JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2009, 57, 8379–8385 8379
DOI:10.1021/jf901618z

DNA Barcoding of Commercially Important Salmon and Trout Species (*Oncorhynchus* and *Salmo*) from North America

ROSALEE S. RASMUSSEN,*[,1] MICHAEL T. MORRISSEY,[1] AND PAUL D. N. HEBERT[2]

[1]Oregon State University Seafood Lab, Department of Food Science and Technology, Oregon State University, 2001 Marine Drive, Room 253, Astoria, Oregon 97103, and [2]Biodiversity Institute of Ontario, Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada

The present study investigated the ability of DNA barcoding to reliably identify the seven commercially important salmon and trout species (genera *Oncorhynchus* and *Salmo*) in North America. More than 1000 salmonid reference samples were collected from a wide geographic range. DNA extracts from these samples were sequenced for the standard 650 bp barcode region of the cytochrome *c* oxidase subunit I gene (COI). DNA barcodes showed low intraspecies divergences (mean, 0.26%; range, 0.04−1.09%), and the mean congeneric divergence was 32-fold greater, at 8.22% (range, 3.42−12.67%). The minimum interspecies divergence was always greater than the maximum intraspecies divergence, indicating that these species can be reliably differentiated using DNA barcodes. Furthermore, several shorter barcode regions (109−218 bp), termed "mini-barcodes", were identified *in silico* that can differentiate all eight species, providing a potential means for species identification in heavily processed products.

KEYWORDS: DNA barcoding; salmonids; COI; species identification; mini-barcode; FISH-BOL

INTRODUCTION

There are seven commercially important salmon and trout species in North America belonging to the genera *Oncorhynchus* and *Salmo*. Chinook salmon (*Oncorhynchus tshawytscha*), sockeye salmon (*Oncorhynchus nerka*), coho salmon (*Oncorhynchus kisutch*), chum salmon (*Oncorhynchus keta*), and pink salmon (*Oncorhynchus gorbuscha*) are wild-harvested, whereas rainbow (steelhead) trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*) are sold only as farm-raised products. The wide variation in quality and availability of these salmonid species leads to substantial market differentials, with average ex-vessel/ex-farm prices per kilogram ranging from U.S. $0.29 for *O. gorbuscha* to U.S. $5.71 for *O. tshawytscha* (*1*). In fact, prices for the highly valued spring chinook reached U.S. $22/kg (whole fish weight) in early 2009 (*2*). After processing, species identification of salmonids becomes difficult because of the similar appearance of fillets from different species. Not surprisingly, given these value differences, the U.S. Food and Drug Administration (FDA) has detected cases of fraud involving the substitution of *O. keta* with *O. gorbuscha*, the substitution of salmon with *O. mykiss*, and the substitution of wild salmon with farmed salmon (http://www.cfsan.fda.gov/~frf/rfeecon.html).

To advance its capacity to detect such substitutions in the marketplace, the FDA is considering the adoption of DNA barcoding as an official regulatory method, a shift that will see the incorporation of DNA barcodes into the *Regulatory Fish Encyclopedia* (*3*). DNA barcoding is a method for species identification that is based on the surveillance of sequence diversity in a 650 bp region of the mitochondrial gene coding for cytochrome *c* oxidase I (COI) (*4*). This gene region generally shows little variation within a species but substantial divergence between species, allowing for species differentiation. To use this approach for species identification, the DNA barcode of an unknown sample is screened against a reference sequence library and a species assignment is made when the query sequence matches just one of the species in the reference library. A reference library of DNA barcodes for all fish species is currently under assembly by the Fish Barcode of Life campaign (FISH-BOL) (*5*). With records now in place for more than 6500 species, barcodes have proven to unambiguously discriminate about 93% of freshwater species and 98% of marine species.

Despite the high potential of DNA barcoding for fish identification, some salmonids may lack the diagnostic sites required for species differentiation. They are a closely related group of anadromous and non-anadromous species with marked intraspecific diversity (*6*), suggesting the possibility of overlap between intra- and interspecific divergences. Furthermore, while rates of nucleotide substitution in mitochondrial (mt) DNA are typically about 2% per million years, mtDNA seems to evolve more slowly in salmonids, at about 1% per million years (*7*). Perhaps as a consequence, recent studies have reported between-species divergence values that are exceptionally low (< 1.0%) for some salmonids (*8*, *9*). Hubert et al. (*9*) did obtain promising results for the seven commercially important salmonid species mentioned above, as all interspecies divergences were greater than 3%, while intraspecific divergences were below 1%; however, their sample sizes were small (2−12 per species), and all specimens were derived from Canadian waters. Moreover, another study

*To whom correspondence should be addressed. Telephone: 503-325-4531 ext. 107. Fax: 503-325-2753. E-mail: rosalee.rasmussen@oregonstate.edu.

