Frequent Assimilation of Mitochondrial DNA by Grasshopper Nuclear Genomes

Douda Bensasson, De-Xing Zhang, and Godfrey M. Hewitt
School of Biological Sciences, University of East Anglia, Norwich, England

Multiple copies of mitochondrial-like DNA were found in the brown mountain grasshopper, *Podisma pedestris* (Orthoptera: Acrididae), paralogous to COI and ND5 regions. The same was discovered using the ND5 regions of nine other grasshopper species from four separate subfamilies (Podisminae, Calliptaminae, Cyrtacanthacridinae, and Gomphocerinae). The extra ND5-like sequences were shown to be nuclear in the desert locust, *Schistocerca gregaria* (Cyrtacanthacridinae), and probably so in *P. pedestris* and an *Italopodisma* sp. (Podisminae). Eighty-seven different ND5-like nuclear mitochondrial pseudogenes (Numts) were sequenced from 12 grasshopper individuals. Different nuclear mitochondrial pseudogenes, if descended from the same mitochondrial immigrant, will have diverged from each other under no selective constraints because of their loss of functionality. Evidence of selective constraints in the differences between any two Numt sequences (e.g., if most differences are at third positions of codons) implies that they have separate mitochondrial origins. Through pairwise comparisons of pseudogene sequences, it was established that there have been at least 12 separate mtDNA integrations into *P. pedestris* nuclear genomes. This is the highest reported rate of horizontal transfer between organellar and nuclear genomes within a single animal species. The occurrence of numerous mitochondrial pseudogenes in nuclear genomes derived from separate integration events appears to be a common phenomenon among grasshoppers. More than one type of mechanism appears to have been involved in generating the observed grasshopper Numts.

Introduction

There are nuclear copies of mitochondrial sequences in a variety of organisms, including insects and other invertebrates, vertebrates, fungi, and plants (reviewed in Zhang and Hewitt 1996a). The occurrence of these sequences (hereafter referred to as Numts, following the abbreviation of Lopez et al. 1994) complicate the employment of mitochondrial DNA as a molecular marker in evolutionary studies. This is particularly so when Numts are present in high copy numbers, because in small organisms such as insects, they make the effective separation of mtDNA from its nuclear paralogs difficult, if not impossible. However, the study of Numts is also of considerable evolutionary interest, because they provide a window on the dynamics of genome evolution and intergenomic interactions.

Hundreds of mitochondrial pseudogenes appear to reside in the human nuclear genome (Fukuda et al. 1985). A similar estimate has been made for the migratory locust, *Locusta migratoria* (Gellissen and Michaelis 1987). In the course of studying postglacial colonization events for the brown mountain grasshopper, *Podisma pedestris*, using mtDNA as a genetic marker, high copy numbers of mitochondrial pseudogenes were encountered. This finding complements earlier observations of Numts in the desert locust *Schistocerca gregaria* (Zhang and Hewitt 1996b). As knowledge about Numts in insects is generally poor (Gellissen et al. 1983; Sunnucks and Hales 1996; Zhang and Hewitt 1996a), we have carried out genomic in situ hybridization (Vaughan, Heslop-Harrison, and Hewitt 1999) and PCR cloning studies of various acridid grasshoppers to investigate the genomic abundance, organization, taxonomic distribution, and evolutionary dynamics of these nuclear pseudogenes.

Presented here are the results of a comparative study of 87 different ND5-like Numt sequences taken from *P. pedestris*, its Italian relative, *Italopodisma* sp., a more distant Japanese grasshopper, *Parapodisma mikiade*, and other acridid species of different subfamilies. We address the following questions: How abundant are these mitochondrial-like sequences in the nuclear genome? How different are they from one another? Could most of them be identical in sequence? How were they produced: by a single major event or through separate integrations? Other aspects revealed by these data, such as rate and patterns of Numt sequence evolution, differences among grasshopper populations in which Numts are held, and ancient population processes, will be discussed in detail elsewhere.

Materials and Methods

Sample Storage and DNA Extraction

Grasshoppers from 10 acridid species were collected mainly from western Europe; collection sites are summarized in tables 1–3. The *P. pedestris* individuals were collected from across their Arctic-Alpine Eurasian distribution. Classification within the genus *Italopodisma* is known to be difficult (Harz 1975) and is based solely on subtle differences in the shape of the male genitalia. For this reason, the *Italopodisma* used in this study are referred to as *Italopodisma* sp. Grasshoppers were stored at −80°C or in absolute ethanol at 4°C. Total genomic DNA was extracted from legs of single grasshoppers by freezing the tissue in liquid nitrogen, grinding it, then incubating it in 0.02 M Tris HCl (pH 8), 0.01 M EDTA, 0.5% SDS, and 50 µg/ml of
Proteinase K overnight at 50°C. DNA was purified from the supernatant using the Wizard DNA Clean-Up System from Promega.

