 doi: 10.1111/j.1755-0998.2008.02350.x

DNA BARCODING

# DNA barcoding of eight North American coregonine species

ORA L. SCHLEI,* ALEXIS CRÊTE-LAFRENIÈRE,† ANDREW R. WHITELEY,† RANDY J. BROWN,‡ JEFFREY B. OLSEN,* LOUIS BERNATCHEZ† and JOHN K. WENBURG*
**US Fish and Wildlife Service, Conservation Genetics Laboratory, MS 331, 1011 E. Tudor Road, Anchorage, AK 99503, USA, †Département de Biologie, Québec Océan, Pavillon Charles-Eugène Marchand, Université Laval, QC, Canada G1 K 4P7, ‡US Fish and Wildlife Service, Fairbanks Fish and Wildlife Field Office, Room 222, 101 12th Ave., Fairbanks, AK 99701, USA*

## Abstract

**Coregonine fishes have a circumpolar distribution in the Arctic and sub-Arctic Northern Hemisphere. This subfamily of Salmonidae consists of three genera: *Prosopium*, *Stenodus* and *Coregonus*, including over 30 species. Many species overlap spatially and are difficult to distinguish based on morphological characteristics, especially as larvae or juveniles. Here we present a method for rapid and cost-effective species identification for representatives of the three genera based on sequence variation at the mitochondrial cytochrome *c* oxidase subunit I gene (COI). We examined eight species common to North America with distributional overlap in Alaska. Mean pairwise sequence divergence for all eight species was 7.04% and ranged from 0.46% to 14.23%. This sequence variation was used to develop a genetic assay based on restriction fragment length polymorphism. In a blind test, this assay provided correct species assignment for 48 of 49 individuals representing all eight species. The single incorrect assignment may reflect hybridization between two closely related species. This DNA barcode-based assay promises to aid fishery managers and researchers by providing a cost-effective alternative to large-scale sequence analysis for identification of North American coregonine fishes.**

*Keywords*: coregonine, cytochrome *c* oxidase I, DNA barcoding, RFLP, Salmonidae, whitefish

*Received 1 May 2008; revision accepted 15 July 2008*

## Introduction

Coregonine fishes occur in the Northern Hemisphere and have a broad circumpolar distribution. They consist of three genera; *Prosopium*, *Stenodus* and *Coregonus* and over 30 species (Bernatchez 2004). Given this species diversity, species identification is often difficult because phenotypic characteristics vary depending on environment, life stage and other factors. This problem is compounded by the fact that migratory behaviour (anadromy/potamodromy) can vary within species and the spatial distributions of some species broadly overlap (Brown *et al*. 2007). As a result, information on basic biology, including abundance and distribution, particularly at the larval and juvenile stages, as well as migratory behaviour, spawning location and population structure, is limited for species in many areas. Due to increased subsistence and commercial demand, research and management applications require correct identification of many individuals at a time and it remains costly to sequence the DNA of all individuals in question. A tool for rapid and cost-effective identification of co-occurring coregonine species would be of great value.

The global bio-identification system that uses the mitochondrial cytochrome *c* oxidase subunit I (COI, the DNA barcode gene region) has proven to be an effective tool in species identification (Hebert *et al*. 2003a, 2003b). The advantages of the COI gene include the availability of robust, universal primers and a more expansive range of genetic signal than other mitochondrial genes (Hebert *et al*. 2003a). The COI gene also exhibits a rate of evolution which appears to consistently provide discrimination at both inter- and intraspecific levels (Hebert *et al*. 2003a). DNA barcoding has successfully been used to date to distinguish among taxa of varying degrees of relatedness (Hebert *et al*. 2003b; Spies *et al*. 2006; Smith *et al*. 2008). This method of DNA barcoding can serve as the foundation for further development of genetic assays such as restriction

Correspondence: Ora L. Schlei, Fax : 907-786-3978 ;
E-mail: ora_schlei@fws.gov

Journal compilation © 2008 Blackwell Publishing Ltd
No claims to original US government works

fragment length polymorphisms (RFLP), single nucleotide polymorphisms (SNP), or macroarrays, which offer a relatively inexpensive and high throughput means for potential species identification.

Previous studies on whitefish have shown promising results using mitochondrial DNA for species identification. Politov *et al*. (2000) utilized the mitochondrial ND-1 gene and allozyme loci and are able to differentiate seven of eight Palearctic coregonine species. Although not based on COI, their technique supports the effectiveness of mitochondrial DNA-based approaches for whitefish species identification.