Published on Web 08/24/2009

reported very high intraspecies divergence (7.3%) in *O. mykiss* ($n = 8$), with one cluster showing greater similarity to *O. kisutch*, raising concerns in relation to the diagnosability of these species through DNA barcoding (*3*). To determine if the DNA barcode region can reliably differentiate commercially important salmon and trout species, a thorough examination of barcode divergence within and between these species is required, including individuals from a wide geographic range.

This study involves the comprehensive analysis of DNA barcode divergences within and among key salmon and trout species (*Oncorhynchus* and *Salmo*). It examines the extent of geographic variation in barcode sequences and the clarity of the barcode gap needed for species identification. In addition, the prospects of delivering species identifications through a smaller segment of the barcode region for use in the case of heavily processed foods were explored *in silico*.

MATERIALS AND METHODS

Sample Collection and Preparation. The primary target species were *O. tshawytscha*, *O. nerka*, *O. kisutch*, *O. keta*, *O. gorbuscha*, *O. mykiss*, and *S. salar*. As well, four subspecies of cutthroat trout (*Oncorhynchus clarkii clarkii*, *O. c. bouvierii*, *O. c. utah*, and *O. c. lewisii*) were screened because of their close relationship with the other taxa and reported hybridization with *O. mykiss* (*10*). Reference tissue and DNA samples were obtained for 1035 specimens from the Alaska Department of Fish and Game Gene Conservation Laboratory, American Gold Seafoods, Casitas Municipal Water District, Clear Springs Foods, Creative Salmon, Marine Harvest Canada, National Marine Fisheries Southwest Fisheries Science Center, Oregon Department of Fish and Wildlife, Marine Fisheries Genetics Laboratory at Hatfield Marine Science Center (Oregon State University), Salmon of the Americas, Washington Department of Fish and Wildlife Molecular Genetics Lab, Pacific Salmon Treaty, and the Washington State General Fund. Samples consisted of fin clips, axillary process clips, scales, heart tissue, muscle tissue, liver tissue, and purified DNA. The purified DNA samples ($n = 71$) were extracted from salmonid specimens using the Qiagen DNeasy Blood and Tissue kit (Valencia, CA) and stored in AE buffer. All other samples were stored frozen, preserved in ethanol, dried, or in lysis buffer. A total of 838 samples from both wild and hatchery stocks were collected from locations in Alaska, Washington, Oregon, Idaho, Utah, and California (**Figure 1**), representing 89 water bodies (i.e., rivers, creeks, lakes, and bays) and 143 specific sites, with an average of 5.9 individuals collected per site. In addition to the sampling locations shown in **Figure 1**, tissue samples ($n = 197$) of *O. mykiss*, *S. salar*, and *O. tshawytscha* were acquired from aquaculture facilities in the United States (Washington and Idaho), Canada (British Columbia), and Chile. After completion of the sample collection, molecular analysis of all samples was carried out at the Canadian Center for DNA Barcoding (CCDB) at the University of Guelph, Ontario, Canada.

DNA Extraction. DNA was extracted from tissue samples using a silica-based automated protocol, as described in Ivanova et al. (*11*). DNA from 94 scale samples was eluted in 30 μL of sterile ddH_2O, while DNA from tissue samples was eluted in 60 μL of sterile ddH_2O. In an attempt to maximize recovery of DNA from salmon scales, an additional 94 scale samples were subjected to a semi-automated, plant-based DNA extraction protocol (*12*). The lysis step was modified to include an overnight incubation at 56 °C with 50 μL of cetyltrimethylammonium bromide (CTAB) buffer and proteinase K (20 mg/mL) instead of tissue disruption with carbide beads. DNA obtained from this protocol was eluted in 50 μL of sterile ddH_2O.

