DNA Sequence, October/December 2004 Vol. 15 (5/6), pp. 374–377

Taylor & Francis
healthsciences

Short Communication

Solving a Sequencing Problem in the Vertebrate Mitochondrial Control Region using Phylogenetic Comparisons

JULIE FEINSTEIN[a,*] and JOEL CRACRAFT[b]

[a]*Ambrose Monell Collection for Molecular and Microbial Research, Department of Invertebrate Zoology, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, USA;* [b]*Department of Ornithology, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, USA*

(Received 16 April 2004)

The mitochondrial control region (mtCR) of the bird-of-paradise, *Phonygammus keraudrenii*, the Trumpet Manucode, contains a unique arrangement of homopolymers and short tandem repeats. Homopolymers occur within a few hundred bases of each other, trapping sequence information between unsequenceable barriers. A comparative strategy, involving other manucode species, allowed the prediction of primer sites in the inaccessible region. The method is suggested for similar sequencing problems.

Keywords: *Phonygammus keraudrenii*; Manucodes; Bird-of-paradise; Avian mitochondrial control region; Homopolymers

Phonygammus (Manucodia) keraudrenii, the Trumpet Manucode, is a relatively drab member of the family Paradisaeidae, the birds-of-paradise. A glossy crow-like bird, *P. keraudrenii* lives in the forests of New Guinea and surrounding islands and on the Cape York Peninsula in Australia. The bird is rarely seen, keeping mainly to the canopy and feeding at fig trees in the forest interior. It is more often heard, and is especially noteworthy for the vocalizing mechanism or "trumpet" formed by a distinctively modified trachea more than three times the length expected in a bird of its size. The long trachea is coiled and lies just below the skin above the breastbone. It is thought to enhance the projection of mating and contact calls. *P. keraudrenii* is the sister taxon to four birds in the genus *Manucodia*, where it has been placed by some authors (Frith and Beehler, 1998). Others recognize multiple geographic races of *P. keraudrenii*, the sole species in the genus *Phonygammus*. Cracraft (1992) described several distinct phylogenetic species among these forms. It is an essential organism in any effort to elucidate taxonomic relationships within the birds-of-paradise.

Accordingly, *P. keraudrenii* was included in a molecular study of the systematics of the Paradisaeidae undertaken by our laboratory. Various nuclear and mitochondrial loci were targeted, including the mitochondrial control region (mtCR). The mtCR is a regulatory locus of variable length, commonly about 1500 base pairs (bp) in vertebrates. Certain features of its primary structure recommend it for phylogenies of closely related species and for population studies (Baker and Marshall, 1997). Highly conserved sequence regions, allowing for the placement of universal primers, alternate with rapidly evolving regions where phylogenetically informative variation may occur. From an experimental perspective, the mitochondrial origin of the locus assures high copy numbers per cell and consequently easy amplification for direct PCR sequencing. However, some sequencing problems are occasionally encountered. A cytosine homopolymer region may occur close to the 5′ end of the locus, possibly performing a structural function related to the formation of a hairpin loop. With present fluorescent sequencing technology, homopolymers approaching 10 cytosine/guanine (CG) residues, or

*Corresponding author. Tel.: +1-212-769-5663. Fax: +1-212-496-3380. E-mail: jfstein@amnh.org

