ORIGINAL ARTICLE

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

WILEY

Mind the *numt*: Finding informative mitochondrial markers in a giant grasshopper genome

Ricardo J. Pereira[1] 🆔 | Francisco J. Ruiz-Ruano[2,3,4] 🆔 | Callum J.E. Thomas[1] |
Mar Pérez-Ruiz[5] | Miguel Jiménez-Bartolomé[5] | Shanlin Liu[6] | Joaquina de la Torre[5,7] |
José L. Bella[5,7]

[1]Division of Evolutionary Biology, Faculty of Biology II, Ludwig- Maximilians-Universität München, Planegg-Martinsried, Germany

[2]Department of Genetics, University of Granada, Granada, Spain

[3]Department of Ecology and Genetics – Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden

[4]Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden

[5]Departamento de Biología (Genética), Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

[6]Department of Entomology, College of Plant Protection, China Agricultural University, Beijing, China

[7]Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Madrid, Spain

Correspondence
Ricardo J. Pereira, Division of Evolutionary Biology, Faculty of Biology II, Ludwig-Maximilians-Universität München, Grosshaderner Strasse 2, Planegg-Martinsried 82152, Germany.
Email: ricardojn.pereira@gmail.com

Present address
Mar Pérez-Ruiz, Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen, Copenhagen N, Denmark
Miguel Jiménez-Bartolomé, Institute of Environmental Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

Funding information
H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 658706; Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: PID2019-104952GB-I00/AEI/10.13039/501100011033

Abstract

The barcoding of the mitochondrial *COX1* gene has been instrumental in cataloguing the tree of life, and in providing insights in the phylogeographic history of species. Yet, this strategy has encountered difficulties in major clades characterized by large genomes, which contain a high frequency of nuclear pseudogenes originating from the mitochondrial genome (numts). Here, we use the meadow grasshopper (*Chorthippus parallelus*), which possesses a giant genome of ~13 Gb, to identify mitochondrial genes that are underrepresented as numts, and test their use as informative phylogeographic markers. We recover the same full mitochondrial sequence using both whole genome and transcriptome sequencing, including functional protein-coding genes and tRNAs. We show that a region of the mitogenome containing the *COX1* gene, typically used in DNA barcoding, has disproportionally higher diversity and coverage than the rest of the mitogenome, consistent with multiple insertions of that region into the nuclear genome. By designing new markers in regions of less elevated diversity and coverage, we identify two mitochondrial genes that are less likely to be duplicated as numts. We show that, while these markers show high levels of incomplete lineage sorting between subspecies, as expected for mitochondrial genes, genetic variation reflects their phylogeographic history accurately. These findings allow us to identify useful mitochondrial markers for future studies in *C. parallelus*,

Pereira and Ruiz-Ruano have contributed equally to this work.

Contributing authors: Francisco J. Ruiz-Ruano (fjruizruano@gmail.com), Callum J.E. Thomas (callumthomas@outlook.com), Mar Pérez-Ruiz (marperez89@gmail.com), Miguel Jiménez-Bartolomé (miguel.jimenez-bartolome@boku.ac.at), Shanlin Liu (shanlin1115@gmail.com), Joaquina de la Torre (joaquina@uam.es), José L. Bella (bella@uam.es)

[Correction added on 15 February 2022 after first publication: the article category has been corrected to 'Original Article' in this version.]

----

an important biological system for evolutionary biology. More generally, this study exemplifies how non-PCR-based methods using next-generation sequencing can be used to avoid numts in species characterized by large genomes, which have remained challenging to study in taxonomy and evolution.

1 | INTRODUCTION

The use of mitochondrial DNA has revolutionized the fields of phylogeography and taxonomy (Avise, 2000; Zhang & Hewitt, 1996). Due to the high copy number of mitochondria relative to the nuclear genome, its stable preservation in museum samples, and its fast rate of evolution, the mitochondrial genome has become the molecular marker of choice to catalogue biodiversity. In particular, the mitochondrial *cytochrome c oxidase subunit I gene* (*COX1* or *COI*) has become the most commonly used "DNA barcoding gene" in animal taxonomy because it shows high rates of sequence change between most animal groups while simultaneously showing constrained evolution within species (Hebert et al., 2003). Although this gene has played a pivotal role in global efforts to identify and map the distribution of species (Hebert et al., 2009), these efforts have encountered difficulties in some of the most important branches of the animal tree of life (Buhay, 2009). This is due to abundant nuclear copies of mitochondrial DNA (termed "numts" (Lopez et al., 1994)), which can lead to the inadvertent co-amplification or preferential amplification of nuclear pseudogenes when amplified using PCR-based methodologies (Calvignac et al., 2011; Hazkani-Covo et al., 2010). Because mitochondrial and nuclear markers have different rates of evolution and modes of inheritance, numts can lead to erroneous estimates of coalescent times and intraspecific variation, making studying the evolutionary history of many organisms challenging using standard PCR-based methods (Bensasson et al., 2001).

Gene transfer from the organelle to the nuclear genome has been a key process in the evolution of the mitochondrial and chloroplast genomes and in the evolution of eukaryotic life (Gould et al., 2008; Kleine et al., 2009; Rand et al., 2004). This gene transfer is still an ongoing evolutionary process. Studies in humans and closely related species suggest that the rate of numt insertion is relatively high (~5.1–5.6×10^{-6} per germ cell per generation (Bensasson et al., 2003; Ricchetti et al., 2004)), yet only ~80% of numts are shared with chimpanzees (Hazkani-Covo & Graur, 2006), and less to more distantly related species. This decay of numts is believed to be caused by pseudogenization upon the arrival into the nucleus due to differences between the nuclear and the mitochondrial genetic codes (Perna & Kocher, 1996) as well as rapid methylation (Huang et al., 2005). Nevertheless, the rate of decay of numts varies widely across taxa, and recent studies sequencing eukaryotic genomes show that the frequency of numts is positively correlated with genome size (Bensasson et al., 2001; Hazkani-Covo et al., 2010; Kaya & Çıplak, 2018). PCR-based barcoding efforts using the *COX1* mitochondrial gene in organisms with large genomes, such as some invertebrate clades, have resulted in the co-amplification and sequencing of numts with stop codons (Buhay, 2009), leading to inaccurate sequence data. Although several laboratory strategies have been suggested to avoid amplifying numts (Ibarguchi et al., 2006), these tasks are often prohibitive for a large number of samples, limiting research programs at several significant animal clades.

Grasshoppers have some of the largest genomes among insects (Tsutsui et al., 2008), with genome sizes varying from ~6 Gb in *Locusta migratoria* (Wang et al., 2014) to ~17 Gb in *Podisma pedestris* (Westerman et al., 1987). Similar to other invertebrates with large genomes, the insertion of numts has been rampant during orthopteran diversification (Song et al., 2014). For example, numts paralogous to the *COX1* and *ND5* mitochondrial genes have been found in *Podisma pedestris*, with 87 *ND5*-derived numts resulting from at least 12 separate integrations into the nuclear genome (Bensasson et al., 2000). Although this has posed important challenges to studies using these genes (Moulton et al., 2010; Song et al., 2008), it is yet unclear whether other mitochondrial genes are less likely to have been transferred to the nuclear genome and thus constitute more reliable loci for use in PCR-based methodologies.

