Evidence for recombination in scorpion mitochondrial DNA (Scorpionida: Buthidae)

Benjamin Gantenbein1,*,†, Victor Fet2, Iris A. Gantenbein-Ritter1 and François Balloux1

1Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK
2Department of Biological Sciences, Marshall University, Huntington, WV 25755-2510, USA

There has been very little undisputed evidence for recombination in animal mitochondrial DNA (mtDNA) provided so far. Previous unpublished results suggestive of mtDNA recombination in the scorpion family Buthidae, together with cytological evidence for a unique mechanism of mitochondrial fusion in that family, prompted us to investigate this group in more details. First, we sequenced the complete mtDNA genome of Mesobuthus gibbosus, and chose two genes opposing each other (16S and coxI). We then sequenced 150 individuals from the natural populations of four species of Buthidae (Old World genera Buthus and Mesobuthus). We observed strong evidence for widespread recombination through highly significant negative correlations between linkage disequilibrium and physical distance in three out of four species. The evidence is further confirmed when using five other tests for recombination and by the presence of a high amount of homoplasy in phylogenetic trees.

Keywords: cytochrome oxidase subunit I; linkage disequilibrium; mitofusion; mitochondrial DNA; scorpions; Buthidae

1. INTRODUCTION
Mitochondrial DNA (mtDNA) is widely used for phylogeographic and phylogenetic studies for two principal reasons. First, mtDNA is a compact double-stranded circular DNA molecule (approximately 15–16 kb) and second, recombination (the exchange of homologous DNA sequences between different chromosomes) is thought to be absent in animals owing to the clonal inheritance of mtDNA. Numerous phylogenetic studies rely on mtDNA genes (both protein-coding and ribosomal RNA genes) amplified through conserved polymerase chain reaction (PCR) primers (Simon et al. 1994). Most models of genetic divergence and phylogenetic methods assume absence of recombination. Thus, major conclusions drawn from these analyses may be inaccurate since tree-building algorithms are misled by recombination (Avise 1994; Schierup & Hein 2000; Posada & Crandall 2002).

Despite the fact that animal mitochondria undoubtedly have the necessary toolkit of enzymes to recombine (Thyagarajan et al. 1996; Eyre-Walker et al. 1999; Yaffe 1999; Sokas et al. 2003), evidence for recombination in animal mtDNA is scarce (Rokas et al. 2003). Evidence based on different indirect methods was found in humans, mice (Mus musculus), flatfish (Platichthys spp.), frogs (Rana spp.) and bivalve mussels (Mytilus spp.). However, only in the last case is there a direct experimental confirmation for homologous recombination owing to biparental inheritance (Awadalla et al. 1999; Ladoukakis & Zouros 2001). Whether human mtDNA is really recombining has been questioned since Awadalla et al. (1999) used a potentially uncertain measure of linkage disequilibrium (LD) in their test for recombination (Ingman et al. 2000; Kivisild & Villems 2000; Kumar et al. 2000) and there is a possibility of sequencing errors in their data (Macaulay et al. 1999; Wiu 2001; Innan & Nordborg 2002).

While working on mtDNA phylogeographies of Old World buthid scorpions (genera Buthus and Mesobuthus; Gantenbein & Largiadér 2002, 2003; Gantenbein et al. 2003), results were repeatedly obtained that led to the suspicion that there might be recombination in the mtDNA of some of these species. These suspicions were further fuelled by cytological evidence, reporting the fusion of mitochondria into a ring (mitofusion) in the Buthidae scorpion family. The unique cytological phenomenon of mitofusion together with the preliminary statistical evidence for recombination prompted the investigation of this issue in more detail.

Here, we present strong statistical evidence for recombination in newly sequenced populations of west and Central Asian scorpions of the genus Mesobuthus and in previously published scorpion populations (Gantenbein & Largiadér 2002, 2003; Gantenbein et al. 2001, 2003), sequenced at two mitochondrial loci, the large ribosomal RNA unit (16S) and a fragment of the cytochrome oxidase subunit I (coxI). In order to precisely estimate the physical distance between the two sequenced loci, which is required for the correlation between genetic LD and distance, we sequenced the complete mitochondrial genome of the scorpion Mesobuthus gibbosus, which will be published elsewhere (Gantenbein in progress).