PCR Amplification. Polymerase chain reactions (PCRs) were carried out using a Mastercycler EP Gradient (Eppendorf, Brinkman Instruments, Inc., Westbury, NY). The total reaction volume was 12.5 μL and included the following components: 6.25 μL of 10% trehalose, 2.0 μL of ddH_2O, 1.25 μL of 10× PCR buffer [10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, and 0.1% Triton X-100], 0.625 μL of $MgCl_2$ (50 mM), 0.125 μL of each primer cocktail (0.01 mM), 0.0625 μL of dNTPs (10 mM), 0.0625 μL of *Taq* DNA polymerase (New England Biolabs, Ipswich, MA), and 2.0 μL of template DNA. A set of fish primer cocktails (C_FishF1t1 and C_FishR1t1) with M13 tails was used under the following reaction conditions: 94 °C for 2 min; 35 cycles of 94 °C for 30 s, 52 °C for 40 s and 72 °C for 1 min; and a final extension step of 72 °C for 10 min (*13*). In cases where C_FishF1t1 and C_FishR1t1 failed to generate an amplicon, an additional primer cocktail (C_VF1LFt1 and C_VR1LRt1) was used in combination with M13 tails under the following reaction conditions: 94 °C for 1 min; five cycles of 94 °C for 30 s, 50 °C for 40 s, and 72 °C for 1 min; 35 cycles of 94 °C for 30 s, 54 °C for 40 s, and 72 °C for 1 min; and a final extension step of 72 °C for 10 min (*13*). All primer cocktails are described by Ivanova et al. (*13*). PCR products were separated on 2% agarose gels using an E-Gel96 precast agarose electrophoresis system (Invitrogen, Carlsbad, CA). Images were photographed under UV light with an AlphaImager 3400 imaging system (Alpha Innotech Corp., San Leandro, CA) and processed with Invitrogen E-editor software.

Sequencing. PCR products were sequenced bidirectionally with BigDye Terminator version 3.1 cycle sequencing kit (Applied Biosystems, Inc., Foster City, CA) on an ABI 3730XL DNA analyzer capillary sequencer (Applied Biosystems, Inc.). Contiguous read lengths and trace scores were generated for all sequences using Applied Biosystems sequence scanner software, version 1.0. Sequences were assembled and edited using CodonCode Aligner, version 2.0.6. All sequences were aligned in MEGA, version 3.1 (*14*), before uploading to the Barcode of Life Data System [BOLD; (*15*)].

Mini-barcodes *in Silico* Test. A total of 11 mini-barcode regions (107−218 bp) were analyzed *in silico* based on previously identified segments of the full-length barcode (*16*, *17*). Barcode sequences that were obtained in the current study were selected for mini-barcode analysis according to the following criteria: (1) original barcode sequence greater than 500 bp and (2) no gaps in the mini-barcode region. All suitable barcode sequences were examined for genetic distances in the mini-barcode region, as described in Data Analysis.

Data Analysis. The sampling locations for wild and hatchery specimens examined in this study were mapped with ESRI ArcMap 9.2 software (Environmental Systems Research Institute, Inc., Redlands, CA). Genetic distances among barcode and mini-barcode sequences were quantified using the Kimura two-parameter (K2P) distance model (*18*) through the BOLD online interface (www.barcodinglife.org). Barcode haplotypes were identified using sequence identity matrices generated in BioEdit sequence alignment editor, version 7.0.9 (*19*). Neighbor-joining trees (*20*) were generated in MEGA, version 4.0 (*21*), using the K2P distance model for all representative haplotypes of the full data set. All codon positions were included, and all positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. Branch support was assessed with bootstrap analysis (1000 replicates) with sequences from *S. salar* used to root the tree. In cases where only one individual displayed a specific haplotype, the trace files for that sequence were double-checked to ensure that no errors were made in base-calling. Regression analyses were carried out with SPSS 13.0 for Windows to determine the relationships between the number of individuals sampled per species and (1) the number of haplotypes, (2) the mean intraspecies divergence, and (3) the maximum intraspecies divergence. Significance levels were set at $p < 0.05$.

RESULTS AND DISCUSSION

Barcode Recovery. Partial or full barcode sequences (302−652 bp) were obtained from 934 of the 1035 individuals (GenBank accession numbers FJ998606−FJ999539; see Appendix S1 in the Supporting Information). Sequences greater than 500 bp in length were recovered from 924 individuals (89%), and barcodes greater than 600 bp were recovered from 874 individuals (84%). Amplification and sequencing failures may be due to factors such as the presence of PCR inhibitors, primer mismatches, or DNA degradation (*13*). Many of the unsuccessful samples in this study consisted of degraded tissue or scales, which contain known PCR inhibitors (i.e., mucopolysaccharides). The number of sequences greater than 500 bp recovered per species ranged from 47 (*O. gorbuscha*) to 216 (*O. mykiss*), with an average of 132 individuals per species (**Table 1**). No insertions, deletions, or stop codons were observed in these sequences, indicating that all barcodes represent the functional mitochondrial COI sequence.

