Canadian Technical Report of
Fisheries and Aquatic Sciences 2487

2003

DESIGN OF PRIMERS FOR POLYMERASE CHAIN REACTION AMPLIFICATION OF MITOCHONDRIAL DNA OF COREGONINE FISHES

by

T.L. Hisanaga, R. Banjo, T.J. Carmichael, and J.D. Reist

Central and Arctic Region
Fisheries and Oceans Canada
Freshwater Institute
501 University Crescent
Winnipeg, Manitoba R3T 2N6

Cat. No. Fs 97-6/2487E ISSN 0706-6457

Correct citation for this publication:

Hisanaga, T.L., R. Bajno, T.J. Carmichael, and J.D. Reist. 2003. Design of primers for polymerase chain reaction amplification of mitochondrial DNA of Coregonine fishes. Can. Tech. Rep. Fish. Aquat. Sci. 2487: iv + 22 p.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Hisanaga, T.L., R. Banjo, T.J. Carmichael, and J.D. Reist. 2003. Design of primers for polymerase chain reaction amplification of mithochondrial DNA of Coregonine fishes. Can. Tech. Rep. Fish. Aquat. Sci. 2487: iv + 22 p.

Primers for the polymerase chain reaction were designed to amplify the entire mitochondrial genome of various Coregonines. The mitochondrial genome of *Coregonus lavaretus* (Genbank accession no. NC_002646) was used as the template for primer design. The mitochondrial genomes for a variety of fish determined to be representative of the Coregonines (Appendix 1) were successfully amplified. This basic step provides the basis for sequencing the complete mitochondrial genome of Coregonine fishes, thus allowing for taxonomic and phylogenetic studies.
Keywords: PCR product, mitochondrial molecule, coding and noncoding regions, sequence alignment, taxonomic identification.

Key words: PCR, direct sequencing, taxonomic identification, phylogenetic relationships, Coregonus, Prosopium, Salvenlinus.

RÉSUMÉ

Hisanaga, T.L., R. Banjo, T.J. Carmichael, and J.D. Reist. 2003. Design of primers for polymerase chain reaction amplification of mithochondrial DNA of Coregonine fishes. Can. Tech. Rep. Fish. Aquat. Sci. 2487: iv + 22 p.

On a conçu des amorces de la réaction en chaîne de la polymérase en vue d'amplifier le génome mitochondrial intégral de divers Corégonidés en se servant du génome mitochondial de *Coregonus lavaretus* (numéro d'accession de GenBank NC002646) comme matrice. On a ainsi réussi à amplifier les génomes mitochondriaux de divers poissons identifiés comme étant des Corégonidés (annexe 1). Cette étape fondamentale servira d'assise au séquençage du génome mitochondrial intégral de poissons de cette famille, ce qui permettra de mener des études taxinomiques et phylogénétiques.

Mots clés : PCR, séquençage direct, identification taxinomique, relations phylogénétiques, Coregonus, Prosopium, Salvelinus.

INTRODUCTION

Mitochondrial DNA (mtDNA) is maternally inherited DNA that replicates independent of nuclear DNA, and is not subject to recombination. The rate of mutation of mtDNA is constant and 5-10 times greater than most regions of nuclear DNA (Moritz *et al.*, 1987), rendering it extremely useful as a molecular chronometer in phylogenetic studies. Furthermore, consistent differences in mtDNA sequences across taxa, if they occur, may be useful for identifying closely related taxa. Single base changes can be detected among taxa using such techniques as restriction fragment analysis and direct sequencing. As well, comparisons can be made if additions and deletions of bases of significant size exist. These can be detected by simply determining the length of a polymerase chain reaction product. The utility of these techniques for elucidating differences in mtDNA among species, and for using such information to determine taxonomic uniqueness and relationships is well established (Bernatchez *et* al., 1991, Brzuzan, 2000, Fournier-Lockwood *et al*., 1993, Reist *et al*., 1998,).

The polymerase chain reaction (PCR) results in rapid, exponential amplification of minute amounts of DNA, allowing for sufficient quantities of DNA to be analyzed using the sequencing and related techniques (Baumforth *et al*., 1998). The PCR process requires short oligonucleotide primers to both start and direct the amplification of a region of interest. In order to achieve high fidelity amplification, the primers must: 1) bind to the target DNA sequence with high specificity, 2) bind at strictly one site along the target (i.e., be unique over the entire mitochondrial genome), and 3) anneal to the target DNA at a fairly high temperature (Baumforth *et al*., 1998, Beckenbach, 1991). This will prevent the amplification of artifacts while ensuring an abundance of the desired product. It is also desirable to design primers that bind to areas which are highly conserved areas, making them useful for PCR amplification in a wide variety of species. Primers are designed for both the plus and minus strands of the mitochondrial DNA molecule. This brackets the desired region, and ensures the PCR product is uniform in size so that when the PCR product is later used as a template for DNA sequencing, the integrity of the sequence achieved is greater, and a greater number of nucleotides can be accurately determined.

The mitochondrial DNA of the Salmonidae, of which the Coregonines form a distinct subfamily, is circular and approximately 17 000 nucleotide pairs in length, although considerable variation in length exists among and sometimes within species (Brzuzan, 2000). Mitochondrial DNA codes for ribosomal RNAs and transfer RNAs used in the mitochondrion, and contains only 13 recognizable genes that code for polypeptides (Beckenbach, 1991). Each of the genes code for a different polypeptide that makes up part of the electron transport chain in the inner mitochondrial membrane: NADH dehydrogenase subunits 1-6 and 4L, cytochrome oxidase subunits 1-3, cytochrome b, and ATPases 6 and 8. There also exists a region of higher variability within the mitochondrial genome, the control region, which is also known as D-loop (Lee *et al*., 1995). Since this region contains the origin of replication, it is shown to be more liable to contain tandemly repeated sequences (Brzuzan, 2000), and a greater number of substitutions as well as insertions and deletions compared to the remaining regions of mtDNA, such as the gene-coding regions (Lee *et al*., 1995).

Requirements envisioned as part of the impending species-at-risk legislation for confirming distinctness and identification of fish taxa, particularly Coregonines that exhibit some levels of risk (Sue Cosens, Freshwater Institute,501 University Crescent, Winnipeg, MB, Canada, R3T 2N6, personal communication), suggested the need for applying genetic techniques such as direct sequencing that have recently become available. The framework for sequencing studies is outlined in Figure 1 with the portion customized in this report indicated. Previous work regarding direct sequencing of mtDNA conducted within the Freshwater Institute Science Laboratory fish genetics laboratory focussed primarily upon a small portion of the entire mitochondrial genome, approximately 400 bases of the D-loop (Brown-Gladden *et al*., 1995, Reist *et al*., 1998). This work demonstrated that variation in the D-loop region was reasonable for studies among major lineages of Coregonines (e.g., whitefishes versus ciscoes). However, initial work within the cisco group indicated that D-loop variation was insufficient to differentiate lineages among the more closely related Coregonines (e.g., lake ciscoes versus shortjaw ciscoes). Thus, this study was conducted to develop basic techniques to be used in subsequent mtDNA sequencing studies of cisco taxa, especially those designated to be at some level of risk. The aims of this study were to: 1) design appropriate primers for PCR amplification of Coregonine mtDNA genomes, and 2) test their suitability of amplifying various Coregonines by amplifying the mitochondrial genomes of Coregonines chosen to represent all major lineages within the group.

MATERIALS AND METHODS

Polymerase chain reaction primers were designed using MacVector 7.0 (Accelrys Inc., Madison, WI.), based upon the *Coregonus lavaretus* mitochondrial genome (Genbank accession no. NC_002646) as a template (Miya and Nishida, 2000). When designing the primers with the MacVector program, the following constraints were applied based upon established rules (Baumforth *et al*., 1998, Kampke *et al*., 2001, and Rychlik, 1995). Primers were to be free of any significant complementarity and secondary structures such as hairpin loops. Desired primer length was between 18-24 nucleotides, with a similar GC ratio to that of the target (~50% GC). Primer melting temperatures for the two primers were to be as similar to one another as possible, and unique within the target DNA.

The mitochondrial genomes for *Coregonus lavaretus*, *Oncorhynchus mykiss* (Zardoya, 1995), *Salvelinus fontinalis* Mitchill (Dorion *et al*., 2002), *Salvelinus alpinus* (Dorion *et al*., 2002),and *Salmo salar* (Hurst *et al*., 1999) were aligned using MacVector. Highly conserved regions among the aforementioned teleosts were identified as being preferable for the binding locations of PCR primers. Once the primers were chosen, they were again analysed using MacVector to determine whether they could potentially bind to other locations along the mitochondrial genomes of the aforementioned Salmonids. Any primers exhibiting the ability to bind to and amplify sequences other than the desired targets were rejected. These primers were then tested on representative fish species (Appendix 1).