The first objective of this study is to develop an RFLP assay based on the sequence variation of COI for identification of eight coregonine species common to Arctic and sub-Arctic North America that overlap in Alaska; inconnu (*Stenodus leucichthys*), round whitefish (*Prosopium cylindraceum*), pygmy whitefish (*Prosopium coulterii*), broad whitefish (*Coregonus nasus*), Arctic cisco (*Coregonus autumnalis*), Bering cisco (*Coregonus laurettae*), humpback whitefish (*Coregonus pidschian*) and least cisco (*Coregonus sardinella*). The second objective is to evaluate the performance of this assay using a blind test.

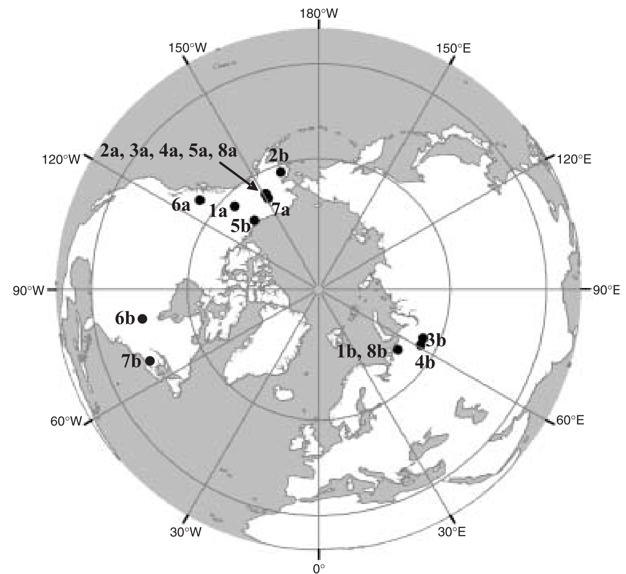


**Fig. 1** Locations of coregonine sample collections: *Coregonus autumnalis* (1a, Peel River, Canada; 1b: Pechora River, Russia); *C. laurettae* [2a, Yukon River (Rapids) Alaska, USA; 2b, Yukon River (Marshall) Alaska, USA]; *C. pidschian* [3a, Yukon River (Rapids) Alaska, USA; 3b: Lyapin River, Russia]; *C. nasus* [4a, Yukon River (Rapids) Alaska, USA; 4b, Sos´va River, Russia]; *C. sardinella* [5a, Yukon River (Rapids) Alaska, USA; 5b, Shingle Point, Canada]; *Prosopium coulterii* (6a, Chapman Lake, Canada; 6b, Lake Superior, Canada); *P. cylindraceum* (7a, Kanuti River, Alaska, USA; 7b, Musquacook Lake, Maine, USA); *Stenodus leucichthys* [8a, Yukon River (Rapids), Alaska, USA; 8b, Pechora River, Russia].

## Materials and methods

### *Samples and DNA sequencing*

Whole fin and muscle tissue samples from 16 individuals representing two specimens for each of the eight species were collected from North American and Russian locations (Fig. 1) and preserved in 95% ethanol. We followed taxonomic designations in accordance with Robins *et al*. (1991), but were aware that taxonomic confusion of whitefish species often produces inconsistencies with regard to nomenclature, especially in Eurasian populations (Bernatchez 2004).

Genomic DNA was isolated from each sample using DNeasy Tissue Kits (QIAGEN, Inc.) following their standard procedure. A 650-bp segment of the COI gene was sequenced using primers FishF1 and FishR1 from Ward *et al*. (2005). Polymerase chain reaction (PCR) was conducted on a PerkinElmer Cetus 480 DNA Thermal Cycler in 50-μL reactions containing 5 μL of genomic DNA (10–50 ng), 5 μL of 10× PCR buffer [500 mM KCl, 100 mM Tris-HCl (pH 9.0), 1.5 mM $MgCl_2$, 1% TritonX-100], 4 μL of dNTP mix (2.5 mM of each dNTP), 2 μL (20 pmol) of each primer and 1 U of *Taq* polymerase and 34 μL of distilled deionized water. The thermal regimen consisted of an initial cycle at 95 °C for 300 s followed by 35 cycles of denaturation at 95 °C for 60 s, annealing at 51 °C for 60 s and 72 °C extension for 90 s with a final hold at 4 °C. Amplification products were purified using the QIAquick Gel Extraction Kit (QIAGEN, Inc.) after electrophoresis on 1.2% low-melting point agarose gels. Amplicons were labelled using the BigDye Terminator version 3.0 Cycle Sequencing Kit (Applied Biosystems, Inc.) and sequenced in the forward direction using an ABI 3730xl capillary sequencer.

