Rapid concerted evolution in animal mitochondrial DNA

Andrey Tatarenkov* and John C. Avise

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA

Recombinational genetic processes are thought to be rare in the uniparentally inherited mitochondrial (mt) DNA molecules of vertebrates and other animals. Here, however, we document extremely rapid concerted microevolution, probably mediated by frequent gene conversion events, of duplicated sequences in the mtDNA control region of mangrove killifishes (Kryptolebias marmoratus). In local populations, genetic distances between paralogous loci within an individual were typically smaller (and often zero) than those between orthologous loci in different specimens. These findings call for the recognition of concerted evolution as a microevolutionary process and gene conversion as a likely recombinational force in animal mtDNA. The previously unsuspected power of these molecular phenomena could greatly impact mtDNA dynamics within germ cell lineages and in local animal populations.

Keywords: duplicated control regions; mangrove killfish; orthologous; paralogous; Kryptolebias marmoratus

1. INTRODUCTION

Mitochondrial (mt) DNA is probably the single most popular and powerful molecule for a wide range of phylogenetic, phylogeographic and population genetics studies. It is thus both surprising and lamentable that phylogenetic, phylogeographic and population genetic patterns for animal mtDNA.

Intracellular genetic processes in animal mtDNA remain poorly understood (Ballard & Whitlock 2004; Ballard & Rand 2005; Kmiec et al. 2006). Nowhere is this more obvious than in ongoing controversies about whether genetic recombination might play some role in the microevolution of this otherwise asexually transmitted genome (McVean 2001). Here, we document extreme concerted microevolution, probably indicative of rapid gene conversion, between two duplicated mtDNA control regions (CR1 and CR2) in local populations of the killfish Kryptolebias marmoratus. The unexpectedly rapid pace of this recombinational process has potential broad implications for intracellular population dynamics and population genetic patterns for animal mtDNA.

Portions of the mtDNA control region are duplicated in many animal taxa (Arndt & Smith 1998; Black & Roehrdanz 1998; Campbell & Barker 1999; Shao & Barker 2003; Ogoh & Ohmiya 2004), and several reports have implicated concerted evolution for those regions at the levels of species or distinct subspecies (Kumazawa et al. 1996; Eberhard et al. 2001; Abbott et al. 2005; Shao et al. 2005; Ogoh & Ohmiya 2007). Lee et al. (2001) first reported duplicated control regions in K. marmoratus, and, based on high sequence similarity between CR1 and CR2 in the single specimen examined, they also noted the possibility of concerted evolution. However, their study could not rule out an alternative hypothesis that the duplication was recent and the two copies had not yet accumulated mutational differences. To demonstrate concerted evolution at the microevolutionary level, it is necessary to (i) identify polymorphic duplicated sequences in a sample of conspecific individuals, preferably within local populations and (ii) show that the genetic differences between paralogous loci in the same individuals are unambiguously smaller than the differences between orthologous loci in different individuals. Here, we fulfil both of these criteria using complete (or nearly so) CR1 and CR2 sequences from 88 specimens.

2. MATERIAL AND METHODS

Samples were collected from the following locales (with abbreviations indicated in parentheses): Twin Cays, Belize (TC); Exuma and San Salvador Islands, Bahamas (EL, SS); Everglades National Park near Flamingo in Monroe County, FL (EP); Shark River also in Monroe County, FL (SR); Charlotte County, FL (CC); St Lucie County, FL (SL); Marco Island in Collier County, FL (MI); and No Name Key in the Florida Keys (NK). Additional details on capturing are provided in Mackiewicz et al. (2006). Following Lee et al. (2001), we refer to CR1 as the 887 bp non-coding region between the 5’ ends of loki coding for tRNA-Pro and tRNA-Phe, and CR2 as a 795 bp non-coding region between the 3’ end of the loki coding for tRNA-Leu(UUR) and the 5’ end of ND1. For each fish specimen, we sequenced two disjunct mtDNA regions: (i) ‘Cytb-trnP-CR1’, which spans 1243 bp (positions 16 060–17 300 on GenBank Accession AF283503) and (ii) ‘trnL1-CR2-ND1’, which spans 871 bp (positions 2791–3662 on AF283503). A large segment (735 bp) of CR2 displays high overall sequence similarity (greater than 99%) to CR1, and only this region is the focus of our analysis. Region (i) was amplified with primers 5’-TCGCCCTTACTGGCCTCAAATTCT-3’ and 5’-TTTAAGCTACACAGGCCCTAAGTTC-3’. Region (ii) was amplified with primers 5’-AACGCTCTGTT-TAGGGTGCCAGA-3’ and 5’-AGGAAAGCAAGGCTGAAAGGA-3’. GenBank accession numbers for the nucleotide sequences analysed in this study are EF202348–EF202523.

