Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses)

Masaki Eda1*,†, Masaki Kuro-o2, Hiroyoshi Higuchi3, Hiroshi Hasegawa4 and Hiroko Koike1

1Graduate School of Social and Cultural Studies, Kyusyu University, Moto-oka 744, Nishi District, Fukuoka City, Fukuoka Prefecture, 819-0395, Japan
2Department of Biology, Hirosaki University, Bunkyo-cho 3, Hirosaki City, Aomori Prefecture, 036-8561, Japan
3School of Agriculture and Life Sciences, University of Tokyo, Yayoi 1-1-1, Bunkyo District, Tokyo, 113-8657, Japan
4Department of Biology, Toho University, Miyama 2-2-1, Funabashi City, Chiba Prefecture, 274-8510, Japan

(Received 17 October 2009, accepted 31 March 2010)

Although the tandem duplication of mitochondrial (mt) sequences, especially those of the control region (CR), has been detected in metazoan species, few studies have focused on the features of the duplicated sequence itself, such as the gene conversion rate, distribution patterns of the variation, and relative rates of evolution between the copies. To investigate the features of duplicated mt sequences, we partially sequenced the mt genome of 16 Phoebastria albatrosses belonging to three species (P. albatrus, P. nigripes, and P. immutabilis). More than 2,300 base pairs of tandemly-duplicated sequence were shared by all three species. The observed gene arrangement was shared in the three Phoebastria albatrosses and suggests that the duplication event occurred in the common ancestor of the three species. Most of the copies in each individual were identical or nearly identical, and were maintained through frequent gene conversions. By contrast, portions of CR domains I and III had different phylogenetic signals, suggesting that gene conversion had not occurred in those sections after the speciation of the three species. Several lines of data, including the heterogeneity of the rate of molecular evolution, nucleotide differences, and putative secondary structures, suggests that the two sequences in CR domain I are maintained through selection; however, additional studies into the mechanisms of gene conversion and mtDNA synthesis are required to confirm this hypothesis.

Key words: concerted evolution, control region, Diomedeidae, gene conversion, gene duplication, mitochondrial DNA

INTRODUCTION

Whereas the arrangement of genes in the metazoan mitochondrial (mt) genome is variable, its content is highly conserved (Boore, 1999). The tandem duplication of genes and subsequent (random) loss of one copy of each duplicate has been proposed as a mechanism of gene rearrangement in the mt genome (Moritz and Brown, 1987; Boore, 2000; Lavrov et al., 2002). This model predicts that one copy of each duplicated gene should disappear rapidly due to a loss of functionality caused by disruption of the secondary structure or anticodon (Moritz and Brown, 1986; Boore, 2000). In contrast, copies of small multigene families in the nuclear genome are maintained by gene conversion (Nielsen et al., 2003; Nei and Rooney, 2005). In this process, one sequence is replaced by another related sequence; however, the precise mechanism is unclear (Walsh, 1987; Innan, 2004). Gene conversion also appears to occur in duplicated mtDNA sequences. This is based on observations of mtDNAs containing two identical or nearly identical copies that are shared with sister species, and which are more similar to each other than to the other copy in a different species (Shao et al., 2005a). Sequence duplications in...
the mt genome frequently involve the control region (CR), which is associated with the initiation of transcription and mtDNA replication. Duplicated and concertedly evolved CRs have been reported in snakes (Kumazawa et al., 1996, 1998), sea cucumbers (Arndt and Smith, 1998), ticks (Black and Roebrdaanz, 1998; Campbell and Barker, 1999; Shao et al., 2005a), birds (Eberhard et al., 2001; Abbott et al., 2005; Jouventin et al., 2006; Gibb et al., 2007; Smith et al., 2007; Lawrence et al., 2008; Cho et al., 2009; Gomez-Diaz et al., 2009), fish (Lee et al., 2001; Tatarenkov and Avise, 2007), thrips (Shao et al., 2003), lizards (Kumazawa and Endo, 2004; Amer and Kumazawa, 2005), a sea firefly (Ogoh and Ohmiya, 2004), cephalopods (Yokobori et al., 2004), a frog (Sano et al., 2005), mites (Shao et al., 2005b, 2006), and turtles (Parham et al., 2006a, 2006b). Some of the duplications involve structural genes or tRNAs (Kumazawa et al., 1998; Campbell and Barker, 1999; Eberhard et al., 2001; Yokobori et al., 2004; Abbott et al., 2005; Shao et al., 2005b; Parham et al., 2006b; Gibb et al., 2007; Cho et al., 2009). Duplicated and concertedly evolved structural genes, or tRNAs without CR involvement, are rare but have been reported in a chameleon (Townsend and Larson, 2002) and a pentatomid (Lavrov et al., 2004).