### *COI sequencing and RFLP assay development*

Sequences were aligned using BioEdit 7.0.9.1 (Hall 1999); (Table 1). Each unique COI sequence was submitted to GenBank (accession nos EU202649–EU202658, EU427541; Table 2). The sequence for each species was then screened using the freeware Sequence Analysis version 1.5.9 (http://informagen.com/SA/) to identify enzyme recognition sites and determine the length (bp) of the restriction fragments. Results were then compared across species to estimate sequence variation and to identify restriction enzymes that produced species-specific restriction fragments. Pairwise nucleotide-sequence divergences were calculated for all samples using the Kimura 2-parameter (K2P) model contained in the REAP version 4.0. software (McElroy *et al*. 1992). These pairwise values were used to generate a neighbour–joining tree using the program Neighbour in the computer software PHYLIP version 3.5 c (Felsenstein 1993).

Journal compilation © 2008 Blackwell Publishing Ltd
No claims to original US government works

**Table 1** Variable nucleotide sites in 650-bp sequences of cytochrome *c* oxidase subunit I in eight species of whitefish. Dots indicate nucleotide identity with the *Coregonus autumnalis* sequence

Nucleotide position

| Species | 4 | 45 | 54 | 66 | 75 | 84 | 87 | 88 | 93 | 96 | 102 | 114 | 120 | 123 | 129 | 132 | 141 | 156 | 177 | 190 | 195 | 198 | 204 | 207 | 210 | 213 | 219 | 222 | 228 | 234 | 237 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Coregonus autumnalis* | T | A | A | A | A | T | T | C | G | T | G | A | C | G | C | C | G | T | C | T | T | A | A | C | G | C | C | G | T | A | G |
| *Coregonus laurettae* | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *Coregonus pidschian* | . | . | . | . | . | . | . | . | . | . | . | R | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . |
| *Coregonus nasus* | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | C | . | . |
| *Coregonus sardinella* | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . |
| *Prosopium coulterii* | Y | C | . | . | G | C | C | T | A | . | . | G | T | A | . | . | . | C | . | . | C | C | G | T | . | A | T | A | . | T | A |
| *Prosopium cylindraceum* | . | . | G | G | G | C | . | . | . | C | A | . | . | . | T | T | A | . | A | C | . | C | G | . | A | A | T | A | C | . | A |
| *Stenodus leucichthys* | . | G | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . |

| Species | 243 | 246 | 252 | 255 | 258 | 264 | 267 | 270 | 276 | 282 | 283 | 291 | 294 | 297 | 309 | 315 | 321 | 330 | 333 | 336 | 339 | 342 | 345 | 348 | 354 | 357 | 366 | 375 | 378 | 384 | 387 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *C. autumnalis* | T | G | T | G | C | C | G | C | C | C | C | G | C | A | T | C | A | C | C | C | T | G | A | C | C | C | G | C | T | T | T |
| *C. laurettae* | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . |
| *C. pidschian* | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . |
| *C. nasus* | . | . | . | . | . | . | C | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . |
| *C. sardinella* | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *P. coulterii* | C | A | C | A | T | G | . | . | T | . | . | A | T | G | C | T | R | T | T | G | . | A | G | R | A | T | A | T | . | C | . |
| *P. cylindraceum* | C | . | C | A | T | . | A | . | T | K | Y | A | . | . | C | . | . | G | T | . | A | A | . | . | . | T | . | T | C | . | C |
| *S. leucichthys* | C | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . |

| Species | 396 | 399 | 405 | 411 | 414 | 420 | 423 | 426 | 447 | 453 | 456 | 468 | 471 | 474 | 477 | 480 | 489 | 492 | 501 | 504 | 507 | 510 | 511 | 516 | 522 | 525 | 526 | 528 | 529 | 532 | 534 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *C. autumnalis* | C | C | T | T | C | C | A | A | C | T | C | A | T | T | C | G | C | T | C | A | C | C | T | T | A | C | C | T | C | C | G |
| *C. laurettae* | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . |
| *C. pidschian* | . | . | . | . | . | . | G | G | . | . | Y | . | . | . | . | . | . | . | . | G | . | . | . | . | . | X | . | . | . | . | . |
| *C. nasus* | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . |
| *C. sardinella* | . | . | . | . | . | . | . | G | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *P. coulterii* | . | T | G | Y | A | T | . | T | T | C | T | . | C | C | . | A | G | C | A | . | T | T | C | . | . | A | T | A | . | . | A |
| *P. cylindraceum* | A | . | . | . | . | T | . | . | . | . | T | . | C | C | G | . | A | C | T | . | . | T | . | C | C | . | . | . | T | . | A |
| *S. leucichthys* | . | . | . | C | . | . | . | G | . | . | T | G | . | . | . | . | . | . | T | . | . | . | . | . | . | A | . | . | T | . | . |

