Avian mtDNA primers

Michael D. Sorenson, Boston University
msoren@bu.edu

2/14/2003

This document provides a set of primer sequences for amplifying and sequencing the entire avian mitochondrial DNA. Most of the primers were originally developed as part of earlier studies (see refs) but many have since been revised or replaced with primers in different locations. All of the primers have been designed to work with most or all birds and most perform very well in this regard.

This document is intended as a resource for those interested in using mtDNA for avian population genetics and systematics, whether that work involves sequencing complete mitochondrial genomes or single genes. A second goal is to encourage the use of genes other than cytochrome *b*, which is a poor choice for studies in which only one mitochondrial gene will be sequenced.

If I'm going to sequence only one gene, which one should I choose?

ND2!

Why not cyt *b*?

All mitochondrial protein genes have similarly high rates of substitution at 3rd codon positions, but they vary greatly in the rate of amino acid substitution. Therefore, all mt protein genes provide similar information (per base pair sequenced) for very closely related taxa (e.g., individuals within a species) that differ primarily at degenerate 3rd codon positions. For somewhat more distantly related taxa (including species within a genus, for example, and everything else up to and including comparisons between avian orders), more variable mt genes accumulate more informative variation at 1st and 2nd codon positions, whereas 3rd positions increasingly accumulate multiple substitutions (i.e., homoplasy) in all genes. **With cyt *b*, you get all of the noise and none of the signal! — with other mt genes, you get a little signal along with the noise.**

Why ND2?

In terms of amino acid sequence, ND2 is the 3rd most variable gene after ATPase 8, which is very short (~165-168 bp) and therefore provides relatively little information, and ND6, which is also relatively short (~519-522 bp) and is generally more difficult to amplify and sequence, given its unusual base composition and location near the control region. In contrast, the complete ND2 gene can be amplified in either one or two pieces with primers that have worked well on essentially all birds (L5216-H6313 for the whole gene; L5216-H5766 and L5758-H6313 for two pieces). We routinely amplify the gene in two pieces and run four sequencing reactions (both strands of each PCR product).

What are all those Y's and R's in the primer sequences?

Many of the primers listed below have several and as many as six or more "degenerate" positions. What this means is that there are actually different versions of the primer that have

alternative nucleotides at each position indicated by a Y, R, M, etc. (see Table below). If you specify a **Y** at a particular position, then both **C** and **T** nucleotides are added at that position when the primer is synthesized, such that half the primer molecules have **C** and half have **T**.

If a primer has 2 four-fold degenerate sites, 2 three-fold sites, and 3 two-fold sites, your stock solution and every PCR will have 1152 different versions of the primer! At first glance, this might seem like an undesirable approach – after all, doesn't it mean that only 1 out of 1152 primer molecules will match the product perfectly? This is true, but it doesn't mean you get less product. In the first cycle of PCR, primer molecules that match the template perfectly and those that mismatch by a base or two will get the reaction going. In subsequent cycles, primer molecules that mismatch by two or three or four bases will be able to anneal to products previously generated by primer molecules mismatching by one or two or three bases (and so on, such that after several cycles all the different primer versions are involved in generating product). Empirically, there seems to be no reduction at all in the amount of PCR product obtained when using these degenerate primers as compared to a primer that matches the template perfectly.

OK, but why not use a primer that matches perfectly?

Primers that match perfectly are OK when you already know the sequence and are sure that it does not vary among the individuals/species that you want to sequence. If you're doing a systematics study and working on different species then you need a primer that will work with on a variety of organisms. In general, one looks for a sequence region that is conserved among taxa and designs primers in those locations.

Within protein-coding sequences, however, 3rd codon positions almost always vary among taxa even if the amino acid sequence is completely conserved. When designing primers in these regions, there are two options: 1) Use the consensus sequence from an alignment of various taxa that have already been sequenced. Doing this results in a primer that generally has no more than a few mismatches in comparison to the sequence of a particular taxon. 2) Build in degenerate sites to accommodate the variation likely to be present in the taxa you're working on. Doing this results in a primer that has at least some molecules that match every taxon perfectly. If the logic above is correct, then the primer with degenerate sites will yield product more consistently.