Enrichment of mtDNA

The grasshoppers used in this study are relatively small, so not enough tissue was available from single individuals for a complete purification of mtDNA. DNA was enriched for mtDNA relative to total DNA using the protocol described in Zhang and Hewitt (1996b), which was adapted from Lansman et al. (1981). This protocol was modified for Podisma and Italopodisma in that DNA was not purified by CsCl/ethidium bromide gradient ultracentrifugation, as DNA yields were prohibitively low.

To check whether the protocol had yielded DNA which was mostly mitochondrial, the ND5 region was amplified from the enriched mtDNA and from total DNA of the same individual and digested with AluI, Sau3AI, and DraI. Digestion with these enzymes of the PCR products amplified from total genomic DNA had previously revealed more than one type of PCR product. If after completion of mitochondrial enrichment, only one type of PCR product is observed after digestion with these enzymes, this suggests that the enrichment has been successful and the extra types of PCR product digest were nuclear. Restriction analysis showed the successful enrichment for mtDNA (fig. 1). The PCR product was also cloned, and several clones were sequenced. Only one type of sequence was common among the clones after enrichment (table 4).

Nuclear and mitochondrial DNA were purified for Schistocerca gregaria (desert locust) as in Zhang and Hewitt (1996b).

PCR Amplification

ND5 PCR (688 bp)

The primer N5-J-6579, designed in R. Harrison’s laboratory for use on the flea and described by Simon et al. (1994), was extended and adapted for use in Orthoptera in our laboratory. Its sequence is now based on L. migratoria, but the position implied in its name refers to the 3’ base position in Drosophila yakuba as in the convention of Simon et al. (1994): N5-J-6578 (Imelda) (5’-ACTCACTCTCAACCAGATCAA-3’). A second primer was designed for the reverse direction: N5-N-7225 (Ferdinand) (5’-ACTCATGCTTTATTTAAGGCTTTA-3’). These primers always amplified well together and matched the template DNA of S. gregaria, Chortippus parallelus (the meadow grasshopper), P. pedestris, and Italopodisma sp., except for one mismatch for Ferdinand in Podisma and Italopodisma (5’-…CAITTA-3’).

Conditions when using BIOTAQ polymerase (Bioline) were as follows: 1 U of enzyme (in 50-µl reactions), 0.3 µM of each primer, 200 µM of each dNTP, and 1 × KCl buffer (Bioline). Cycling parameters were 94°C for 4 min; 35 cycles of 94°C for 40 s, 52°C for 1 min, and 72°C for 1 min 20 s; and 72°C for 7 min.

Conditions when using a higher fidelity polymerase, Pfu DNA polymerase (Stratagene), were as follows: 3.5 U of enzyme (in 50-µl reactions), 0.6 µM of each primer, 144 µM of each dNTP, and 1 × Pfu reaction buffer (Stratagene). Cycling parameters were the same as for the Taq PCR except with the annealing temperature at 50°C and 25 cycles.

COI PCR (324 bp)

The primers C1-J-1763 (UEA3) and C1-N-2087 (UEA4) (Lunt et al. 1996; Zhang and Hewitt 1997) were

Table 1

Summary of Evidence for Multiple COI-like Types Within Single Podisma pedestris Individuals, Revealed Through Restriction Enzyme Digestion

<table>
<thead>
<tr>
<th>Country</th>
<th>Individual</th>
<th>Digests Which Showed</th>
<th>&gt;1 COI Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>FIN4</td>
<td>MboII, FokI</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>MAD5</td>
<td>XbaI, HphI</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>CEN1</td>
<td>MboII</td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Summary of the Number of Different ND5-like Types (Digestotypes) Observed in Each Individual After Restriction Enzyme Digestion

<table>
<thead>
<tr>
<th>SUBFAMILY</th>
<th>SPECIES</th>
<th>LOCATION</th>
<th>ALU1 DIGESTS</th>
<th>SAU3AI DIGESTS</th>
<th>DRAI DIGESTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NO. OF POPULATIONS</td>
<td>NO. OF INDIVIDUALS TESTED</td>
<td>NO. OF DIGESTS IN EACH INDIVIDUAL</td>
<td>NO. OF INDIVIDUALS TESTED</td>
</tr>
<tr>
<td>Podisminae</td>
<td>Podisma pedestris</td>
<td>Many</td>
<td>22</td>
<td>89</td>
<td>2–3</td>
</tr>
<tr>
<td></td>
<td>Italopodisma sp.</td>
<td>Apennines</td>
<td>3</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Cophopodisma pyrenea</td>
<td>Pyrenees</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Parapodisma mikado</td>
<td>Japan</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Ognvia longipennis</td>
<td>Japan</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Primnoa hayachinensis</td>
<td>Japan</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Calliptaminae</td>
<td>Calliptamus sp.</td>
<td>Pyrenees</td>
<td>1</td>
<td>2</td>
<td>3–4</td>
</tr>
<tr>
<td>Gomphocerinae</td>
<td>Chorthippus parallelus</td>
<td>U.K.</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Arcrypta sp.</td>
<td>Pyrenees</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 3
Summary of ND5-like Sequences from Individuals Whose ND5-like PCR Products (Amplified from Total Genomic DNA) Were Cloned and Sequenced