The majority of the primer pairs had an optimal annealing temperature between 50-60 ^{o}C. However, the annealing temperatures actually used to amplify the PCR products were not always those determined by the MacVector program to be optimal.

The primer annealing temperatures were altered to accommodate for maximum use of the 40 block thermal cyclers. This allowed for multiple reactions using various primer pairs to be carried out in a single thermal cycler, greatly reducing the need for multiple thermal cyclers and time. In each case, the annealing temperature was rounded off within a five degree range from the optimal temperature.

The primers were manufactured by Invitrogen Canada Co. (Burlington, ON.) as oligonucleotides of standard desalted deoxynucleotides which are lyophilized, then resuspended in sterile distilled water. Total DNA (mitochondrial and genomic) was isolated from tissue[1] using DNeasy tissue kits (Qiagen, Mississauga, ON.) as per manufacturer's recommendations, and stored at –20^{o}C. Polymerase chain reaction mixtures were prepared containing 0.8 mM forward and reverse primers, 0.1 mM dNTP (except where otherwise indicated in Appendix 2), 1 X *Taq* polymerase buffer with 1.5 mM $MgCl_2$ (Roche Diagnostics, Laval, QC.), 10-60 ng of DNA, and 0.5 U *Taq* polymerase (Roche Diagnostics, Laval, QC.) in a 100 μl reaction volume. Amplification of mtDNA was performed in a Techne Genius thermal cycler (Fisher Scientific, Ottawa, ON.) with heated lid as per the thermal cycling parameters in Appendix 2. Successful amplification of the various mtDNA portions was verified by running 5 μl of the PCR product and 1 KB plus ladder (Invitrogen Canada Co., Burlington, ON.) on a 1.0% agarose gel, stained with ethidium bromide (1 mg/ml) and photographed on an ultraviolet transilluminator with Polaroid film.

RESULTS AND DISCUSSION

Polymerase chain reaction primers were designed to accomplish the following goals: 1) to amplify gene coding regions of the Coregonine mtDNA in their entirety, 2) complement existing mtDNA primers (Bernatchez and Danzmann, 1993, Miller *et al*., 1998, Nielsen *et al*., 1998) currently used in the FWI fish genetics programs, and, 3) result in mtDNA PCR products that would be suitable for using as templates in automated sequencing.

Since one of the most important factors affecting the quality of PCR is the choice of primers, a few basic parameters were adhered to when designing the primer pairs. The primer should only bind to the target site and not hybridize at any other site. This would result in non-specific amplification, a potential mixed pool of PCR products, and a greatly reduced amount of desired product (Rychlik, 1995). The primers must also be free of significant complimentarity at their 3' ends otherwise primer-dimers may occur, which could lead to a decrease in product due to a lack of primers. Another feature to be avoided is the potential formation of a hairpin loop by primers. Should this occur the primer would become unavailable to bind to the target site. As a result, an uneven primer ratio would become established within the reaction, and assymetrical PCR would occur, or worse yet, no product at all would result (Rychlik, 1995).

Although degenerate primers are just as easy and cheap to produce as unique primers, the design of such primers were selected against. When degeneracy is too high, unrelated sequences may be amplified as well as the desired sequence (Linhart and Shamir, 2002). By avoiding PCR primers that are degenerate, a second purpose

[1] Tissues can be muscle, fin clips, adipose fin, or any tissue removed from either live-released or dead-sampled fish and preserved frozen, in ethanol, or in a NaCl-saturated 20% dimethyl sulphoxide (1:2 v/v).

the PCR primers may serve is to function as sequencing primers in automated sequencing reactions, since PCR and sequencing primers share many characteristics (Gerischer and Durre, 2001).

An overview of primer locations and the corresponding areas amplified is given in Figure 2, and a more detailed map of primer locations with respect to gene-coding regions is shown in Figure 3. The underlined bases indicate where the primers are located, and not necessarily their exact sequences. Since only the plus strand is shown, reverse primers are the reverse and complement of those bases underlined. Also, individual mismatches between bases underlined and the primers listed in Appendix 2 occur as the depicted sequence is that of *Coregonus lavaretus* and the primers were designed with the intent of amplifying a variety of Coregonines.

Due to the number of existing primers for the D-loop and surrounding regions (Bernatchez and Danzmann, 1993, Miller *et al*., 1998), no attempt was made to design primers to amplify an intact D-loop region. Rather, primers were designed that by default were able to amplify two separate yet overlapping regions of D-loop. This was done for two reasons. First, the D-loop is often amplified in its entirety using the CytB and HN20 primers with only limited success among different samples and different species. Second, these two primers result in a PCR product of an additional 500 bases that is already amplified when targeting the 12S rRNA region, one of the two separate yet overlapping regions of D-loop. If the D-loop and the 12S rRNA gene are each to be amplified for sequencing purposes, the redundant 500 bases amplified will result in an unnecessary overlap, resulting in both wasted supplies and time.

As previously mentioned, the actual annealing temperatures for the PCR primer pairs employed in the thermal cycling parameters was not always that determined to be optimal by the MacVector program. In the interest of maximizing space in the thermal cyclers, as well as time, the annealing temperatures were “rounded off”. No difference was observed between the optimal annealing temperatures and those actually used (data not shown). This allowed for multiple polymerase chain reactions, with different PCR primer pairs to be performed in the same thermal cycler, significantly reducing the need for multiple machines.

The mitochondrial genome was successfully amplified with all of the primers for all Coregonines examined with the exception of *Coregonus muksun* (Table 1). The ND5, ND6, ATPase 6, ATPase 8, and COII regions were not successfully amplified for *C. muksun*. This may be due to the quality of the DNA samples for this fish rather than the primers not being able to bind and amplify through these regions, as the ND5/ND6 primers were able to produce a PCR product for the more distantly related *Salvelinus alpinus*. Wherever possible, primers resulted in PCR products that overlapped by at least fifty base pairs, which will facilitate the alignment of sequences obtained for the different mtDNA fragments.

CONCLUSION

Sequencing of mtDNA for studies investigating taxonomic distinctness, identification and interrelationships requires adequate amounts of DNA to be targeted and amplified using the polymerase chain reaction. For Coregonine fishes, the primers

designed and tested herein allow for the successful amplification of the entire mitochondrial genome, and thus provides the basis for such studies.

ACKNOWLEDGEMENTS

This study was supported by funding from the Department of Fisheries & Oceans Canada through its internal Species At Risk programme. The following individuals are thanked for providing or aiding in the collection of samples: K. Alt, J. Babaluk, D. Bodaly, F. Decicco, P. Gudkov, J. Johnson, G. low, L. Murray, M. Papst, Y. Reshetnikov, M. Steinhilber, T. Todd, J. Vuorinen

REFERENCES

Baumforth, K.R.N., P.N. Nelson, J.E. Digby, J.D. O'Neil, and P.G. Murray. 1998. The polymerase chain reaction. J. Clin. Pathol: Mol. Pathol. 51: 1-10.

Beckenbach, A.T. 1991. Rapid mtDNA sequence analysis of fish populations using the polymerase chain reaction (PCR). Can. J. Fish Aquat. Sci. 48: 95-98.

Bernatchez, L., and R.G. Danzmann. 1993. Congruence in control region sequence and restriction site variation in mitochondrial DNA of Brook charr (*Salvelinus fontinalis* Mitchill). Mol. Biol. Evol. 10: 1002-1014.

Bernatchez, L, F. Colombani, and J.J. Dodson. 1991. Phylogenetic relationships among the subfamily Corigoninae as revealed by mitochondrial DNA restriction analysis. J. Fish Biol. 39: 283-290.

Brown-Gladden J.G., L.D. Postma-Maiers, T.J. Carmichael, and J.D. Reist. 1995. Mitochondrial DNA control region sequence variation in arctic char (*Salvelinus alpinus*). Can. Data Rep. Fish. Aquat. Sci. 968: 16p.

Brzuzan, P. 2000. Tandemly repeated sequences in mtDNA control region of whitefish, *Coregonus lavaretus*. Genome 43: 584-587.

Dorion, S., L. Bernatchez, and P.U. Blier. 2002. Comparative genomic analysis of the potential adaptive value of arctic charr mtDNA introgression in brook charr populations (*Salvelinus fontinalis* Mitchill). Mol. Biol. Evol. 19: 1902-1909.

Fournier-Lockwood, S., R.E. Dillinger Jr., T.P. Birt, J.M. Green, and T.P. Snyder. 1993. Phylogenetic relationships among members of the Coregoninae inferred fromdirect sequencing of PCR-amplified mitochondrial DNA. Can. J. Fish Aquat. Sci. 50: 2112-2118.

Gerischer, U., and P. Durre. 2001. Primer design and primer directed sequencing. Methods. Mol. Biol. 167: 39-51.