| Species | 537 | 543 | 546 | 549 | 552 | 558 | 567 | 570 | 573 | 576 | 579 | 582 | 585 | 588 | 589 | 591 | 606 | 612 | 615 | 618 | 621 | 627 | 630 | 633 | 642 | 645 | 648 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *C. autumnalis* | C | T | T | C | A | A | C | G | A | C | A | C | G | T | T | A | T | A | G | C | G | T | G | C | A | T | C |
| *C. laurettae* | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *C. pidschian* | Y | . | . | . | . | . | . | . | . | . | . | . | . | . | C | G | . | . | . | . | . | . | R | . | . | C | . |
| *C. nasus* | T | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | A | . | . | C | . |
| *C. sardinella* | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | C | . | . | . | A | . | A | . | G | C | . |
| *P. coulterii* | T | . | C | A | G | G | G | A | . | T | . | T | C | . | C | . | . | . | A | A | . | C | A | T | G | C | . |
| *P. cylindraceum* | T | C | . | G | G | . | A | A | G | T | G | . | A | C | C | . | . | . | A | A | . | C | A | T | R | C | A |
| *S. leucichthys* | T | . | . | . | . | . | . | . | . | . | . | . | . | . | C | . | . | G | . | . | A | . | A | . | . | C | . |

Shaded sites indicate polymorphisms within species; K = G or T, Y = C or T, R = A or G, X = C or G.

© 2008 The Authors
No claims to original US government works

**Table 2** Whitefish species (common and scientific names), number of individuals sequenced (*n*) for cytochrome *c* oxidase subunit I and GenBank accession nos. All samples are available from L. Bernatchez. Figure 1 locations are shown in parenthesis following the GenBank number for species with multiple sequences

| Common name | Scientific name | *n* | GenBank Accession no. |
|---|---|---|---|
| Arctic cisco | *Coregonus autumnalis* (Pallas, 1776) | 2 | EU202649 |
| Bering cisco | *Coregonus laurettae* Bean, 1881 | 2 | EU202650 |
| Humpback whitefish | *Coregonus pidschian* (Gmelin, 1788) | 2 | EU202651 (3b) EU427541 (3a) |
| Broad whitefish | *Coregonus nasus* (Pallas, 1776) | 2 | EU202652 |
| Least cisco | *Coregonus sardinella* Valenciennes, 1848 | 2 | EU202653 |
| Pygmy whitefish | *Prosopium coulterii* (Eigenmann & Eigenmann, 1892) | 2 | EU202654 (6b) EU202655 (6a) |
| Round whitefish | *Prosopium cylindraceum* (Pallas, 1784) | 2 | EU202656 (7b) EU202657 (7a) |
| Inconnu | *Stenodus leucichthys* (Güldenstädt, 1772) | 2 | EU202658 |

Restriction digests were performed for each enzyme in the diagnostic suite according to manufacturer specifications (New England BioLabs, Inc.). A group of four to six individuals (different from those sequenced) from each species was used to verify that the assay was robust and repeatable. These samples were obtained in Alaska from mature individuals with distinctive, species-specific morphologies. DNA was isolated from fin tissue from each individual using DNeasy tissue kits. Amplification of the COI gene was conducted using the primers FishF1 and FishR1 (Ward *et al*. 2005) and the thermal regimen described above. PCR was conducted in 50-μL reactions with 5 μL 10× buffer, 2.5 mM $MgCl_2$, 1 U of *Taq* polymerase (Novagen, Inc.), 0.4 mM dNPTs, 0.1 μM of each primer, 3 μL of 30 ng template DNA, and 33.8 μL of deionized water. Products from the digests were loaded on 2.5% high-resolution agarose gels polymerized with ethidium bromide and run at 100 V for 2.5 to 3 h with 100–1000 bp molecular ladder (Bio-Rad Laboratories, Inc.) positioned in the first, middle and last lane of each gel. Visualization of each digest was conducted after electrophoresis with ultraviolet light to confirm fragment sizes.