<table>
<thead>
<tr>
<th>Individual</th>
<th>Label</th>
<th>No. of Clones Sequenced</th>
<th>No. of mtDNA Clones</th>
<th>No. of Different Numts</th>
<th>No. of Different Numts with Frameshifts</th>
<th>% Divergence of Numts from mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>FrA</td>
<td>15</td>
<td>4</td>
<td>9</td>
<td>4</td>
<td>0.2–7.8</td>
</tr>
<tr>
<td></td>
<td>FrB</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0.9–5.8</td>
</tr>
<tr>
<td></td>
<td>FrC</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3.9</td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrenees</td>
<td>SpC</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>0.2–12.5</td>
</tr>
<tr>
<td>South Picos</td>
<td>SpB</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>0–4.7</td>
</tr>
<tr>
<td>North Picos</td>
<td>SpA</td>
<td>13</td>
<td>4</td>
<td>9</td>
<td></td>
<td>0.5–9</td>
</tr>
<tr>
<td>Podisma pedestris total</td>
<td></td>
<td>65</td>
<td>28</td>
<td>36</td>
<td>15/27 (56%)b</td>
<td>0–125</td>
</tr>
<tr>
<td>Italopodisma total</td>
<td></td>
<td>34</td>
<td>4</td>
<td>24</td>
<td>19</td>
<td>0–2–7</td>
</tr>
<tr>
<td>Italopodisma total</td>
<td></td>
<td>49</td>
<td>7</td>
<td>35</td>
<td>29/35 (83%)</td>
<td>0.2–7</td>
</tr>
<tr>
<td>Podisma mikadoa</td>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
<td>4.8–8.7</td>
</tr>
<tr>
<td>Schistocerca gregariac</td>
<td></td>
<td>23</td>
<td>9</td>
<td>13</td>
<td>7</td>
<td>0.5–1.4</td>
</tr>
<tr>
<td>Chorthippus parallelusc</td>
<td></td>
<td>6</td>
<td>—</td>
<td>4</td>
<td>3</td>
<td>7a</td>
</tr>
<tr>
<td>Grand total</td>
<td></td>
<td>148</td>
<td>47</td>
<td>87</td>
<td>54/79 (68%)b</td>
<td>0–12.5</td>
</tr>
</tbody>
</table>

*Includes data from ND5-like sequences in which frameshifts could not always be resolved with certainty.
*b Percentages were calculated after excluding data from individuals whose frameshift data were uncertain.
*c From purified nuclear DNA.
*d mtDNA sequence was not confirmed.

used. Conditions were as follows: 2 U of Taq DNA polymerase (Promega) in 50-μl reactions, 0.3 μM of each primer, 200 μM of each dNTP, 1 × reaction buffer (Promega), and 1 mM MgCl2. Cycling parameters were same as for the ND5 PCR with Taq except with the annealing temperature at 43°C.

**Universal PCR**

The modified M13 universal primer (5′-CGAGCGTGAAGCTTGTAAAACGAGGCCAG-3′) and the M13 reverse primer (extended to 5′-GACGATTGACCAGCTGATGACC-3′) were used to amplify the cloned ND5 insert from colonies dissolved in DNA-free water. Conditions and cycling parameters used were same as for the ND5 Taq PCR except with the annealing temperature at 55°C.

**Cloning and Sequencing of the ND5 PCR Products**

Taq PCR products were cloned using the pMOS Blue T-Vector cloning kit from Amersham according to the manufacturer’s instructions. *Pfu* PCR products were ligated into *Sma*I (Gibco BRL) cut pUC18 using T4 DNA ligase (Gibco BRL), and Epicurian Coli XL1-Blue competent cells (Stratagene) were transformed accord-