Hurst, C.D., S.E. Batrlett, W.S. Davidson, and I.J. Bruce. 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239: 237-42.

International Union of Biochemistry nomenclature committee. 1985. Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. Eur. J. Biochem. 50: 1-5.

Kampke, T., M. Kieninger, and M. Mecklenburg. 2001. Efficient primer design algorithms. Bioinformatics 17: 214-225.

Lee, W., J. Conroy, W. Huntting-Howell, T.D. Kocher. 1995. Structure function of the Teleost mitochondrial control regions. J. Mol. Evol. 41; 51-66.

Limhart, C., and R. Shamir. 2002. The degenerate primer design problem. Bioinformatics 18: S172-S180.

Miller, S., T. Underwood, and W.J. Spearman. 1998. Genetic assessment of Inconnu (*Stenodus leucichthys*) from the Selawik and Kobuk Rivers, Alaska, using PCR and RFLP analyses. Alaska Fisheries Technical Report, number 48.

Miya M., and M. Nishida. 2000. Use of mitogenetic information in teleostean molecularphylogenetics: a tree based exploration under the Maximum-Parsimony optimality criterion. Mol. Phylogenet. Evol. 17: 437-455.

Moritz, C., T.E. Dowling, and W.M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18: 269-292.

Nielsen, E.E., M.M. Hansen, and K.L.D. Mensberg. 1998. Improved primer sequences for the mitochondrial ND1, ND3/4 and ND5/6 segments in salmonid fishes: application to RFLP analysis of Atlantic salmon. J. Fish. Biol. 53: 216-220.

Reist, J.D., L.D. Maiers, R.A. Bodaly, J.A. Vuorinen, and T.J. Carmichael. 1998. The phylogeny of new- and old-world coregonine fishes as revealed by sequence variation in a portion of the D-loop of mitochondrial DNA. Arch. Hydrobiol. Spec. Issues Advanc. Limnol. 50: 323-339.

Rychlik, W. 1995. Selection of primers for polymerase chain reaction. Mol. Biotechnol. 3: 129-134.

Zardoya, R., A Garrido-Pertierra, and J.M. Bautista. 1995. The complete nucleotidesequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. J. Mol. Evol. 41: 942-951

Table 1. Success of PCR amplification for taxa examined. For an explanation of the taxon acronyms used, refer to Appendix 1.

Taxon	12S rRNA partial D-loop	16S rRNA	ND1	ND2	COI	COII	COIII	COII & COIII ATPase 6&8	ND 3&4	ND 5&6	cyt *b* partial D-loop
ARCS Baicalensis	+	+	+	+	+	+	+	+	+	+	+
PELD	+	+	+	+	+	+	+	+	+	+	+
LVRT Pidschian	+	+	+	+	+	+	+	+	+	+	+
RSKA	+	+	+	+	+	+	+	+	+	+	+
LKWF (Mississippian form)	+	+	+	+	+	+	+	+	+	+	+
LKCS	+	+	+	+	+	+	+	+	+	+	+
LKWF (Beringian form)	+	+	+	+	+	+	+	+	+	+	+
SJCS	+	+	+	+	+	+	+	+	+	+	+
LSCS	+	+	+	+	+	+	+	+	+	+	+
BDWF	+	+	+	+	+	+	+	+	+	+	+
BRCS	+	+	+	+	+	+	+	+	+	+	+
ARCS	+	+	+	+	+	+	+	+	+	+	+
MUKS	+	+	+	+	+	-	+	-	+	-	+
ARCS (Pollan)	+	+	+	+	+	+	+	+	+	+	+
ARCH	+	+	+	+	+	+	poor	-	+	+	+
BVCS	+	+	+	+	+	+	+	+	+	+	+

"Poor" designation indicates weak band intensity (amplification) as revealed by ethidium bromide staining of PCR products run on a 1% agarose gel. The "+" symbol indicates successful amplification of the coding regions (outer strand) of mtDNA as shown in Figure 1. The "-" symbol indicates unsuccessful amplification of the coding regions (outer strand) of mtDNA.

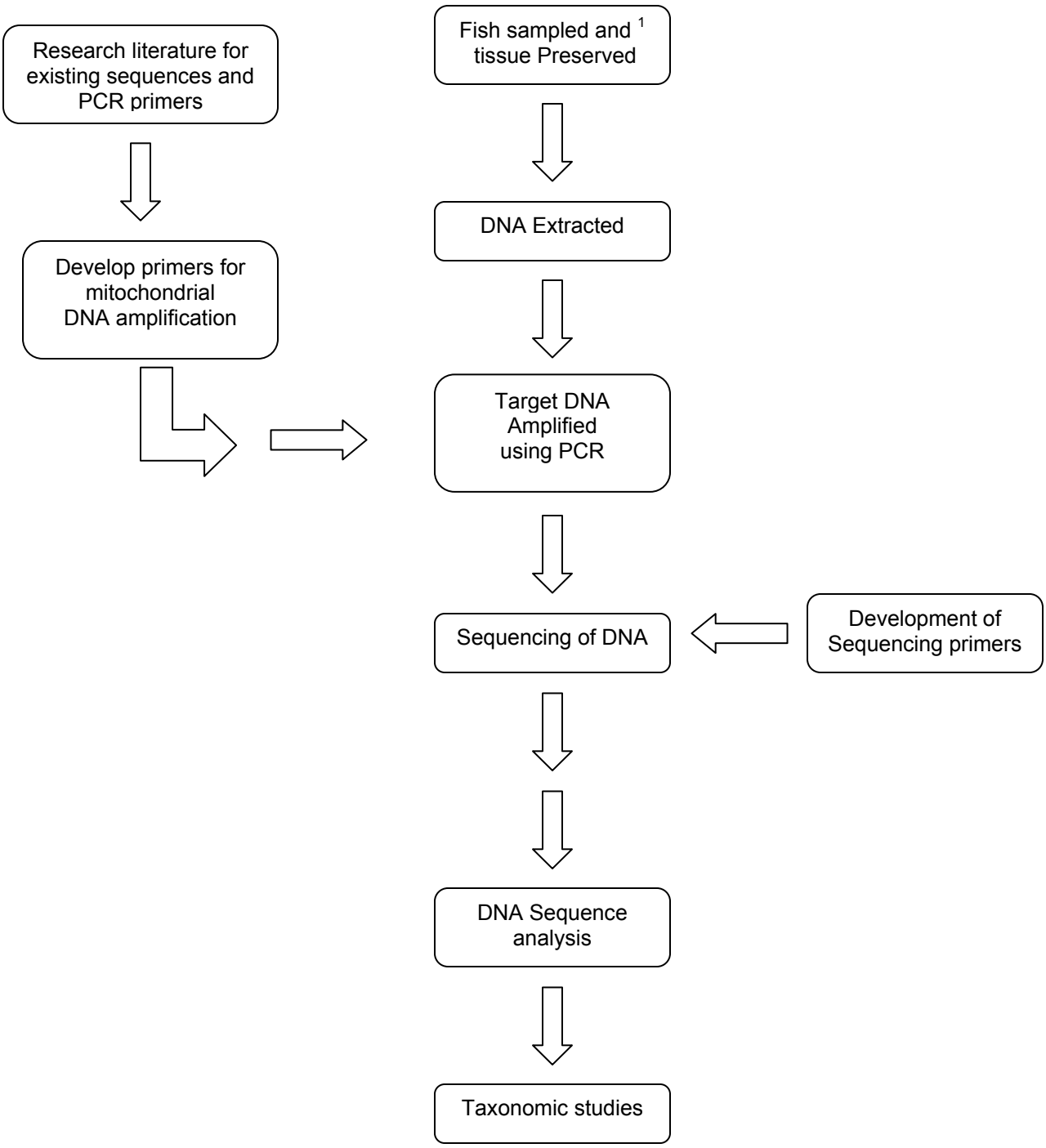

Figure 1. Flow diagram depicting the steps leading to direct sequencing of mitochondrial DNA to obtain taxonomic information. Activity done in this report includes the extraction of DNA, development of PCR primers, and the amplification of mtDNA using PCR.

[1] Tissues can be muscle, fin clips, adipose fin, or any other tissue removed from either live-released or dead-sampled fish and preserved frozen, in ethanol, or in a NaCl-saturated 20% dimethylsulphoxide solution at 1:2 v/v.