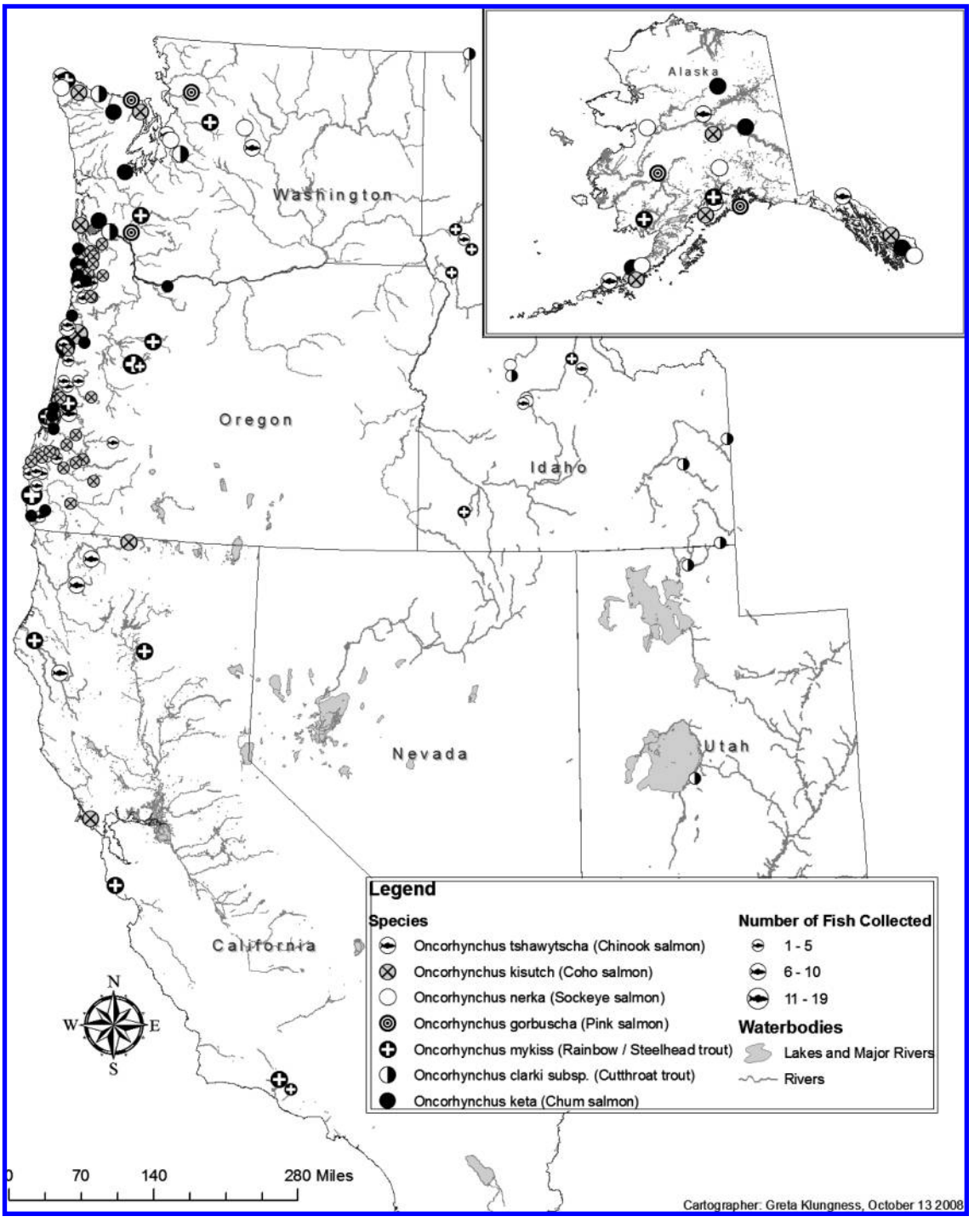

Figure 1. Geographic origins of reference salmonid tissues obtained in this study from wild and hatchery stocks (n = 838). Salmonids from farmed locations are not shown (n = 197). Icons are representative of the collection regions but, in some cases, do not reflect the exact site.

Barcode Divergences and Haplotypes. Genetic divergences were calculated for all COI barcodes with a sequence length greater than 500 bp, the minimum required length for formal barcode status, and less than 1% uncertain base calls (*15*). Regression analyses indicated no significant relationships between the number of individuals analyzed per species and the mean within-species sequence divergence (R^2 = 0.171, p = 0.154), the maximum within-species divergence (R^2 = 0.131, p = 0.189), or the number of haplotypes (R^2 = 0.38, p = 0.051), indicating that sampling efforts on each species were sufficiently comprehensive to provide a good understanding of variation.

