *M. galloprovincialis* specimens studied (44.8%). Hypotheses of heteroplasmy and pseudogenes (*Numts*) were examined. The mtDNA isolation by alkaline lysis and cloning (for three successfully used individuals) both support the hypothesis that the ambiguous sequences amplified were not of mtDNA nature and validate the presence of *Numts* in the nuclear genome of *M. galloprovincialis*. Multiple copies of mtDNA-like sequences were found paralogous to COI, tRNA leucin and COII regions. Phenetic analysis placed different recently diverged mtDNA-like sequences as a close relative of mtDNA sequences, and grouped 10 closely related mtDNA-like sequences as a more basal clade, closer to ancestral states and to *M. sutor*. This result supports that this nuclear family of pseudogenes arose independently of the other events and may represent mitochondrial haplotypes sampled from *M. galloprovincialis* ancestral populations. This is the first time that *Numts* are proved for a longhorned beetle, whereas no *Numts* were found within its sister species *M. sutor*. The incorporation mechanism of *Numts* in unknown for *M. galloprovincialis*, however, excess of ambiguous sites corresponding to synonymous mutations placed on third codon position as well as the absence of *Numts* in *M. sutor*, conducted to the hypothesis of a recent transfer of these *Numts* in the nuclear genome of *M. galloprovincialis*.

Key words: *Monochamus* – mitochondrial DNA – *Numts*

Introduction

Accurrate species delimitation is crucial in biodiversity studies, especially for invasive species recognition and for biological control (Godfrey 2002). Among molecular markers used to test species boundaries, mitochondrial DNA (mtDNA) genes became very popular, through projects such as the global bioidentification mtDNA-based system, namely the 'DNA barcode project' (Hebert et al. 2003). mtDNA has numerous remarkable characteristics like (i) small size followed by simple structure and organization; (ii) high copy number that makes amplification easier; (iii) effective haploidy in DNA sequences, uniparental inheritance (in most cases maternal); (iv) absence of recombination, although some cases of recombination have been found (Ladoukakis and Zouros 2001; Hoarau et al. 2002; Rokas et al. 2003a) and (v) important variability in evolution rates of its regions providing the possibility to conduct studies on various taxonomic levels (Avise et al. 1987; Moritz et al. 1987). Nevertheless, it averred to be a controversial molecular marker because it is also prone to incomplete lineage sorting, retained ancestral polymorphisms and introgression which could lead to species misidentifications (Funk and Omland 2003). Moreover, application of per cent DNA sequence difference widely used for the identification of species could overestimate sequence divergence because of heteroplasmy and pseudogenes. Heteroplasmy is the existence of more than one mtDNA type in a single individual and can be expressed as length or single nucleotide polymorphism. The origins of heteroplasmy can be recombination somatic mutations, doubly uniparental inheritance, paternal leakage or⁄and hybridization. Many cases of heteroplasmy are reported in insects, birds, fish, mussels and seaweeds, lizards and mammals including humans (Solignac et al. 1983; Boyce et al. 1989; Kondo et al. 1990; Crochet and Desmarais 2000; Ladoukakis and Zouros 2001; Katsares et al. 2003; Kvist et al. 2003; Sekiguchi et al. 2003; Coyer et al. 2004). Knowledge about nuclear integrations of mtDNA fragments (mtDNA pseudogenes or *Numts*, Lopez et al. 1994) is particularly poor, especially when their origin and evolutionary patterns are considered. It is generally admitted that *Numts* are non-functional mitochondrial gene copies transported into the nuclear genome under both genetic and environmental influence (Bensasson et al. 2001b). As heteroplasmy, *Numts* amplified simultaneously with mtDNA during polymerase chain reaction (PCR) procedures can bias results of evolutionary studies since most of the time similarity between these sequences makes their identification difficult or even impossible. Numerous cases of *Numts* on fungi, plants and metazoan have been listed by Bensasson et al. (2001b, see also http://www.pseudogene.org, Karro et al. 2007). *Numts* were most commonly found in plants than in animals (Bensasson et al. 2001b; Richly and Leister 2004). In insects, few studies have been reported, e.g. on Homoptera (Sunnucks and Hales 1996), Hymenoptera (Rokas et al. 2003b) and most commonly on Orthoptera (Bensasson et al. 2000, 2001a; Keller et al. 2007). Recently, Coleoptera have been added to the list (Pons and Vogler 2005; Mestrovic et al. 2006; Jordal 2007), but no *Numts* have been reported in Cerambycidae beetles so far. When such mtDNA gene duplications were considered, their origin, evolution, geographical dispersion and population or species dynamics were rarely developed.

The genus *Monochamus* (Coleoptera: Cerambycidae) includes about 150 different species spread all around the world, besides South America. Seven *Monochamus* species are reported so far to be vectors of the Pine Wood Nematode (PWN) *Bursaphelenchus xylophilus* (Steiner and Buhrer, 1934; Nickle, 1970) (Nematoda: Aphelenchoididae), causative agent of the Pine Wilt Disease. Among them *Monochamus galloprovincialis* (Olivier, 1795) has been recently proved as a vector in Portugal (Sousa et al. 2001).

In the course of studying phylogeographic structure and molecular systematics of two European sister species of *Monochamus* (*M. galloprovincialis* and *M. sutor* Linnaeus, 1758) using mtDNA as genetic marker, high number of ambiguous sites was encountered in *M. galloprovincialis* sequences. To overcome the bias of the ambiguities interference in the phylogenetic analyses their nature and origin are precised in this paper. In addition, the rate and patterns of *Numt* sequence evolution in this species as well as *Numt* geographic distribution are discussed.