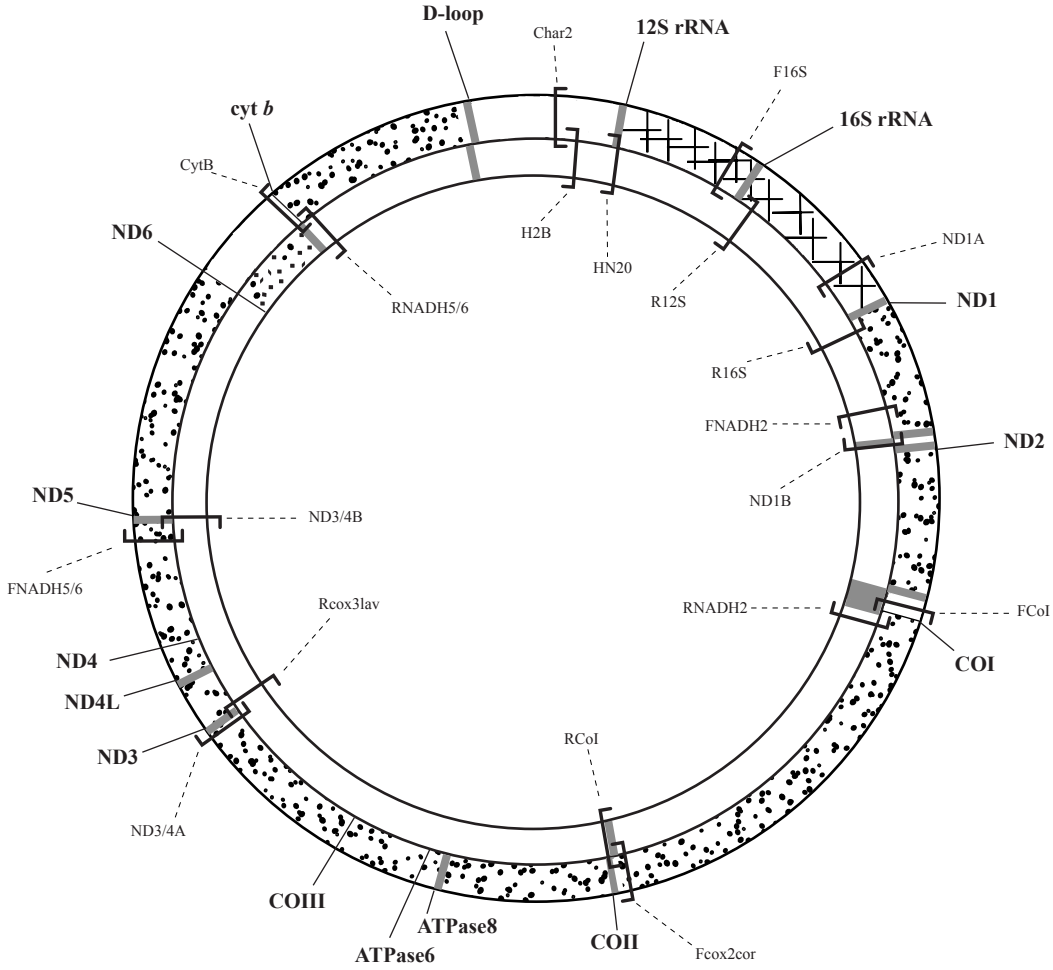

Figure 2. An overview of the mitochondrial DNA of Coregonines (figure no to scale). Gene-coding regions are speckled, rRNA coding regions are cross hatched, tRNA coding regions are gray, and non-coding regions are white. Primers are depicted along the strand and location to which they bind. Primer names are in the non-bold, smaller font (dashed lines); and gene names are in the bold, larger font (solid lines). Forward primers are indicated by "[", and reverse primers by "]". 12S and 16S rRNA are the 12S and 16S ribosomal RNAs; ND1-6 and 4L are the NADH dehydrogenase subunits 1-6 and 4L; COI-III are the cytochrome oxidase subunits 1-3; ATPase6 and 8 are the ATPase 6 and 8 subunits; cyt *b* is cytochrome *b*; and D-loop is the control region.

Figure 3. Locations of primers used in this study, and corresponding gene-coding regions with respect to the *Coregonus lavaretus* mitochondrial genome +strand (Miya and Nishida, 2000). Start of gene-coding regions is indicated by "[" and "]" indicates the end. Gene names are given at the start of the gene, except for ND6 which is located on the complement strand. Primer locations are underlined, and primer names are written above in bold font with arrows indicating direction of primer amplification. Underlined bases are not identical to the primer sequences used in this study; for proper primer sequences, refer to Appendix 2. Primers amplifying in the reverse direction are in most cases, complementary to the –strand. Numbers at the right indicate the position number of the immediately following base in the *Coregonus lavaretus* mtDNA genome.

```
                  ← HN20
gctggcgtag cttaactaaa gcataacact gaagctgtta agatggaccc tagaaagtcc       61
          12S rRNA
cgcaggca[ca aaggcttggt cctgacttta ttatcagctt taactgaact tacacatgca      121
agtctccgca ctcctgtgag gatgccctta atcccctgcc cggggacgag gagctggcat      181
caggcacgcc ccggcagccc aagacgcctt gctaagccac acccccaagg aaactcagca      241
gtgatagata ttaagctata agcgaaagct tgacttagtt aaggttaaga gggccggtaa      301
aactcgtgcc agccaccgcg gttatacgag aggccctagt tgataatcac cggcgtaaag      361
agtggttagg aattatattt aataaagccg aacacccccc tggctgtcat acgcacctgg      421
gggcacgaag ccccactgcg aaagcagctt taatcaccac ctgaacccac gacagctatg      481
atacaaactg ggattagata ccccactatg cctagccgta aactttgatg gaaacataca      541
actaacatcc gccagggaac tacaagcgcc agcttaaaac ccaaaggact tggcggtgcc      601
tcagacccac ctagaggagc ctgttctaga accgataacc cccgttcaac ctcaccacct      661
cttgttttcc ccgcctatat accaccgtcg tcagcttacc ctgtgaagga tttatagtaa      721
gcaaaatggg catgacccaa aacgtcaggt cgaggtgtag cgcatggggt gggaagaaat      781
gggctacatt ctctaaatta gagcattacg aaccacgctg tgaaaccagc gtccgaaggt      841
ggatttagca gtaaacagaa agcagagagt tctcttgaaa ctggctctga ggcgcgcaca      901
caccgcccgt cactctcccc aagttcaatc tacccttcta actaagaagt taaccgaaca      961
aaggggaggc aagtcgtaac atggtaagtg taccggaagg tgcacttgga ataac]cagag     1021
                                  F16S RNA →
tgtagctaag acagaagagc acctcccctta caccgagaag acatccgtgc aaatcgggtc     1081
        16S rRNA
accctga[gct gactagctag cccacacatt tggtctaaca ccacaacata tataccccca     1141
                                                  ← R12S
caaaacttag aattaagtca acaaaccatt tttccccctt agtatgggcg acagaaaagg      1201
gaataattga gcaacagaga aagtaccgca agggagagct gaaagagaac tgaaacaacc      1261
catttaagcc tagaaaagca gagattaaat ctcgtacctt ttgcatcatg atttagccag      1321
caaacccgag caaagagaac tttagttcag gcccccgaaa ctagacgagc tactccggga      1381
cagcctatta tagggccaac ccgtctctgt ggcaaaagag tgggaagagc cccgagtaga      1441
```

ggtgataaac ctatcgagcc tagttatagc tggttgctta ggaaatgaat agaagttcag 1501
ccccctggct ttcttaggac cctaaggtaa aactaacctc gtcccataga aaccaaggga 1561
gttagtcaaa ggaggtacag ctcctttgaa caaggacaca accttaacag gcggctaagg 1621
atcataatta ctaaggtaac ctgttacagt gggcctaaga gcagccacct gcatagaaag 1681
cgttaaagct cagacagaca cgaacctctt attttgataa gaaatcctac cccctaaccg 1741
tactaagccg ttccatgccc ccatggaaga gattatgcta gaatgagtaa taagagggga 1801
caaccctctc ccagcacatg tgtaagtcgg accggacccc ccaccgacaa ataacgaacc 1861
taaaccaaga gggaaatgca ggccagaaga gaaaccgaga aaagcctaca aaactaatcg 1921
ttaaacccac acaggagtgc ccacaaggaa agacccaaag gaagagaagg aactcggcaa 1981

ND1A →

acacaagcct cgcctgttta ccaaaaacat cgcctcttgc aaatcaaagc ataagaggtc 2041
ccgcctgccc tgtgactatg ggtttaacgg ccgcggtatt ttgaccgtgc gaaggtagcg 2101
caatcacttg tcttttaaat gaagacctgt atgaatggca tcacgagggc ttagctgtct 2161
cctcttccaa gtcagtgaaa ttgatctgcc cgtgcagaag cggacataag tacataagac 2221
gagaagaccc tatggagctt tagacaccag gcagatcacg tcaagcaacc ttgagttaac 2281
aagtaaaaac gcagtgaccc ctagcccata tgtctttggt tggggcgacc gcgggggaaa 2341
acaaagcccc catgtggact gggggcactg cccccacagc cgagagctac agctctaagc 2401
accagaattt ctgaccagaa atgatccggc gaacgccgat caacggaccg agttacccta 2461
gggataacag cgcaatcctc tcccagagtc cctatcgacg agggggttta cgacctcgat 2521
gttggatcag gacatcctaa tggtgcagcc gctattaagg gttcgtttgt tcaacgatta 2581
aagtcctacg tgatctgagt tcagaccgga gtaatccagg tcagtttcta tctatgaagt 2641
gatgtttcct agtacgaaag gaccggaaag aaggggccca tgcttaaggc acgccccacc 2701
cccacctgat gaaggcaact aaaacaggca agggggcaca ccaaggtgtg cctgagataa 2761