---

**Table 3**

<table>
<thead>
<tr>
<th>Individual</th>
<th>Label</th>
<th>No. of Clones Sequenced</th>
<th>No. of mtDNA Clones</th>
<th>No. of Different Numts</th>
<th>No. of Different Numts with Frameshifts</th>
<th>% Divergence of Numts from mtDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>French</td>
<td>FrA</td>
<td>15</td>
<td>4</td>
<td>9</td>
<td>4</td>
<td>0.2–7.8</td>
</tr>
<tr>
<td></td>
<td>FrB</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0.9–5.8</td>
</tr>
<tr>
<td></td>
<td>FrC</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3.9</td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrenees</td>
<td>SpC</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>0.2–12.5</td>
</tr>
<tr>
<td>South Picos</td>
<td>SpB</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>0–4.7</td>
</tr>
<tr>
<td>North Picos</td>
<td>SpA</td>
<td>13</td>
<td>4</td>
<td>9</td>
<td></td>
<td>0.5–9</td>
</tr>
<tr>
<td>Podisma pedestris total</td>
<td></td>
<td>65</td>
<td>28</td>
<td>36</td>
<td>15/27 (56%)b</td>
<td>0–125</td>
</tr>
<tr>
<td>Italopodisma total</td>
<td></td>
<td>34</td>
<td>4</td>
<td>24</td>
<td>19</td>
<td>0–2–7</td>
</tr>
<tr>
<td>Italopodisma total</td>
<td></td>
<td>49</td>
<td>7</td>
<td>35</td>
<td>29/35 (83%)</td>
<td>0.2–7</td>
</tr>
<tr>
<td>Podisma mikadoa</td>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
<td>4.8–8.7</td>
</tr>
<tr>
<td>Schistocerca gregaria</td>
<td></td>
<td>23</td>
<td>9</td>
<td>13</td>
<td>7</td>
<td>0.5–1.4</td>
</tr>
<tr>
<td>Chorthippus parallelus</td>
<td></td>
<td>6</td>
<td>—</td>
<td>4</td>
<td>3</td>
<td>7a</td>
</tr>
<tr>
<td>Grand total</td>
<td></td>
<td>148</td>
<td>47</td>
<td>87</td>
<td>54/79 (68%)b</td>
<td>0–12.5</td>
</tr>
</tbody>
</table>

*a Includes data from ND5-like sequences in which frameshifts could not always be resolved with certainty.
*b Percentages were calculated after excluding data from individuals whose frameshift data were uncertain.
*c From purified nuclear DNA.
*d mtDNA sequence was not confirmed.

**Fig. 1.**—Digestypes before and after enrichment. ND5-like PCR products (all approximately 688 bp) were amplified from total genomic DNA (g) and from DNA enriched for mitochondria (m) in three different grasshoppers (ItB, FrA, and ItA). Each ND5-like PCR product was digested with *Alu*I, and the digest was run out on the gel shown here. Only one digestype was found for PCR products amplified from the enriched DNA, whereas at least two types are clearly visible for PCR products from unenriched DNA. The same results were observed after digestion with *Dra*I and *Sau*3AI, suggesting that the enrichment was successful.
ing to the manufacturer’s instructions with this recombinant DNA before plating on standard color selection plates. Colonies were picked into 100 μl DNA-free water, vortexed, heated to 95°C for 2 min, vortexed, then spun for 5 min at about 10,000 × g; 3 μl was used for the subsequent PCR.

The universal PCR products (amplified from single colonies) were purified using the Wizard PCR Purification Kit from Promega and then sequenced using the M13 primers described above. The Thermo Sequenase dye terminator cycle sequencing pre-mix kit (Amersham Pharmacia) was used for sequencing, and samples were run on an Applied Biosystems (ABI) 377 automated DNA sequencer.

For the two individuals whose Taq-amplified PCR products were sequenced (see table 3—SpA and Parapodisma), plasmid DNA was prepared using the QIAGEN Plasmid Midi Kit, and the purified double-stranded plasmid templates were sequenced using the AutoRead sequencing kit (Pharmacia) and run on an Automated Laser Fluorescent A.L.F. DNA sequencer (Pharmacia).

Restriction Analysis

To complement the sequence data, Taq-amplified ND5 PCR products were digested with three different restriction enzymes (AluI, Sau3AI, and DraI), and their digests were separated on agarose gels (see table 2 for list of samples). If an individual harbors more than one type of ND5-like DNA, more than one “allele” may be observed within an individual, and they should agree with those expected from the sequence data. But as the different “alleles” may not correspond to the same mitochondrial or chromosomal locus, they are referred to here as “digestypes”. A 324-bp PCR product, amplified from the COI region, was also digested (with MboII, XbaI, and FokI) to check for multiple digestypes. Digestions were carried out using restriction enzymes from various sources (Gibco BRL, Boehringer Mannheim, Pharmacia, Promega) following manufacturers’ instructions and run on 2%–2.5% agarose gels at 50–75 V.