Materials and Methods

Specimens examined

One hundred and forty adults (71 males and 69 females) and eight larvae of *M. galloprovincialis* were examined in this study. The sampling was performed between May and October for three consecutive years, 2003–2005, using 150 cross van traps placed in 40 localities in France. Three to four traps fitted with ethanol, a-pinene and a mixture of Ips-enol and Ips-dienol were installed in each locality (Koutroumpa, 2007). A total of 117 specimens were sampled in France. In addition, 17 *M. galloprovincialis* captured in Greece, five in Spain, five in Portugal, two in Morocco and two in Austria were added to this study (Table S1). Trapped insects were frozen at −80°C or in absolute ethanol until DNA extraction. The totality of the individuals was sampled from 43 geographic populations [the majority (33) originated from France, see also Table S1). For each population, one to 11 individuals were examined and body parts and DNA were kept as vouchers in the laboratory (Laboratoire de Biologie et des Grandes Cultures, L.B.L.G.C.) in Orleans University in France. In addition, 49 specimens sampled from 14 populations of *M. sutor* species were equally analyzed. Ten of the 14 populations were sampled in France, two in Greece and two in Austria (Table S2).

PCR amplification and sequencing of mtDNA

Both *Monochamus* species' DNA was extracted from one leg (two legs when specimens were small) to permit subsequent morphological observations and to avoid contamination from organisms like the PWN which is found in the tracheas of *Monochamus* (Mamiya 1976).

Extraction and isolation of genomic DNA was performed with the GenElute Mammalian Genomic DNA miniprep kit (Sigma-Aldrich, St Quentin, France) with final elution to 100 μl. A part of the cytochrome oxidase I (COI) and II (COII) genes was amplified by PCR using the Sigma Red Taq package (Sigma-Aldrich, St Quentin, France). One microlitre of DNA extraction, in 50 μl reaction volumes, was used as template for 35 cycles of mtDNA amplification. The reaction was performed in a gradient Peltier thermal cycle (PTC-200; Applied Biosystems, Courtaboeuf, France) under the following conditions: denaturation at 94°C for 1 min, annealing at 45°C for 1 min and extension at 72°C for 1 min 30 s. The initial cycle employed a 3 min denaturation at 95°C and the final cycle had an extension of 72°C for 5 min. PCR products were examined via agarose gel electrophoresis and purified using GenElute PCR Clean-Up kit (Sigma-Aldrich). Sequencing of purified PCR products was performed using BigDye Terminator version 3.1 sequencing kit (Applied Biosystems, Courtaboeuf, France). All samples were sequenced, on both strands, over an 878-bp fragment of COI gene, with primers C1-J-2183a (5′-CAA CAT TTA TTT TGA TTT TTT GG-3′; Simon et al. 1994) and TL2-N-3014 (5′-TCC AAT GCA CTA ATC TGC CAT ATT A-3′; Simon et al. 1994), and equally over a 693-bp of COII with primers C2-J-3138a (5′-AGA GCC TCT CCT TTA ATA GAA CA-3′; Simon et al. 1994) and TK-N-3782 (5′-GAG ACC ATT ACT TGC TTT CAG TCA TCT-3′; Bogdanowicz et al. 1993). Sequencing reactions were purified using Sephadex G-50 (Amersham; Sigma-Aldrich) and detected with an ABI 3100 automatic sequencer. For all individuals, fragments of 792 continuous nucleotides of COI and 521 of COII could be reliably amplified. These fragments of COI and COII genes correspond to base pair positions 441–1319 and 1397–2090, respectively, of the *Drosophila yakuba* reference sequence (Simon et al. 1994).

Sequence electropherograms were examined and corrected using Sequence Navigator 1.0.1 (Applied Biosystems) and were aligned with CLUSTAL W version 1.4 (Thompson et al. 1994) as implemented in BIOEDIT version 4.7.8. Sequences are available in GenBank under the accession numbers EU556542 to EU556700 (Table S3).

Determination of nuclear integrations (*Numts*)

A mixture of two bases was observed at several sequence positions represented by two overlapped peaks on the electropherograms. This phenomenon was confirmed on both strands and by numerous assays of amplifying and sequencing the ambiguous sequences. DNA Strider 1.4f3 (software for DNA and protein sequences analysis) was used to analyse the nucleotide sequences in order to clarify the changes of the amino acids due to these mutations and their codon positions. Codon position bias was tested with a chi-squared test. To investigate on the origin of these double peaks, we performed an isolation of mtDNA by using the Tamura and Aotsuka (1988) alkaline lysis protocol modified by Sunnucks and Hales (1996). Ten individuals were used for this analysis (two fresh larvae, one fresh adult and seven dead adults frozen at −80°C) (for collection sites see Table S1). Fifty micrograms of tissue were taken from each sample tested (three last larval abdominal segments and five legs from adults). After the alkaline lysis, we proceeded in a classic extraction with the Sigma kit, skipping the first part of cell lysis. PCR and sequencing followed, with the mitochondrial primers previously defined. In addition, we proceeded into a PCR with nuclear primers to verify that there was no nuclear contamination. Primers D1F (5′-ACC CGC TGA ATT TAA GCA TAT-3′; Harry et al. 1996) and D3R (5′-TAG TTC ACC ATC TTT CGG GTC-3′; Lopez-Vaamonde et al. 2001) of the 28S gene were used for this verification.

To check whether longer PCR fragment yielded to amplification of both nuclear and mitochondrial DNA, a 1600 bp fragment corresponding to a part of the COI, tRNA leucine and COII genes was amplified and sequenced for eight individuals using a combination of the two more external primers (C1-J-2183a and TK-N-3782). In addition, three resulted amplified individuals were cloned using the StrataClone™ PCR Cloning Kit (STRATAGENE; Zuidoost, Amsterdam, The Netherlands). Clones were screened for inserts of the expected size by PCR amplification with the universal primers M13 and T7. Selected PCR products were purified and sequenced directly for both strands as detailed previously. Clones to be sequenced were taken in random. Clones' sequences are available in GenBank under the accession numbers EU599191 to EU599210. A Neighbor-joining tree based on Kimura 2-P was reconstructed with PAUP using 1486 bp sequences (base pair positions 441–2090 of the *D. yakuba* reference sequence, Simon et al. 1994) of the clones (mtDNA-like sequences), the *M. galloprovincialis* sequences without ambiguities and *M. sutor* as outgroup.