Here, we test this hypothesis using the short-horned meadow grasshopper *Chorthippus parallelus* (Zetterstedt, 1821). This species has recently been ascribed to the genus *Pseudochorthippus* (Defaut, 2012), based on a 653 bp fragment of the *COX1* gene (Vedenina & Mugue, 2011). However, due to lack of follow-up studies using other markers and for consistency with previous evolutionary studies on this system (Butlin, 1998; Hewitt, 1993), here we maintain the historical denomination of *Chorthippus parallelus*. This species is particularly suitable to determine whether some regions of the mitochondrial genome are disproportionally incorporated as numts, since its large nuclear genome of 13.26 ± 0.98 Gb (Lechner et al., 2012) is hypothesized to be tolerant to a high incidence of numts.

Several studies have established a phylogeographic hypothesis for the evolution of this species that is common across many European species (Hewitt, 2000). In short, during the last glacial maximum in Europe (18,000–20,000 years ago) ice sheets extended down across central Europe, isolating the distribution of temperate species to the southern peninsulas of Iberia, Italy and the Balkans (Hewitt, 1993). Geographic isolation in these refugia led to differentiation of the

subspecies *C. p. erythropus* in the Iberian Peninsula, and of the subspecies *C. p. parallelus* in the Italic and/or Balkan peninsulas (Korkmaz et al., 2014). With the warming of the climate starting ~10,000 years ago ice sheets began receding and these subspecies underwent a rapid post-glacial expansion following the expansion of suitable habitat. The two subspecies' ranges finally met in secondary contact along the ridge of the Pyrenees, forming a hybrid zone. This hybrid zone has remained narrow (between 10 and 42 km in a nuclear marker (Vázquez et al., 1994)), suggesting that genetic boundaries between subspecies are maintained by a balance between selection against hybrids and dispersal of parentals into the hybrid zone (Butlin, 1998; Hewitt, 1993). Experimental crosses have demonstrated selection against hybrids due to the interaction between the autosomal background and the X-chromosome (Hewitt et al., 1987), with the contribution of the cytosolic endosymbiont *Wolbachia* (Zabal-Aguirre et al., 2014). The strength of such incompatibilities depends on the maternal genetic background, potentially favoring asymmetric gene flow between subspecies as is observed in the X-chromosome. Although this evolutionary scenario is reflected in nuclear allozymic loci (Butlin, 1998; Hewitt, 1993), phylogeographic studies using the *COX1* gene show a low number of genetic differences between subspecies (Lunt et al., 1998), limiting the scope of evolutionary studies in this species.

In this study, we first use non-PCR-based methods, combining next-generation sequencing data derived from DNA and RNA, to assemble the complete mitogenome of *Chorthippus parallelus*. Second, we use reads from the whole genome to identify mitochondrial genes that are less likely to be represented in numts. Finally, we test if these mitochondrial genes are suitable for phylogeographic studies between these subspecies.

2 | MATERIALS AND METHODS

2.1 | Sampling

Samples for assembling the mitochondrial genomes were collected from localities within the range of the two pure subspecies of *Chorthippus parallelus*, that is outside of the known boundaries of the Pyrenean hybrid zone (Butlin, 1998; Hewitt, 1993). For the whole genome sequencing (WGS) dataset, we collected one male and one female from each subspecies: individuals of *C. p. erythropus* were collected in Escarrilla (Pyrenees, Spain, 42°43′54.1″N, 0°18′39.3″O), and individuals of *C. p. parallelus* were collected in Arudy (Pyrenees, France, 43°7′0″N, 0°25′60.0″W). Individuals were sacrificed and fixed in 100% ethanol. For the RNA sequencing (RNAseq) dataset, we sampled one adult male of *C. p. erythropus* from Escarrilla. This individual was sacrificed and preserved in RNAlater (Sigma). We use the terms "WGS" and "RNAseq" data to refer to the original molecules that were sequenced even though, in both cases, we only analyzed reads with homology to the mitogenome.

Samples for the phylogeographic study were collected throughout the known distribution of *C. p. erythropus* across the Iberian Peninsula, in order to represent as much of the genetic diversity as possible within this subspecies (Bella et al., 2007; Buño et al., 1994). In addition, we sampled *C. p. parallelus* from localities far removed from the Iberian Peninsula (England, Alps, Germany and Slovenia), to account for genetic variation found within the range of this subspecies. We have also sampled a transect across the Pyrenean hybrid zone (Serrano et al., 1996; Vázquez et al., 1994), in order to describe the genetic transitions in mitochondrial markers. Together, the phylogeographic sampling included 43 localities and 133 individuals, with between one and 14 individuals sampled in each locality (see details in Table S1). These samples were preserved in 100% ethanol. All tissue samples were stored at −20°C in a 2-mL centrifuge tube until extraction.

2.2 | Whole genome and transcriptomic sequencing

For DNA extraction, tissues were frozen in liquid nitrogen and subsequently homogenized in a mill (Retsch, Germany). TNES buffer (300 μL; 50 mM Tris–HCl, pH 8.0; 400 mM NaCl; 20 mM EDTA, pH 8.0; 0.5% SDS) containing proteinase K (0.03%w/v) was added to each homogenized sample and incubated at 37°C overnight. Standard phenol-chloroform extraction followed by purification with ethanol precipitation was used to extract DNA. Resulting DNA pellets were dissolved in Tris–EDTA (1 mM Tris–HCl pH 8.0; 1 mM EDTA pH 8.0) and incubated with 5 μg/mL RNase for 2 hr. DNA samples were standardized at a final concentration of 50 ng/μL using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, USA). In all cases, it was previously verified that the individuals were not infected by *Wolbachia*, using a nested PCR amplification of the *16S rRNA* gene, as described by Martínez-Rodríguez et al. (2013). The four samples were individually barcoded and sequenced in one lane of Illumina HiSeq 2000 with 101 bp paired-end sequencing. This generated between 4.5 and 6.4 Gb of reads for each library (accession number for the BioProject: PRJNA679335).

For the RNA extraction, tissue from the whole body was used with the exception of the head, to avoid contaminant RNA from the upper digestive tract and inhibitors associated with eye pigments. Total RNA was extracted using the standard Tri-Reagent protocol (Sigma), and re-suspended RNA pellets were further purified with RNeasy Mini columns (Qiagen). Final sample integrity and quantity were assessed with an Agilent 2100 BioAnalyzer. The mRNA enrichment, library construction, and sequencing were performed by the Beijing Genomics Institute (BGI), using 150 bp paired-end sequencing on an Illumina HiSeq 4000. This gave a final total of ~60 million cleaned reads, approximately 9 Gb of data (accession number for the BioSample: SAMN16264389; Nolen et al., 2020).