Most of these data were obtained as an unexpected by-product of phylogenetic analysis or as a complication of phylogeographic analysis. Although Tatarenkov and Avise (2007) documented the rapid concertedly evolved duplication of sequences in a mtDNA CR, few studies have focused on the features of the duplicated sequences themselves, such as the gene conversion rate, distribution pattern, and relative rate of evolution between copies.

Diomedieidae (albatrosses) can provide insight into the fate of tandemly duplicated sequences in the mt genome. In addition to the usual contents of a metazoan mt genome (i.e., 13 proteins, two rRNAs, 22 tRNAs, and a CR), the mt genomes of Thalassarche cauta and T. melanophris were reported to have a tandemly duplicated sequence consisting of two partial sequences of cytochrome b (Cyt b; degenerated-Cyt b (d-Cyt b) and pseudo-Cyt b (p-Cyt b)), tRNA\textsubscript{Thr}, tRNA\textsubscript{Phe}, NADH dehydrogenase subunit 6 (ND6), tRNA\textsubscript{Glu}, and CR (Abbott et al., 2005; Gibb et al., 2007). In those studies, the copies of p-Cyt b, tRNA\textsubscript{Thr}, tRNA\textsubscript{Phe}, ND6, and tRNA\textsubscript{Glu} were identical or nearly identical, whereas d\_Cyt b was not. Four other Thalassarche albatrosses also had at least two CRs that could be divided into orthologous and paralogous sequences (Abbott et al., 2005).

In this study, we analyzed the mtDNA of Cyt b and 12S rRNA (12S) of three Phoebastria (another genus in Diomedieidae) species (P. albatrus, P. nigripes, and P. immutabilis) to investigate whether the genus also had duplicated sequences in the mt genome. In addition, we analyzed 16 duplicated CR sequences, determined the distribution pattern of variation, estimated the gene conversion rate, and tested the heterogeneity of the rate of molecular evolution among sequences. Finally, we examined gene order differences in the Diomedieidae, and we assessed the similarity and dissimilarity between and within CRs in Phoebastria.

**MATERIALS AND METHODS**

**DNA experiments** Five feathers from *P. albatrus* were collected from chicks on Torishima, an island of Izu island group, Japan (Table 1). The samples were also analyzed as part of a conservation genetic study of the species and selected to represent two distinct clades (Kuro-o et al., 2010). Muscle samples from three *P. immutabilis* individuals and one *P. nigripes* were collected in the northern Pacific and on Chichi-jima, one of the Bonin Islands. The root of each feather was sliced into small fragments with clean scissors. Whole genomic DNA was extracted from the *P. albatrus* feathers sampled from chicks using phenol-chloroform, as described by Kuro-o et al. (2010). An IsoQuick Nucleic Acid Extraction Kit (ORCA Research) was used to extract genomic DNA from the other samples, as described by Eda et al. (2008).