ISSN 1042-5179 print/ISSN 1029-2365 online
DOI: 10.1080/10425170400012909

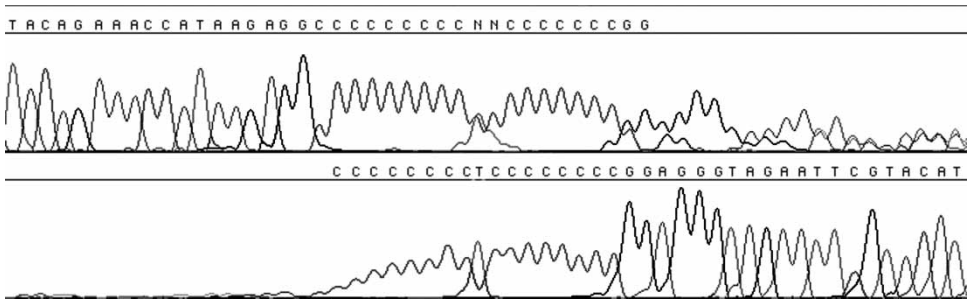

FIGURE 1 Fluorescence chromatograms showing the characteristic homopolymer found at the 5′ end of the avian Control Region. The upper trace was obtained using an L-strand (forward) primer, the lower trace with the H-strand (reverse) primer. The sequence fails after the homopolymer, in both directions. This example is from the mtCR of the Common Tern, *Sterna hirundo* (GenBank Accession # AY597018).

20 adenosine/thymine (AT) residues can cause sequence failure immediately after the homopolymer, probably because of polymerase slippage. Figure 1 illustrates this phenomenon. Obtaining bi-directional double-strand DNA sequences is not possible at sites with these primary structures. Additionally, an AT-rich region exists near the 3′ end of avian mtCRs. It may contain a thymine homopolymer of 16 bases or more. The AT-rich region is sometimes followed by short tandem repeats, typically of CA motifs, precluding the placement of primers. Accurate single-strand reads can be obtained by placing primers on both sides of a homopolymer and sequencing up to and across it, approaching from both sides. The sequence generated in this manner typically shows the same number of nucleotide residues in the homopolymer when read from either direction and often, though not always, is clearly readable up to one base after the homopolymer. Multiple replicate reads across the homopolymer, using different fluorescent chemistries, are sometimes needed to give a clear picture of the sequence. This situation becomes much more difficult when two homopolymers occur in close proximity.

During the sequencing of the mtCR of *P. keraudrenii*, we found an unfortunate arrangement of nucleotides that included all of the aforementioned mtCR problems. We found a 5′ cytosine homopolymer (Fig. 2(b), base 13), a thymine homopolymer in the AT-rich region (Fig. 2(b), base 953), and short tandem repeats at the 3′ end (Fig. 2(b), base 1262–base 1338). The 5′ cytosine homoploymer was interrupted by two interspersed thymine residues, and did not present a sequencing problem. However, another unusual cytosine homopolymer was found (Fig. 2(b), base 1240) between the thymine homopolymer and the repeat region. This configuration