2. MATERIAL AND METHODS
(a) Samples
The DNA sequences from the GenBank of three published
datasets were extracted (Gantenbein & Largiadèr 2002, 2003; Gantenbein et al. 2003), including four different buthid species, Buthus mardechi (Simon 1878; Morocco, North Africa), Mesobuthus caucasicus (Nordmann 1840; Central Asia), M. epeus (C. L. Koch 1839; Anatolia and Central Asia) and M. gibbosus from Central Asia (Turkmenistan, Kazakhstan, Uzbekistan and Tajikistan) were collected between June 2000 and May 2002. New sampling sites and a detailed list of identified haplotypes can be downloaded at http://www.science.marshall.edu/fet/euscorpius/mt_buthid.htm.

(b) DNA methods
DNA was extracted according to a standard phenol–chloroform/ethanol precipitation method of 100% ethanol-stored tissue samples (Gantenbein & Largiadèr 2002). Approximately 200–400 ng total genomic DNA was used for 20 μl PCR reactions as previously reported. We used the same PCR profiles as in Gantenbein et al. (2003) and Gantenbein & Largiadèr (2003). Amplicons were checked for clean single band amplification on 1–2% agarose gel, and subsequently purified as previously described (Gantenbein et al. 2003). Sequencing reactions were carried out using one of the PCR primers (Gantenbein & Largiadèr 2003) and were run on an automated sequencer (either ABI377XL or on a capillary sequencer ABI3730). All sequence traces were inspected manually for sequencing errors. Haplotypes with singletons were reamplified and resequenced from both directions. In order to determine the distance between the two gene fragments (16S and coxI, 461 bp). The samples of M. caucasicus and M. epeus from Central Asia (Turkmenistan, Kazakhstan, Uzbekistan and Tajikistan) were collected between June 2000 and May 2002. New sampling sites and a detailed list of identified haplotypes can be downloaded at http://www.science.marshall.edu/fet/euscorpius/mt_buthid.htm.

Table 1. The total number of sites and the number of segregating sites.

<table>
<thead>
<tr>
<th>species</th>
<th>n</th>
<th>total number of sites</th>
<th>segregating sites (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. o. mardochi</td>
<td>34</td>
<td>360</td>
<td>464</td>
</tr>
<tr>
<td>M. caucasicus</td>
<td>36</td>
<td>349</td>
<td>461</td>
</tr>
<tr>
<td>M. epeus</td>
<td>38</td>
<td>354</td>
<td>461</td>
</tr>
<tr>
<td>M. gibbosus</td>
<td>42</td>
<td>341</td>
<td>461</td>
</tr>
</tbody>
</table>

M. eupeus

(c) Statistical tests for recombination
All sequences were aligned using CLUSTALW1.5 (Thompson et al. 1997) and by eye. The coxI data could be unambiguously aligned because of the presence of an open reading frame. Following Posada & Crandall (2001) and Wiuf et al. (2001), the five following recombination tests were applied: (i) Recombination Detection Program (RDP; Martin & Rybicki 2000), (ii) Geneconv (Sawyer 1989), (iii) Maxchi (Maynard Smith 1992), (iv) Chiamaera (Posada & Crandall 2001; a slight modification of Maxchi) and (v) the homoplasy test (Maynard Smith & Smith 1998). The first four tests were run using RDP (Martin & Rybicki 2000). In RDP, we set the highest acceptable p value to 0.05 (= the probability that sequences could share high identities in potentially recombinant regions by chance alone) and we corrected for multiple comparisons using the standard Bonferroni method for Geneconv, and sequential Bonferroni for RDP, Maxchi and Chiamaera (Sokal & Rohlf 1995; Martin & Rybicki 2000). In RDP p values are binominal and are of the Karlin–Altschul-type in Geneconv, and χ²-distributed in Maxchi and Chiamaera. The significance of χ² peaks in Maxchi and Chiamaera were determined with 1000 permutations. Tests were run on the concatenated 16S and coxI data or on the protein-coding coxI exclusively. The homoplasy test (Maynard Smith & Smith 1998) is based on an estimation of expected homoplasies under clonality and compares this estimate with the true number of homoplasies in a dataset. When computing the homoplasy test, we only considered third coding positions of the coding region (coxI) since non-coding regions characterized by high mutation rate biases can lead to erroneous results when applying this test (Piganeau & Eyre-Walker 2004). As an out-group for the estimation of the effective number of sites, the species Androctonus australis (L., 1758; Buthidae) was included. These tests were run using the Qsisc program available at http://www.biols.susx.ac.uk/home/John_Maynard_Smith/.