2.3 | Assembly of the mitochondria and identification of numts

The mitogenome was assembled for the four WGS libraries individually with NOVOPlasty software (Dierckxsens et al., 2017). The

complete mitogenome of *Chorthippus chinensis* (Liu & Huang, 2008; accession number NC_011095.1) was used as a seed, with a kmer size of 39. The assemblies derived from WGS are expected to include reads largely enriched for the original mitochondrial genes relative to numts, because of the larger amount of mitochondrial DNA sequenced relative to nuclear DNA per cell. However, recently duplicated numts are still expected to map to their original mitochondrial genes due to a lack of accumulated mutations.

The mitochondrial assembly from the RNAseq library was generated using Trinity de novo assembler (Grabherr et al., 2011), which uses Trimmomatic with default parameters for trimming and filtering the raw data. After testing three different kmer sizes, $K = 51$ was chosen as this retrieved the longest scaffolds. The assembled scaffolds were blasted to the mitogenome of *C. chinensis* used for the genomic assemblies above, which produced a single scaffold containing all mitochondrial genes. The assembly derived from RNAseq is expected to include exclusively original mitochondrial genes because numts become quickly pseudogenized and are not transcribed into mRNA.

The mitochondrial assemblies derived from DNA and RNA data were aligned using the MAFFT software (Katoh & Standley, 2013) with LINSI options (see Alignment S1 in Supplementary Information). This alignment was used to test if the gene order and sequence was similar between alternative assembly approaches.

The annotation of the mitogenome was performed using the WGS library of the female individual of *C. p. parallelus* as a reference. The MITOS2 program (Bernt et al., 2013) was used for the annotation of protein-coding genes, tRNAs, and rRNAs. In addition, the tRNA annotation was refined considering its secondary structures with ARWEN (Laslett & Canback, 2008). We then visualized the annotation of the whole mitogenome with OGDRAW (Greiner et al., 2019). The resulting assemblies and annotations were deposited in GenBank with accession numbers MT166298-302.

To test if certain mitochondrial genes are overrepresented in numts, the four genomic samples were used to estimate coverage and diversity across the mitogenome. The reads for each library were mapped against our reference mitogenome, using SSAHA2 (Ning et al., 2001) and considering reads with at least 40 bp and a minimal mapping identity of 80%. Then, coverage and nucleotide diversity per position were calculated for each individual, using the pysamstats software (https://github.com/alimanfoo/pysamstats), and 100 bp non-overlapping windows. The coverage values were normalized by the size of the smallest library (*C. p. erythropus* male; coverage ~0.34×). Mitochondrial genes that are overrepresented in numts are expected to have higher coverage and diversity relative to other mitochondrial genes due to the additional mapping of reads from diploid numts.

2.4 | Phylogeographic analyses

Primers were designed for two mitochondrial genes, *CYTB* and *COX3*, which show lower diversity and coverage, and thus are less likely to be represented in numts. All the 133 samples of the phylogeographic dataset were extracted as described for the WGS data. The samples were amplified and sequenced for both genes with forward and reverse primers (*CYTB*_fwd: CGA ACA CTA CAC GCA AAT GGA GCA; *CYTB*_rev: AGG TTC TTC AAC TGG TCG TTT TCC A; *COX3*_fwd: CCT TGA CCA TTA ACA GGA GCA ATT GGA; *COX3*_rev: TGT CAG TAT CAT GCT GCT GCT TCA A), using a T_m of 68 °C and 69 °C for *CYTB* and *COX3*, respectively. The length of the PCR products generated was 686 bp for *COX3* and 801 bp for *CYTB*. The ends of forward and reverse sequencing reads were trimmed in Geneious (v. 11.1.4). Generally, chromatograms showed single unambiguous peaks, but some positions had a secondary peak with a variable height. A first alignment without ambiguities was produced by keeping the original calls based on the highest peak. A second alignment including ambiguities was produced by calling ambiguities wherever the secondary peak was greater than 50% of the height of the primary peak, using Geneious (v. 11.1.4). Paired forward and reverse sequences were then aligned together for each individual, and where they overlapped, the highest quality base was used to call the final consensus sequence. Consensus sequences or single reads were aligned for each gene and gaps were removed (GenBank accessions for *CYTB*: MW2322466–MW232411, and for *COX3*: MW232097–MW232245; see Table S1 for details). The two mitochondrial genes were trimmed to the same length and concatenated for phylogenetic analyses, since these genes do not recombine and thus share the same genealogical history. The alignments with and without ambiguities were deposited in GenBank are available as Supplementary Information (Alignment S2 and S3, respectively).

To test if the new markers reflect the known phylogeographic history of *C. parallelus*, a phylogenetic network was estimated using the program PHYLOViZ online (https://online.phyloviz.net/). Because network methods do not accommodate ambiguities, we used the alignment without ambiguities to estimate a minimum spanning tree. For visualization purposes, each haplotype was color-coded considering the range of each subspecies, scaling notes and links with a factor of ten. The most divergent haplogroups were identified by varying the number of mutational steps between haplotypes (nLV value), and by choosing the haplogroups that were consistent across a wider range of mutations. The distribution of these haplogroups was then inferred on the basis of the geographic coordinates of the sampling localities.

To confirm that the phylogenetic signal of the network is not confounded by ambiguities caused by sequencing error, heteroplasmy or by a lower representation of these genes in numts, we also estimated a maximum likelihood tree, using the alignment with ambiguities. The *COX3* and *CYTB* sequences taken from the annotated *C. chinensis* genome were included as an outgroup. The optimal substitution models for each alignment were determined using Modeltest (Posada & Crandall, 1998), and a maximum likelihood tree was estimated using a PHYML plugin (Guindon & Gascuel, 2003) under the GTR + Γ model. The support for the inferred topology was estimated using 1,000 bootstrap replicates.

3 | RESULTS

3.1 | Assembly of the mitochondria and identification of numts

Our assemblies based on whole genome sequencing (WGS) show that the complete mitogenome of *C. parallelus* is ~15,620 bp and has an A + T content of ~75.5%. By combining the annotations from MITOS2 and ARWEN, we show that the *C. parallelus* mitogenome consists of 13 protein-coding genes, 22 tRNAs, two rRNAs and the control region (Figure 1). All the 13 protein-coding genes show complete coding sequences, and the tRNAs show their typical secondary structure (Figure S1).

Combining the four assemblies derived from WGS with the assembly derived from RNAseq resulted in an alignment of 15,632 bp (Figure S2). By comparing the assemblies of *C. p. erythropus* derived from both datasets, we observed one single nucleotide polymorphism (SNP) in the *ND4* gene without aminoacidic change, and a 6 bp indel close to the 3' end of the large rRNA gene. The assembly derived from RNAseq lacked most of the control region and the three first tRNAs (length of 14,716 bp), as expected since these regions are not transcribed into mRNA. By comparing the four WGS assemblies from the two subspecies, we found a total of 169 SNPs, none of them adding a stop codon or shift in the open reading frame of the protein-coding genes, but showing some SNPs. Of these, 19 SNPs are consistent with fixation between subspecies, 142 are consistent with polymorphisms within *C. p. erythropus*, 10 are consistent with

polymorphisms within *C. p. parallelus*, and two are consistent with shared polymorphisms between subspecies. However, a geographically broader sampling is required to confirm the fixation of these SNPs.