Results

Only mtDNA from fresh material could be unambiguously separate from nuclear DNA using the alkaline lysis protocol. For the three fresh individuals submitted to the alkaline lysis for mtDNA isolation, amplification with nuclear primers failed and double peaks totally disappeared when sequencing with mtDNA primers, whereas double peaks persisted for the seven remaining ethanol preserved or frozen individuals .

For the three individuals (GLor3, GLor6, GAud4 see Table S1 for details) whose part of the COI, tRNA leucin and COII genes were amplified, cloning resulted in multiple mtDNA-like sequences, 7, 8 and 5 clones respectively. Pairwise comparison of the mtDNA-like sequences showed two to 44 differences, when the complete data set (COI, tRNA leucin and

COII) was considered. In the case of two clones of the individual GLor6 (GCL2.4 and GCL2.3), we obtained clean reads which exhibited a deletion of 13 nucleotides at the beginning of the COII gene.

Among the 148 *M. galloprovincialis* individuals analysed for the two separate genes, 146 and 136 were successfully amplified and sequenced for a 792 bp fragment of the COI and a 521 bp fragment of the COII genes respectively. The alignment of the sequences obtained revealed no deletions or insertions but 19 polymorphic sites for the COI and 24 for the COII gene (Tables S2 and S4). Ambiguities were found in 14 and 17 of these polymorphic sites, for each gene respectively. Ambiguous sites were observed in half of the *M. galloprovincialis* individuals; 75 and 61 (51.4% and 44.8%) individuals showed ambiguities for the COI and the COII respectively. Mean ambiguous sites number found per individual was 1.72 (0.2%) and 2.77 (0.5%) for the two genes respectively. No ambiguities were observed in the *M. sutor* sequences obtained, whatever the gene studied.

For substitution location and event as well as for ambiguity frequency on each position and transition/transversion rate see Table S4. Most of the polymorphic sites, i.e. 12 (63.2%) and 18 (75%) for the COI and the COII, respectively, correspond to synonymous mutations. Even though sequence analysis revealed that six (31.6%) of the ambiguous positions for the COI fragment and six (25%) for the COII were non-synonymous mutations, no codon stops were formed for any of the ambiguous positions in the two fragments when they were translated with the mitochondrial genetic code. Synonymous to non-synonymous substitutions bias was five times more important for ambiguous sequences than for mitochondrial ones (10/1 for mtDNA and 20/10 for ambiguous sequences) Transitions were more abundant than transversions for both genes studied. mtDNA substitutions were predominantly observed at third codon positions (90%), less at second positions (10%) and never at the first ones. Considering the ambiguous sequences, most of the substitutions were also found at third positions (67.7%), sometimes at first positions (19.4%) and rarely at second positions (12.9%). When examining the codon position bias in ambiguous sequences, the deficit of second position and excess of third positions led to a pattern significantly different from *M. galloprovincialis* mtDNA codon position pattern ($\chi^2 = 6.37$, 2 d.f.) at $p = 0.05$ but not significantly different at $p = 0.025$. When random codon position bias (1 : 1 : 1) was tested significantly different pattern was observed ($\chi^2 = 16.71$, 2 d.f., $p = 0.01$).

Substitutions states of *M. sutor* sequences equivalent to each of the polymorphic sites (ambiguous or not) found for *M. galloprovincialis* were also added in Table S4, for comparison and further discussion on the evolutionary events.

Five of the 43 substitutions (positions 632, 751, 1122, 1478 and 1537) were observed in both *M. galloprovincialis* and *M. sutor* species. Consequently, as no *Numts* were found for *M. sutor*, the ancestral mitochondrial origin is reinforced, these changes may have occurred prior to the transposition.

Five other ambiguities (positions 1100, 1149, 1165, 1178 and 1179) were always observed at polymorphic sites in *M. galloprovincialis* sequences but only one of the two possible states was found in *M. sutor* sequences. So the former hypothesis is invalid for these cases. These mutations could be considered as real *Numts*' mutations because, even though no codon stops are formed, four out of five mutations changed the amino acid and three of these changes implemented biochemical change as well. Moreover, they were found only in one or two sequences and so they could be occasional *Taq* polymerase misincorporation.

The 33 remaining substitutions (Table S4) appeared on polymorphic sites, and for a large number of *M. galloprovincialis* sequences they were characterized by double peaks, contrary to *M. sutor* sequences that were monomorphic (presenting only one of the two states proposed for *M. galloprovincialis*). In this case, it is difficult to generate hypotheses on the time of their apparition but we could say that they seem to appear after the separation of the two species.

Fig. 1 presents the geographical distribution of individuals with and without ambiguous sites observed in each gene fragment, COI and COII, for each sampled population of *M. galloprovincialis*. A great variability was observed within populations of *M. galloprovincialis*. Distinct haplotypes reached the number of 9 per population (in the average, 1.2 for the COI and 1.7 for the COII). It is shown that *Numts* have a large and heterogeneous repartition. Interestingly, populations from which individuals did not exhibit *Numts* are located in the south of the sampling area, beneath the latitude of 45°58′, with the exception of Austrian populations found at maximum latitude of 47°53′. For each haplotype, in both COI and COII genes, variable nucleotide positions were showed (Table S4).