### *Testing RFLP assay*

The accuracy of the RFLP assay was examined using a blind test with fifty individuals representing all eight species. Each species was represented by six individuals with the exception of *Coregonus pidschian*, which had eight. Samples were from the same populations as the group used in the assay development, but were not the same individuals used in the development. They were provided randomly without labels to the researcher conducting the assay as described in the development phase above. Each individual was assigned a species designation based on the results of RFLP analysis, and then these designations were compared to the known identities of the blind samples in order to quantify the accuracy of the test.

## Results and discussion

### *Sequence analysis*

Eleven COI sequences were identified from the eight species: five species each with a single sequence and three species, *Prosopium coulterii*, *Prosopium cylindraceum* and *Coregonus pidschian*, each with two sequences (Table 2). A total of 132 mutations were found in 120 segregating sites and polymorphisms within species occurred at 12 sites (Table 1).

In general, estimates of sequence divergence decreased progressively from the intergeneric to interspecific and then intraspecific levels; however, some variation was observed within taxonomic levels. For species with two different sequences, the individual values for each unique comparison were first averaged. Mean intergeneric sequence divergence varied sixfold and was highest (mean: 12.89%, STD: ± 1.02%; range: 11.77–14.23%, Table 3) between the *Prosopium* and *Coregonus*/*Stenodus* groups. This level of genetic distinction is consistent with the mitochondrial RFLP results in Bernatchez *et al*. (1991) showing obvious taxonomic separation between *Prosopium* and the two other genera (Fig. 2). A much lower intergeneric sequence divergence of 2.94% (STD: ±0.19%; range: 2.68–3.17%) occurred between *Stenodus leucichthys* and *Coregonus* spp. At the species level, the greatest divergence observed was 11.98% (STD: ±0.34%; range: 11.60–12.35%) in the *Prosopium* group between *P. coulterii* and *P. cylindraceum*. Within *Coregonus*, interspecific sequence divergence was 1.56% (STD: ±0.53%; range: 0.46–2.12%). Finally, the estimates of intraspecific variation ranged from 0.00% in five species to 0.77% in *C. pidschian*, similar to the levels that have been reported for COI in a variety of vertebrates, including fishes (Hebert *et al*. 2003a; Ward *et al*. 2005; Spies *et al*. 2006).

There were two instances in which interspecific sequence divergence was very low and comparable to the observed

Journal compilation © 2008 Blackwell Publishing Ltd
No claims to original US government works

**Table 3** Kimura 2-parameter (K2P) calculated pairwise nucleotide differences (below diagonal) and percentage sequence divergence (above diagonal) in 650-bp sequences of cytochrome *c* oxidase subunit I from eight species of whitefish*†

| | *C. aut* | *C. lau* | *C. pid* 1 | *C. pid* 2 | *C. nas* | *C. sar* | *P. cou* 1 | *P. cou* 2 | *P. cyl* 1 | *P. cyl* 2 | *S. leu* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *C. aut* | — | 0.46 | 1.88 | 1.72 | 1.72 | 1.56 | 14.33 | 14.13 | 11.87 | 11.70 | 3.01 |
| *C. lau* | 3 | — | 2.04 | 2.20 | 1.88 | 2.04 | 14.33 | 14.13 | 12.07 | 11.89 | 3.17 |
| *C. pid* 1 | 12 | 13 | — | 0.77 | 0.77 | 1.56 | 13.93 | 13.73 | 12.42 | 12.24 | 2.84 |
| *C. pid* 2 | 11 | 14 | 5 | — | 1.25 | 1.72 | 13.91 | 13.72 | 12.63 | 12.45 | 3.17 |
| *C. nas* | 11 | 12 | 5 | 8 | — | 1.40 | 13.52 | 13.33 | 11.86 | 11.68 | 2.84 |
| *C. sar* | 10 | 13 | 10 | 11 | 9 | — | 13.72 | 13.52 | 12.05 | 11.49 | 2.68 |
| *P. cou* 1 | 83 | 83 | 81 | 81 | 79 | 80 | — | 0.62 | 12.35 | 11.79 | 13.55 |
| *P. cou* 2 | 82 | 82 | 80 | 80 | 78 | 79 | 4 | — | 12.16 | 11.60 | 13.75 |
| *P. cyl* 1 | 70 | 71 | 73 | 74 | 70 | 71 | 73 | 72 | — | 0.46 | 12.05 |
| *P. cyl* 2 | 69 | 70 | 72 | 73 | 69 | 68 | 70 | 69 | 3 | — | 11.87 |
| *S. leu* | 19 | 20 | 18 | 20 | 18 | 17 | 79 | 80 | 71 | 70 | — |