Sequence Alignment and Phylogenetic Analysis

The sequences used in this study have been entered into GenBank (accession numbers AF085501–AF085599). All of the sequences were determined from both strands, but in some clones the overlap was less than 80%. As the sequences involved were not very divergent, they were aligned and checked by eye in MacClade, version 3.01 (Maddison and Maddison 1992). PAUP*, version 4.0b2 (Swofford 1998), and MEGA, version 1.0 (Kumar, Tamura, and Nei 1993), were used for pairwise comparison of sequences and for phylogenetic construction.

Pairwise distances were estimated using the F84 model (Felsenstein 1984). This measure allows for considerable base frequency bias and for a transition: transversion ratio bias, which may be expected when handling insect mtDNA sequences. Our sequences are very AT-rich. A neighbor-joining approach was used to construct a tree from these distances in order to illustrate approximately how diverged different ND5-like regions are from each other. Gaps were treated as missing data and distributed proportionally to unambiguous changes. Bootstrap values calculated were based on 1,000 replicates. The DNA regions included in the analysis may be evolving at different rates, and some are apparently under selective constraints, while others are not; thus, many of the assumptions made when estimating genetic distances are violated. Therefore, this tree may not represent a true phylogeny of the nuclear and mitochondrial sequences it includes.

Deduction of Mitochondrial DNA Sequences

Mitochondrial DNA was successfully enriched for one P. pedestris individual, two Italopodisma individuals and one S. gregaria individual. Amplifications of the ND5 region from the (unenriched) DNA of these individuals yielded more than one type of ND5-like sequence. In contrast, after the enrichment for mitochondrial DNA, only one type of ND5 sequence was abundant in the PCR product (fig. 1 and table 4). Under this circumstance, and because no unusual traits were observed in the sequence (e.g., frameshift mutations, stop codons), this sequence was assumed to be the mitochondrial sequence in each of these individuals.

PCR products from the total DNA of eight ethanol-preserved grasshoppers (one P. mikado, one C. parallelus, and six P. pedestris individuals) were also cloned.

Table 4

Frequencies of the Proposed mtDNA Sequence Before and After Enrichment

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>INDIVIDUAL</th>
<th>TOTAL DNA</th>
<th>MITOCHONDRIAL ENRICHED DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italopodisma</td>
<td>ItA</td>
<td>4/34 (12%)</td>
<td>11/12 (92%)</td>
</tr>
<tr>
<td></td>
<td>ItB</td>
<td>3/15 (20%)</td>
<td>5/5 (100%)</td>
</tr>
<tr>
<td>Podisma pedestris</td>
<td>FrA</td>
<td>4/15 (27%)</td>
<td>4/6 (67%)</td>
</tr>
<tr>
<td>Schistocerca gregaria</td>
<td>FrA</td>
<td>9/23 (39%)</td>
<td>2/2 (100%)</td>
</tr>
</tbody>
</table>

- *In the French individual (FrA), there was still only one copy of a second ND5 type which was also present in the total DNA. It may be that this second type represents mtDNA (heteroplasmacy) in this individual. This second, possibly mitochondrial, DNA does not change the outcomes of this study, as it differs from the most common mtDNA by only 1 nt.
- Assuming that the second ND5 type observed after enrichment is nuclear.
- In Zhang and Hewitt (1996b), it is demonstrated that the S. gregaria mtDNA sample was pure enough that only mtDNA would be detectable by PCR amplification.
and sequenced. Because they were stored in ethanol and their mitochondria were therefore not well preserved (Dowling et al. 1996), their mtDNA sequences could not be identified directly by mtDNA enrichment. However, for *P. pedestris* and *P. mikado*, the most likely mtDNA sequence candidates could be deduced. In general, a copy was assumed to be mitochondrial (or at least very recently mitochondrial) if it contained no frameshifts or stop codons, if there were several identical copies of it present within a grasshopper, and if there were no other likely candidates.

**Determination of Separate Nuclear Integration Events**

A pair of nuclear mitochondrial pseudogenes could have descended from the same mitochondrial immigrant; for example, they could be the result of an intragenomic (nuclear) duplication or amplification event. If this is the case, and if mitochondrial sequences do not retain their function after their nuclear integration (a reasonable assumption for animal Numts discussed in Brennicke et al. [1993] and Gellissen and Michaelis [1987]), then they will have diverged from each other under no selective constraints. The lack of selective constraints during nuclear sequence divergence should ensure that there is no significant bias in favor of the accumulation of changes at otherwise selectively constrained sites (e.g., at third positions of codons). If significant bias among codon positions is observed in the changes accumulated during the evolutionary divergence of two pseudogenes, this implies that the pseudogenes are descended from different (selectively constrained) mitochondrial immigrants. Therefore, if significant codon position bias is observed in the differences between two Numts, this implies that they originated from separate transfers from mitochondria to nucleus.