Additionally, LD between polymorphic sites using two different measures of LD, 〈D〉 (Lewontin 1964) and R² (Hill & Robertson 1968) was estimated, the latter being more robust to variation in mutation rates (Awadalla et al. 1999; Kivisild & Villems 2000; Kumar et al. 2000; Meunier & Eyre-Walker 2001; Wiuf 2001; Innan & Nordborg 2002; Piganeau & Eyre-Walker 2004). LD was estimated using the LDhat computer package (available at http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat.html). 〈D〉 and R² versus physical distance was then correlated. A spacer fragment, was included corresponding to 5038 bp in M. gibbosus, thus, the spacer fragment was approximated to ~5 kb for all correlations in all four species (16S-5 kb-coxI). The correlation was also repeated with maximum power going the longer way around the mitochondrial genome with a ~10 kb spacer fragment (coxI-10 kb-16S). Significance of the Pearson correlation coefficient was assessed by using the one-tailed Mantel test computed over 10,000 permutations.

3. RESULTS
The DNA sequence datasets for the four species contain considerable polymorphism in terms of the total number
of segregating sites (table 1). The significant decline of LD versus distance is regarded as strong evidence for recombination (Awadalla & Charlesworth 1999). When using \( R^2 \) highly significant correlation was obtained in all four species \((p < 0.001; \text{table } 2)\). Since \( R^2 \) is dependent on allele frequencies, the analysis was additionally restricted to sites at which both alleles were present at a frequency greater than 10%. It has been argued that this restriction limits the analysis to alleles, which tend to be older and are more likely to show evidence for past recombination events (Awadalla et al. 1999). Using this subset of polymorphisms, significant correlations were still observed for three out of four species, with the correlation in \( M. \) caucasicus becoming non-significant (table 2). Performing the correlation with \( |D'| \), the correlations remain highly significant only for \( B. \) mardochei (both when considering all polymorphisms or only those above 10%), and marginally so only for polymorphisms in frequencies above 10% in \( M. \) gibbosus (table 2). Conversely, the homoplasys test finds strong evidence for recombination only in \( M. \) caucasicus (table 3). The best evidence for recombination is thus found in the two species \( B. \) mardochei and \( M. \) caucasicus (tables 2 and 3), whereas in the two other species \((M. \) epeus and \( M. \) gibbosus) the evidence is somewhat conflicting. In \( M. \) epeus in particular, only the correlation between \( R^2 \) and physical distance is significant.

Testing for recombination using the general RDP, Geneconv, Maxchi and Chimaera, on a concatenated ‘supergene’ of the 16S and \( CoxI \) data recovered many putative recombinant ‘daughter’ haplotypes between ‘parental’ haplotypes in the \( B. \) mardochei sequence alignment. An example is given in figure 1, with the RDP and the Geneconv test provide strong support for two ‘breakage points’ at sequence positions 319 and 685: The pairwise identity in the regions 1–319 and 686–823 bp is relatively high between BomAA1a and BomAA5a (haplotypes sampled from two populations from the Moroccan Anti-Atlas) relative to a third sequence (BomHA5a, a sequence from the Moroccan Haut-Atlas; figure 1a). In the middle piece (320–685 bp), however, the two haplotypes BomAA5a and BomHA5a share many polymorphic sites, thus being more similar than the two haplotypes BomAA1a and BomAA5a. ‘Breakage points’ similar to those shown in figure 1 are found for the three haplotypes BomAS1a (As Saghir region, Morocco), BomHA1a and BomHA1b. When only the protein-coding region was tested using the same four recombination tests, no potential recombination events were found using RDP and Geneconv, whereas the algorithms Maxchi and Chimaera found 12 and 8 putative recombination events in \( B. \) mardochei, 18 and 1 in \( M. \) caucasicus, 2 and 1 in \( M. \) epeus, and finally, no recombination events in \( M. \) gibbosus.