← R16S

cggcgc**]**gcta aggtggcaga gcccggtaat tgcggaaggc ctaagccctt tttctcagag 2821

ND1

gttcaaaccc tctccttagc t**[**atgatcacg accctaatca cccacgttat taatcccctt 2881
gcctacatcg tgcccgttct cctggcagtt gctttcctca ccctactaga acggaaagtc 2941
ctcgggtata tacaacttcg gaaaggacct aacattgtcg gcccctatgg gttacttcaa 3001
cctattgcag atggggttaa gctctttatt aaagaaccga ttcgaccgtc tacctcctcc 3061
cccttcctat tcctcgctac accaatactt gccttaacac tcgcacttac cttatgggcc 3121
cccatgccca ttccctaccc cgtcactgat ctaggcctag gggtactctt tgtccttgcc 3181
ttgtcaagcc ttgccgtata ttcaattctt ggctcaggct gggcctctaa ttctaagtat 3241
gctttaatcg gggcccttcg agccgtagca caaactattt cctacgaagt aagcctaggc 3301
cttatcttac ttagcgtgat tattttcaca ggcggtttta cacttcagac tttcaatgtt 3361
gctcaagaaa gcatctgatt gctcgttcca gcctgacccc ttgctgccat atgatatatt 3421

```
tcaacgctag ctgagacaaa ccgtgccccc tttgacctca cagaggggga atcagaacta       3481
gtctctgggt ttaatgtaga atacgccgga ggacccttcg ccctttttt tctggcggag       3541
tatgccaaca tcctcctcat aaatacgctc tcaaccatcc tatttctggg ggcatcacat       3601
                                                    FNADH2 →
atccccgcct ttcccgaact aacagccata aatctaataa caaaaagccgc cctactatcc      3661
gtagtctttt tatgagtgcg agcctcatac ccccgctttc ggtacgacca gctcatacac       3721
cttgtttgaa aaagcttcct acctataacc ctggcacttg tcctatgaca ccttgcgctc       3781
ccaatcgcac tagccggcct accaccacag ctttaa]tact gcaggaattg tgcctgaatg      3841
tttaagggcc accttgatag cgtggctgat aggggttcaa gtcccctcaa ttctagagag      3901
aaggggctcg aacccatcct caagagatca aaactcttgg tgcttccact acaccacttt       3961
                ← ND1B
ctagtaaggt cagctaaata agcttttggg cccatacccc aaatatgttg gttaaaatcc       4021
            ND2
ttccctcact a[atgaaccct tatgtactca ccatcttgct ttcaagcctc ggcctgggca      4081
cagtcctcac ctttgccagc tcccactgac ttcttgcatg aataggactt gaaatcaaca       4141
cccttgccat cattcctctc atagcacggc aataccaccc ccgagcagtt gaagctacaa       4201
caaaatactt cttaacacaa gccaccgctg cagccataat cctattcgct agcaccacta       4261
atgcttgact ggttggagaa tgggatattc aacaattaac ccacccaatt gcagttacaa       4321
ccgccatgct ggcccttgca cttaaggtag gcctggcacc ggtacacttt tgactcccag       4381
aagttctcca aggtctagac ctcaccaccg ggttaatcct ttcaacctga caaaagcttg       4441
cgcccttttgc acttatgctc caggtagccc caaccgttaa ctcctcccta attgtcacac      4501
tgggactcct atcaactctc gttgggggct gggggggact taaccaaacc caactacgta       4561
agatcctagc atactcctca atcgcccacc ttgggtgaat agtgctgatt atacaattcg       4621
cgccttccct caccctcctg agcctgatca tatatattat cataacatct tcagccttta       4681
taacactaaa aactaataac tccctaacca tcaatgccct tgcaatctcc tgaactaaag       4741
caccaaccct ggccgcacta gccatcctag tccttctatc acttgggggc ctcccgcctc       4801
tctcaggttt tatacctaaa tgactaattc tacaagaact gacaaagcaa gggctgccaa       4861
tatccgccac actagctgcc atgacagccc ttctcagcct atacttctac ctacgactat       4921
gttacgccat aaccttgact atctgaccca ataccctcgc cgccactacc ccctgacgac       4981
tggactttac ccacaccacc ctgccgctat cgcttgtcac tatgttagcc ctaggcctac       5041
ttccccttac cccagccgta gctgccttac taaacttata a]caagggctt aggatagcac      5101
taagaccaag aaccttcaaa gttctaagcg ggagtgagaa tctcccagcc cttgttaaga       5161
cttgcgggac tctatcccac atcttctgaa tgcaacccag acactttaat taagctaaag       5221
cctttctagg tgggaaggcc tcgatcctac aaactcttag ttaacagcta agcgctctat       5281
ccggcgagca tccacctact ttcccccgcc accggggtgg cgaggcgggg gaaagccccg       5341
gtaggctgtt agcctacttc ttcaggtttg caatctgaca tgtaagtaca ccacaaggct       5401
```

FCo1 →

tgataaggag agggtttaaa cctctgttca tggggttaca atccaccgct taactctcag 5461

COI

ccaccttacc t[gtggcaatc acacgatgat ttttctcaac caaccacaaa gacattggca 5521

ccctttattt agtatttggt gcctgagccg gaatagtcgg cacagcccta agccttttaa 5581

tccgagcgga actaagccaa cccggggctc ttctgggggа tgatcagatt tataatgtaa 5641

← RNADH2

tcgtcacggc ccacgccttc gttatgattt tctttatagt tatgccaatt atgattggag 5701

gctttggaaa ctgattaatc ccacttataa tcggggcccc cgacatggca tttccccgaa 5761

tgaataatat gagcttttgg ctccttcccc cctcctttct ccttctcctg gcctcgtccg 5821

gagttgaagc cggtgccggc acaggatgaa cagtctaccc ccctctggca ggcaaccttg 5881

cccacgcagg agcctccgtc gatttaacta ttttctccct ccacttagct ggtatttcct 5941

ctatcttggg ggccgttaat tttattacaa ccattattaa catgaaaccc ccagctattt 6001

cccagtatca aacccctctt tttgtctggg ccgtcttaat taccgcagtg cttctactgc 6061

tttcccttcc tgtcctagca gcaggtatta ccatgctact cacagaccgg aatctgaaca 6121

ccactttctt tgacccagcg ggcgggggag atccaatcct gtatcaacac ctcttctgat 6181

tctttggtca tccggaagtc tacattctaa ttctccccgg ttttggtatg atctcccaca 6241

ttgttgcata ctactccggc aaaaaagaac ccttcgggta tatgggaata gtctgagcta 6301

tgatagccat tggactacta ggctttatcg tctgggccca ccatatgttt actgtcggga 6361

tggatgttga cactcgtgcc tactttacat ctgccactat gatcattgcc attcctacgg 6421

gtgtaaaagt gtttagctgg ttagccacat tgcatggcgg ttcaatcaaa tgggaaacgc 6481

cacttctttg ggccctgggg tttattttcc tatttacagt gggaggactg acaggcattg 6541

tcctagcaaa ttcctccctg gacatcgtcc ttcatgacac ctactacgta gtcgcccact 6601

tccactacgt tctatcaata ggagctgttt tcgccattat aggcgctttc gtgcactgat 6661

tccccctatt caccggatat actcttcaca gcacatgaac taaaatccac tttggaatta 6721

tgtttattgg cgtaaattta accttcttcc cccagcattt cctaggcctt gcgggaatac 6781

cacgacggta ctctgattac cccgatgcct acacactctg aaacactgtc tcttcaatcg 6841

ggtccctcat ctccctcgtt gctgtaatta tattcttatt tatcctttga gaagcctttg 6901

ccgccaaacg ggaggtcgca tcaattgagc tgacctcaac aaacgtagag tgactacacg 6961

ggtgtcctcc gccctatcac acattcgagg aaccggcgtt tgtacaagta caagcagcct 7021

Fcox2cor →

aa]cgagaaag ggaggaattg aacccccatg tgctggtttc aagccaaccg cataaccact 7081

ctgccacttt cttccataag acactagtaa aactagtata ttacactgcc ttgtcaaggc 7141

COII ← **RCoI**

aaaattgtgg gttaaaaccc cgcgtgtctt gagcacttag ctaca[atggc acatccctca 7201
caactaggat tccaagacgc ggcctcacct gtaatagaag aacttcttca cttccacgac 7261
catgctctta tgattgttct tcttatcagc acactagtgc tttatatcat cgtagcaata 7321
gtctccacta aacttactaa caaatatatc ctcgattctc aagaaattga gatcgtttga 7381
actgtccttc cagcagttat tcttattctt atcgctctcc cctccctccg aattctctat 7441
cttatagacg aaattaacga cccccacctt accatcaaag caatgggaca ccagtgatac 7501
tgaagttacg aatacactga ctacgaggac ctgggctttg actcctacat aatccccacc 7621
gtggaatctc caatccgagt tctagtctca gctgaagatg tccttcattc ctgagctgtc 7681
ccttctttag gtgtaaaaat agacgccgtc cccggacgtt taaatcaaac agcctttatt 7741
gcctctcgac ccggggtatt ctacggacag tgttctgaaa tttgcggtgc taaccacagc 7801
ttcatgccca tcgttgtcga agcagtgccc cttgaacact tcgagaaatg atccactata 7861
atacttgaag atgcct]cact aagaagctaa atagggaata gcgttagcct tttaagctaa 7921