As expected for mitochondrial genome sequencing, all individuals show very high coverage: averaging 500× for three individuals and 1,000× for a deeply sequenced sample (Figure 2). Notably, the region between 1,100 and 4,600 bp shows a coverage 1.61 to 2.92 times higher than the rest of the mitogenome. The same pattern is seen in average nucleotide diversity, where this region is 2.50 to 4.46 times higher than the rest of the mitogenome (Table S2).

3.2 | Phylogeographic analyses

The trimmed alignment of the *CYTB* gene had 652 bp and 32 SNPs. We find that six of these SNPs were never ambiguous, 26 contained an ambiguity in at least one individual, and none was consistently ambiguous across more than half of the individuals. On average, there were some ambiguities called for 11% of all individuals. The trimmed alignment of the *COX3* gene had 504 bp and 55 SNPs. Of these, 12 SNPs were never ambiguous, 43 contained one ambiguity in at least one individual, and two were ambiguous in more than half of the total number of individuals. On average, there were some ambiguities called in 12% of all individuals. We did not observe a bias on ambiguities regarding certain SNPs or geographic location for either gene.

FIGURE 1 Annotation of the reference mitochondrial genome of *Chorthippus parallelus parallelus*
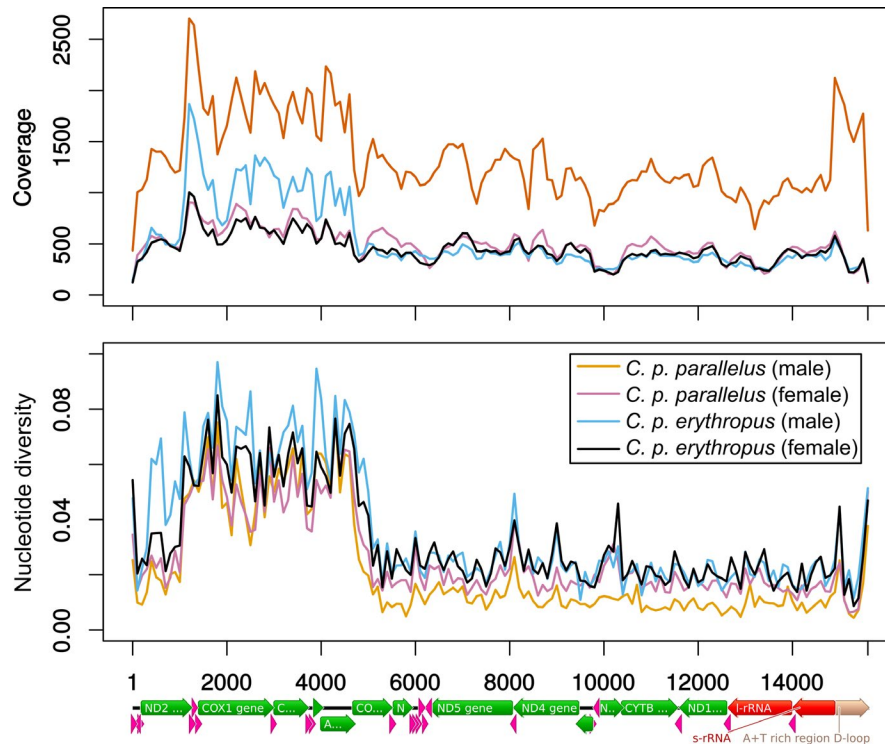
FIGURE 2 Nucleotide diversity and coverage along the four WGS references of the *C. p. parallelus* mitogenome
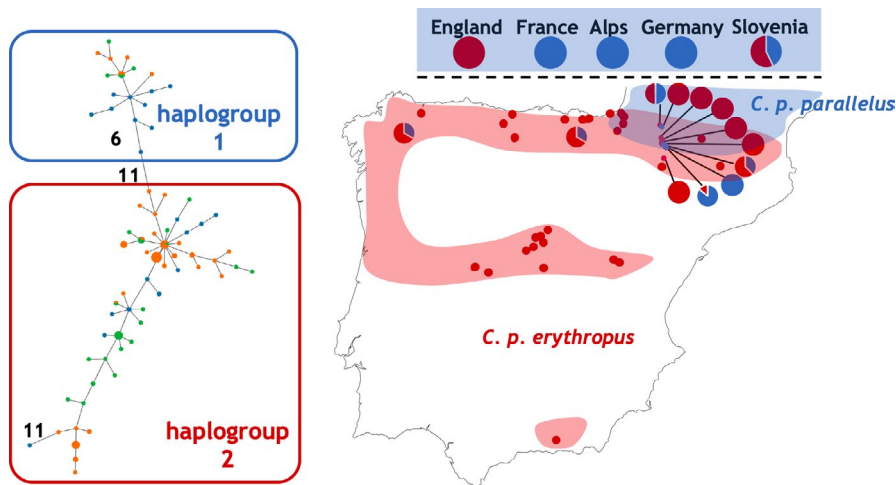

FIGURE 3 Distribution of two haplogroups of the concatenated *CYTB* and *COX3* mitochondrial genes in the two subspecies of *Chorthippus parallelus*. The two haplogroups were defined based on haplotypes that differed between three to ten mutations (see Figure S3 for details). Colors represent individuals collected in the range of pure *C. p. parallelus* (in blue), of pure *C. p erythropus* (in orange), and within the boundaries of the hybrid zone (in green). Branch lengths are not to scale, but the number of mutations are denoted at branches representing more than five mutational steps. Colored areas on the map show the known distribution of the subspecies. Dots show the sampling localities and are colored according to the haplogroup detected. Localities from the hybrid zone and from where both haplogroups co-occur are amplified for visualization purposes

The phylogenetic network showed two highly divergent haplogroups that are separated by 11 mutations (Figure 3) and that remained distinct considering nLV thresholds between three to 10 mutations (Figure S3). Both haplogroups contain samples from both subspecies at different frequencies. Haplogroup 1 is most common in *C. p. parallelus*, including the samples from Slovenia, Italy, Alps, and Germany. Haplogroup 2 is most common in *C. p. erythropus*, including the samples from South and Central Spain. Individuals from the hybrid zone have both haplogroups. Across the hybrid zone, the frequency of the haplogroups does not appear to vary clinally. The northmost

locality of the transect (Arudy, $n = 2$) has both mitochondrial haplotypes, followed by localities ($n = 1$–3) where we only found haplogroup 2, typical of *C. p. erythropus*, including in the locality at the crest of the Pyrenees (Portalet, $n = 14$). The following southern localities (Corral de Mulas, $n = 8$; Escarrilla, $n = 7$) have both haplogroups or haplogroup 1, typically found in *C. p. parallelus* (Sallent de Gállego, $n = 3$). Localities further south generally have haplogroup 2, typical of *C. p. erythropus*. Notably, some localities far removed from the hybrid zone (i.e. Slovenia, Lugo and Álava) contain both haplogroups, and the three individuals sampled in England contain only haplogroup 2.

The maximum likelihood tree containing ambiguities confirmed the results of the network with the mitochondrial haplotypes falling into two major clades (Figure S4). One clade (bootstrap = 61) contains mainly samples of the subspecies *C. p. parallelus*, including the samples from central Europe. The other clade (bootstrap = 70) contains mainly samples of *C. p. erythropus* from central and southern Spain. Hybrids are distributed among both clades, as well as the three other localities removed from the hybrid zone.