The average intraspecies variation (**Table 1**) ranged from a low of 0.04% in *O. keta* (maximum of 0.47%) to a high of 1.09% in *O. clarkii* (maximum of 1.96%). The restricted genetic divergence in *O. keta* supports a prior report of very low mtDNA diversity in this species from 42 populations [(n = 788) (*22*)]. Relatively high levels of genetic divergence among cutthroat trout have also been found previously (*6*). An analysis of data from the "Barcoding of Canadian Freshwater Fishes" project on BOLD (*9*) also revealed higher divergence among 12 individuals of *O. clarkii* from Canada (mean intraspecies divergence of 0.97%, maximum of 1.87%) than for other *Oncorhynchus* species.

Table 1. Salmonid Species Collected and Sequenced for the DNA Barcode Region[a]

species	number of individuals collected	number of individuals sequenced (>500 bp)	mean intraspecies divergence (%) ± SD
O. tshawytscha	229	212	0.38 ± 0.23
O. nerka	81	67	0.40 ± 0.34
O. keta	119	90	0.04 ± 0.08
O. kisutch	156	146	0.19 ± 0.17
O. gorbuscha	50	47	0.31 ± 0.22
O. mykiss	219	216	0.14 ± 0.12
S. salar	116	87	0.29 ± 0.29
O. clarkii subspp.[b]	65	59	1.09 ± 0.72

[a] Intraspecies genetic divergences based on the K2P model are reported in terms of mean ± standard deviation for barcodes greater than 500 bp (n = 924). [b] *O. c. clarkii*, *O. c. bouvierii*, *O. c. utah*, and *O. c. lewisii*.

When *O. clarkii* was excluded from the data set, the highest mean intraspecies divergence for the seven target salmonid species dropped to 0.40% (*O. nerka*), indicating that the COI barcode region is highly conserved among these species. This conclusion agrees with Hubert et al. (*9*), whose data show mean intraspecies divergence values ranging from 0% in *O. keta* (n = 2) and *O. gorbuscha* (n = 8) to 0.57% in *O. nerka* (n = 4). In contrast, a previous study examining the potential use of DNA barcodes for regulatory purposes reported a maximum of 7.3% intraspecies divergence in *O. mykiss*, an exceptionally high value (*3*). The authors suggested that this case may have been due to the mislabeling of a tissue sample from *O. kisutch* as *O. mykiss*. The mean intraspecies divergence for *O. mykiss* (n = 216) in the current study was very low, at 0.14% (maximum of 0.62%), supporting the suggestion that the deeply divergent sequence in the previous study was not derived from *O. mykiss*.

Each COI barcode haplotype (n = 78) encountered in this study was restricted to a single species (**Figure 2**), but the number of haplotypes per species ranged from a low of 3 (*S. salar*) to a high of 16 (*O. mykiss*), with an average of 10. While some haplotypes were widespread throughout the sampling range, many were restricted to a particular region. For example, 7 of the 11 haplotypes in *O. gorbuscha* were unique to Alaska, 2 were unique to Washington State, and the remaining 2 were detected in individuals from both states. For most species, the majority of individuals belonged to one or two haplotypes, while the remaining haplotypes were rare. For example, 74 individuals of *O. keta* belonged to one haplotype (HAP22), while the other 7 haplotypes for this species were observed in only 1−6 individuals. Similarly, almost half of the individuals of *O. mykiss* shared a haplotype (HAP54) that was detected in all collection states, while 11 haplotypes were unique to 1−9 individuals in Washington, Oregon, Idaho, or California. Interestingly, some fish from aquaculture broodstocks of *O. mykiss* exhibited haplotypes (HAP52, HAP57, HAP58, and HAP59) that were not detected in the wild, but other aquaculture fish from the same source shared haplotypes with wild stocks. Barcodes for *O. tshawytscha* showed a slightly different trend, with 8 haplotypes that contained more than 10 individuals each. Two of those haplotypes were unique to Oregon (HAP06 and HAP08), while one was unique to Alaska (HAP05). *O. nerka* showed a similar trend, with most samples distributed among 3 haplotypes (HAP15, HAP16, and HAP17). HAP15 (n = 17) and 2 other haplotypes were unique to Alaska, whereas 2 haplotypes were unique to Oregon, and 1 was unique to Idaho. Among the 10 haplotypes for *O. kisutch*, 6 were unique to Alaska, Washington, Oregon, or California. A previous study based on restriction site variation reported 3 COI/COII haplotypes for *O. kisutch* (n = 70) in