Each of the 61 nuclear mitochondrial pseudogenes shown in figure 2 was compared with every other ND5-like pseudogene (a total of 1,769 pairwise comparisons). The number of nucleotide differences between each pair were counted separately using MEGA, version 1.0 (Kumar, Tamura, and Nei 1993), for the first, second, and third codon positions. For each pairwise comparison, a χ² test was performed to test if the pair of Numts showed significant codon position bias (df = 2, P < 0.05) in the differences between them. These pairwise comparisons generate a list of Numts which could not have descended from the same mitochondrial immigrants. This list can be used to infer the minimum number of separate integration events required to explain the data.

**Results**

More than One Type of COI and ND5 in a Grasshopper

Individuals of the brown mountain grasshopper (*P. pedestris*) each possess many different mtDNA-like sequences. This applies at least to the COI (table 1) and ND5 (table 2) mtDNA regions, and the latter has been studied in depth. Multiple ND5-like sequences were found in the other grasshoppers of the tribe Podismina (*Italopodisma* sp., *Cophopodisma pyrenea*, and *P. mikado*), in other species of the same subfamily (Podisminae) (*Ognevia longipennis* and *Primnoa hayachinea*), in grasshoppers from other subfamilies (the desert locust *S. gregaria* [Cyrtacanthacridinae]), in a *Calliptamus* sp. (*Calliptaminae*), and in two species of the subfamily Gomphocerinae: the meadow grasshopper *C. parallelus* and an *Arcyptera* sp. (tables 2 and 3).

The possibility of the cross-contamination of samples was thoroughly controlled for, as was the possibility of somatic differences (Hadler, Daniel, and Pratt 1971; Nielsen et al. 1994; Liang 1996; Hadler, Devadas and Mahalingam 1998) in ND5-like regions (data not shown). The results from sequencing and restriction enzyme approaches are in close agreement (data not shown). In particular, the majority of sequences obtained were from PCR products which had been amplified with *Pfu* DNA polymerase, a high-fidelity enzyme. Controls were carried out to check the error rate, which was shown not to be significant (data not shown). It is therefore unlikely that the multiple different mtDNA-like regions observed within single individuals are artifacts of the techniques employed.

Extra ND5 Types Are Not Heteroplasmic mtDNAs

Nuclear and mitochondrial DNA were purified for *S. gregaria* (Zhang and Hewitt 1996b). After this separation, many different ND5-like regions (many of which contained frameshift mutations) had been amplified from the purified nuclear DNA, and only one sequence (with no frameshift mutations) appeared to be amplified from the purified mtDNA. This suggests that the extra ND5-like sequences are nuclear in *S. gregaria*.

After enriching *P. pedestris* and *Italopodisma* DNA for mtDNA, the DNA amplified was mostly of one type, whereas multiple ND5-like regions had been amplified before enrichment (fig. 1 and table 4). This separation of the extra ND5-like sequences (successful for *P. pedestris* and two of the *Italopodisma*) implies that the extra ND5 types observed are not mitochondrial. Although heteroplasmity may occur in grasshoppers, it has
not been detected here and therefore cannot explain the many extra ND5-like sequences reported in this study.

Although it is possible that these extra ND5-like sequences exist as extrachromosomal, episomal DNA (Sunnucks and Hales 1996), several lines of evidence suggest they are of nuclear origin. The extra ND5-like copies are nuclear in *S. gregaria*, as are the extra 12S rRNA, D-loop, and tRNA<sup>Asp</sup> regions observed by Zhang and Hewitt (1996b). Fluorescent in situ hybridization (FISH) data have shown COI and control region-like sequences to map to the telomeres in *S. gregaria*. COI-like sequences are centromeric in *C. parallelus*, and preliminary results for *Italopodisma* suggest a dispersed distribution of numerous COI-like sequences in the nuclear genome (Vaughan, Heslop-Harrison, and Hewitt 1999). There are thought to be hundreds of tRNA<sub>Leu</sub>, 12S, and 16S rRNA-like Numts in *L. migratoria* (Orthoptera, Acrididae) (Gellissen and Michaelis 1987). All of these studies suggest that nuclear locations for mitochondrial-like sequences are common in grasshoppers.

In addition, most of the extra ND5-like copies were lost after amplification of a large fragment of the mtDNA molecule by long-PCR (data not shown). Studies of nonmitochondrial ND5-like sequence evolution in *P. pedestris* and *Italopodisma* sp. confirm that the extra ND5-like regions are evolving in a manner that would be expected in the nucleus (data not shown). Some of the extra ND5-like regions appear to have escaped the mitochondria before *P. pedestris* diverged from *Italopodisma* sp. (fig. 2). This suggests that the extra ND5-like regions can persist in grasshoppers over long periods of evolutionary time, which is most likely if they are stably integrated into the nuclear genome.