4. DISCUSSION

(a) Correlation of LD versus distance

Our mtDNA data show a clear decline of LD versus physical distance, a pattern that is typical for nuclear recombining genes (Awadalla & Charlesworth 1999).
Figure 1. Test results for recombination in *B. mardochei* scorpion mtDNA of a concatenated 16S and *coxI* 'supergene' using (a) the general recombination method (RDP), (b) the Geneconv method, (c) the Maxchi and (d) the Chimaera test statistics. The *x*-axis gives the nucleotide positions of the alignment, whereas the *y*-axis represents the particular test statistics for each test. Peaks in the log $P$ of $\chi^2$ values in the Maxchi and Chimaera tests mark potential points of recombination. Dashed lines represent $p$ value cut-offs: uncorrected (lower line) and corrected for multiple comparisons (upper line) at the 0.05 level.
It has been shown that decline of LD versus distance can also be reached with unusual models of mutation rates such as ‘mutational hot spots’ or adaptive substitution (McVean 2001; Innan & Nordborg 2002). However, we do not believe that the observed decline is owing to mutational hot spots or to adaptive substitution since the decline of LD versus distance is also observed but less statistically supported if only protein-coding sequences or synonymous third codon sites are compared (correlation coefficients are significant for $R^2$ in B. mardochei: $-0.070 \pm 0.004$, $p = 0.001$, including all sites and $-0.053 \pm 0.049)$, including only alleles with a frequency greater than 10%, and in M. epeus: $-0.139 \pm 0.01$, including all sites, all other tests are non-significant at the 0.05 level). Moreover, mutational hot spots are less likely to occur in protein-coding regions owing to codon usage bias (Piganeau & Eyre-Walker 2004). Comparing evolutionary rates of these two genes, we estimated an approximate 1% net number of ‘variations’ under the taxon B. occitanus, thus underlining their close phylogenetic relationship. A recent phylogeographic study on this species on four out of five subspecies using the same two mitochondrial makers as in this study revealed a total of 11 highly divergent clades (with approximately 12–15% sequence divergence between these clades; Gantenbein & Largiader 2003). Furthermore, the most parsimonious tree in this study suffered from a low consistency index (CI = 0.44) and a relatively high number of homoplasies (table 3). Eight out of these 11 Moroccan mtDNA clades belong to the recently elevated species B. mardochei. Nuclear allozyme data and nuclear ITS-1 sequence data completely disagreed with mtDNA sequence data with respect to the Moroccan populations (Gantenbein & Largiader 2003; Gantenbein 2004): nuclear data indicated a common gene pool for all the Moroccan samples, suggesting high gene flow among these population groups. We interpret this strong discrepancy between nuclear and mitochondrial data in the Moroccan samples (all other European and Tunisian clades being congruent in phylogenetic analysis) as evidence that the genetic population structure of Buthus prevents the extinction of these phylogenetically old and diverse mtDNA lineages, and that these Moroccan population groups ‘hybridize’. While hybridization between subspecies might itself increase mtDNA recombination, the higher divergence between sequences will obviously ease its detection through statistical means.

(c) Mitofusion and heteroplasmy in buthid scorpions

Although these results show indirect evidence for recombination, here, a clear link is seen to early cytological studies on the spermatogenesis of buthid scorpions. The studies by Wilson et al. (Wilson 1916, 1931; Wilson & Pollister 1937) were the first to observe two very different mechanisms of mitochondrial segregation in scorpions during early meiosis—fusion of mitochondria into a ring in the primary spermatocytes in the family Buthidae (genus Centruroides) as opposed to a non-buthid species (genus Opisthacanthus; family Liochelidae), which does not have such fusion. This ring is pulled apart during the first meiotic division, with a half-ring moving to each pole (figure 2). The half-ring straightens into a rod which divides so that each secondary spermatocyte receives two mitochondrial rods. During the second meiotic division, these rods divide again, and each spermatid receives two rods, which later join to form a new fused structure posterior to the nucleus called Nebenkern (German for ‘next to the nucleus’). This phenomenon has also been observed in the European buthid species, B. occitanus (Tuzet 1938) and in the Chinese buthid, Mesobuthus martensii (Sato 1940). However, such mitochondrial fusion before the first meiotic division was not observed in other scorpion families which appear to have random distribution of separate mitochondria: recorded evidence includes species of Hadrurus (Caraboctonidae), Euscorpius (Euscorpiidae), Opisthacanthus (Liochelidae), Vaejovis (Vaejovidae) and Heterometrus (Scorpionidae; Sokolow 1913; Bhattacharya & Gatenby 1924; Gatenby 1925). Electron microscopy supports the existence of a unique pattern limited to Buthidae. There are two elongated mitochondria in the middle piece of the buthid sperma-
It was later shown that both types of recombination events can occur in the same cell (heteroplasmy). In mammals, it is if at least two types of mtDNA molecules (haplotypes) are found in the same cell (heteroplasmy). In mammals, it is known that there are different mechanisms acting to prevent recombination of mtDNA (Rokas et al. 2003). Up to 100 paternal mitochondria enter the egg but are destroyed during the first few hours following fertilization. Several mechanisms ensure that the zygote contains only maternal mitochondria. However, it has been reported that the maternal inheritance of mitochondria in mammals seems to be merely a quantitative phenomenon, and situations of paternal leakage have been reported in mice (Gyllensten & Wilson 1987; Gyllensten et al. 1991). Failure of the recognition mechanisms for paternal mitochondria seems to be more common in hybrid populations, in which the specificity in the recognition process is relaxed, as was shown in drosophilids (Kondo et al. 1990).