ATPase 8

agattggtgg cccccaacca cccctagtga c[atgccccaa ctcaaccccg ccccctgatt 7981

← **Rcox2lav** ATPase 6

tgccatcttg gtattctcgt gactggtttt cctaactgtt attcccccc[a aagtccttg]g 8041
ccacaccttc acaaatgagc ctacctcaca aagcactgaa aaagctaaac ctgaaccctg 8101
aaactgacca tgacactaag cttctttgac caatttataa gccccacata cctgggcatc 8161
ccactcattg ctgtagcact aaccctccca tgaattctct tcccaacccc ctctgcccgg 8221
tgactaaaca accgccttat cacactacaa gggtggttta tcaaccgatt cacccaacag 8281
cttcttctcc ccttgaacct aggaggccat aagtgagcag taatgctgac ctctttaata 8341
ctattcctaa ttaccctgaa tatattaggc cttcttccat acaccttcac cccgaccacg 8401
cagctctccc taaatatggg gcttgcagtt ccactatgac ttgcaacagt aattatcggt 8461
atacgaaacc aaccaactgc cgccctggga cacctcttac ctgaaggaac ccctgtccca 8521
cttatcccgg ttcttatcat catcgaaaca attagcctct tcatccgccc ccttgctcta 8581
ggcgtacggc ttacagccaa ccttacggca ggccaccttc taattcaact aattgcaaca 8641
gcagcctttg ttcttctacc cctgatacca acagtggcaa tccttactgc tattgtccta 8701
ttcctgctta cccttcttga gatcgccgtt gccataattc aagcctacgt cttcgtcctc 8761

Fcox3cor → COIII

ctattaagcc tttacctaca agaaaacgtc ta[a]tggcaca ccaagcacac gcataccaca 8821
tggtcgatcc aagcccctga cccttaaccg gcgcaattgc cgccctttta ctcacatcag 8881
gcactgcagt ctgattccac ttccactcgc tcacactcct agctatggga aatattctta 8941
tacttctcac tatataccaa tgatgacgag atattattcg agagggcaca ttccaaggac 9001
accacacgcc ccccgtccaa aaaggcctac gctacggcat agttctattt atcacctccg 9061
aagtattctt tttcttgggt ttcttttggg ccttctatca ttctagtctt gcccccacac 9121

```
ccgaactagg gggctgctga ccccccacgg gcattattac tcttgacccc tttgaagtcc      9181
cgctactgaa caccgcagtc ctcctagcat ctggtgttac cgttacatga gcccaccaca      9241
gcattatgga aggtgaacga aaacaggcca tccaatctct caccctaact attttactgg      9301
gattctactt cacctttctc caaggtatgg agtactacga agcgcccttc acaatcgctg      9361
acggcgtata tggctccact ttctttgtag ccacaggctt ccacggccta cacgtaatta      9421
ttggctccac cttcctagcc gtctgtctgc ttcgacagat tcaatatcac tttacatcag      9481
aacatcactt tggctttgaa gctgccgcct gatattgaca ctttgtggac gtagtctgac      9541
                                                      ND3/4A →
tattccttta cgtctctatc tactgatgag gctcata]gtc tttctagtat taatgcgtat      9601
                                                    ND3
aagtggcttc caatcacccg gtcttggtta aaacccaagg aaagata[atg aacttaatca      9661
                                                 ← Rcox3lav
caacaattgt taccatcacc atcgcactat ccatagtact ggctactgtt tctttctgac      9721
taccacagat cacaccggac gcagagaaac tgtcccccta tgagtgcgga tttgaccccc      9781
taggatctgc ccgactacct ttctccctgc gctttttcct tatcgccatt ttatttcttt      9841
tatttgacct agagatcgcc cttcttctac ccctcccatg aggagaccaa ctagacaccc      9901
ccaccctaac acttgcttga tccgccgccg tccttacctt gctcaccctt ggcttgattt      9961
acgagtgaac ccaaggaggc ctagaatggg ctgaat]aggc agttagtcca aaaataagac     10021
                                                   ND4L
ccttgatttc ggctcaaaag accatggttt aagtccatga ccgccttt[atg acaccagtac     10081
acttcagctt tacctcagcc tttgttctag gactgatagg gctcgcattc caccgcaccc      10141
accttctctc agcccttctt tgcctagaag gaataatact ctctctattt attgccctat      10201
ccctctgggc acttcaaata gaagcaaccg gttattcggt agcccccatg ctattgctag      10261
ccttctcagc ctgtgaagcc agcgcaggcc tcgccctact agtagcaact gcacggacac      10321
                                          ND4
acggtacgga ccgcctccaa agcctcaacc ttctcca[atg ttaa]aaattc tcatcccaac     10381
actaatgcta tttcccacaa tctggcttag tcctgcaaaa tgactatgag caacatcaac      10441
agcccaaagc ctacttattg cactagcaag cttatcctgg ctcaagtggt catcagaaac      10501
cggctgaacc tcctccaacc tttatttagc cacagacccc ttgtcgacgc ccctcctagt      10561
attaacctgc tgactacttc ccttaataat cctcgcaagt cagaaccaca tcaacccaga      10621
gcccccttaat cgccaacgga cctacatctc tctcttagtc tctcttcaga tattcttaat     10681
cttagcattt ggtgccacag agatcattat gttttatatc atatttgaag ccacacttct      10741
tccaactcta attatcatca cccggtgggg caaccagaca gaacgcctca gtgctggcac      10801
ctacttctta ttctacacct tagccggctc cctgcccctc cttgtagccc tcctcctcct      10861
acaaaatgac aatgggacac tgtccatgct gacactgcaa tatacacagc ccctccacct      10921
tctaacatga ggagataagc tatggtgagc tgcctgtctc ctagcctttc ttgtaaaaat      10981
```

```
acccctatac ggcgtccacc tctgactccc caaagcccat gtagaagccc caatcgcagg      11041
ctctatggtc ttagctgccg tccttcttaa gctgggcgga tacggcataa tacgaataat      11101
aattatgcta gacccccta ccaaggaact agcatatcca ttcattgtcc tagccctctg      11161
aggtattatt ataaccggat ccatttgcct ccgtcaaacg gacctgaaat cactgatcgc      11221
ttactcctcc gtaggacata taggattggt cgcaggggc attctagttc aaacaccctg      11281
gggatttact ggcgcaatta tccttataat tgcacacggc cttgcctctt cagcactatt      11341
ctgcctagca aatacaagct acgaacgaac acacagccgt accatacttc tagcccgagg      11401
gatacaaata attctccctc taataaccac ctggtgattt gtagccagtt tagccaacct      11461
agccctccca cctcttccta atctaatggg agaactaata atcatcacca ccatatttaa      11521
ctgatcacac tgaactcttc tcctcacagg agtcggaacg ctgattacag ccagctattc      11581
cctctattta ttcctaataa cccaacgagg gcccctacct tctcatatta ttgcccttga      11641
acccacccac acccgtgagc acctgcttat caccctacat cttatcccca tcatcctcct      11701
aatcctaaaa ccggagctca tgtgaggctg atgtttct]gt agatatagtt taactaaagc      11761
                                                  FNADH5/6 →
attagattgt gattctaaag acagaggtta aagccctctt atccaccgag agaagtctgt      11821
tgacagtagg gactgctaat cttctacccc ctcggttaaa ctccgtggtt cactcgtgct      11881
                          ← ND3/4B
tctaaaggat aatagctcat ccgttggtct taggaaccaa agactcttgg tgcaactcca      11941
            ND5
agtagcagct [atgcacccaa ccacactcat cttaagctca acccttctaa tgatcttggc      12001
acttcttatt tatccccttt tgaccaccct tgaccccaac caacgacccg gaaactgagc      12061
cctcacccac gttaagactt ccgtcaagat ggcttttctg gtaagcctac tccccctatt      12121
cattttctt gaccagggga cggagacaat tgtcaccaac tggcaatgaa tgaacacctc      12181
aacctttgat gtaaacctca gctttaaatt tgaccactac tccatcatct tcacccctat      12241
tgccctctat gtaacctgat caattcttga gtttgcatca tgatatatac atgctgatcc      12301
caacatgaac cgattcttta aataccttct cctgtttttg gtagccataa tcgttctagt      12361
gactgctaac aatatgttcc aactgtttat cggctgagaa ggtgttggta ttatatcatt      12421
cctccttatc gggtgatggt acggccgagc cgatgccaac acagctgcca tacaagccgt      12481
tgtgtataac cgagtcggag atattgggct tatcttaagc atggcttgat ttgcaacaaa      12541
cctcaactct tgagaaattc aacaaatgtt cgcctcatca aaagaccttg acctaacact      12601
cccactcata ggcctcatct tagccgccac tggcaaatca gcacaatttg gacttcatcc      12661
ctggctccct tccgcaatgg aaggtcctac gccggtatct gccctgctgc actctagcac      12721
catagttgtt gcaggcattt tcctactaat ccgactccac ccccttatgg agaacaacca      12781
aacagccctt accacctgct tgtgccttgg tgccctaacc acgctattca ccgctacttg      12841
cgccctgaca caaaacgaca tcaaaaaaat cgtcgcattc tccacatcca gtcaactagg      12901
actgataatg gtcaccatcg gacttaacca gccacagcta gccttccttc acatctgtac      12961
```

```
ccacgccttc tttaaagcca tactattcct gtgttcagga tcaatcattc acagcctaaa      13021
cgatgagcag gacatccgaa aaatgggggg tatacacaac ctcaccccct tcacctcttc      13081
ctgccttaca attggtagcc tcgcacttac cggcaccccc ttcttggcag gattcttctc      13141
caaagatgct attattgaag ccttgaatac atctcacctc aacgcctgag ccctaaccct      13201
taccctactg gccacctctt ttacagccgt atacagcctc cgagttgtgt atttcgtatc      13261
tatgggacac cctcgtttta cagccatctc accaattaat gaaaataacc cctccgtcat      13321
taacccgatc aagcgattag cctggggaag cattgttgca ggccttctga tcacctcaaa      13381
cttcctcccc tctaaaaccc ccgttataac aatacccccc gccctaaagc tagccgccct      13441
cctggtcact attttaggtc ttcttgttgc actagaactt gcgtccttga ccagcaagca      13501
atttaagacc acacccaacc ttgttacaca caacttctcc aacatattag gctttttccc      13561
cgccattgtc caccgactag cccccaagct taacctgacc ctcggtcagg ccattgccag      13621
ccaaatagta gatcaaacat ggtttgaaaa gattggacca aaaggagttg tatccaccca      13681
tctgcccata attacaacaa ccagcaacat gcaacaaggg atgattaaaa catacctcac      13741
cctattcttc ctttcaacaa ccctagctgt cttattaacc tcaac[ctaa]a ctgcccgaag    13801
ggctccccga ctaagacccc gagtcaactc tagtaccaca aatagtgtta gcaagagaac      13861
ccatgcacac accaccaaca gccccccacc ggaagagtat attagggcca ccccgcttgt      13921
gtcccctcgc acaacagaaa actccttaaa ctcatccacc gcaactcatg aagtctcgta      13981
tcacccaccc caaaatcaac ccgccaccag agccaccccc accacataag ccaccacata      14041
acctaaaact gaacgatcac ctcaagactc aggaaacggc tcggcagcta aagcagctga      14101
gtaagcaaac actacaagta tccccccccaa ataaatcaag aatagtacta aagacaagaa      14161
cgacccccca tgccccacca aaacaccaca gcccacgccc gctgccacta ccaaccccag      14221
ggcagcaaag tagggagcag gatttgatgc aacagctaca agacccaaaa ccaaccccaa      14281
                        ND6 (complement)
tagaaataaa gacacaagat aagtcat]aat tcctgctcgg actctaaccg aaactaatga      14341
            CYTB →                          cytb
cttgaaaaac caccgttgtt attcaactac aagaaccta[a tggctaacct ccgaaaaacc      14401
            ← RNADH5/6
cacccctcc taaagattgc taatgacgca ctagtcgatc ttccagcgcc ctcaaacatc       14461
tcagtgtgat gaaactttgg ctcacttctg ggcttatgtt tagccaccca aattctcaca       14521
ggacttttcc tggccatgca ctacacttct gacatctcaa cagccttctc ctccgtatgc       14581
catatttgcc gagatgtcag ctacggctga cttatccgaa atattcacgc caatggagca       14641
tctttcttct ttatctgcat ttatatacac attgcccgag gactttacta tggctcctac       14701
ctatacaaag aaacctgaaa tattggggtt gtccttctac tccttacaat aataactgcc       14761
tttgtgggct acgttcttcc atgaggacaa atatctttct gaggtgcaac agtcatcaca       14821
aacctccttt ctgccgtgcc ctacgtaggg ggcgcccttg tgcagtgaat ctgaggtggg       14881
ttttccgtag ataatgccac cctaacacgg ttttttgcct ttcacttctt attcccattt       14941
```