Fig. 2 presents a phenetic analysis of nuclear and mitochondrial DNA sequences. Eight different mtDNA-like sequences were grouped together with *M. galloprovincialis* mtDNA sequences, whereas a large group of more closely related mtDNA-like sequences formed a separate clade. This nuclear family of pseudogenes are basal to all *M. galloprovincialis* mtDNA haplotypes and branche off after the split with *M. sutor*.

Discussion

Our findings, especially after mtDNA isolation by alkaline lysis, support the hypothesis that the multiple sequences amplified for *M. galloprovincialis* individuals are not of mtDNA nature. The fact that for fresh individuals only one sequence was amplified after mtDNA isolation argues against heteroplasmy and confirms the existence of mtDNA transpositions to the nucleus (i.e. *Numts*). Nevertheless heteroplasmy cannot be definitively excluded because seven individuals tested with this method maintained the double peaks in their sequences, even if we could consider this fact as a contamination. Using purified mtDNA has been recommended to check that amplified sequence is not of nuclear origin, especially for recent transfer events (Smith et al. 1992), but despite the fact that more fresh individuals would better support our findings, testing the totality of the individuals sampled had to be abandoned due to inadequate conditions of material conservation for alkaline lysis procedures. Bensasson et al. (2000) and Ibarguchi et al. (2006) reported that mitochondria of specimens conserved in alcohol are not well preserved and consequently chromosomal contamination is possible during alkaline lysis. As also proved by our results material preservation is indeed of great importance for the success of the mtDNA isolation.

Other cause of double peaks on sequence electropherograms, such as error produced during PCR and sequencing, cannot be neglected. We are aware that no polymerase is

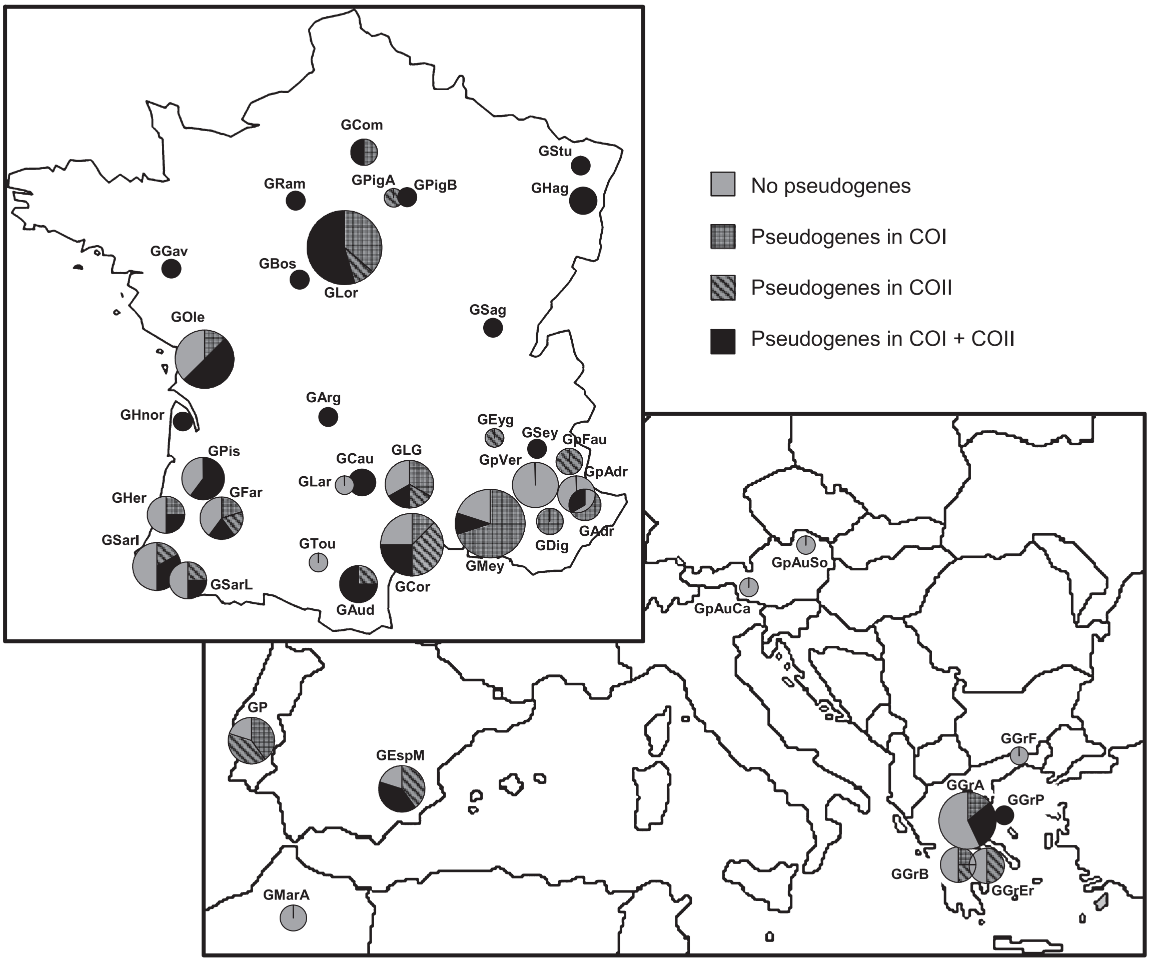

Fig. 1. Sampling localities of *Monochamus galloprovincialis* and presentation of *Numts* geographic distribution according to their presence in the COI, COII or both genes. Population codes were given in Table S1

error-free, but technical artifacts cannot occur almost exclusively on point mutation sites (hot spots, 57% COI and 100% COII). Besides, artifacts cannot be responsible for all of the observed ambiguities, because error produced by using the Taq Red polymerase is insignificantly small (Frey and Frey 2004) and we used the BigDye Terminator for sequencing which is considered as one of the most performing. The results were reproducible and similar outcomes occur in both strands sequenced. For more information on the evaluation of PCR and automated DNA sequencing components see Hancock et al. (2005), Keohavong and Thilly (1989), Nollau and Wagener (1997) and Brandstätter et al. (2005).