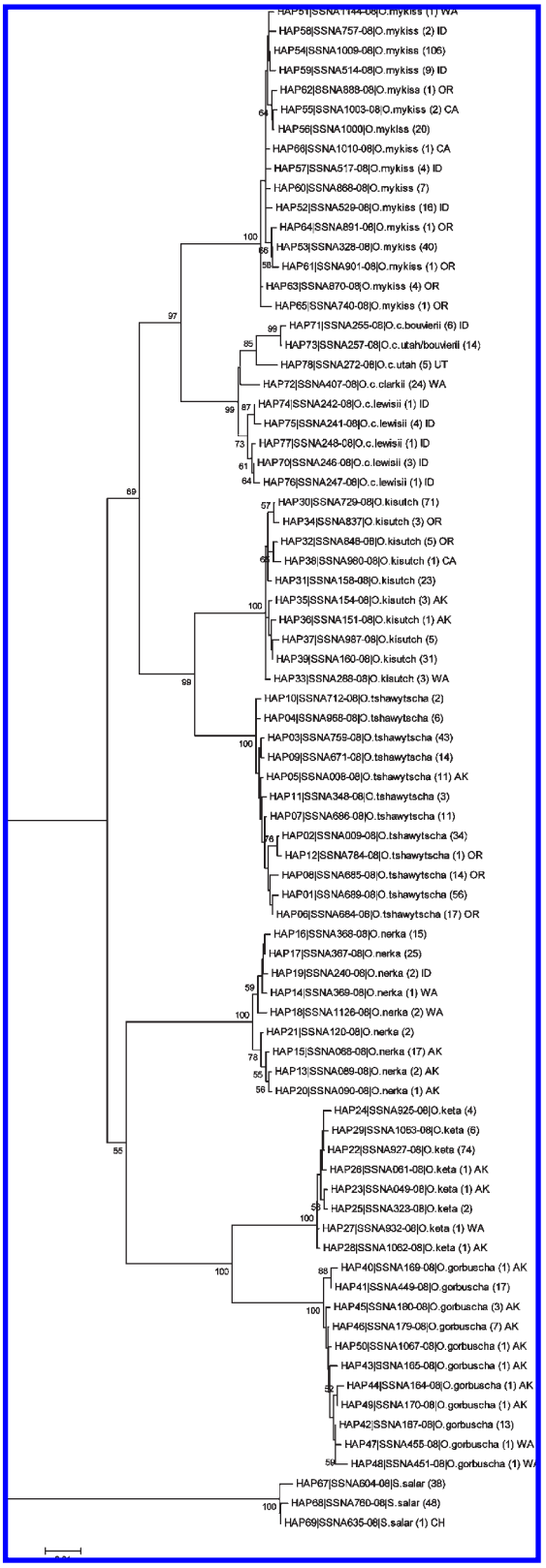

Figure 2. K2P neighbor-joining consensus tree of all salmonid COI barcode haplotypes (n = 78) identified in this study. Bootstrap values greater than 50 are shown (1000 replicates). The tree is drawn to scale, and units are the number of base substitutions per site. Branch labels include haplotype number, BOLD sample number, species, and number of individuals with this haplotype. In cases where the haplotype was found to be unique to one geographic region, the abbreviation for that region is also given (AK, Alaska; WA, Washington; OR, Oregon; CA, California; ID, Idaho; UT, Utah; CH, Chile).

Table 2. Summary of the K2P Genetic Distances for All Barcodes Obtained in This Study Greater than 500 bp[a]

comparisons within	number of comparisons	mean	minimum	maximum	SE
species	68920	0.263	0	1.955	0.001
genus, between species	284687	8.224	3.419	12.671	0.004
family, between genus	72819	15.653	13.446	19.716	0.003

[a] Data are from 924 individuals representing 8 salmonid species and 2 genera (*Salmo* and *Oncorhynchus*).

Alaska (*23*), but this study revealed 5 haplotypes in this state. Each of the four subspecies of *O. clarkii* included in the present analysis (*O. c. clarkii*, *O. c. bouvierii*, *O. c. utah*, and *O. c. lewisii*) had at least one haplotype that was not present in the other subspecies. The only shared haplotype among *O. clarkii* subspecies was HAP73, found in both *O. c. bouvierii* and *O. c. utah* (collected from the Bear River drainage). Previous reports also have indicated that populations of *O. c. utah* from the Bear River drainage are more closely related genetically to *O. c. bouvierii* than to other populations of *O. c. utah* (*24*, *25*).

As shown in **Table 2**, the mean divergence between species within the same genus was 8.22% (range, 3.42−12.67%), a value 32-fold greater than the mean intraspecies divergence (0.26%) for the species examined in this study. The mean intraspecies divergence found in this study was slightly lower than previous fish barcoding studies, which have reported mean conspecific divergences of 0.30% (range, 0−7.42%), 0.39% (range, 0−14.08%), and 0.99% (0.19% when possible misidentifications were omitted) for 194 Canadian fish species (*9*), 207 Australian fish species (*26*), and 72 U.S. commercial fish species (*3*), respectively. The mean congeneric divergence between species was similar to previous studies, which have reported values of 8.29−9.93% (25−27-fold greater than the conspecific divergences) (*9*, *26*, *27*). The mean divergence between the *Oncorhynchus* and *Salmo* genera (15.65%) was also in agreement with previous values of mean divergence between fish genera within the same family (15.38−15.46%) (*9*, *26*).