Received 6 February 2007
Accepted 23 April 2007

doi:10.1098/rspb.2007.0169
Published online 8 May 2007

* Author for correspondence (tatarenk@uci.edu).
Table 1. Evidence for concerted evolution in CR1 and CR2 sequences. (Shown is the common sequence in most specimens and observed departures from it at the indicated loci in specified individuals at each of 27 polymorphic nucleotide positions. Each letter ‘d’ is a 1 bp deletion. Cases of intra-individual genetic identity for CR1 and CR2 are shown in bold.)

<table>
<thead>
<tr>
<th>Specimens (and loci)</th>
<th>47 specimens (CR1, CR 2); EL6 (CR1); CC5 (CR1); MI1 (CR1); MI4 (CR1); MI5 (CR1); TC2 (CR2); TC5 (CR2); SR6 (CR1, CR2); SR7 (CR1, CR2); SR9 (CR1, CR2); SR2 (CR1, CR2); SR3 (CR1, CR2); SR8 (CR1, CR2); SR11 (CR1, CR2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>111222333344444555567 11145012180000003368034563 01466414441234673765914135 specimens and loci</td>
</tr>
</tbody>
</table>

We aligned CR1 and CR2 sequences visually, with assistance from CLUSTAL V (default parameters) as implemented in MEGALIGN (LASERGENE v. 6, DNASTAR, Inc.). We employed the p distance as the measure of sequence divergence (Nei & Kumar 2000). Other distance measures were unnecessary or inappropriate, respectively, because the sequences were so similar and because we also included indels as characters (these would require subjective weight assignments in more complex distance metrics). Gene conversion rates were estimated by the method of Eberhard et al. (2001), which assumes that genetic distances between pairs of sequences scale with time to the most recent gene conversion events. The haplotype network was constructed using the method of Templeton et al. (1992) implemented in software TSC v. 1.21 (Clement et al. 2000).

3. RESULTS AND DISCUSSION

Alignments of the 735 bp sequences from CR1 and CR2 revealed 27 variable nucleotide positions, including five 1 bp deletions (table 1). A striking result is the near identity of CR1 and CR2 within individuals, despite extensive sequence differences (at as many as nine nucleotide positions) between CR1 and CR2 in different specimens. At nucleotide position 663, for example, most individuals displayed an adenine in both CR1 and CR2, but three individuals displayed a cytosine at both loci and one showed a guanine at both CR1 and CR2. Qualitatively similar outcomes were apparent at each of 26 other polymorphic nucleotide sites surveyed (table 1; figure 1). Overall, this pattern of concerted microevolution is also quantitatively evidenced by a 10-fold higher mean genetic distance (p) between orthologues in different individuals (p = 1.9 × 10^-3) vis-à-vis paralogues within individuals (p = 1.7 × 10^-4).

CR1 and CR2 did differ in sequence within 11 out of 88 specimens, but never by more than one mutation step (table 1). Furthermore, each of these intraspecific sequence discrepancies was confined to a single population, the only possible exception being a 1 bp deletion in CR2 in specimens EL6 and CC5. The finding that most intraspecimen differences between paralogues are confined to a local population further indicates that such discrepancies are short lived due to rapid concerted evolution at these two loci.

Following Eberhard et al. (2001), we provisionally estimated the pace of gene conversion by assuming a standard vertebrate rate for CR evolution (10% change per lineage per million years) as applied to the mean within-individual genetic distance between CR1 and CR2. By this approach, the mean time (± 95% CI) between gene conversion events is 850 ± 479 years or about eightfold shorter than the mean time-interval between successive nucleotide substitutions within either CR region. Although the precision of this estimate is debatable, the fact remains that the data evidence a stunning pace of concerted evolution in a molecule that is normally assumed to evolve without appreciable recombinant processes (Wolstenholme 1992; Birky 2001).