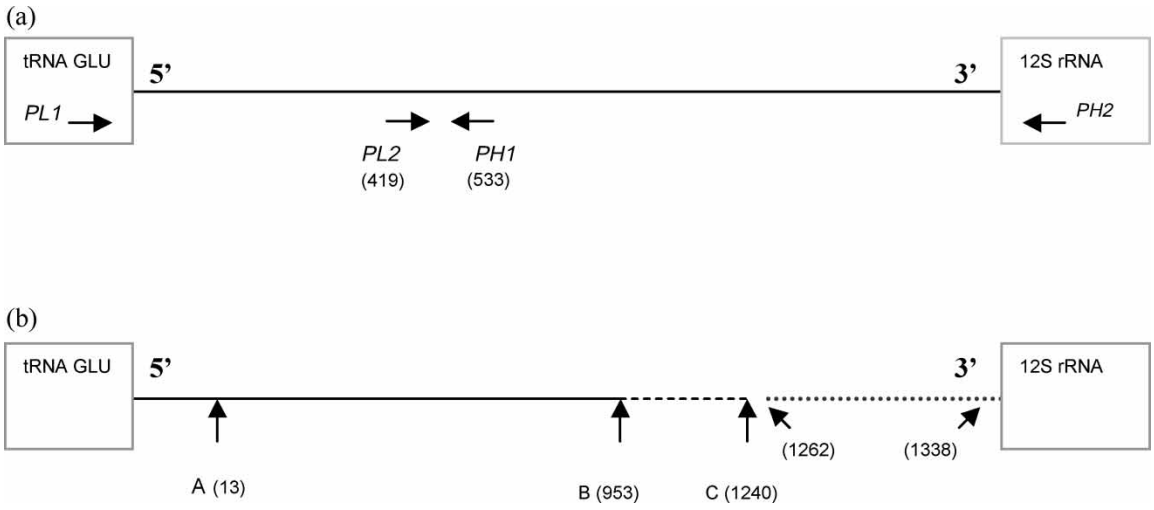

FIGURE 2 (a) Standard primer strategy for sequencing the mtCR of birds-of-paradise. PL1 (5′- CCAAGATCTACGGCCTGAAAAGCC-3′) is paired with PH1 (5′-TGAAATAGGAACCAGAGG-3′), PL2 (5′-AGCCAGAGAACCTGGTTATC-3′) is paired with PH2 (5′-GGTACCATCTTGGATCTTCAGTG-3′). Numbers in parentheses indicate position of 3′ base of the primer in the sequence of *M. atra*. (b) Features in the mtCR of *P. keraudrenii*. Base sequences are reported for individuals #1 and #2. Numbers in parentheses indicate base positions in the completed sequence of individual #2: A. cytosine homopolymer at position 13, $^{\#1\&\#2}$CCCCCCCCTTCCCCCCCC, B. thymine homopolymer (base 953), $^{\#1}$TTTTTTTTTTTTTT, $^{\#2}$TTTTTTTTTTTTTTTT, C. cytosine homopolymer (base 1240), $^{\#1}$CCCCCCCCCCTTACCC, $^{\#2}$CCCCCCCCTTACCCCC. Between B and C (dashed line) is a 300 bp gap, left by the standard sequencing strategy. From 1262 to 1338 (dotted line), are short tandem repeats. In *P. keraudrenii* the repeat sequence is: $^{\#1}$AAAAA (1 ×) CAAAAA (12 ×), $^{\#2}$CAAAAAAA (1 ×) CAAAAA (6 ×) CAAAAAAAAA (1 ×) CAAAAA (3 ×).

resulted in a rare arrangement in which closely spaced homopolymers formed a barrier around an internally bracketed sequence region (Fig. 2(b), dashed line from B to C). We usually sequence the mtCR with two sets of universal primers (Fig. 2(a)) positioned in conserved regions of the tRNA glutamine, the mtCR and the 12S rRNA gene. For the mtCR of *P. keraudrenii*, this primer strategy approaches the homopolymer at position 953 from upstream, and the homopolymer at position 1240 from downstream and results in sequence failure at the homopolymers from both directions. Comparison with the sequences of other birds-of-paradise suggested that the gap was about 300 bp long. Loci at which DNA sequence information is trapped between closely spaced unsequenceable barriers has been seen in our laboratory only twice in eight years. During this time, we sequenced numerous individuals of hundreds of avian taxa at many loci. The condition seems, therefore, to be quite rare.

To access the area between the homopolymers in the mtCR of *P. keraudrenii*, we used information from closely related species to predict priming sites inside the unknown area. Of the four other living manucodes, one, *Manucodia jobiensis*, the Jobi Manucode, is so rare that only old museum study skins are available. Tissue samples for DNA extraction were obtained for the other three: *M. atra*, the Glossy-mantled Manucode, *M. chalybata*, the Crinkle-collared Manucode, and *M. comrii*, the Curl-crested Manucode. DNA sequences were generated for these species, as described in Fig. 3, and deposited in GenBank: *M. atra*, Accession #