Duplication of the mt genes in the *Phoebastria* albatrosses was identified by amplification of the region between the Cyt b and 12S genes using three sets of primers: A (CytL246.dio and 12sR1), B (cbF1 and 12sR2), and C (Leon2.dio and Hcyt1.dio) (Abbott et al., 2005; Eda et al., 2006, 2008). If the region was duplicated as in the *Thalassarche* albatrosses (Abbott et al., 2005; Gibb et al., 2007), i.e., approximately 2,000 base pairs (bp) of a sequence duplication, we expected that the products for set A, B, and C would be approximately 5,700, 4,600, and 950 bp in length, respectively. If there was no sequence duplication in the region, we expected that the products to be approximately 3,700 bp (primer set A), 2,600 bp (set B), and not present (set C). Each amplification was carried out in a 25 μl reaction volume consisting of 1 μl of extracted DNA, 0.2 μM each nucleotide, 0.1 μM each primer, 0.2 mg/ml bovine serum albumin, and 1.25 U of ExTaq polymerase, hot-start version (Takara). The PCR program consisted of an initial denaturation at 94°C for 1 min, followed by 40 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 10 min (for sets A and B) or 60 s (for set C), with a final extension at 72°C for 5 min. The products were sequenced in both directions using a CEQ 2000 DNA Analysis System (Beckman Coulter) by primer walking, following purification of the template with a PCR Product Pre-sequencing Kit (USB). Using five partially overlapping primer sets (Set D, L14863 and H15149; Set E, CytL246.dio and Hcyt1.dio; and Set F, cbF1 and Hcyt2.dio, in addition to Sets B and C) and 16 primers (nine reported previously and seven newly designed), mtDNA sequences were determined from the 3' end of Cyt b through the 3' end of 12S, except for poly-Cs
MtDNA duplication in *Phoebastria* albatrosses

in the C-stretch of CR domain I and the repetitive sequences in CR domain III (Table 2 and Fig. 1a) (Kocher et al., 1989; Nunn et al., 1996; Abbott et al., 2005; Eda et al., 2006, 2008).

To further investigate the evolution of the duplicated CR sequences, two *P. albatrus* feathers were collected from carcasses on Torishima and Minami-kojima in the Senkaku Islands, one *P. immutabilis* feather was collected from a carcass on the northern Pacific coast of Japan, and four *P. nigripes* feathers were collected from chicks on Muko-jima in the Bonin Islands (Table 1). Whole genomic DNA was extracted using an IsoQuick

---

**Table 1.** Specimen list for this study. Species name, locality, sequenced region, and DDBJ/EMBL/GenBank Accession No. are included

<table>
<thead>
<tr>
<th>Specimen name</th>
<th>Species name</th>
<th>Locality</th>
<th>Sequenced region</th>
<th>DDBJ/EMBL/GenBank Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palb1</td>
<td><em>Phoebastria albatrus</em></td>
<td>Torishima Island</td>
<td>Whole of the duplicates</td>
<td>AB276044</td>
</tr>
<tr>
<td>Palb2</td>
<td>(Clade 1)</td>
<td>Torishima Island</td>
<td>Whole of the duplicates</td>
<td>AB276045</td>
</tr>
<tr>
<td>Palb3</td>
<td></td>
<td>Torishima Island</td>
<td>Whole of the duplicates</td>
<td>AB276046</td>
</tr>
<tr>
<td>Palb4</td>
<td></td>
<td>Torishima Island</td>
<td>Control regions</td>
<td>AB276044</td>
</tr>
<tr>
<td>Palb5</td>
<td><em>P. albatrus</em> (Clade 2)</td>
<td>Torishima Island</td>
<td>Whole of the duplicates</td>
<td>AB276047</td>
</tr>
<tr>
<td>Palb6</td>
<td></td>
<td>Torishima Island</td>
<td>Whole of the duplicates</td>
<td>AB276047</td>
</tr>
<tr>
<td>Palb7</td>
<td></td>
<td>Torishima Island, Senkaku Islands</td>
<td>Control regions</td>
<td>AB276052, AB276053</td>
</tr>
<tr>
<td>Pimu1</td>
<td><em>P. immutabilis</em></td>
<td>Northern Pacific</td>
<td>Whole of the duplicates</td>
<td>AB276048</td>
</tr>
<tr>
<td>Pimu2</td>
<td></td>
<td>Northern Pacific</td>
<td>Whole of the duplicates</td>
<td>AB276049</td>
</tr>
<tr>
<td>Pimu3</td>
<td></td>
<td>Northern Pacific</td>
<td>Whole of the duplicates</td>
<td>AB276050</td>
</tr>
<tr>
<td>Pimu4</td>
<td></td>
<td>Northern Pacific</td>
<td>Control regions</td>
<td>AB276054, AB276055</td>
</tr>
<tr>
<td>Pnig1</td>
<td><em>P. nigripes</em></td>
<td>Chichi-jima Island, Bonin Islands</td>
<td>Whole of the duplicates</td>
<td>AB276051</td>
</tr>
<tr>
<td>Pnig2</td>
<td></td>
<td>Muko-jima Island, Bonin Islands</td>
<td>Control regions</td>
<td>AB276056, AB276057</td>
</tr>
<tr>
<td>Pnig3</td>
<td></td>
<td>Muko-jima Island, Bonin Islands</td>
<td>Control regions</td>
<td>AB276058, AB276059</td>
</tr>
<tr>
<td>Pnig4</td>
<td></td>
<td>Muko-jima Island, Bonin Islands</td>
<td>Control regions</td>
<td>AB276060, AB276061</td>
</tr>
<tr>
<td>Pnig5</td>
<td></td>
<td>Muko-jima Island, Bonin Islands</td>
<td>Control regions</td>
<td>AB276062, AB276063</td>
</tr>
</tbody>
</table>