A high incidence of concerted evolution in mitochondrial sequences has both practical and conceptual implications. Pragmatically, it means that the two CR regions should be of equal, but redundant, utility in...
thousands of mtDNA molecules and belonging to a local
despite having developed from a zygote that had many
single predominant mtDNA genotype (i.e. homoplasmic),
most animal species is why each individual often displays a
phenomenon of homoplasmy. A long-standing puzzle for
evolution might possibly have any bearing on the
routinely (present study) at least raises the possibility that
recombinational machinery in animal mtDNA exists
structure; Eberhard
the duplicate loci within the three-stranded D-loop
hypothetical mechanism involves recombination between
nation is confined to this small section of mtDNA (one
duplicated CR sequences. Perhaps homologous recombi-
on the other hand, it applies at face value only to the
other: gene turnover via frequent duplications and
losses of sequences, or, more likely in this case, some form
of gene conversion (non-reciprocal genetic exchange)
per se (Kumazawa et al. 1996, 1998).
Previous evidence for recombination in animal
mtDNA, e.g. from subtle patterns of linkage disequilibria
across the molecule (Awadalla et al. 1999; Eyre-Walker et al. 1999),
has been indirect and hotly contested (McVean 2001; Kmiec et al. 2006).
Our current evidence for recombinational processes is far more compelling, but,
on the other hand, it applies at face value only to the
duplicated CR sequences. Perhaps homologous recombi-
nation is confined to this small section of mtDNA (one
hypothetical mechanism involves recombination between
parental and nascent DNA strands during replication of the
duplicate loci within the three-stranded D-loop
structure; Eberhard et al. 2001). But the mere fact that
recombinational machinery in animal mtDNA exists
(Thyagarajan et al. 1996) and can apparently function
routinely (present study) at least raises the possibility that
the phenomenon has much wider ramifications.
We wonder, for example, whether concerted micro-
evolution might possibly have any bearing on the
phenomenon of homoplasmy. A long-standing puzzle for
most animal species is why each individual often displays a
single predominant mtDNA genotype (i.e. homoplasmic),
despite having developed from a zygote that had many
thousands of mtDNA molecules and belonging to a local
population of animals in which mtDNA sequence
variation is normally extensive (Avise 2000). A traditional
explanation for rapid evolutionary transitions through
transient heteroplasmic states is that mtDNA numbers
undergo population bottlenecks in the germline cells that
are precursor to each mtDNA-rich oocyte (Chapman et al.
1982; Laipis et al. 1988). However, direct empirical
evidence for such mtDNA bottlenecks in germline cells
is meagre, and another hypothetical process (that could
operate in conjunction with the first) is rapid, biased gene
conversion such that particular mtDNA molecules often
convert others to their own sequence within germline cells.
We also wonder whether concerted evolution in
mtDNA might have a bearing on mtDNA repair
mechanisms, which are generally thought to be deficient
compared with those in the cell nucleus. But if gene
conversion is widespread and routine, perhaps the
underlying mechanisms play a role in mtDNA repair as
well. The current data do not address whether the inferred
gene conversion events are intramolecular (within one
.circular mtDNA molecule) or intermolecular (between
different mtDNA molecules), but if the latter is true, there
are certainly many mtDNA templates within any cell that
could, in principle, serve as useful substrates for the
recombinational repair of sequence defects.
These possibilities are admittedly highly speculative,
but they do highlight the potential significance of the
discovery that concerted microevolution does occur and
can be extremely rapid in animal mtDNA. Nevertheless,
our current findings are unlikely to challenge conventional
wisdom about the utility of mtDNA for phylogeographic
or phylogenetic inferences about matriline. For that to be
true, routine interparent mtDNA recombination
(following paternal leakage) would have to be demon-
strated as well, and numerous studies seem to indicate that
such processes are at most very rare.
We thank Bruce J. Turner, D. Scott Taylor, William P. Davis,
William Dunson, John Grizzle and Carole McIvor for sample
collections. Walter Fitch made helpful comments on the
manuscript. Mark Mackiewicz developed primers for CR2,
which he kindly shared. The work was supported by funds
from the University of California at Irvine to J.C.A.

REFERENCES
Abbott, C. L., Double, M. C., Trueman, J. W. H., Robinson, A.
& Cockburn, A. 2005 An unusual source of apparent
mitochondrial heteroplasm: duplicate mitochondrial
control regions in Thalassarche albatrosses. Mol. Ecol. 14,
Arndt, A. & Smith, M. J. 1998 Mitochondrial gene
rearrangement in the sea cucumber genus Cucumaria.
Avise, J. C. 2000 Phylogeography: the history and formation of
Linkage disequilibrium and recombination in hominid
1126/science.286.5449.2524)
of mitochondrial DNA and its phylogenetic implications.
annurev.ecolsys.36.091704.175513)
Ballard, J. W. O. & Whitlock, M. C. 2004 The incomplete
(doi:10.1046/j.1365-294X.2003.02063.x)