Unexpected was the extremely high number of *M. galloprovincialis* individuals exhibiting *Numts* and the high number of different mtDNA like clones within individuals. Interestingly, no *Numts* have been found in the Cerambycidae so far. Only three reports of *Numts* in Coleoptera were found in the literature, one from Scolytinae beetles (Jordal 2007), one sequence from *Tribolium brevicornis* (LeConte, 1859) (Mestrovic et al. 2006) and finally, Pons and Vogler (2005) revealed a very high percentage of mitochondrial rRNA pseudogenes in the tiger beetles (Coleoptera, Carabidae, Cicindelinae). According to these authors, the lack of pseudogene descriptions from Coleoptera is probably due to their paucity, rather than to the scanty attention paid to them. Furthermore, it is generally admitted that organisms with small genomes, like Coleoptera (Pons and Vogler 2005) and chicken (Pereira and Baker 2004), rarely harbour *Numts* or other pseudogenes compared to organisms with large nuclear genomes, like grasshoppers (Bensasson et al. 2000) and humans (Mourier et al. 2001; Woischnik and Moraes 2002). But, Richly and Leister (2004) showed that there is no correlation between abundance of *Numts* and size or density of the genomes. Anyhow, even if the whole genome sequence and size are not yet known for *M. galloprovincialis*, its *Numts* pattern seems to fit the one proposed for the Coleoptera mentioned above.

Monochamus galloprovincialis Numts have all the characteristics of a very recent event. The majority of the ambiguous sites correspond to synonymous, third codon position, mutations for both mitochondrial fragments. No *Numts* were found for *M. sutor*, the sister species of *M. galloprovincialis,* whereas the genetic distance between the two species is very small (3.7% at most, F. A. Koutroumpa, C. Lopez-Vaamonde, D. Rougon, C. Bertheau, U. Tomiczek, F. Lieutier and G. Roux-Morabito, in prep.), suggesting recent speciation. Moreover, a recent phylogenetic study of the genus *Monochamus* using the COI gene did not mentioned any paralogous sequences of mtDNA, even in other closely related species (Cesari et al. 2005) but only few specimens per species were considered in this study. When closely related species were studied for pseudogenes, heterogeneity in the per cent divergence between *Numts* and current mtDNA sequences has