**Table 2.** Primers used in this study. The primers used for sequencing and primers with two annealing sites due to duplications are shown in a and b, respectively. Abbreviations: *ND5* and *6*, *NADH* dehydrogenase subunits 5 and 6; *Cyt b*, cytochrome b; *CR*, control region; *12S*, *12S* ribosomal RNA

<table>
<thead>
<tr>
<th>Primer</th>
<th>Location</th>
<th>No. in Fig. 1</th>
<th>Sequence 5'-3'</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>L14863</td>
<td><em>ND5</em></td>
<td>1</td>
<td>TTTGCCCTATCTATCTCCCAT</td>
<td>Nunn et al. (1996)</td>
</tr>
<tr>
<td>Lcyt246.dio</td>
<td>Cyt b</td>
<td>2</td>
<td>CCTCCACGC3AAACGGAG</td>
<td>Eda et al. (2006)</td>
</tr>
<tr>
<td>H15149</td>
<td>Cyt b</td>
<td>3</td>
<td>GCCCCTCCAGAATGATTTTGCTCTCA</td>
<td>Kocher et al. (1989)</td>
</tr>
<tr>
<td>Hcyt2.dio</td>
<td>Cyt b</td>
<td>4</td>
<td>GAGCCCGATTCGTTGGAAGAGG</td>
<td>Eda et al. (2008)</td>
</tr>
<tr>
<td>cbF1b</td>
<td>Cyt b</td>
<td>5</td>
<td>ATGAATCGGCAGCCAACCCTAG</td>
<td>Abbott et al. (2003)</td>
</tr>
<tr>
<td>Hcyt1.dio</td>
<td>Cyt b</td>
<td>6</td>
<td>TAAGTATTTTGTCTTAGGGCTG</td>
<td>Eda et al. (2008)</td>
</tr>
<tr>
<td>LND6.dio</td>
<td><em>ND6</em></td>
<td>7</td>
<td>CAGCTTCCCAAGCCCTGAC</td>
<td>This study</td>
</tr>
<tr>
<td>LGlu.dio</td>
<td><em>tRNA</em></td>
<td>8</td>
<td>CCTGAAAAAACCAGCGCTTGGTCTT</td>
<td>This study</td>
</tr>
<tr>
<td>HGlut1.dio</td>
<td><em>tRNA</em></td>
<td>9</td>
<td>TAGTTGAAGTCACGGCGCGG</td>
<td>This study</td>
</tr>
<tr>
<td>Leon2.dio</td>
<td>CR</td>
<td>10</td>
<td>TAGGAAAAACATTTATGCTGACGGACGAC</td>
<td>This study</td>
</tr>
<tr>
<td>H454.grb</td>
<td>CR</td>
<td>11</td>
<td>GCCCTGACATAGGAAACCAAGG</td>
<td>Baba et al. (2005)</td>
</tr>
<tr>
<td>Leon4.dio</td>
<td>CR</td>
<td>12</td>
<td>CTAAGACTTGAGCATAACATGGCGGTTGC</td>
<td>This study</td>
</tr>
<tr>
<td>Heo4d.io</td>
<td>CR</td>
<td>13</td>
<td>GCGCCATAATTATGCACATATGCGG</td>
<td>This study</td>
</tr>
<tr>
<td>Heo5.dio</td>
<td>CR</td>
<td>14</td>
<td>GCCTAGGGGAGTTTCTTTAATGGG</td>
<td>This study</td>
</tr>
<tr>
<td>12sR1</td>
<td><em>12S</em></td>
<td>16</td>
<td>TAAAGCCTTTGCTCGTAGTTTC</td>
<td>Abbott et al. (2005)</td>
</tr>
<tr>
<td>12sR2</td>
<td><em>12S</em></td>
<td>15</td>
<td>GCCTAACTAAGTCAAGCTTCG</td>
<td>Abbott et al. (2003)</td>
</tr>
</tbody>
</table>
Nucleic Acid Extraction Kit, and two PCR fragments (for CR1, LGlu.dio and Hcyt2.dio, Set G; for CR2 of *P. albatrus*, LGlu.dio and Hcon4.dio, Set H; for CR2 of *P. nigripes* and *P. immutabilis*, LGlu.dio and Hcon5.dio, Set I) were used to determine each of the duplicate CRs after the C-stretch in domain I. All sequences were aligned with *Thalassarche* sequences (DDBJ/EMBL/GenBank accession nos. AY158677 and DQ029001-DQ029010) using Clustal-X in MEGA 4.0 (Tamura et al., 2007). The genes were identified by sequence similarity with *T. melanophris*. All of the sequences have been deposited in DDBJ/EMBL/GenBank under accession nos. AB276044-AB276063.