4 | DISCUSSION

4.1 | Standard barcoding genes have multiple copies to the nuclear genome

The *COX1* gene has played a pivotal role in global efforts to document and map biodiversity. Although such efforts have been successful and valuable in many taxa, it has been most challenging in those characterized by larger genomes, arguably due to their frequent involvement in pseudogenes or numts (Buhay, 2009). Here, we test if *COX1* and other genes are preferentially involved in numts in the giant genome of the grasshopper *Chorthippus parallelus*.

Combining genomic and transcriptomic reads, we assembled five high coverage (>200×) mitochondrial genomes with a total size (~15,620 bp) and A + T content (~75.5%), similar to what was reported in other orthopteroid mitogenomes (Erler et al., 2010; Liu & Huang, 2010; Yin et al., 2012). Our reference shows the same gene number, order, and orientation found in Acrididae (Song et al., 2015), without stop codons in the 13 protein-coding genes. The analysis of the mitogenome in RNA is recommended to reduce the noise generated by numts (Bertheau et al., 2011). We found scarce differences between the *C. p. erythropus* assemblies derived from WGS and RNAseq datasets (one SNP and a 6 bp deletion in a rRNA), suggesting that our assembled mitochondrial genomes are unlikely to be influenced by numts and thus provide a reliable reference for the mitogenome of *C. parallelus*.

By mapping WGS reads against our reference mitogenome, we observed a similar pattern of coverage and diversity throughout most of the mitogenome, with the exception of the region between 1,100 and 4,600 bp, where coverage was around two times higher and nucleotide diversity was around three times higher (Figure 2 and Table S2). Notably, this mitochondrial region encompasses the gene *COX1*, which is commonly used in barcoding projects (Hebert et al., 2003), along with the *COX2*, *ATP8*, and *ATP6* genes.

Because mitochondria are exclusively maternally inherited, usually there is only one mitochondrial haplotype per individual. However, mitochondrial polymorphisms within individuals can occur due to numts, or due to heteroplasmy. Heteroplasmy occurs because mitochondria compete for transmission during cell division, and sometimes more than one lineage can survive in a cell line or in an individual. Recognizing the presence of heteroplasmy is important because it can distort phylogenetic inference (Klucnika & Ma, 2019). Once thought to be rare, heteroplasmy is increasingly being described in many animals, including humans (Naeem & Sondheimer, 2019), crustaceans (Rodríguez-Pena et al., 2020), and ants (Meza-Lázaro et al., 2018). Although heteroplasmy is also known to occur in *C. parallelus* (Zhang et al., 1995), it cannot explain our results because (a) heteroplasmy involves the whole mitochondrion and thus, cannot result in localized variation of coverage, and (b) SNPs between alternate heteroplasmic variants are expected to be randomly located along the mitogenome and thus, would not increase diversity in only one mitochondrial region. This pattern is instead more consistent with the hypothesis that this region of the mitogenome has been duplicated multiple times into the nuclear genome (i.e. as numts). As a result, these numt sequences may be mapping to their original mitochondrial genes. This result coincides with the previous observation of multiple copies of the mitochondrial *COX1* gene in close association to repetitive centromeric sequences of the *C. parallelus* evidenced by in situ hybridization (Vaughan et al., 1999). This is also consistent with previous studies using Sanger sequencing of the genes *COX1*, *COX2* and *ATP8* that showed abnormally high rates of sequence variation in this species (Szymura et al., 1996).

Given an estimated genome size of 13.26 ± 0.98 Gb (Lechner et al., 2012) and our sequencing effort, we expect to achieve below 1× coverage for nuclear genes and above 300× coverage for mitochondrial genes (Table S2). Our observation of a 1.5- to 2-fold increase in coverage in a single mitochondrial region suggests that this region has been copied into the nuclear genome hundreds of times. However, estimating the exact number of duplications and their timing will require a high-quality reference nuclear genome, long read sequencing, and much higher sequencing effort than presented here.

4.2 | Other mitochondrial genes are phylogenetically informative

The gene *COX1* has been particularly instrumental not only for barcoding the animal tree of life (Hebert et al., 2003), but also for phylogeographic studies. Recent barcoding projects in grasshoppers (Hawlitschek et al., 2016), as well as earlier phylogeographic studies on *Chorthippus parallelus* (Lunt et al., 1996, 1998), have found few phylogenetically informative SNPs in the *COX1* gene. Here, we used a phylogeographic approach to understand if alternative mitochondrial genes that do not appear to be overrepresented in numts are phylogenetically informative for *C. parallelus*.

We designed new primers for the *CYTB* and *COX3* genes that, according to our hypothesis, are less likely to be duplicated as numts

(Figure 2). When comparing 145 individuals, we found a modest amount of variability: 5% for *CYTB* and 8.5% for *COX3*. These values are similar to those reported in *COX1* studies of *C. parallelus* (5.3% (Lunt et al., 1998)). Contrary to previous attempts to amplify and sequence *COX1* in *C. parallelus*, these two genes show sequences with single peaks in almost all positions and no stop codons. Using a stringent threshold for detecting ambiguities, we detect at least one ambiguity in 11%–12% of the individuals. Yet, these ambiguities tend to occur rarely across individuals and thus can be explained by sequencing error and by heteroplasmy. We tested if the reduced number of ambiguities detectable in these gene sequences results in a different phylogenetic signal from the most commonly sequenced nucleotide. By comparing a phylogenetic network that does not accommodate ambiguities (Figure 3) with a maximum likelihood tree that accommodates ambiguities (Figure S4), we show that the phylogenetic relationships between samples are identical. We thus conclude that, in these two mitochondrial genes, ambiguities potentially caused by sequencing error, heteroplasmy or by recently duplicated numts do not change the phylogenetic signal significantly. Our observation of an absence of stop codons and rarity of double peaks that do not affect phylogenetic inference, does not preclude these genes from being involved in numts (see Haran et al., 2015). However, our results are consistent with a lower duplication of these mitochondrial genes into the nuclear genome relative to the barcoding *COX1* gene, providing a more reliable PCR-based approach for mitochondrial studies in this species.

We further investigated how geographic patterns of this genetic variability reflect the known phylogeographic history of *C. parallelus*. In particular, we were interested in signals of population contraction in the southern European peninsulas during the Pleistocene glacial periods, which is expected to result in changes of haplotype frequency between subspecies, and expansion and the secondary contact in the Pyrenees during the last post-glacial period, which is expected to result in genetic admixture (Butlin, 1998; Hewitt, 1993). Our results show that the subspecies *C. p. parallelus* and *C. p. erythropus* do not form reciprocally monophyletic clades (Figure 3 and Figure S4). One haplogroup is found more frequently within the European subspecies *C. p. parallelus*, and however, it also occurs at low frequency in the Iberian *C. p. erythropus* in localities far removed from the hybrid zone and therefore not affected by introgression (Figure 3). Conversely, the other haplogroup is more frequently found within *C. p. erythropus* and also occurs in geographically distant population of *C. p. parallelus* in England and Slovenia. These results are consistent with a large amount of incomplete lineage sorting between subspecies, which can result from a relatively recent divergence time and from large effective population sizes. Incomplete lineage sorting has also been reported in previous phylogeographic studies of *C. parallelus* using the *COX1* gene (Lunt et al., 1998), suggesting that this evolutionary process is important in shaping the current genetic variability found in the mitochondria.