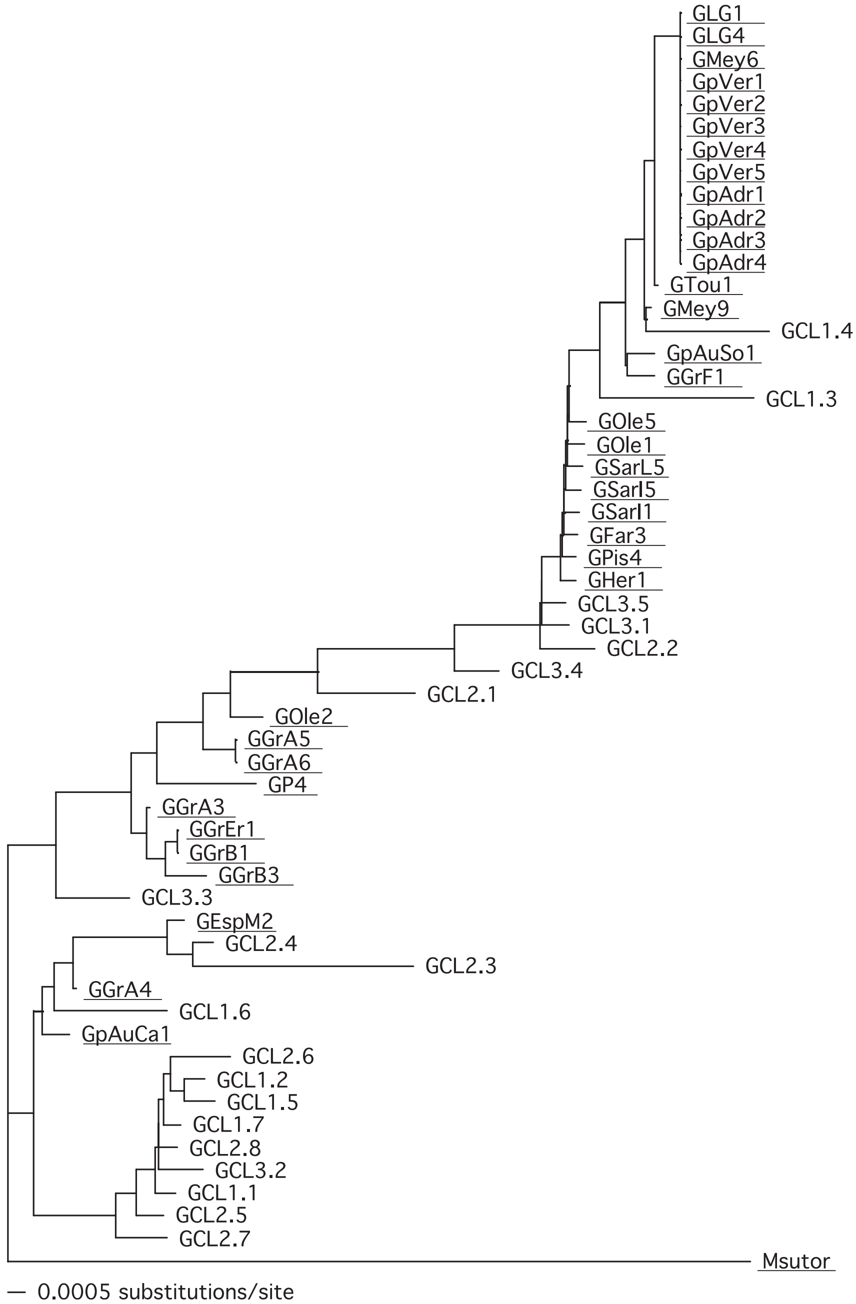

Fig. 2. Neighbor-joining tree (based on Kimura 2-P distances) of mtDNA-like sequences and mtDNA sequences (individuals without ambiguities) of *Monochamus galloprovincialis*, with *Monochamus sutor* as outgroup. The mtDNA sequences are underlined, and mtDNA-like sequences are labelled as the individual they were taken from (mtDNA-like sequence number 1 = individual GLor3, number 2 = GLor6; number 3 = GAud4) followed by the number of the clone that was sequenced, e.g. GCL1.6 is a sequence found in GLor3, and its sequence is given by clone 6

already been noticed. For example, the ND5 like *Numts* diverged at 1.4% from mtDNA sequences in *Schistocerca gregaria*, while the divergence reached 12.5% in *Podisma pedestris* (Bensasson et al. 2000). Nevertheless, the high number of *Numts* detected in *M. galloprovincialis* compared to the absence of such sequences in *M. sutor* suggests differences in the ability of different genomes to gain or lose *Numts* (Bensasson et al. 2000).

Phenetic reconstruction shows that mtDNA-like sequences observed in this species might not result from a unique mitochondrial transposition in the nuclear genome in ancestral populations, but from at least nine independent transfer events, i.e. eight recently integrated in nucleus and one family of pseudogenes which are closely related. This family formed a more basal clade closer to ancestral states than to *M. galloprovincialis* mtDNA sequences and closer to *M. sutor* mtDNA sequence. This result supports that this family of nuclear pseudogenes is likely to have descended from a single mitochondrial immigrant which has been amplified and diverged as its transfer. Independent mitochondrial origins of nuclear pseudogenes within insect species have already been described. For example, it was established that there have been at least 12 separate mtDNA integrations into the grasshopper *P. pedestris* nuclear genomes (Bensasson et al. 2000). Likewise, a family of almost identical *Numts* has been described into *Italopodisma* spp., in addition to multiple horizontal transfer events which could also explain *Numts* diversity within individuals (Bensasson et al. 2000). Multiple independent transposed sequences of mitochondrial sequences have also been observed among species, such in Aphids of the genus *Sitobion* (Sunnucks and Hales 1996) and in seven species of diving ducks (Sorenson and Fleischer 1996). However, further study is needed to explain better the events involved in transfer of mitochondrion to nucleus in *M. galloprovincialis*.

It is not well known which factors determine whether *Numts* are common in a species (Bensasson et al. 2000). Likewise *Numts* geographic distribution within a species and their frequencies within populations have been rarely studied. Pons and Vogler (2005) found that 59% of the *Rivancindela* individuals studied presented *Numts* and 7 out of 65 populations had zero, one or two individuals without *Numts*. In our sampling, also, 22 out of 43 populations had individuals without *Numts*. Even though, we strongly believe the lack of *Numts* from these individuals we could not exclude a PCR failure due, for example, to *Numts*' mutations on primers annealing sites. When considering the geographic distribution of *M. galloprovincialis Numts* in the southern area of this species repartition, a wide geographic distribution was observed (Fig. 1), whereas populations from the North of France recorded no individuals without *Numts*. Even though, our sampling is not representative of the whole distribution of *M. galloprovincialis*, our results raise questions about the apparition and evolutionary patterns of these transposed sequences in this species. It has been shown that unfavourable environmental factors and different origins of mitochondrial stress could favour mtDNA escape into the cytoplasm (Bensasson et al. 2001b; Woischnik and Moraes 2002). *Monochamus galloprovincialis* is considered to have a mediterranean origin (Hellrigl 1971; Sama 2002) and unfavourable environmental conditions (as lower temperature or/and absence of the original hosts *P. pinaster* and *P. halepensis* in northern regions) could be a pitfall for *M. galloprovincialis* individuals. However, we cannot rule out the hypothesis that factors such as distance isolation, founder effect or past climatic oscillations with contrasted patterns of recolonization may have affected the observed distribution of *M. galloprovincialis Numts*. Larger sampling is clearly needed in order to test these hypotheses.

Using mtDNA as a molecular marker in evolutionary studies becomes problematic when *Numts* are involved, especially for assessment of species boundaries (Pons and Vogler 2005; Teletchea et al. 2005; Park et al. 2007), through universal identification systems like the DNA-based barcoding system, recently proposed by Hebert et al. (2003). Usually, when ambiguous sites are detected, they are withdrawn from analyses (Brandstätter et al. 2005; Mestrovic et al. 2006). However, the study of *Numts* provides a simultaneous view on the dynamics of mtDNA and nuclear DNA evolution and interactions (Bensasson et al. 2000) and consequently is of great evolutionary interest. Moreover, *Numts* detection can vary within species, populations and, quite often, tissues of the same individual should be considered when dealing with them (Bensasson et al. 2001a,b, Ibarguchi et al. 2006; this study). A warning is given about recently translocated genes which, due to less or lack of divergence, are at highest risk of being amplified together or instead of their mitochondrial counterparts and in this case discrimination from them is very hard and sometimes impossible. So far, phylogenetic studies involving both *Numts* and mtDNA sequences showed contrasted results, in some cases placing the *Numts* sequences relative to their mtDNA counterparts (e.g. in Sorenson and Fleischer 1996), and in other cases grouping paralogous nuclear sequences with more than one taxa (*Aythya valisineria* and *A. ferina* nuclear sequence, Sorenson and Fleischer 1996; *P. pedestris Numts* sequences distributed within Podisma family, Bensasson et al. 2000).