**Data analysis** To detect heterogeneity in the polymorphic pattern along the CR sequence, RSW 1.1 (Trueeman, 1998) was used to examine each species or distinct clade of *P. albatrus*. In the analysis, groups with bootstrap scores exceeding 70% were set to indicate potential phylogenetic signals in the sequences (details in Supplementary Methods online). Because two types of polymorphic pattern that separated the CR sequence into four sections (Sections A–D) were detected, additional analyses were conducted for each section. MEGA 4.0 was used to generate the most parsimonious (MP) trees using equal-weighted parsimony with close neighbour-interchange searches. The reliability of each specific grouping was assessed using 1,000 bootstrap replicates. Gene conversion and the mutation rate parameters (*C* = 2*Nc* and *θ* = 2*Nμ*, where *N*, *c*, and *μ* are the population size, gene conversion rate per site per generation, and mutation rate per site per generation, respectively) were estimated using the following formulae from Innan (2004):
MtDNA duplication in Phoebastria albatrosses

\[ C = \frac{(\pi_w - 2D_{nm})}{2(\pi_w - \pi_w)} \]

\[ \theta = \frac{(\pi_w + 2D_{nm})}{2L} \]

where \(\pi_w\), \(\pi_n\), \(D_{nm}\), and \(L\) are the average numbers of pairwise nucleotide differences (nucleotide diversity) within each copy, average number of pairwise nucleotide differences between two copies, sum of the adjusted level of linkage disequilibrium, and total number of nucleotides in each copy, respectively. R 2.9.1 (R-Development-Core-Team, 2009) was used to calculate 95% confidence intervals for the estimated parameters using bootstrap replicates.