*For species with multiple sequences, the average percentage sequence divergence was used when computing overall intergeneric and interspecific sequence divergence. For example, the percentage sequence divergence between *Coregonus pidschian* and *Coregonus nasus* is (0.77 + 1.25)/2 = 1.01. †*C. pid* 1 and 2, *P. cou* 1 and 2, *P. cyl* 1 and 2 correspond to locations 3b, 3a, 6b, 6a, 7b and 7a, respectively, on Fig. 1. *C. aut, Coregonus autumnalis; C. lau, Coregonus laurettae; C. pid, Coregonus pidschian; C. nas, Coregonus nasus; C. sar, Coregonus sardinella; P. cou, Prosopium coulterii; P. cyl, Prosopium cylindraceum; S. leu, Stenodus leucichthys.*

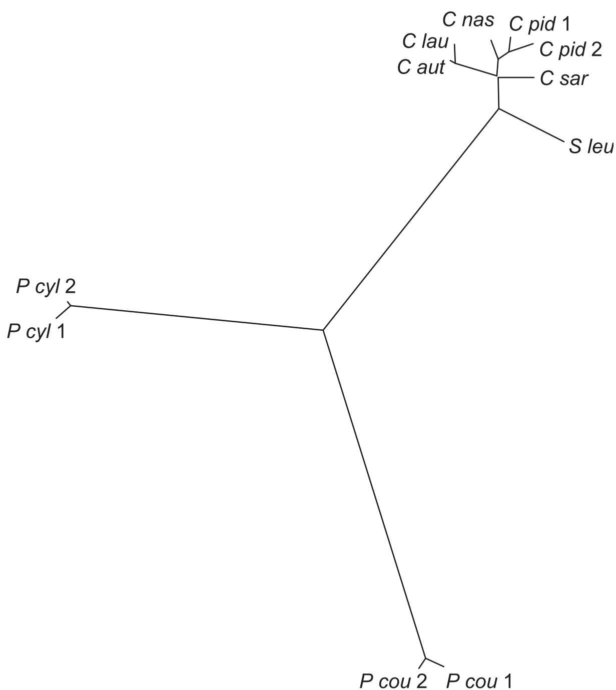


**Fig. 2** A neighbour-joining tree of cytochrome *c* oxidase subunit I sequence divergences (K2P) in eight coregonine species as shown in Table 3.

intraspecific sequence divergence. The first instance involved *Coregonus autumnalis* and *Coregonus laurettae* (mean sequence divergence = 0.46%). This result is not surprising given that the two species are closely related and their taxonomic separation has received scrutiny in the past (e.g. Bickham *et al*. 1997). Bickham *et al*. (1997) report *C. laurettae* genes in two of 49 *C. autumnalis* individuals, suggesting ongoing hybridization between the two species. The second instance involved *C. nasus* and *C. pidschian*. Sequence divergences were 0.77% and 1.25% for Lyapin River (Russia) and Yukon River (Alaska, USA) *C. pidschians*, respectively, with the lower value identical to the intraspecific value between the two *C. pidschian* samples (Table 3). Collectively, these results indicate a close relationship (Fig. 2) between the two species, consistent with previous studies (Bernatchez *et al*. 1991). The lower level of divergence between *C. nasus* and *C. pidschian* in Russia may limit the applicability of the COI gene to distinguish between the two species in that region.

### *RFLP assay development*

Restriction site mapping revealed an average of 126 enzymes with recognition sites in the COI amplicon for all species and the number of recognition sites averaged 291 across all the enzymes. Five potentially diagnostic enzymes (*Rsa*I, *Hae*III, *Bse*RI, *Bsm*I *and Ase*I) were used in a stepwise RFLP assay. One enzyme, *Ase*I, showed promise for isolating *P. cylindraceum* and *S. leucichthys*, but was dropped during initial testing since the diagnostic fragment lengths were indistinguishable from other fragments on an agarose gel. Of the remaining enzymes, the four-step assay began with *Rsa*I, which isolated *P. coulterii* and *S. leucichthys* (Table 4). The remaining species formed a single group (*Rsa*I-3) lacking the *Rsa*I restriction site. Second, *Hae*III isolated *P. cylindraceum* and *C. sardinella*, leaving four species in two groups with different fragment lengths. Third, *Bse*RI