This work highlights the importance for evolutionary research of studying transposed mtDNA and the great risk of ignoring them in phylogenetic and phylogeographic studies. We further encourage scientists to treat pseudogenes as important elements to every biological study rather than obstacles that should be eliminated from analyses.

Acknowledgements

We would like to thank Jerome Rousselet, Christian Stauffer, Wolfgang Arthofer and François Hericourt for helpful discussion and suggestions as well as Emmanuelle Magnoux for her assistance with DNA sequencing and cloning.

Résumé

Incorporation de fragments mitochondriaux dans le génome nucléaire (Numts) du longicorne Monochamus galloprovincialis (Coleoptera Cerambycidae).

Une étude de l'ADNmt, basée sur 43 populations (33 échantillonnées en France) de *Monochamus galloprovincialis*, vecteur du nématode du pin *Bursaphelenchus xylophilus,* et 14 populations de l'espèce sœur *M. sutor* a été réalisée. Le séquençage de 792 pb du gène COI et 521 pb du COII a révélé que près de la moitié des individus analysés (44,8%) présentaient des ambiguïtés sur plusieurs sites nucléotidiques. Les hypothèses de la présence d'hétéroplasmie ou de pseudogènes mitochondriaux (*Numts*) ont été examinées. La séparation de l'ADNmt par lyse alcaline et le clonage (utilisés avec succès sur trois individus) confortent tous les deux l'hypothèse de l'origine non mitochondriale de ces séquences ambiguës et valident la présence de *Numts* dans le génome nucléaire de *M. galloprovincialis*. De multiples copies similaires a des séquences mitochondriales se sont révélées des paralogues des gènes COI, tRNA leucine et COII. Une analyse phénétique a placé les paralogues qui ont divergé récemment près des sequences mitochondriaux et a groupé dix autres paralogues à un clade plus basal et plus près des statuts ancestraux et à *M. sutor*. Ce résultat montre que cette famille de séquences nucléaires de pseudogènes est apparue indépendamment des autres événements et représente probablement des haplotypes mitochondriaux des populations ancestrales de *M. galloprovincialis*. C'est la première fois que des *Numts* sont démontrés chez un longicorne malgré leur absence chez *M. sutor*. Le mécanisme de l'incorporation de *Numts* est inconnu pour *M. galloprovincialis*, cependant, l'excès de sites ambigus correspondants à des mutations synonymes en troisième position du codon chez *M. galloprovincialis* ainsi que l'absence de *Numts* chez *M. sutor*, conduisent à l'hypothèse d'une apparition récente de *Numts* dans le génome nucléaire de *M. galloprovincialis*.

References

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst **18:**489–522.

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

Bensasson D, Petrov DA, Zhang D, Hartl DL, Hewitt GM (2001a) Genomic gigantism: DNA loss is slow in mountain grasshoppers. Mol Biol Evol **18:**246–253.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001b) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol **16:**314–321.

Bogdanowicz SM, Wallner WE, Bell J, O'Dell TM, Harrison RG (1993) Asian gypsy moths (Lepidoptera, Lymantriidae) in North America – evidence from molecular data. Ann Entomol Soc Am **86:**710–715.

Boyce TM, Zwick ME, Aquadro CF (1989) Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123:**825–836.

Brandstätter A, Sänger T, Lutz-Bonengel S, Parson W, Béraud-Colomb E, Wen B, Kong QP, Bravi CM, Bandelt HJ (2005) Phantom mutation hotspots in human mitochondrial DNA. Electrophoresis **26:**3414–3429.

Cesari M, Marescalchi O, Francardi V, Mantovani B (2005) Taxonomy and phylogeny of European *Monochamus* species: first molecular and karyological data. J Zool Syst Evol Res **43:**1–7.

Coyer JA, Hoarau G, Stam WT, Olsen JL (2004) Geographically specific heteroplasmy of mitochondrial DNA in the seaweed, *Fucus serratus* (Heterokontophyta: Phaeophyceae, Fucales). Mol Ecol **13:**1323–1326.

Crochet PA, Desmarais E (2000) Slow rate of evolution in the mitochondrial control region of gulls (Aves: Laridae). Mol Biol Evol **17:**1797–1806.

Frey JE, Frey B (2004) Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? Hereditas **140:**92–98.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol S **34:**397–423.

Godfrey HCJ (2002) Challenges for taxonomy. Nature **417:**17–19.

Hancock DK, Tully LA, Levin BC (2005) A standard reference material to determine the sensitivity of techniques for detecting low-frequency mutations, SNPs, and heteroplasmies in mitochondrial DNA. Genomics **86:**446–461.

Harry M, Solignac M, Lachaise D (1996) Adaptative radiation in the Afrotropical region of the Paleotropical genus *Lissocephala* (Drosophilidae) on the pantropical genus *Ficus* (Moraceae). J Biogeogr **23:**543–552.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond **270:**313–321.

Hellrigl KG (1971) La bionomie des espèces de *Monochamus* (Coleoptera, Cerambycidae) et leur importance pour la sylviculture et l'économie du bois. Redia **52:**367–511.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. Mol Biol Evol **19:**2261–2264.

Ibarguchi G, Friesen VL, Lougheed SC (2006) Defeating *Numts*: semi-pure mitochondrial DNA from eggs and simple purification methods for field-collected wildlife tissues. Genome **49:**1438–1450.

Jordal BH (2007) Reconstructing the phylogeny of Scolytinae and close allies: major obstacles and prospects for a solution. USDA For Serv Proc **45:**3–9.

Karro JE, Yan Y, Zheng D, Zhang Z, Carriero N, Cayting P, Harrison P, Gerstein M (2007) Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. Nucleic Acids Res **35:**D55–D60.