We tested the heterogeneity of the rate of molecular evolution in Section B of the CR1 and CR2 sequences in two ways. First, to test the null hypothesis that sequence divergence over the entire tree can be described using a global molecular clock, the likelihood ratio test was performed with MolecularClock.bf in HyPhy (Pond et al., 2005). In the analysis, significance of the test was given by its asymptotic \(P\) value. HKY85 model was selected by Model-Test (Posada and Crandall, 1998) and assumed in the analysis. Second, to evaluate whether one copy of the CR Section B evolves more quickly than the other, we calculated the number of substitutions between each of the CRs for each of the Phoebastria samples and each of the CRs for T. cauta (outgroup) using MEGA 4.0. Then the number of sites for CR1 and CR2 were compared using the Wilcoxon signed-rank test with the continuity correction to determine whether one of the CR sequences tended to yield more nucleotide substitutions. The tests were performed with R 2.9.1.

RESULTS AND DISCUSSION

For each of the samples, the length of the obtained product was longer than 6,000 bp for primer set A, approximately 3,000 bp for set B, and approximately 1,250 bp for set C. The length of each product differed from those expected in Thalassarche albatrosses (Abbott et al., 2005; Grib et al., 2007), i.e., approximately 5,700 bp for set A, approximately 4,600 bp for set B, and approximately 950 bp for set C, or those expected with no sequence duplication, i.e., approximately 3,700 bp for set A, approximately 2,600 bp for set B, and not present for set C. This suggests that Phoebastria albatrosses have duplicated sequences, but that the content of duplication is more or less distinct from that of Thalassarche albatrosses.

When the mtDNA sequence from the 3' end of Cyt b through 12S was examined for the three species, we found approximately 2,350 bp of tandemly duplicated sequence between the first CR (CR1) and tRNA\(^{Thr}\) (Fig. 1a). The duplicated sequence for the species began with a partial copy of Cyt b (Cyt b-2; the last 549 bp) and was followed by copies of tRNA\(^{Thr}\), tRNA\(^{Pro}\), ND6, tRNA\(^{Glu}\), and the CR. Therefore, the gene order for the three species was Cyt b-1, tRNA\(^{Thr}\)-1, tRNA\(^{Pro}\)-1, ND6-1, tRNA\(^{Glu}\)-1, CR1, Cyt b-2, tRNA\(^{Thr}\)-2, tRNA\(^{Pro}\)-2, ND6-2, tRNA\(^{Glu}\)-2, CR2, tRNA\(^{Glu}\), and 12S. In the CR sequences of the products produced using primer set A, a number of double peaks were consistently observed. This would arise from products included two CRs and the two CRs were not identical at some sites. Primer set B did not reveal any sign of sequence duplication in the Phoebastria albatrosses. This would arise from annealing of primer cbF1 into Cyt b-2.

The three Phoebastria species shared the same tandem duplication from part of Cyt b to part of the CR. This suggests that the duplication occurred in the common ancestor of the three Phoebastria species. However, the duplicate partial copy of Cyt b through tRNA\(^{Glu}\), or more than 1,300 bp, within each individual was identical, with the exception of one P. albatrus that had a single nucleotide difference in tRNA\(^{Glu}\). The duplicates include identical partial Cyt b sequences that lack an initiation codon and that are therefore functionless, even though non-functional copies are expected to degenerate and disappear rapidly under the tandem duplication and random loss model (Moritz and Brown, 1986; Boore, 2000). These findings suggest that gene conversions have occurred multiple times and maintained two identical or nearly identical copies, in accordance with previous studies (Abbott et al., 2005; Shao et al., 2005a).

Although the duplicated content in the Cyt b sequence in Phoebastria is different from Thalassarche, the duplication might have originated in a common ancestor of the two genera. In addition to the usual Cyt b sequence, Phoebastria albatrosses have a partial copy of the Cyt b (Cyt b-2), while T. cauta has two partial copies of the Cyt b sequence: d- and p-Cyt b (Abbott et al., 2005). Large insertions and/or deletions (indels) terminate concerted evolution by gene conversion, since indels prevent a gene conversion from being initiated and extended (Walsh, 1987; Teshima and Innan, 2004). Large deletion(s) could occur in T. cauta between d- and p-Cyt b and prevent the gene conversion of d-Cyt b. As a result, d-Cyt b degenerated under the neutral selection model. By contrast, no large indels seemed to have occurred in Cyt b-2 in Phoebastria albatrosses. Consequently, gene conversion has been maintained Cyt b-2 as identical to Cyt b-1.