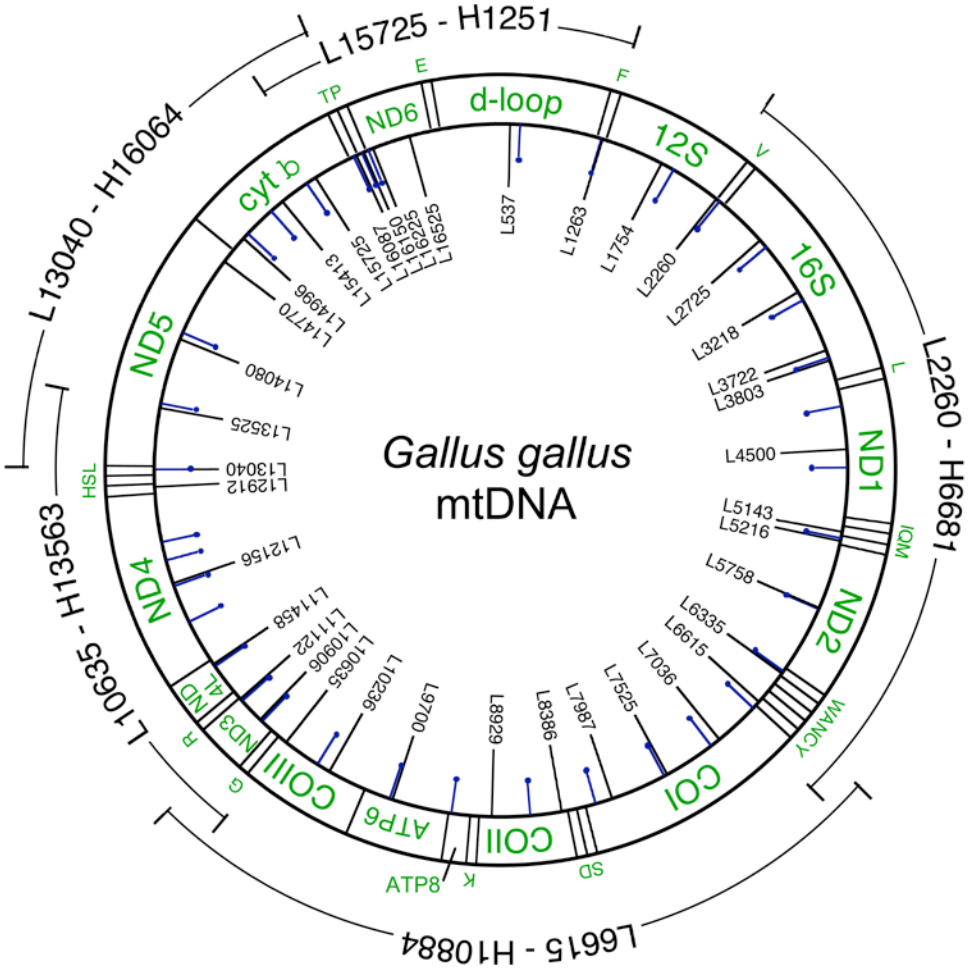
For studies of mtDNA, the degenerate primer has an additional important advantage – it is less likely to preferentially amplify nuclear copies of the mtDNA. Because nuclear copies (or numts) generally have a slower rate of sequence evolution than the mtDNA copy, they often are more similar to ancestral sequences than the extant mtDNA sequence. Because consensus primers tend to approximate ancestral sequences, they may preferentially amplify numts and lead to the unintended inclusion of one or more numt sequences in a mtDNA data set. This can be a serious problem (see Sorenson & Fleischer 1996; Sorenson & Quinn 1998). In contrast, a primer with degenerate sites will have primer molecules that match the mtDNA and any nuclear copies in a particular organism equally well. In this situation, the most numerous copy (generally the mtDNA copy when working with avian tissues other than blood) will predominate in the PCR product.