Two recent papers provide strong support for actual recombination in human mtDNA. A recent reevaluation of newly added/corrected data found clear indirect evidence for recombination in human mtDNA using LD versus distance correlations (Piganeau & Eyre-Walker 2004). Moreover, a patient having paternal and maternal mtDNA (heteroplasmic) in his muscle tissue was recently screened for recombined haplotypes using single template PCR (Kraytsberg et al. 2004). A frequency of approximately 0.7% of recombinant genotypes was reported, providing direct proof that human mtDNA is able to recombine. A recent compilation on published mtDNA sequence data from various vertebrates and invertebrates also showed indirect evidence for recombination in primates, insects and crustaceans (Piganeau et al. 2004).

The finding of strong negative correlation between LD and distance, together with the unique cytological evidence for mitofusion in early spermatogenesis, provides a strong suggestion for recombination in buthid scorpions. A combination of cytological methods to document scorpion spermatogenesis with controlled crossing experiments of populations of known haplotypes with subsequent genotype screening of offspring would provide a powerful direct approach to address recombination in buthid scorpion mtDNA (Ladoukakis & Zouros 2001; Rokas et al. 2003). If mtDNA recombination were directly linked to the phenomenon of mitofusion, non-buthids (e.g. Euscorpiidae, Scorpionidae) should not show any indirect evidence for recombination. Such a major difference between buthid and non-buthid scorpions would not be entirely unexpected as Buthidae represent a separate, ancient lineage within the scorpion family Buthidae (Ladoukakis et al. 2003). Two recent papers provide strong support for actual recombination in human mtDNA. A recent reevaluation of newly added/corrected data found clear indirect evidence for recombination in human mtDNA using LD versus distance correlations (Piganeau & Eyre-Walker 2004). Moreover, a patient having paternal and maternal mtDNA (=heteroplasmic) in his muscle tissue was recently screened for recombined haplotypes using single template PCR (Kraytsberg et al. 2004). A frequency of approximately 0.7% of recombinant genotypes was reported, providing direct proof that human mtDNA is able to recombine. A recent compilation on published mtDNA sequence data from various vertebrates and invertebrates also showed indirect evidence for recombination in primates, insects and crustaceans (Piganeau et al. 2004).

The finding of strong negative correlation between LD and distance, together with the unique cytological evidence for mitofusion in early spermatogenesis, provides a strong suggestion for recombination in buthid scorpions. A combination of cytological methods to document scorpion spermatogenesis with controlled crossing experiments of populations of known haplotypes with subsequent genotype screening of offspring would provide a powerful direct approach to address recombination in buthid scorpion mtDNA (Ladoukakis & Zouros 2001; Rokas et al. 2003). If mtDNA recombination were directly linked to the phenomenon of mitofusion, non-buthids (e.g. Euscorpiidae, Scorpionidae) should not show any indirect evidence for recombination. Such a major difference between buthid and non-buthid scorpions would not be entirely unexpected as Buthidae represent a separate, ancient lineage within the scorpion phylogeny (Soleglad & Fet 2003) and possess a number of other unique, derived evolutionary mechanisms (such as mammal-specific neurotoxins).

B.G. was supported by an SNF-IHP grant (83-EU-065528), The Royal Society (R-36579) and the ‘Basler Stiftung für experimentelle Zoologie’ (c/o Eidgenössisches Tropeninstitut). Field trips of B.G. were supported by the commission for travel grants of the Swiss Academy of Natural Science (SANW). The field trip to Central Asia in March–May 2002 was supported by the National Geographic Society Research and Exploration grant 7001-01 to V.F. We thank Alexander V. Gromov for his help in the field and additional specimens of Mesobuthus. We also thank Gwenael Piganeau, Michael Gardener and Adam Eyre-Walker for their helpful correspondence and for granting access to their unpublished manuscript that included evidence for recombination in scorpions based on a subset of M. euripus. We are particularly grateful to
Bill Amos who let us use his sequencer free of charge to finish the project when our finances had run out. Lily Jin ran three gels for us on the ABI377. B.G. would like to acknowledge Matt E. Braunwalder for his skilled guidance on scorpion biology and for leading him into the field of ‘speriology’. One anonymous reviewer and Adam Eyre-Walker significantly improved an earlier version of the manuscript.

REFERENCES


As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.