gttattgcag ctgcaacagt catccacctc ctcttccttc atgaaactgg gtctaacaac 15001
ccagcaggga tcaactctga tgccgataaa atctcattcc acccctactt ctcatacaaa 15061
gacctgttgg gatttgtggc tatactacta ggactaacat ccttggcact atttgcaccc 15121
aacctcttgg gggacccaga taattttaca ccggccaacc cactagtcac cccgccccac 15181
atcaagcctg agtgatactt cctcttcgcc tacgcaatcc tacgatcgat ccccaacaag 15241
ctaggcgggg tcctcgccct gctattctct atcctcgtac tcatggtcgt ccccattcta 15301
catacctcta agcagcgagg actaaccttc cggcccctta cgcaattctt attctgaacc 15361
ctagtagcag acatgctcat cctcacctga attggcggta tacctgtaga acacccccttc 15421
atcatcatcg gccaagtcgc ctcagttatc tacttcacca tcttcctagt cttggccccc 15481
ttagccggat gagctgagaa taaagccctc gaatgagcct] gccctagtag ctcagtgtaa 15541
gagcgccggt cttgtaatcc ggaggccgga ggttaaaccc ctccctagtg ctcagagaga 15601
ggagatttta actcccaccc ttaactccca aagctaagat tctaaattaa actaccctct 15661

D-loop

g[acgccgcag tgtacatggt aaatcatata tccccataca gtgtgtgtat attacaccac 15721
tatgtataat attgcatatt atgtactgac ccatatatta ttattgcacg taggtagtac 15781
atcctatgta ttatcaacat aagtgatttt aagccctcat acatcagtac tgttccaagg 15841
tttacataag caagactcgg ataatcacca acggaaccgt tctaacctga ttaattgcta 15901

Char2 →

aacaacaaac ctccaactaa cacgggctcc gtctttaccc accaaatttc agcatcggtc 15961
ccgtttaatg tagtaagaac cgaccaacga tttattagta ggcatactct taatgatggt
16021 **← H2B**

cagggtcaaa tatcgtatta ggtagcatct cgtgaattat tacttgcatc tggttcctat 16081
ttcatgggct atccttaaga aaccacccccc tgaaagccga atgtaatgca tctggttaat 16141
ggtgtcaacc ttactgttcg ttacccacct agccgggcgt tctcttatat gcatagggtt 16201
ctccttttttt ttttttcctt tcagcttgca tatacaagtg cacaccgaga agtctaacaa 16261
ggtcgaacta gatcttggtc tccagcggac ccaataataa tggcggaatg atattctata 16321
aagaattgca taattgatat caagtgcata aggtcagttt ctttcctcac agatacctaa 16381
gatctccccg gcttttgcgc ggctaaaccc ccctaccccc ctacgctgag cgatccttat 16441
tattcctgtc aaaccccaaa accaggaagt ctcgatagcg ctattaccca tcaaaccata 16501
cattaacaaa ctttggcacc gacaatccta ttatcaaagc caccccttaa ttaaagtgta 16561
cactaaaact ttttattata cattaataaa ctttattact tacaaacttt ggcaccgaca 16621
atcctattat caaagccacc ccttaattaa agtatacatt aataaaattt ttgttatact 16681
taacaaactt tggcaccgac aaccctatca tgaaggccac tcctggttaa aatatat]

Appendix 1. Fish taxa used in this study.