IUPAC Codes:	
R = A or G	B = C or G or T
Y = C or T	D = A or G or T
S = G or C	H = A or C or T
W = A or T	V = A or C or G
K = G or T	N = any base
M = A or C	

Suggested primer pairs for PCR and sequencing. (L and H numbers refer to strand and position of 3' base in the mtDNA sequence of *Gallus gallus*, Desjardins & Morais 1990):

Forward	Reverse	Notes
L1263 (tPhe)	H1859 (12S)	
L1754 (12S)	H2294 (tVal)	
L2260 (12S)	H2891 (16S)	
L2725 (16S)	H3292 (16S)	
L3218 (16S)	H3784 (16S)	
L3722 (16S)	H4170 (ND1)	
L3803 (16S)	H4644 (ND1)	
L4500 (ND1)	H5201 (tMet)	
L5143 (tGln)	H5766 (ND2)	
L5758 (ND2)	H6681 (COI)	1
L6615 (tTyr)	H7122 (COI)	
L7036 (COI)	H7548 (COI)	
L7525 (COI)	H8121 (COI)	
L7987 (COI)	H8628 (COII)	
L8386 (COII)	H9235 (ATP8/ATP6)	
L8929 (COII)	H9726 (ATP6)	
L9700 (ATP6)	H10343 (COIII)	
L10236 (COIII)	H10884 (ND3)	
L10635 (COIII)	H11837 (ND4)	2
L11458 (ND4L)	H12344 (ND4)	2
L12156 (ND4)	H13047 (tLeu)	2
L13040 (tLeu)	H13563 (ND5)	
L13525 (ND5)	H14127 (ND5)	
L14080 (ND5)	H15049 (cyt *b*)	2
L14770 (ND5)	H15295 (cyt *b*)	3
L14996 (cyt *b*)	H15646 (cyt *b*)	
L15413 (cyt *b*)	H16064 (tThr)	
L15725 (cyt *b*)	H1251 (tPhe)	4
	or H1530 (12S)	
L16087 (tThr)	H16137 (tPro)2	5

[1]Use H6313 and L6335 as internal sequencing primers to get complete sequence of both strands.
[2]Regions in which taxon-specific primers may be needed to obtain complete sequences of both strands.
[3]L14770 does not work consistently with all birds – this is a difficult region.
[4]XL PCR from cytochrome b to tPhe or 12S – a good start for getting into the control region of a new taxon.
[5]This combination may work to amplify the control region in birds with tPro/ND6/tGlu moved to the 3' end of the control region (see Mindell et al. 1998).


Gallus gallus
mtDNA
L15725 - H1251
L2260 - H6681
L6615 - H10884
L10635 - H13563
L13040 - H16064
d-loop
12S
16S
ND1
ND2
COI
COII
ATP8
ATP6
COIII
ND3
ND4L
ND4
ND5
cyt b
ND6
TP
E
F
V
L
IQM
WANCY
SD
K
G
R
HSL
L537
L1263
L1754
L2260
L2725
L3218
L3722
L3803
L4500
L5143
L5216
L5758
L6335
L6615
L7036
L7525
L7987
L8386
L8929
L9700
L10236
L10635
L10906
L11122
L11458
L12156
L12912
L13040
L13525
L14080
L14770
L14996
L15413
L15725
L16087
L16225
L16525

The figure on the preceding page includes suggested primer pairs for XL-PCR. If co-amplification of nuclear copies is a problem, amplifying larger fragments is one approach to getting a clean mtDNA product that sometimes works.

References:

Desjardins, P., & R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome. *Journal of Molecular Biology* 212: 599-634.

Johnson, K.J. & M.D. Sorenson. 1998. Comparing molecular evolution in two mitochondrial protein coding genes (Cytochrome *b* and ND2) in the dabbling ducks (Tribe: Anatini). *Molecular Phylogenetics and Evolution* 10: 82-94.