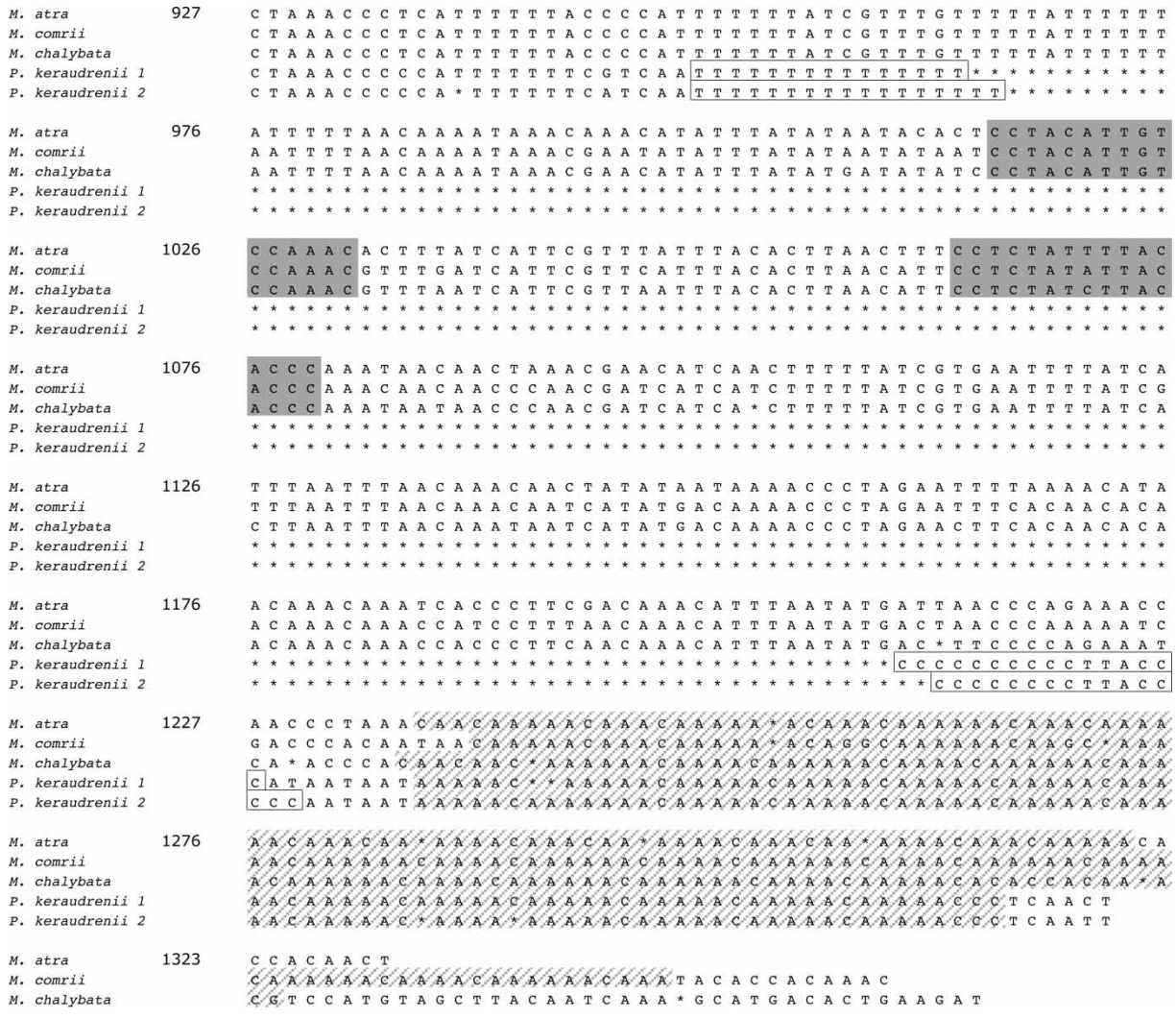

FIGURE 3 Partial alignment of mtCR sequences of three *Manucodia* species with fragments of two *P. keraudrenii* individuals. Sequences were generated by direct sequencing of PCR products amplified with the primers in Fig. 2(a), following standard protocols. Sequencing was done on an ABI 377 automated fluorescence sequencer using dRhodamine chemistry (ABI, Foster City, CA) following the manufacturer's recommendations. Sequences were edited with Sequencher (GeneCodes, Ann Arbor, MI). The alignment begins at base 927 of *M. atra*. Homopolymers are shown in boxes. Sequence regions conserved among *M. atra, M. comrii* and *M. chalybata* are shown with dark shading. A forward and a reverse primer were designed in each dark-shaded area: PKL 3 and 4 (forward primers) and PKH1 and 2 (reverse primers). Repeat regions are shown with light diagonal shading.