© 2008 The Authors
No claims to original US government works

**Table 4** Summary of the stepwise RFLP assay for identification of eight coregonine species based on mtDNA sequence variation in the cytochrome *c* oxidase subunit I gene. Enzymes providing positive species identification are denoted with a Y. The test result shows the proportion of individuals correctly assigned to species in a blind test. Combined fragment lengths reflect the length of the COI gene (650 bp) plus the 56 bp flanking region

| Step: enzyme | Group | Species | Cut sites | COI fragments (bp) | species ID | Test result |
|---|---|---|---|---|---|---|
| 1: *Rsa*I | *Rsa*I-1 | *Prosopium coulterii* | 2 | 343–233–130 | Y | 6/6 |
| | *Rsa*I-2 | *Stenodus leucichthys* | 1 | 552–154 | Y | 6/6 |
| | *Rsa*I-3 | *Prosopium cylindraceum* | 0 | | | |
| | | *Coregonus sardinella* | 0 | | | |
| | | *Coregonus autumnalis* | 0 | | | |
| | | *Coregonus laurettae* | 0 | | | |
| | | *Coregonus nasus* | 0 | | | |
| | | *Coregonus pidschian* | 0 | | | |
| 2: *Hae*III | *Hae*III-1 | *P. cylindraceum* | 5 | 295–162–110–81–42–16 | Y | 6/6 |
| | *Hae*III-2 | *C. sardinella* | 5 | 251–152–87–81–75–60 | Y | 5/5* |
| | *Hae*III-3 | *C. autumnalis* | 5 | 295–152–87–81–75–16 | | |
| | | *C. laurettae* | 5 | 295–152–87–81–75–16 | | |
| | *Hae*III-4 | *C. nasus* | 5 | 173–152–141–87–78–75 | | |
| | | *C. pidschian* | 5 | 173–152–141–87–78–75 | | |
| 3a: *Bse*RI | *Bse*RI-1 | *C. autumnalis* | 0 | | Y | 6/6 |
| | *Bse*RI-2 | *C. laurettae* | 1 | 378–328 | Y | 5/6† |
| 3b: *Bsm*I | *Bsm*I-1 | *C. nasus* | 1 | 449–257 | Y | 6/6 |
| | *Bsm*I-2 | *C. pidschian* | 0 | | Y | 8/8 |

*No PCR product was observed for one individual after two attempts at amplification so the individual was dropped from the test.
†One putative *Coregonus laurettae* was twice identified as a *Coregonus autumnalis* in separate assays following independent DNA extraction.

identified *C. autumnalis* and *C. laurettae*, while fourth, *Bsm*I distinguished *C. nasus* and *C. pidschian*.

### *RFLP assay testing and application*

The final suite of four enzymes was tested on 50 individuals from eight species of whitefish whose identity was unknown to the researcher. Forty-nine of the 50 individuals were successfully amplified and 48 of those were correctly assigned to species. One putative *C. laurettae* was twice identified as a *C. autumnalis* in separate assays following independent DNA extraction. These were the same two species that exhibited an interspecific divergence equivalent to the lowest observed intraspecific divergence. The putative *C. laurettae* observed in the blind test may be an example of a stray *C. autumnalis* in the Yukon River (Alaska, USA), misidentified as a *C. laurettae*, or a hybrid from a previous straying event (e.g. Bickham *et al*. 1997). It should also be noted that the assay correctly identified all *C. pidschian* and *C. nasus* samples, which represent the other example of low interspecific sequence variation. The blind test was conducted on *C. pidschian* and *C. nasus* samples solely of Alaskan origin. The individuals used in sequencing were of both Alaskan and Russian origin; their differing levels of interspecific divergence with *C. nasus* may reflect geographical variation in *C. pidschian*. Further clarification on this subject could be provided by applying the species identification assay to more *C. pidschian* and *C. nasus* samples from a variety of Russian locations.

We were able to correctly identify seven of the eight species with 100% success; *P. coulterii*, *S. leucichthys*, *P. cylindraceum*, *C. sardinella*, *C. nasus C. pidschian* and *C. autumnalis*. These results are consistent with those of Politov *et al*. (2000) where the use of mitochondrial ND-1 RFLP and allozyme assays provide accurate identification of *C. autumnalis, C. sardinella, C. nasus* and *S. leucichthys*. Although these four species were analysed in both studies, it should be noted that the Politov *et al*. (2000) study does not include *C. laurettae* which may hybridize with *C. autumnalis*. For example, one of the six *C. laurettae* samples was misassigned as a *C. autumnalis* in our blind test. The results from our study and that of Bickham *et al*. (1997) suggest that where the two species overlap, they should conservatively be regarded as a single group for species identification when using COI. In addition, further investigation into hybridization in whitefish, particularly in *C. autumnalis* and *C. laurettae* is needed. Such a study will require nuclear markers to confirm hybridization (Campton 1988).