Johnson, K.J. & M.D. Sorenson. 1999. Phylogeny and biogeography of the dabbling ducks (Genus: *Anas*): A comparison of molecular and morphological evidence. *The Auk* 116: 792-805.

Mindell, D.P., M.D. Sorenson & D.E. Dimcheff. 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Molecular Biology and Evolution* 15: 1568-1571.

Mindell, D.P., M.D. Sorenson & D.E. Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. *Proceedings of the National Academy of Science USA* 95: 10693-10697.

Mindell, D.P., M.D. Sorenson, D.E. Dimcheff, M. Hasegawa & T. Yuri. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Systematic Biology* 48: 138-152.

Sorenson, M.D. & R.B. Payne. 2001. A single, ancient origin of obligate brood parasitism in African finches: implications for host-parasite coevolution. *Evolution* 55: 2550-2567.

Sorenson, M.D. & R.C. Fleischer. 1996. Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proceedings of the National Academy of Science USA* 93:15239-15243.

Sorenson, M.D. & T.W. Quinn. 1998. Numts: A challenge for avian systematics and population biology. *The Auk* 115: 214-221.

Sorenson, M.D., A. Cooper, E. Paxinos, T.W. Quinn, H.F. James, S.L. Olson & R.C. Fleischer. 1999. Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. *Proceedings of the Royal Society of London, Series B* 266: 2187-2194.

Sorenson, M.D., J.C. Ast, D.E. Dimcheff, T. Yuri & D.P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution* 12: 105-114.

Bird mtDNA primers:

L0537 CCT CTG GTT CCT CGG TCA G	H0614 GGR AAR ATG CCG CGA TYA CG
L1263 YAA AGC ATG RCA CTG AA	H1251 TCT TGG CAT CTT CAG TGC CRT GC
L1267 YAA AGC ATG RCA CTG AAG HYG	H1530 GTG GCT GGC ACA RGA TTT ACC
L1512 TAA GCA ATG AGT GHA ARC TYG ACT TAG	H1806 GTT TYA AGC GTT KGY GCT CGT A
L1754 TGG GAT TAG ATA CCC CAC TAT G	H1859 TCG DTT RYA GRA CAG GCT CCT CTA
L1843 AAA CYC TAA GGA CYT GGC GG	H1918 GAC GGC GGT ATR TAG GCT G
L1936 CAG CCT AYA TAC CGC CGT C	H1993 DDG CTA TAC CTY GAC CTG TC
L2010 TAR HAM GAC AGG TCR AGG TAT AGC	H2084 NTT TAC TDC TAA ATC CDC CTT
L2260 CAW GGT AAG YRT ACY GGA ARG TG	H2294 TYT CAG GYG TAR GCT GAR TGC TT
L2725 CGA GCY KGG TGA TAG CTG G	H2891 TGR TGG CTG CTT DAR GGC CHA C
L3218 CGM CTG TTT ACC AAA AAC ATM RCC	H3292 TGA TTR YGC TAC CTT YGC ACG G
L3722 GGT TTA CGA CCT CGA TGT TGG	H3784 CGG TCT GAA CTC AGA TCA CG
L3803 CTA CGT GAT CTG AGT TCA GAC CG	H4170 CCW ACN AYR TTD GGN CCY TTD CG
L3827 GCA ATC CAG GTC GGT TTC TAT C	H4644 TCR AAD GGG GCD CGG TTW GTY TC
L4500 GTN GCM CAA ACN ATY TCH TAY GAA G	H5191 GGG GTA TGG GCC CGA TAG C
L5143 GAA CCT ACA CAR AAG RGA TCA AAA C	H5201 CCA TCA TTT TCG GGG TAT GG
L5216 GGC CCA TAC CCC GRA AAT G	H5766 RGA KGA GAA RGC YAG GAT YTT KCG
L5758 GGN GGN TGA ATR GGN YTN AAY CAR AC	H6305 GGC TTT GAA GGC YCT TGG TC
L6335 GCC TTC AAA GCC TTA AAY AAG AG	H6313 ACT CTT RTT TAA GGC TTT GAA GGC
L6615 CCY CTG TAA AAA GGW CTA CAG CC	H6681 GGT ATA GGG TGC CRA TRT CTT TGT G
L7036 GGN ACN GGN TGA ACH GTN TAY CC	H7122 ATN GTD GTR ATR AAR TTR ATD GCH CC
L7525 GTN TGR GCH CAY CAY ATR TTY AC	H7548 GTD GCN GAN GTR AAR TAD GCT CG
L7594 ATA ATY ATY GCH ATY CCN ACN GG	H8121 GGG CAG CCR TGR ATT CAY TC
L7987 TCH GAY TAY CCW GAY GCN TAY AC	H8628 TCR TAG STT CAR TAT CAY TGR TGN CC
L8386 GCN TCV TCN CCH ATY ATR GAA GA	H9235 TCR AAG AAG CTT AGG TTC ATG GTC A
L8929 GGH CAR TGY TCA GAR ATY TGY GG	H9726 AGR TGN CCD GCT GTD AGR TTN GC
L9700 GAA ACH AYM AGY CTN CTN ATY CGN CC	H10343 TGD GCT CAD GTD ACD GTN ACN CC
L10236 TTY TGA GCN TTY TTY CAY TCH AG	H10884 GGR TCR AAN CCR CAY TCR TAN GG
L10635 CAY CAY TTY GGN TTY GAA GCH GC	H11100 TCT GCY CAY TCT ARK CCT CCY TG
L10647 TTY GAA GCH GCH GCH TGA TAY TG	H11442 CGA ACC CAC GGC TCC GAC CAC CTA