Acronym	Nominate taxon (common name)	Fish ID[1]	Sample Location	Location codes[1]
LKCS	*Coregonus artedi* (Lake cisco)	45769 45770	Lake Athapapaskow (S. Basin), Manitoba	00-21-02
SJCS	*Coregonus zenithicus* (Shortjaw cisco)	46317 46318	George Lake, Manitoba	00-50-01
LSCS	*Coregonus sardinella* (Least cisco)	43297 43298	North Killeak Lake, Alaska	96-24-05
ARCS	*Coregonus autumnalis baicalensis* (Baikal cisco)	44418 44419	Lake Baikal, Russia	98-19-01
ARCS	*Coregonus autumnalis* (pollan) (Irish pollan)	36303 36319	Loch Neagh, Ireland	91-29-01
ARCS	*Coregonus autumnalis* (Arctic cisco)	32051 32052	Atkinson Point, Northwest Territories	88-38-01
LVRT	*Coregonus lavaretus pidschian* (European whitefish)	71080 71081	Kuluymbe River, Northern Siberia	00-61-01
PELD	*Coregonus peled* (Peled)	71270 71271	Khantaiskoe Resevoir, Northern Siberia	00-60-01
LKWF	*Coregonus clupeaformis* Mississippian form (Lake whitefish)	28308 28309	Big Black River, Manitoba	89-30-01
LKWF	*Coregonus clupeaformis* Beringian form (Lake whitefish)	37215 37216	Klutina River, Alaska	91-12-02
BDWF	*Coregonus nasus* (Broad whitefish)	41306 41307	Little Chicago, Mackenzie River, Northwest Territories	95-05-01
MUKS	*Coregonus muksun* (Muskun)	M1 M2	Middle Ob River, Siberia	
BRCS	*Coregonus laurettae* (Bering cisco)	36812 36813	Yukon River, Alaska	91-32-01
RSKA	*Coregonus lavaretus* Reeska form (Reeska)	36071 36702	Lake Inari, Finland	91-26-01
BVCS	*Prosopium gemmifer* (Bonneville cisco)	47289 47290	Bear Lake, Utah	00-69-01
ARCH	*Salvelinus alpinus* (Arctic char)	70444 70445	Taserssuit Lake, Greenland	00-40-01

1. Fish identification (ID) number and location code cross-reference the sample to the (FWI AFEAR) internal fish sample database.

Appendix 2. Thermal cycling programs for amplification of mitochondrial DNA[a].

1. 12S rRNA, partial d-loop
   Forward primer: CHAR2 5'-CAA AAC TCC AAC TAA CAC GG-3'
   Reverse primer: R12S 5'-AAC CTC TGA GAA AAA GGG C-3'
   Thermal cycling program: -2 min. @ 95°C
   -30 sec @ 94 °C, 35 sec @ 50 °C, 2 min 30sec @ 70 °C for 30 cycles
   Results in a product ~2000 bases in length.

2. 16S rRNA
   Forward primer: F16S 5'-ACA CCG AGA AGA CAT CCG TG-3'
   Reverse primer: R16S 5'-AAC CTC TGA GAA AAA GGG C-3'
   Thermal cycling program: -2 min @ 95 °C
   -45 sec @ 95 °C, 1 min @ 60 °C, 2 min @ 72 °C for 30 cycles
   Results in a product ~1800 bases in length.

3. ND1
   Forward primer: ND1A 5'-GCC TCG CCT GTT TAC CAA AAA CAT-3'
   Reverse primer: ND1B 5'-GGT ATG GGC CCG AAA GCT TA-3'
   (Nielsen *et al.*, 1998)
   Thermal cycling program: -2 min @ 95 °C
   -45 sec @ 95 °C, 30 sec @ 50 °C, 2min 30 sec @ 70 °C for 8 cycles
   -45 sec @ 95 °C, 30 sec @ 53 °C, 2min 30 sec @ 70 °C for 21 cycles
   Results in a product ~2000 bases in length.

4. ND2
   Forward primer: FNADH2 5'-AAA AGC CGC CCT ACT TAC CG-3'
   Reverse primer: RNADH2 5'-ATC ATA ACG AAG GCG TGG GC-3'
   Thermal cycling program: -2 min @ 95 °C
   -30 sec @ 94 °C, 35 sec @ 55 °C, 2 min 30 sec @ 72 °C for 30 cycles
   Results in a product ~2000 bases in length.

5. COI
   Forward primer: FCoI 5'-ATG GGG TTA CAA TCC ACC GC-3'
   Reverse primer: RCoI 5'-GTG AGG GAT GTG CCA TTGTAG C-3'
   Thermal cycling program: -2 min @ 95 °C
   -30 sec @ 94 °C, 35 sec @ 55 °C, 2 min 30 sec @ 72 °C for 30 cycles
   Results in a product ~1800 bases in length.

6. COII
   Forward primer: Fcox2cor 5'-CAA GCC AAC CGC ATA ACC-3'
   Reverse primer: Rcox2lav 5'-ACG AGA ATA CCA AGA TGG-3'
   Thermal cycling program: -2 min @ 95 °C
   -45 sec @ 95 °C, 1 min @ 55 °C, 2 min @ 72 °C for 30 cycles
   Results in a product ~1100 bases in length.

7. COIII
   Forward primer: Fcox3cor 5'-CCT TTA CCT ACA AGA AAA CG-3'
   Reverse primer: Rcox3lav 5'-CAG AAA GAA ACA GTA GCC AG-3'
   Thermal cycling program: -2 min @ 95 °C
   -45 sec @ 95 °C, 1 min @ 55 °C, 2 min @ 72 °C for 30 cycles
   Results in a product ~950 bases in length.
   Note: Can alternatively use the following program (8) to amplify Cox2, Cox3 and everything in between.

8. COII, COIII, ATPase6, ATPase8
   Forward primer: Fcox2cor 5'-CAA GCC AAC CGC ATA AAC-3'
   Reverse primer: Rcox3lav 5'-CAG AAA GAA ACA GTA GCC AG-3'
   Thermal cycling program: -2 min @ 95 °C
   -45 sec @ 95 °C, 1 min @ 55 °C, 2 min @ 72 °C for 30 cycles
   Note: Must use 4 *u*l of dNTPs/sample to be amplified due to product length.
   Results in a product ~2700 bases in length.

9. ND3, ND4
   Forward primer: ND3/4A 5'-TTA ATA CGT ATA AGT GAC TTC CAA-3'
   Reverse primer: ND3/4B 5'-TTT TGG TTC TA AGA CCA ATG GAT-3'
   (Nielsen *et al.*, 1998)
   Thermal cycling program: -2min @ 95 °C
   -45 sec @ 95 °C, 30 sec @ 50 °C, 2min 30 sec @ 70 °C for 8 cycles
   -45 sec @ 95 °C, 30 sec @ 53 °C, 2 min 30 sec @ 70 °C for 21 cycles
   Results in a product ~2300 bases in length.

10. ND5, ND6
    Forward primer: FNADH5/6 5'-CTT ATC CAC CGA GAG AAG TCT GTT G-3'
    Reverse primer: RNADH5/6 5'-TGC GTC ATT AGC AAT CTT TAG GAG-3'
    Thermal cycling program: -2min @ 95 °C
    -30 sec @ 94 °C, 35 sec @ 55 °C, 2 min 30 sec @ 70 °C for 30 cycles
    Note: Must use 4 *u*l of dNTPs/sample to be amplified due to product length.
    Results in a product ~2600 bases in length.

11. cyt *b*, partial D-loop
    Forward primer: CytB 5'-GAA AAA CCA YCG TTG TWA TTC AAC T-3'
    (Miller *et al.*, 1998)
    Reverse primer: H2B 5'-GCC AGG AAT AAT TCA CGA G-3'
    Thermal cycling program: -2min @ 95 °C
    -1 min @ 94 °C, 1 min @ 48 °C, 1 min 30 sec @ 72 °C for 32 cycles
    Results in a product ~1700 bases in length.

Note: The d-loop has been amplified in two separate yet overlapping segments (1 & 11), but can also be amplified intact using the following program (12) with limited success.

12. D-loop
    Forward primer: LN20 5'-ACC ACT AGC ACC CAA AGC TA-3'
    Reverse primer: HN20 5'-GTG TTA TGC TTT AGT TAA GC-3'
    (Bernatchez and Danzmann, 1993)
    Thermal cycling program: -4min @ 95 °C
    -1 min @ 94 °C, 1 min @ 48 °C, 1 min 30 sec @ 72 °C for 32 cycles
    Note: Must use 4 *u*l of dNTPs/sample to be amplified due to product length. Results in a product ~1150 bases in length.

---

[a] All nucleotides are given according to their IUB codes (IUB nomenclature committee, 1985).