As indicated in the K2P neighbor-joining tree (**Figure 2**), there was clear separation between species (99−100% bootstrap values) with no shared or overlapping barcodes. The nearest neighbor distances (i.e., minimum divergence between species) for the eight salmonids in this study ranged from 3.42% between *O. tshawytscha* and *O. kisutch* to 13.45% between *S. salar* and *O. nerka*. The barcode data from *S. salar* was also compared to the closely related *Salmo trutta* using samples from the Canadian freshwater fishes project. There was no overlap between the two species, and the minimum divergence was 7.28%. Within the genus *Oncorhynchus*, all nearest neighbor values were under 5.0%, with the exception of *O. nerka*, whose nearest neighbor (*O. kisutch*) was 8.13% away. These values mirror those found for the same species from Canadian waters (range within *Oncorhynchus*, 3.8−8.36%; 14.35% between *S. salar* and *O. nerka*) (*9*). A neighbor-joining tree illustrating the combined data from these two projects is available as Appendix S2 in the Supporting Information. Most of the nearest neighbor distances within the genus *Oncorhynchus* were lower than the average value (7.5%) reported for 194 Canadian freshwater fish species (*9*). Despite the low divergences of the *Oncorhynchus* species, the high ratio of congeneric to conspecific divergence (> 30-fold) ensured effective barcode-based species differentiation. Overall, the intra- and interspecific nucleotide divergence values found here are similar to those found in previous studies investigating mtDNA divergence among the Pacific salmonids (*28*−*34*). For example, Thomas et al. (*34*) found relatively low intraspecific divergence (< 1%) and slightly higher interspecific divergences, ranging

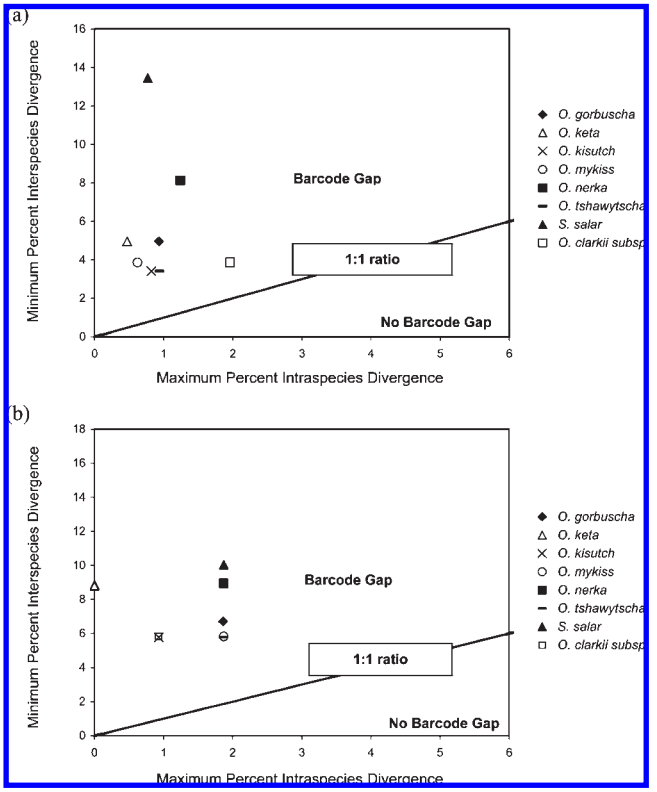

Figure 3. DNA barcode gaps for salmonid sequences obtained in this study with (a) COI barcodes greater than 500 bp (n = 924) and (b) COI mini-barcode 109-5 (n = 923). A data point above the 1:1 ratio line represents a species with a barcode gap (i.e., the species can be identified through DNA barcoding).

from 2.46% for *O. kisutch* and *O. tshawytscha* up to 6.68% for *O. kisutch* and *O. keta*, in an analysis of mtDNA restriction site cleavage for six *Oncorhynchus* species.