Bird mtDNA primers (cont.):

L10906 CCN TAY GAR TGY GGN TTY GAY CC	H11837 AGR GTD GCY TCR AAT GHR ATR TAR AA
L11122 CAR GGA GGM YTA GAR TGR GCA GA	H12134 GGW GCY TCT ACR TGD GCT TTD GG
L11458 TCH ACH CGN ACN CAY GGY TCH GA	H12344 CTA TRT GRC TDA CDG ADG ART ADG C
L12156 CCH AAA GCH CAY GTW GAA GCH CC	H12488 ATT CGG CTG TGT GTY CGY TC
L12912 TAG AYT GTG AYT CTR AAA AYA GRA G	H13047 CTT YYA YTT GGA KTT GCA CCA A
L12976 CAA GAA CTG CTA AYT CYT GCA TCT G	H13563 TGN AGD GCD GCD GTR TTD GC
L13040 ATC CRT TGG TCT TAG GAR CCA	H14127 CCT ATT TTT CGR ATR TCY TGY TC
L13525 GCT GAG ARG GHG TDG GMA TYA TRT C	H15049 GTR TCN GCD GTR TAR TGY ATD GC
L14080 TCA ACN CAY GCM TTC TTY AAR GC	H15295 CCT CAG AAK GAT ATY TGN CCT CAK GG
L14770 TAG GNC CNG ARG GNY TNG C	H15646 GGN GTR AAG TTT TCT GGG TCN CC
L14996 AAY ATY TCW GYH TGA TGA AAY TTY GG	H16064 CTT CAN TYT TTG GYT TAC AAG RCC
L15413 GGG GGW TTY TCM GTN GAY AAY CC	H16137 ARA ATR YCA GCT TTG GGA GYT GG
L15725 AAR CCM GAA TGR TAY TTY CTW TTY GC	H16191 TCT CGD GGG GCD ATT CGG GC
L16087 TGG TCT TGT AAR CCA AAR ANY GAA G	
L16150 CCT CYA YCW CCA RCT CCC AAA GC	
L16206 TAA ACH GCC CGA ATH GCC CC	
L16225 CCG AGA CAA CCC ACG CAC AAG	
L16525 ACA AAC ACC ACC ARC ATH CCH CC	