The two CRs in the three Phoebastria species were not identical in sequence; however, the first 1,024 bp were readily aligned. In the domain I, both CRs included C-stretch and termination-associated sequences (TAS) which are involved in the termination of DNA synthesis (Figs. 1b and 2) (Doda et al., 1981; Sbisa et al., 1997). Also included were F-, D-, B- and Bird similarity boxes in the domain II and the conserved-sequence block-1 (CSB-1) located near the origin of heavy-strand replication, which is involved in the transition from RNA transcription to DNA replication (Ruokonen and Kvist, 2002), in
domain III. After that, however, they were completely different. CR1 had an additional 15 bp and was followed by Cyt b-2 while CR2 had several hundred additional nucleotides composed primarily of repetitive sequences.

Two types of polymorphic site were detected in the three Phoebastria species using RSW (Supplementary Figs. 1–5). For each taxon, CR1 and CR2 within an individual were mostly different at one type of polymorphic site, while they were mostly identical at the other type of polymorphic site (Fig. 2). The former type of polymorphic site was concentrated in the middle part of domain I (Section B) and the last part of the duplicated domain III (Section D), and was separated from the latter type of polymorphic site (Section C). Because the duplicated
sequences within each individual had evolved in concert from the partial Cyt b to the first part of the CR domain I as well as Section C, we named this section Section A, although no informative sites were found in the CR sequence of this section in each taxon. The length of each section differed among the taxa (Table 3, Fig. 1b). Sections B and C would correspond to Sections A (orthologous portion) and B (paralogous portion) in Thalassarche albatrosses, respectively (Abbott et al., 2005). Opposite evolutionary patterns between Sections A and
C and Sections B and D were clearly shown in the MP trees (Fig. 3). In the trees for Sections A (data not shown) and C (Fig. 3b), the duplicate sequences within the individuals were more closely related to one another than to other orthologous sequences. In the Section C tree, monophyletic relationships among the sequences for *P. albatrus* (85%), *P. nigripes* (98%), *P. immutabilis* (66%), and the *Phoebastria* albatrosses (99%) were relatively well supported by bootstrap resampling, although reciprocal monophyly of the two clades in *P. albatrus* was not supported in the topology. When duplicate sequences from the *Thalassarche* che albatrosses, except *T. melanophris*, were removed and longer sequences were used to construct the Section C tree, this pattern was consistent, and the reciprocal monophyly of the two clades in *P. albatrus* was supported (data not shown). In the Section A tree, the monophyly of each species and a distinct clade in *P. albatrus* was supported by more than 99 and 80%, respectively.

By contrast, in the trees for Section B (Fig. 3a) and D (data not shown), orthologous sequences were more closely related to one another than to duplicate sequences within individuals. In the Section B tree, the monophyletic relationship of the *Phoebastria* and *Thalassarche* CRs (58%), *Phoebastria* CR2 (66%), and *Thalassarche* CR1 and CR2 (81 and 98%, respectively) were relatively well supported, while *Phoebastria* CR1 was not monophyletic. Within *Phoebastria* CR2, a monophyletic relationship for *P. nigripes* and *P. immutabilis* was supported by bootstrap resampling (89 and 97%, respectively), whereas the monophyly of *P. albatrus* was not fully supported (49%) and two distinct clades within *P. albatrus* were not reciprocally monophyletic. Within *Phoebastria* CR1, a monophyletic relationship for each species was not supported. In the Section D tree, the monophyly of CR1 and CR2 was relatively well supported by bootstrap resampling (93 and 59%, respectively), although monophyly was not indicated in each species.