Katsares V, Apostolidis A, Triantafyllidis A, Kouvatsi A, Triantaphyllidis C (2003) Development of mitochondrial DNA primers for use with Homarid lobsters. Mar Biotechnol **5:**469–479.

Keller I, Bensasson D, Nichols RA (2007) Transition–transversion bias is not universal: a counter example from grasshopper pseudogenes. PLoS Genetics **3:**185–191.

Keohavong P, Thilly WG (1989) Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci U S A **86:**9253–9257.

Kondo R, Satta Y, Matsuura ET, Ishiwa H, Takahata N, Chigusa SI (1990) Incomplete maternal transmission of mitochondrial DNA in drosophila. Genetics **126:**657–663.

Koutroumpa FA (2007) Biologie et phylogeographie de Monochamus galloprovincialis (Coleoptera, Cerambycidae) vecteur du nematode du Pin. Thesis, University of Orléans, France, pp 188.

Kvist L, Martens J, Nazarenko AA, Orell M (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Mol Biol Evol **20:**243–247.

Ladoukakis ED, Zouros E (2001) Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. Mol Biol Evol **18:**1168–1175.

Lopez JV, Yuhki N, Mascuda R, Modi W, O'Brien SJO (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.

Lopez-Vaamonde C, Rasplus JY, Weiblen GD, Cook JM (2001) Molecular phylogenies of fig wasps: partial cocladogenesis of pollinators and parasites. Mol Phylogenet Evol **21:**55–71.

Mamiya Y (1976) Pine wilt disease caused by the Pine wood nematode, *Bursaphelenchus lignicolus*, in Japan. Jnp Agr Res Q **10:**206–212.

Mestrovic N, Mravinac B, Plohl M, Ugarcovic D, Bruvo-Madaric B (2006) Preliminary phylogeny of *Tribolium* beetles (Coleoptera: Tenebrionidae) resolved by combined analysis of mitochondrial genes. Eur J Entomol **103:**709–715.

Moritz C, Dowling TE, Brown WN (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann Rev Ecol Syst **18:**269–292.

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Mol Biol Evol **18:**1833–1837.

Nickle WR (1970) A taxonomic review of the genera of the Aphelenchoidea (Fuchs, 1937) Thorne, 1949 (Nematoda: Tylenchida). J Nematol **2:**375–392.

Nollau P, Wagener C (1997) Methods for detection of point mutations: performance and quality assessment. Clin Chem **43:**1114–1128.

Park J-M, Manen J-F, Schneeweiss GM (2007) Horizontal gene transfer of a plastid gene in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). Mol Phylogenet Evol **43:**974–985.

Pereira SL, 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 **4:**1–8.

Pons J, Vogler AP (2005) Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudogene in a recent radiation of tiger beetles. Mol Biol Evol **22:**991–1000.

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

Rokas A, Ladoukakis E, Zouros E (2003a) Animal mitochondrial DNA recombination revisited. Trends Ecol Evol **18:**411–417.

Rokas A, Melika G, Abe Y, Nieves-Aldrey J-L, Cook JM, Stone GN (2003b) Lifecycle closure, lineage sorting, and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. Mol Phylogenet Evol **26:**36–45.

Sama G (2002) Atlas of the Cerambycidae of Europe and the Méditerranean Area. Vit Kabourek, Zlín, pp 173.

Sekiguchi K, Kasai K, Levin BC (2003) Inter- and intragenerational transmission of a human mitochondrial DNA heteroplasmy among 13 maternally-related individuals and differences between and within tissues in two family members. Mitochondrion **2:**401–414.

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.

Smith MF, Thomas WK, Patton JL (1992) Mitochondrial DNA-like sequence in the nuclear genome of an akodontine rodent. Mol Biol Evol **9:**204–215.

Solignac M, Monnerot M, Mounolou JC (1983) Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. Proc Natl Acad Sci USA **80:**6942–6946.

Sorenson MD, Fleischer RC (1996) Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proc Natl Acad Sci U S A **93:**15239–15243.

Sousa E, Bravo MA, Pires J, Naves P, Penas AC, Bonifacio L, Mota MM (2001) *Bursaphelenchus xylophilus* (Nematoda; Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera; Cerambycidae) in Portugal. Nematology **3:**89–91.

Steiner G, Buhrer EM (1934) *Aphelenhoides xylophilus*, n. sp. a nematode associated with blue-stain and other fungi in timber. J Agric Res **48:**949–951.

Sunnucks P, Hales DF (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.
Teletchea F, Maudet C, Hanni C (2005) Food and forensic molecular identification: update and challenges. Trends Biotechnol **23:**359–366.
Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673–4680.
Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res **12:**885–893.

Authors' addresses: Fotini A. Koutroumpa (for correspondence) and François Lieutier, Laboratoire de Biologie des Ligneux et des Grandes Cultures, UPRES-EA-1207, Université d'Orléans, BP6759, Rue de Chartres, 45067 Orléans cedex 2, France. E-mail: fotini.koutroumpa@univ-orleans.fr; francois.lieutier@univ-orleans.fr; Géraldine Roux-Morabito, INRA, Centre d'Orléans, Unité de Zoologie Forestière, BP20619 Ardon, 45166 Olivet cedex, France. E-mail: geraldine.roux-morabito@orleans.inra.fr

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. *Monochamus galloprovincialis* population characteristics and collectors' names.

Table S2. *Monochamus sutor* population characteristics and collectors' names.

Table S3. Haplotypes' (HT) variable nucleotide sites of *Monochamus galloprovincialis*. Variable sites position is given according to the alignment with the *Drosophila yakuba* reference sequence (Simon et al. 1994) and their codon position is noted. For ambiguous substitutions both states are shown. Haplotypes' correspondence to the sampled populations is given in Table S1.

Table S4. Position and characteristics of polymorphic sites of *Monochamus galloprovincialis* sequences and comparison with *Monochamus sutor* sequences.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article