Numts Observed in Different Acridid Species Are the Result of Independent Migrations to the Nucleus

It is clear from the distance tree (fig. 2), that the nuclear ND5-like pseudogenes observed in acridid species from different acridid subfamilies (the Podisminae, the Cyrtacanthacridinae, and the Gomphocerinae) did not result from a single mitochondrial invasion of the nuclear genome in ancient acridid history. The Numts of *S. gregaria* (Cyrtacanthacridinae) and *C. parallelus* (Gomphocerinae) are distinct from those of the Podisminae and appear to have arisen relatively recently in the evolution of these species.

Independent Mitochondrial Origins of the Numts Observed Within Species and in Single Individuals

Many of the pseudogenes observed within single individuals displayed significant codon position bias (more changes at third and first than at second codon positions) when compared with each other (1,450 of the 1,769 pairwise comparisons). This implies that they are descended from significantly diverged mtDNA sequences, and therefore from separate mitochondrial migrations to the nucleus. At least 12 independent mitochondrial to nucleus transfer events are required to explain the codon position biases revealed through the pairwise comparisons of *P. pedestris* pseudogenes, two for *Italopodisma*, two for the three sequenced pseudogenes of *P. mikado*, three for *S. gregaria*, and the five different ND5-like Numts sequenced for *C. parallelus* originate from at least three significantly diverged mtDNAs.

If two pseudogenes have separate mitochondrial origins but the two mtDNAs they were descended from were similar, there will not be enough differences between them for any resulting codon position bias to be significant. In addition, if many nuclear mutations have accumulated in a Numt, they will swamp selectively constrained changes between it and other Numts. Therefore, the number of independent integration events has probably been underestimated.

A Nuclear Family of Pseudogenes in *Italopodisma*

There is a large group of *Italopodisma* pseudogenes which are closely related (fig. 3). Examination of their aligned sequences reveals that these copies differ from each other only by small insertions, deletions (indels), and unique substitutions. These are not the result of polymerase error; the clones that were sequenced from *Italopodisma* are from a *Pfu*-amplified PCR product, and the *Pfu* error rate was empirically found to be too low to be significant for the purpose of this study (data not shown). There is no significant codon position bias among the pseudogenes belonging to this family; the pooled numbers of changes among the 30 different pseudogene sequences of this family are 12 at first codon positions, 19 at second codon positions, and 12 at third codon positions, and this does not significantly differ from a 1:1:1 ratio ($\chi^2$ test: df = 2, $P = 0.3$). Therefore, this family of nuclear pseudogenes is likely to have descended from a single mitochondrial immigrant which has been amplified since it escaped from the mitochondria.

The differences between the pseudogenes belonging to the nuclear family and the pseudogenes which are much more diverged from the mtDNA sequence are not randomly distributed; most lie at the third positions of codons (data not shown). This and the strong bootstrap support for the branch leading to the cluster of the nuclear family (fig. 3) suggest that the nuclear family of Numts arose independently of the events that generated any of the other Numts observed.

Discussion

Multiple ND5-like sequences were encountered in all 10 grasshopper species studied, representing four different acridid subfamilies (Podisminae, Calliptaminae, Cyrtacanthacridinae and Gomphocerinae). This supplement updates observations of nuclear mitochondrial pseudogenes in another acridid species, *L. migratoria* (subfamily: Oedipodinae) (Gellissen et al. 1983). The evidence presented here suggests that in *P. pedestris*, *Italopodisma*, and *S. gregaria*, at least, the extra ND5-like sequences are nuclear. This conclusion is also supported by the nuclear localization of other mtDNA-like sequences in *S. gregaria* (Zhang and Hewitt 1996b; Vaughan, Heslop-Harrison, and Hewitt 1999), *C. par-
allelus, and Italopodisma (Vaughan, Heslop-Harrison, and Hewitt 1999).

This study demonstrates the use of a novel approach for distinguishing Numts with separate mitochondrial origins. This is done by examining the codon position bias in the differences revealed through pairwise comparisons of the sequences (where they show a reasonable degree of divergence). Such an approach could be used to establish the independent origins of other types of DNA which lose their functions upon nuclear incorporation (“dead-on-arrival sequences” as described in Graur, Shuali, and Li 1989), e.g., processed pseudogenes or non-LTR retrotransposable elements.

The many different Numts observed in different grasshopper subfamilies appear to be the result mainly of horizontal transfer events from mitochondria to nuclei which have occurred since these grasshopper species diverged from each other. Similar observations of independent mitochondrial origins of nuclear pseudogenes among related species have been made for aphids of the Sitobion genus (Sunnucks and Hales 1996) and for diving ducks of the Aythyini tribe (Sorenson and Fleischer 1996).