The differences in the MP trees between Sections A and C and Sections B and D suggest that the former are under concerted evolution while the latter are not. The MP trees for Sections A and C indicate that gene conversion in these sections is so frequent that most of the latest substitutions are shared between the copies. By contrast, the MP trees for Sections B and D indicate that the last gene conversion occurred after divergence from the common ancestor of the *Thalassarche* (i.e., the common ancestor for the three *Phoebastria* species studied). The estimated rate of gene conversion emphasizes the differences between Sections A and C and Sections B and D. Considering 95% confidence intervals, the rates for the former and latter were, respectively, more than and
less than an order of $1.0 \times 10^{-2}$, except for *P. immutabilis* Section A (0.000–∞) in which only two polymorphic sites were present (Table 3). The observed gene conversion rate for Sections A and C suggest that duplicated sequences in mtDNA might not disappear, even if they lose functionality due to disruption of the secondary structure or anticodon.

What makes the gene conversion rates and evolutionary patterns different between Sections A and C and Sections B and D? Large indels, which terminate gene conversion, are a possible explanation; however, large indels were not detected near those sections. The accumulation of neutral point mutations and natural selection resulting in neo- or sub-functionalisation has been proposed as other hypotheses (Walsh, 1987; Lynch and Force, 2000; Innan, 2003; Teshima and Innan, 2004). Unfortunately, it is unclear what degree and type of sequence similarity are required for successful gene conversion. The frequency of homologous recombination in prokaryotes has been suggested to be linearly related to the degree of sequence similarity (Walsh, 1987). Conjugational recombination between genomes that are 20% divergent is essentially nonexistent under natural conditions (Rayssiguier et al., 1989). Assuming these figures to be a limit for homologous recombination and gene conversion, sequence differences within Section D for each individual (30.0–54.5%) may be sufficient to prevent successful gene conversion. On the other hand, the sequence differences within Section B (8.0–12.0%) are not likely to be high enough.

The selection hypothesis calls for one type of sequence to be selected over another in different copies. Wilcoxon signed-rank tests and the branch lengths in the Section B tree clearly indicate that CR1 is evolving more slowly than CR2 (using *T. cauta* CR1 as the outgroup, W* = 91, P < 0.005; using *T. cauta* CR2 as the outgroup, W* = 136, P < 0.001). In addition, the tree as a whole does not obey the global molecular clock ($2.0L = 145.667$, $P < 0.0001$). By contrast, the mutation rates for the CRs were predicted to be similar, since nucleotide diversity was similar in each copy (for CR1, 0.8–2.7; for CR2, 0.4–3.3, Table 3). These data suggest that the point mutations occurring in CR1 and CR2 are not neutral, and that CR1 is under stronger constraint than CR2. It is important to note that CR1 sequences have not replaced CR2s through gene conversion since at least the time of the common ancestor of the *Phoebastria*, despite the frequent gene conversion occurring outside Section B. This suggests that CR2 is also under constraint and that some functional differences exist between the two CRs.

A possible explanation for the functional difference between the two CRs might be the termination of mtDNA synthesis. Section B includes TAS elements (Doda et al., 1981). Sequence differences between the CRs led to different secondary structures, and the structure of CR1 appears to be much more conservative than that of CR2 (data not shown). This difference may halt DNA synthesis at CR1 but not at CR2. In other words, replacing the CR2 sequence with CR1 through gene conversion may stop DNA synthesis at the sequence portion that was originally CR2, and this may be disadvantageous for mtDNA synthesis. TAS elements show a variable degree of sequence conservation (Foran et al., 1988), but the specific mechanism through which they halt DNA synthesis is unknown (Doda et al., 1981). Additional studies on the mechanisms of gene conversion and mtDNA synthesis are required to confirm the selection hypothesis.

Kazuto Kawakami provided the Black-footed Albatross samples from the Bonin Islands. Hidenori Tachida, Hideki Innan, Shin Nishida, and two anonymous reviewers clarified the strengths and weaknesses of this study. Ginger Clark, Martin Lovatt, and Raoul Weston corrected the English draft. This study was supported financially in part by a Grant-in-Aid for JSPS Fellows to ME from the Japan Society for the Promotion of Science (No. 16-6316).

**REFERENCES**


MtDNA duplication in *Phoebastria* albatrosses