The localities sampled across the hybrid zone contain both haplogroups (Figure 3), consistent with a post-glacial expansion from the Balkans and central Iberia followed by secondary contact at the crest of the Pyrenees, as previously suggested by nuclear markers. Nevertheless, based on our current sampling, the transition between the two haplogroups does not appear to be clinal. It is important to note that the current sampling density varied largely across the hybrid zone (between one and 14 individuals per site) and thus sampling stochasticity might preclude finding an existing clinal transition in the mitochondria, as was found in previous studies with nuclear markers using a population level sampling (Butlin, 1998; Hewitt, 1993).

Several published studies of *C. parallelus* suggest that selection favors the asymmetric inheritance of several co-inherited cytoplasmatic factors in the same direction observed in our study. First, experimental crosses involving sequential mating of single females with males of both subspecies show that there is an excess of pure progeny, particularly when *C. p. parallelus* is the mother (Bella et al., 1992), perhaps favoring the introgression of *C. p. parallelus* mitochondrial haplotypes into the hybrid zone. Second, experimental crosses performed with hybrid individuals carrying different cytoplasmatic strains of *Wolbachia* show asymmetric *Wolbachia*-incompatibilities, which could favor crosses in the direction of *C. p. erythropus* and prevent those in the opposite direction (Martínez-Rodríguez & Bella, 2018; Zabal-Aguirre et al., 2010, 2014). Third, patterns of cytogenetic introgression across the hybrid zone show that X-linked markers characteristic of *C. p. parallelus* have moved five to 15 km south from the center of the hybrid zone into the genomic background of *C. p. erythropus* (Ferris et al., 1993; Serrano et al., 1996; Vázquez et al., 1994), suggesting that the X-chromosome or any other co-inherited cytoplasmatic factors are under asymmetric introgression. The same demographic or selective processes that could be favoring the asymmetric introgression of X-linked markers or of *Wolbachia* are also expected to lead to sweeps in the mitochondrial haplogroup associated with *C. p. parallelus*. This could explain the presence of the haplogroup typical of *C. p. parallelus* in the southern part of the transect. However, future studies using a population level sampling are needed to clarify this hypothesis.

4.3 | Concluding remarks

Taken together, our results show that, in a grasshopper species characterized by a gigantic genome, numts are, in fact, common. Numts frequently appear to originate predominantly from a contiguous mitochondrial region of 3,500 bp. This region contains the entirety of the *COX1* gene, elucidating why barcoding projects reliant on this marker have remained challenging in some species complexes of grasshoppers, including in closely related species that are challenging to distinguish based on morphology or bioacoustics (Hawlitschek et al., 2016). We show that alternative mitochondrial markers, such as *CYTB* and *COX3* genes, are less duplicated as numts, and can provide useful information about the phylogeographic history of closely related subspecies, without requiring strenuous laboratorial protocols to avoid the co-amplification of numts (Ibarguchi et al., 2006; Moulton et al., 2010; Song et al., 2008). More generally, our study shows how next-generation

sequencing methods can be used to identify mitochondrial genes that are useful for phylogeographic analyses in species that have remained largely unstudied until now.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject of their manuscript.

DATA AVAILABILITY STATEMENT

The raw data that support the findings of this study are openly available in NCBI at www.ncbi.nlm.nih.gov/, BioProject: PRJNA679335, BioSample: SAMN16264389, and in the GenBank accessions MW232097–MW232245, MW2322466–MW232411. The processed data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Ricardo J. Pereira https://orcid.org/0000-0002-8076-4822
Francisco J. Ruiz-Ruano https://orcid.org/0000-0002-5391-301X

REFERENCES

Avise, J. C. (2000). *Phylogeography. The history and formation of species.* Harvard University Press.

Bella, J. L., Butlin, R. K., Ferris, C., & Hewitt, G. M. (1992). Asymmetrical homogamy and unequal sex-ratio from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies. *Heredity*, *68*, 345–352. https://doi.org/10.1038/hdy.1992.49

Bella, J. L., Serrano, L., Orellana, J., & Mason, P. L. (2007). The origin of the *Chorthippus parallelus* hybrid zone: Chromosomal evidence of multiple refugia for Iberian populations. *Journal of Evolutionary Biology*, *20*(2), 568–576. https://doi.org/10.1111/j.1420-9101.2006.01254.x

Bensasson, D., Feldman, M. W., & Petrov, D. A. (2003). Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *Journal of Molecular Evolution*, *57*(3), 343–354. https://doi.org/10.1007/s00239-003-2485-7

Bensasson, D., Zhang, D. X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends in Ecology & Evolution*, *16*(6), 314–321. https://doi.org/10.1016/s0169-5347(01)02151-6

Bensasson, D., Zhang, D. X., & Hewitt, G. M. (2000). Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, *17*(3), 406–415. https://doi.org/10.1093/oxfordjournals.molbev.a026320

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bertheau, C., Schuler, H., Krumböck, S., Arthofer, W., & Stauffer, C. (2011). Hit or miss in phylogeographic analyses: The case of the cryptic NUMTs. *Molecular Ecology Resources*, *11*(6), 1056–1059. https://doi.org/10.1111/j.1755-0998.2011.03050.x

Buhay, J. E. (2009). "COI-like" sequences are becoming problematic in molecular systematic and DNA barcoding studies. *Journal of Crustacean Biology*, *29*(1), 96–110. https://doi.org/10.1651/08-3020.1

Buño, I., Torroja, E., López-Fernández, C., Butlin, R. K., Hewitt, G. M., & Gosálvez, J. (1994). A hybrid zone between two subspecies of the grasshopper *Chorthippus parallelus* along the Pyrenees: The west end. *Heredity*, *73*, 625–634. https://doi.org/10.1038/hdy.1994.170

Butlin, R. K. (1998). What do hybrid zones in general, and the *Chorthippus parallelus* zone in particular, tell us about speciation. *Endless Forms: Species and Speciation*. Oxford University Press.