The goal of this study was to develop a genetic tool for direct application in future management and research of whitefish throughout Alaska. Increasing demand on the coregonine fishery resources in Alaska has spurred the need for more information about juvenile abundance and distribution as well as migration and stock assessment, for

Journal compilation © 2008 Blackwell Publishing Ltd
No claims to original US government works

which little or no data exist (Brown *et al*. 2007). With the exception of one aberrant sample, the species identification assay developed in this study allows for highly accurate identification of eight whitefish species in the absence of other whitefish species. Development of additional assays may be required to distinguish these eight species in geographical regions where other coregonine species occur. Future work will require assessment of COI diversity at the local population level and should include both larger sample sizes and additional blind tests for an RFLP assay. Complications due to hybridization should also be carefully considered (Moritz & Cicero 2004).

## Acknowledgements

Funding was provided by the US Fish and Wildlife Service (USFWS), Office of Subsistence Management, under agreement number 05–702. Samples were provided by Margaret F. Docker, Department of Zoology, University of Manitoba, Canada; Ken Harper, USFWS Kenai Fish and Wildlife Field Office, Kenai, Alaska; Craig R. Primmer, Department of Biological and Environmental Sciences, University of Helsinki, Finland; Dmitry S. Sendek, State Research Institute on Lake and River Fisheries, St. Petersburg, Russia; Julie Turgeon, Département de Biologie, Université Laval, Canada. We thank Mike Canino for insightful comments on an earlier draft of this manuscript.

## References

Bernatchez L (2004) Ecological theory of adaptive radiation: an empirical assessment from Coregonine fishes (*Salmoniformes*). In: *Evolution Illuminated: Salmon and Their Relatives* (eds Hendry AP, Stearns SC). Oxford University Press, New York.

Bernatchez L, Colombani F, Dodson JJ (1991) Phylogenetic relationships among the subfamily Coregoninae as revealed by mitochondrial DNA restriction analysis. *Journal of Fish Biology*, **39**(Suppl. A), 283–290.

Bickham JW, Patton JC, Minzenmayer S, Moulton LL, Gallaway BJ (1997) Identification of Arctic and Bering ciscoes in the Colville River delta, Beaufort Sea Coast, Alaska. *American Fisheries Society Symposium*, **19**, 224–228.

Brown RJ, Bickford N, Severin K (2007) Otolith trace element chemistry as an indicator of anadromy in Yukon River drainage coregonine fishes. *Transactions of the American Fisheries Society*, **136**, 678–690.

Campton D (1988) Natural hybridization and introgression in fishes: methods of detection and genetic interpretations. In: *Population Genetics and Fishery Management* (eds Ryman N, Utter F). University of Washington Press, Seattle.

Felsenstein J (1993) *PHYLIP version 3.5c. Distributed by the author*. Department of Genetics, University of Washington, Seattle, Washington.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, **270**(Suppl. 1), S96–S99.

McElroy D, Moran P, Bermingham E, Kornfield I (1992) REAP, An integrated environment for the manipulation and phylogenetic analysis of restriction data. *Journal of Heredity*, **83**, 157–158.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**, 1529–1531.

Politov DV, Gordon NY, Afanasiev KI, Altukhov YP, Bickham JW (2000) Identification of palearctic coregonid fish species using mtDNA and allozyme genetic markers. *Journal of Fish Biology*, **57**, 51–71.

Robins CR, Bailey RM, Bond CE *et al*. (1991) *Common and Scientific Names of Fishes from the United States and Canada*, 5th ed., Special Publication 20. American Fisheries Society, Bethesda, Maryland.

Smith P, McVeagh S, Steinke D (2008) DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology*, **72**, 464–471.

Spies IB, Gaichas S, Stevenson DE, Orr JW, Canino MF (2006) DNA-based identification of Alaska skates (Amblyraja, Bathyraja and Raja: Rajidae) using cytochrome c oxidase subunit I (COI) variation. *Journal of Fish Biology*, **69**(Suppl. B), 283–292.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1847–1857.

© 2008 The Authors
No claims to original US government works