Independent horizontal transfer events also explain much of the Numt diversity observed within individuals. There have been at least 12 separate integrations of mtDNA into the nuclear genomes of P. pedestrises. This is the highest reported frequency of mitochondrial integration in a single species. Podisma pedestrises individuals are not unusual among grasshoppers. Every individual of the other grasshopper species studied in depth (Italopodisma sp., P. mikado, C. parallelus, and S. gregaria) harbors Numts apparently arising from significantly different mitochondrial ancestors.

The work presented here suggests that more than one mechanism exists in grasshoppers for the generation of Numt sequences. The Italopodisma individuals studied show evidence of multiple horizontal transfer events from mitochondrion to nucleus; however, a family of almost identical Numts also exists (figs. 2 and 3). Given the patterns of nucleotide substitutions observed in these sequences, it is likely that this family of Numts was generated by amplification of a single type of mtDNA-like sequence. Such amplification events have been described for cats (Lopez et al. 1994) and humans (Hu and Thilly 1995). In cats, the amplification event is thought to have occurred prior to the nuclear integration of the mitochondrial-like DNA in question (Lopez et al. 1994), while in humans, amplifications are thought to have occurred since nuclear integration (Hu and Thilly 1995).
Further work is needed to explain how and where the family of Numts reported here arose. At least in grasshoppers, more than one process appears to be involved in the generation of numerous mitochondrial-like sequences in the nuclear genome: intergenomic horizontal transfer (see above), which involves the migration of DNA from mitochondrion to nucleus and is responsible for the independence of mitochondrial origins among Numts, and posttransfer amplification, which is probably a nuclear event (or a cytoplasmic episomal event), responsible for the repetitive presence of some Numt sequences.

The ND5-like Numts observed in S. gregaria are only up to 1.4% diverged from the current mtDNA sequence, while in P. pedestris this divergence can be as high as 12.5% (table 3 and fig. 2). Such heterogeneity has been noticed in other animal species, but never before in species which are this closely related (Zhang and Hewitt 1996a). This suggests heterogeneity among these species in the frequency of Numt generation, possibly reflecting differences in the ability of different genomes to gain or lose Numts.

Grasshoppers are not unique in the persistence of mtDNA-like sequences in their nuclear genomes. Numts are common in other species groups, for example, in hominoids and Old World monkeys (Collura and Stewart 1995; van der Kuyl et al. 1995), in birds (Quinn and White 1987; Arctander 1995; Sorensen and Fleischer 1996; Kidd and Friesen 1998; K. Nielsen, personal communication; N. Harvey, personal communication), and in three species of Sitobion aphid (Sunnucks and Hales 1996). However, the stable nuclear absorption of mtDNA sequences does not appear to be universal; no Numts have been reported in Plasmodium falciparum, Caenorhabditis elegans, or Drosophila, although these organisms are well studied (Blanchard and Schmidt 1996). The occurrence of Numts is not phylogenetically continuous in felines (Lopez et al. 1994), and they also appear to be absent from the nuclear genomes of aphid species not belonging to the genus Sitobion (Sunnucks and Hales 1996).

At present, it is not known which factors ultimately determine whether Numts are common in a species. Interestingly, some of the organisms which lack Numts (e.g., the aforementioned P. falciparum, C. elegans, and Drosophila) have small nuclear genomes, while species in which many Numts have been found (e.g., the grasshoppers studied here) possess much larger nuclear genomes. When more data on the taxonomic distribution of Numts become available, we will be able to explore the possibility of a correlation between these two phenomena.

Acknowledgments

We would like to thank Jon Bridle, Angus Davison, Brent Emerson, Nic Flanagan, Ito Gen, Jaime Gosalvez, Henry VIII School Coventry, Pekka Käär, Phil Mason, Richard Nichols, C. H. F. Rowell, Teresa Stow, Bill Sutherland, Jacek Szymura, and Helen Vaughan for traveling to, and collecting samples from, the far-flung reaches of P. pedestris’ arctic-alpine distribution. Fiammetta Santucci, Brent Emerson, Kamal Ibrahim, Steve Jones, Helen Vaughan, and members of the lab provided helpful discussion. We are also grateful to four anonymous reviewers for their useful comments on an earlier version of this manuscript. This work was partly funded by an NERC studentship and the University of East Anglia.

LITERATURE CITED


Harz, K. 1975. The Orthoptera of Europe II. Dr. W. Junk, the Hague, the Netherlands.


Frequent mtDNA Migration to Grasshopper Nuclei


RICHARD THOMAS, reviewing editor

Accepted November 8, 1999