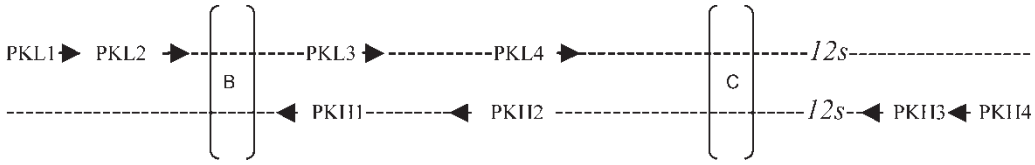

FIGURE 4 PCR primer strategy developed to amplify the sequence region trapped between homopolymers B (as is Fig. 2(b)), at position 953 of *P. keraudrenii* #2, and C (as in Fig. 2(b)), at position 1240. PKL1 (5′-CGCAAACTTGACACTGATGC-3′), PKL2 (5′-AGGAATTGCTACCTAAACCC-3′), PKH3 (5′-CCGTCTTGACATCTTCAGTG-3′) and PKH4 (5′-TGTTTGTAGCAGCCGTCTTG-3′) were designed from three *Manucodia* species and two *P. keraudrenii* individuals. PKL3 (5′-CCTACATTGTCCAAAC-3′), PKL4 (5′-CCTCTATHTTACACCC-3′), PKH1 (5′-GTTTGGACAATGTAGG-3′) and PKH2 (5′-GGGTGTAADATAGAGG-3′) were based on three *Manucodia* species only. Primer pairings (see text) allowed for multiple overlapping single-strand reads approaching from both sides of each homopolymer.

AY597014, *M. chalybata*, Accession # AY597012, *M. comrii*, Accession # AY597013.

Complete mtCR sequences of the three *Manucodia* species were aligned with incomplete sequence fragments of *P. keraudrenii*. Figure 3 shows the sequence gap in two individuals of *P. keraudrenii*, bridged by sequences of the three *Manucodia* species. Areas of conserved nucleotide sequences among the *Manucodia* species were found within the 300 bp sequence gap (Fig. 3). Areas shared by *P. keraudrenii* with the *Manucodia* species were found outside the sequence gap (not shown). Primers (Fig. 4) were designed from the conserved sequence areas. PCR products across the homopolymers were successfully amplified and sequenced from *P. keraudrenii* #2 with the primer combinations: PKL1 with PKH1 and PKH2, PKL2 with PKH1 and PKH2, PKL3 with PKH3 and PKH4, PKL4 with PKH3 and PKH4. The permutation of primer pairings allowed multiple overlapping reads, albeit in only one direction on each side of each homopolymer. In all cases, the sequence was cleanly readable up to the homopolymer, and collapsed or became unusable immediately after. Sequence results from two individuals of *P. keraudrenii* are reported. Initially, a very small sample of individual #1, was obtained. This sample was entirely consumed during the project and could not be replaced. A second individual #2, was used to complete the study. The partial sequence of #1 was used in the alignment that helped solve the sequencing problem under discussion. The section of #1 that is bracketed by homopolymers was not sequenced. The partial sequences of *P. keraudrenii* #1 were deposited in GenBank, Accession # AY597015 and # AY597016, along with the complete sequence of *P. keraudrenii* #2, GenBank, Accession # AY597017.

The close placement of homopolymers is rare, but when it does occur it poses a difficult problem of inaccessible information trapped between unsequenceable barriers. The method used here allowed us to complete the sequencing of a difficult locus for a taxon essential to our study. It is applicable to any similar sequencing problem and requires only the existence of closely related species from which speculative primers could be designed.

References

Baker, A.J. and Marshall, D. (1997) "Mitochondrial control region sequences as tools for understanding evolution", In: Mindell, D.P., ed., *Avian Molecular Evolution and Systematics* (Academic Press, London), pp 51–79.

Cracraft, J. (1992) "The species of the birds-of-paradise (Paradisaeidae): applying the phylogenetic species concept to a complex pattern of diversification", *Cladistics* **8**, 1–43.

Frith, C.B. and Beehler, B.M. (1998) *The Birds of Paradise* (Oxford University Press, Oxford).