Calvignac, S., Konecny, L., Malard, F., & Douady, C. J. (2011). Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*, *11*(2), 246–254. https://doi.org/10.1016/j.mito.2010.10.004

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18. https://doi.org/10.1093/nar/gkw955

Erler, S., Ferenz, H.-J., Moritz, R. F. A., & Kaatz, H.-H. (2010). Analysis of the mitochondrial genome of *Schistocerca gregaria gregaria* (Orthoptera: Acrididae). *Biological Journal of the Linnean Society*, *99*(2), 296–305. https://doi.org/10.1111/j.1095-8312.2009.01365.x

Ferris, C., Rubio, J. M., Serrano, L., & Gosálvez, J. (1993). One way introgression of a subspecific sex chromosome marker in a hybrid zone. *Heredity*, *71*, 119–129. https://doi.org/10.1038/hdy.1993.115

Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid evolution. *Annual Review of Plant Biology*, *59*(1), 491–517. https://doi.org/10.1146/annurev.arplant.59.032607.092915

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, *47*(W1), W59–W64. https://doi.org/10.1093/nar/gkz238

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Haran, J., Koutroumpa, F., Magnoux, E., Roques, A., & Roux, G. (2015). Ghost mt DNA haplotypes generated by fortuitous NUMT s can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research*, *53*(2), 109–115. https://doi.org/10.1111/jzs.12095

Hawlitschek, O., Morinière, J., Lehmann, G. U. C., Lehmann, A. W., Kropf, M., Dunz, A., Glaw, F., Detcharoen, M., Schmidt, S., Hausmann, A., Szucsich, N. U., Caetano-Wyler, S. A., & Haszprunar, G. (2016). DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with focus on Austria, Germany and Switzerland. *Molecular Ecology Resources*, *39*(213), 7–17. https://doi.org/10.1111/1755-0998.12638

Hazkani-Covo, E., & Graur, D. (2006). A comparative analysis of numt evolution in Human and Chimpanzee. *Molecular Biology and Evolution*, *24*(1), 13–18. https://doi.org/10.1093/molbev/msl149

Hazkani-Covo, E., Zeller, R. M., & Martin, W. (2010). Molecular poltergeists: Mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genetics*, 6(2), e1000834–e1000911. https://doi.org/10.1371/journal.pgen.1000834

Hebert, P. D. N., Dewaard, J. R., & Landry, J. F. (2009). DNA barcodes for 1/1000 of the animal kingdom. *Biology Letters*, 6, 359–362. https://doi.org/10.1098/rsbl.2009.0848

Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings Biological Sciences / the Royal Society*, 270(suppl_1), 1–4. https://doi.org/10.1098/rsbl.2003.0025

Hewitt, G. M. (1993). After the ice: *parallelus* meets *erythropus* in the Pyrenees. *Hybrid Zones and the Evolutionary Process*. Oxford University Press.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 450, 907–913. https://doi.org/10.1038/35016000

Hewitt, G. M., Butlin, R. K., & East, T. M. (1987). Testicular dysfunction in hybrids between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Biological Journal of the Linnean Society*, 31(1), 25–34. https://doi.org/10.1111/j.1095-8312.1987.tb01978.x

Huang, C. Y., Grünheit, N., Ahmadinejad, N., Timmis, J. N., & Martin, W. (2005). Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to Angiosperm nuclear chromosomes. *Plant Physiology*, 138(3), 1723–1733. https://doi.org/10.1104/pp.105.060327

Ibarguchi, G., Friesen, V. L., & Lougheed, S. C. (2006). Defeating numts: Semi-pure mitochondrial DNA from eggs and simple purification methods for field-collected wildlife tissues. *Genome*, 49(11), 1438–1450. https://doi.org/10.1139/g06-107

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kaya, S., & Çıplak, B. (2018). Possibility of numt co-amplification from gigantic genome of Orthoptera: Testing efficiency of standard PCR protocol in producing orthologous COI sequences. *Heliyon*, 4(11), e00929–e1026. https://doi.org/10.1016/j.heliyon.2018.e00929

Kleine, T., Maier, U. G., & Leister, D. (2009). DNA transfer from organelles to the nucleus: The idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology*, 60(1), 115–138. https://doi.org/10.1146/annurev.arplant.043008.092119

Klucnika, A., & Ma, H. (2019). A battle for transmission: The cooperative and selfish animal mitochondrial genomes. *Open Biology*, 9(3), 180267. https://doi.org/10.1098/rsob.180267

Korkmaz, E. M., Lunt, D. H., Çıplak, B., Değerli, N., & Başıbüyük, H. H. (2014). The contribution of Anatolia to European phylogeography: The centre of origin of the meadow grasshopper, *Chorthippus parallelus*. *Journal of Biogeography*, 41(9), 1793–1805. https://doi.org/10.1111/jbi.12332

Laslett, D., & Canback, B. (2008). ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, 24(2), 172–175. https://doi.org/10.1093/bioinformatics/btm573

Lechner, M., Marz, M., Ihling, C., Sinz, A., Stadler, P. F., & Krauss, V. (2012). The correlation of genome size and DNA methylation rate in metazoans. *Theory in Biosciences*, 132(1), 47–60. https://doi.org/10.1007/s12064-012-0167-y

Liu, N., & Huang, Y. (2010). Complete mitochondrial genome sequence of *Acrida cinerea* (Acrididae: Orthoptera) and comparative analysis of mitochondrial genomes in Orthoptera. *Comparative and Functional Genomics*, 2010(6), 1–16. https://doi.org/10.1155/2010/319486

Liu, Y., & Huang, Y. (2008). Sequencing and analysis of complete mitochondrial genome of *Chorthippus chinensis* Tarb. *Chinese Journal of Biochemistry and Molecular Biology*, 24, 329–335.

Lopez, J. V., 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. *Journal of Molecular Evolution*, 39(2), 174–190. https://doi.org/10.1007/bf00163806

Lunt, D. H., Ibrahim, K. M., & Hewitt, G. M. (1998). mtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity*, 80, 633–641. https://doi.org/10.1046/j.1365-2540.1998.00311.x

Lunt, D. H., Zhang, D. X., Szymura, J. M., & Hewitt, G. M. (1996). The insect cytochrome oxidase I gene: Evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*, 5(3), 153–165. https://doi.org/10.1111/j.1365-2583.1996.tb00049.x

Martínez-Rodríguez, P., & Bella, J. L. (2018). *Chorthippus parallelus* and *Wolbachia*: Overlapping orthopteroid and bacterial hybrid zones. *Frontiers in Genetics*, 9, 188–214. https://doi.org/10.3389/fgene.2018.00604

Martínez-Rodríguez, P., Hernández-Pérez, M., & Bella, J. L. (2013). Detection of *Spiroplasma* and *Wolbachia* in the bacterial gonad community of *Chorthippus parallelus*. *Microbial Ecology*, 66, 211–223. https://doi.org/10.1007/s00248-013-0226-z

Meza-Lázaro, R. N., Poteaux, C., Bayona-Vásquez, N. J., Branstetter, M. G., & Zaldívar-Riverón, A. (2018). Extensive mitochondrial heteroplasmy in the neotropical ants of the Ectatomma ruidum complex (Formicidae: Ectatomminae). *Mitochondrial DNA A DNA Mapp Seq Anal.*, 29(8), 1203–1214. https://doi.org/10.1080/24701394.2018.1431228

Moulton, M. J., Song, H., & Whiting, M. F. (2010). Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: A case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources*, 10(4), 615–627. https://doi.org/10.1111/j.1755-0998.2009.02823.x

Naeem, M. M., & Sondheimer, N. (2019). Heteroplasmy shifting as therapy for mitochondrial disorders. *Advances in Experimental Medicine and Biology*, 1158, 257–267. https://doi.org/10.1007/978-981-13-8367-0_14

Ning, Z. M., Cox, A. J., & Mullikin, J. C. (2001). SSAHA: A fast search method for large DNA databases. *Genome Research*, 11(10), 1725–1729. https://doi.org/10.1101/gr.194201

Nolen Z. J., Yildirim B., Irisarri I., Liu S., Groot Crego C., Amby D. B., Mayer F., Gilbert M. T. P., Pereira R. J. (2020). Historical isolation facilitates species radiation by sexual selection: Insights from *Chorthippus* grasshoppers. *Molecular Ecology*, 29(24), 4985–5002. https://doi.org/10.1111/mec.15695

Perna, N. T., & Kocher, T. D. (1996). Mitochondrial DNA: Molecular fossils in the nucleus. *Current Biology*, 6(2), 128–129. https://doi.org/10.1016/s0960-9822(02)00441-4

Posada, D., & Crandall, K. A. (1998). Modeltest: Testing the model of DNA substitution. *Bioinformatics*, 14(9), 817–818. https://doi.org/10.1093/bioinformatics/14.9.817

Rand, D. M., Haney, R. A., & Fry, A. J. (2004). Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*, 19(12), 645–653. https://doi.org/10.1016/j.tree.2004.10.003

Ricchetti, M., Tekaia, F., & Dujon, B. (2004). Continued colonization of the human genome by mitochondrial DNA. *PLoS Biology*, 2(9), e273–e312. https://doi.org/10.1371/journal.pbio.0020273

Rodríguez-Pena, E., Verísimo, P., Fernandez, L., Gonzalez-Tizon, A., Barcena, C., & Martínez-Lage, A. (2020). High incidence of heteroplasmy in the mtDNA of a natural population of the spider crab Maja brachydactyla. *PLoS One*, 15(3), e0230243. https://doi.org/10.1371/journal.pone.0230243

Serrano, L., de la Vega, C. G., Bella, J. L., López-Fernández, C., Hewitt, G. M., & Gosálvez, J. (1996). A hybrid zone between two subspecies of *Chorthippus parallelus*. X-chromosome variation through a contact zone. *Journal of Evolutionary Biology*, 9(2), 173–184. https://doi.org/10.1046/j.1420-9101.1996.9020173.x

Song, H., Amédégnato, C., Cigliano, M. M., Desutter-Grandcolas, L., Heads, S. W., Huang, Y., Otte, D., & Whiting, M. F. (2015). 300 million years of diversification: Elucidating the patterns of orthopteran evolution based on comprehensive taxon and gene sampling. *Cladistics*, *31*(6), 621–651. https://doi.org/10.1111/cla.12116

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. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13486–13491. https://doi.org/10.2307/25464070

Song, H., Moulton, M. J., & Whiting, M. F. (2014). Rampant nuclear insertion of mtDNA across diverse lineages within Orthoptera (Insecta). *PLoS One*, *9*(10), e110508–e110514. https://doi.org/10.1371/journal.pone.0110508

Szymura, J. M., Lunt, D. H., & Hewitt, G. M. (1996). The sequence and structure of the meadow grasshopper (*Chorthippus parallelus*) mitochondrial srRNA, ND2, COI, COII ATPase8 and 9 tRNA genes. *Insect Molecular Biology*, *5*(2), 127–139. https://doi.org/10.1111/j.1365-2583.1996.tb00047.x

Tsutsui, N. D., Suarez, A. V., Spagna, J. C., & Johnston, J. S. (2008). The evolution of genome size in ants. *BMC Evolutionary Biology*, *8*(1), 64–69. https://doi.org/10.1186/1471-2148-8-64

Vaughan, H. E., Heslop-Harrison, J. S., & Hewitt, G. M. (1999). The localization of mitochondrial sequences to chromosomal DNA in orthopterans. *Genome*, *42*(5), 874–880. https://doi.org/10.1139/g99-020

Vázquez, P., Cooper, S. J. B., Gosálvez, J., & Hewitt, G. M. (1994). Nuclear DNA introgression across a Pyrenean hybrid zone between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Heredity*, *73*(4), 436–443. https://doi.org/10.1038/hdy.1994.191

Vedenina, V., & Mugue, N. (2011). Speciation in Gomphocerine grasshoppers: Molecular phylogeny versus bioacoustics and courtship behavior. *Journal of Orthoptera Research*, *20*, 109–125. https://doi.org/10.1665/034.020.0111

Wang, X., Fang, X., Yang, P., Jiang, X., Jiang, F., Zhao, D., Li, B., Cui, F., Wei, J., Ma, C., Wang, Y., He, J., Luo, Y., Wang, Z., Guo, X., Guo, W., Wang, X., Zhang, Y., Yang, M., … Kang, L. (2014). The locust genome provides insight into swarm formation and long-distance flight. *Nature Communications*, *5*, 2957. https://doi.org/10.1038/ncomms3957

Westerman, M., Barton, N. H., & Hewitt, G. M. (1987). Differences in DNA content between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity*, *58*, 221–228. https://doi.org/10.1038/hdy.1987.36

Yin, H., Zhi, Y., Jiang, H., Wang, P., Yin, X., & Zhang, D. (2012). The complete mitochondrial genome of *Gomphocerus tibetanus* Uvarov, 1935 (Orthoptera: Acrididae: Gomphocerinae). *Gene*, *494*(2), 214–218. https://doi.org/10.1016/j.gene.2011.12.020

Zabal-Aguirre, M., Arroyo, F., & Bella, J. L. (2010). Distribution of *Wolbachia* infection in *Chorthippus parallelus* populations within and beyond a Pyrenean hybrid zone. *Heredity*, *104*(2), 174–184. https://doi.org/10.1038/hdy.2009.106

Zabal-Aguirre, M., Arroyo, F., García-Hurtado, J., de la Torre, J., Hewitt, G. M., & Bella, J. L. (2014). *Wolbachia* effects in natural populations of *Chorthippus parallelus* from the Pyrenean hybrid zone. *Journal of Evolutionary Biology*, *27*(6), 1136–1148. https://doi.org/10.1111/jeb.12389

Zhang, D. X., & Hewitt, G. M. (1996). Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, *11*(6), 247–251. https://doi.org/10.1016/0169-5347(96)10031-8

Zhang, D.-X., Szymura, J. M., & Hewitt, G. M. (1995). Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, *40*(4), 382–391. https://doi.org/10.1007/BF00164024

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Table S1. List of specimens used in the phylogeographic study.

Table S2. Maximum likelihood tree of *Chorthippus parallelus* for the concatenated alignment of the mitochondrial genes.

Figure S1. Secondary structure predicted for tRNAs of the *C. p. parallelus* mitogenome.

Figure S2. Graphical alignment of the five reference mitogenomes of *C. parallelus*.

Figure S3. Two mitochondrial haplogroups identified considering mutation thresholds (nLV) between three (top) and ten (bottom) mutations.

Figure S4. Maximum likelihood tree of *Chorthippus parallelus* for the concatenated alignments of the mitochondrial genes *CYTB* and *COX3*.

Alignment S1. Alignment for the assembled mitogenomes.

Alignment S2. Alignment for the *CYTB* and *COX3* genes, including ambiguity codes.

Alignment S3. Alignment for the *CYTB* and *COX3* genes, without ambiguity codes.