The interspecies divergence values found in this study can be used to estimate the divergence rate of the barcode region among the Pacific salmonids. Speciation of *O. keta*, *O. nerka*, and *O. gorbuscha* and speciation of *O. tshawytscha* from *O. kisutch* is believed to have occurred at least 6 million years ago (*7*,*35*). The average interspecies divergence values within these groups were 8.27 and 4.31%, respectively. If the estimated speciation times are correct, the average barcode sequence divergence rates are 1.38% per million years for *O. keta*, *O. nerka*, and *O. gorbuscha* and 0.72% per million years for *O. tshawytscha* and *O. kisutch*. These rates are in general agreement with the previously estimated mtDNA divergence rate of approximately 1% per million years for some Pacific salmonids (*7*).

Barcode Gaps. To determine if barcode gaps are present between the salmonid species examined in this study, the relationships between inter- and intraspecies divergences were compared for each species. A graphic representation was created by plotting the minimum interspecies divergence on the y axis and the maximum intraspecies divergence on the x axis (**Figure 3**). The line on the graph represents cases of a 1:1 ratio between these two values. Data points above the line represent species that may be differentiated through DNA barcoding, while those falling below it represent species that cannot be differentiated through DNA barcoding. As shown in **Figure 3**a, all salmonid species examined in this study fell above the line, indicating that they can be differentiated using DNA barcodes.

Mini-barcodes. Full-length DNA barcodes have been used to successfully identify fish species in a variety of commercial fish

Table 3. Mini-barcode Regions Examined in This Study and Salmonid Species Exhibiting Barcode Gaps in These Regions[a]

mini-barcode	position[b]	salmonid sequences analyzed (*n*)	salmonid species with barcode gap[c]
universal mini-barcode	nt 1−127	822	OKi, OG, OKe, ON, SS
109-1	nt 1−109	822	OT, OKi, OKe, ON, SS
109-2	nt 110−218	921	OT, OKi, OM, ON, SS
109-3	nt 219−327	924	OG, OKe, OM, ON, SS, OC
109-4	nt 328−436	924	OT, OKi, OG, OKe, OM, ON, SS, OC
109-5	nt 437−545	923	OT, OKi, OG, OKe, OM, ON, SS, OC
109-6	nt 546−652	807	OG, OKe, ON, SS
109-4 + 109-5	nt 328−545	923	OT, OKi, OG, OKe, OM, ON, SS, OC
218-1	nt 1−218	822	OT, OKi, OG, OKe, OM, ON, SS
218-2	nt 219−436	924	OT, OKi, OG, OKe, OM, ON, SS, OC
218-3	nt 437−652	807	OT, OKi, OG, OKe, OM, ON, SS, OC

[a] The mini-barcode regions selected for analysis were originally described in Hajibabaei et al. (*16*) and Meusnier et al. (*17*). [b] Relative to the 5′ end of the full-length barcode region. [c] OT, *O. tshawytscha*; OKi, *O. kisutch*; OG, *O. gorbuscha*; OKe, *O. keta*; OM, *O. mykiss*; ON, *O. nerka*; SS, *S. salar*; OC, *O. clarkii* subspp.

products, including fresh, smoked, and cooked fish (*36*, *37*). However, it is often impossible to recover a full-length DNA barcode from heavily processed products, such as canned fish, because of DNA degradation (*38*). The use of shorter barcode sequences, "mini-barcodes", has been proposed as a way to enable DNA barcode analysis of degraded samples (*16*, *17*). Previously identified mini-barcode regions were examined for their ability to differentiate commercially important salmon and trout species (**Table 3**). Among the ∼100 bp mini-barcodes, barcode gaps were present for 4−8 of the salmonid species. The mini-barcodes 109-4 and 109-5 had the ability to differentiate all eight salmonid species, with 109-5 providing slightly greater diagnostic power (**Figure 3b**). Because of their diagnostic capabilities, these two 109 bp mini-barcodes were combined as a 218 bp region for comparison to previously identified 218 bp mini-barcodes (*16*). Among the 218 bp regions examined, 218-2, 218-3, and 109-4 + 109-5 showed barcode gaps for all eight species, whereas 218-1 produced barcode gaps for all species except *O. clarkii*. In a comparison of barcode gap charts for the 218 bp regions, both 218-3 and 109-4 + 109-5 exhibited the strongest species resolution. Interestingly, the mini-barcode gaps produced by the analysis of 109-5 were comparable in diagnostic strength to the 218 bp mini-barcodes, indicating that a 109 bp mini-barcode region is sufficient for species differentiation in this case. Overall, the mini-barcodes 109-5, 218-3, and 109-4 + 109-5 show the best diagnostic capabilities for the reliable identification of all eight salmon and trout species examined in this study.

Summary and Conclusions. A comprehensive analysis of DNA barcode sequence divergences in commercially important species of North American salmon and trout species revealed mean within-species divergences that were all below 1%. No cases of shared haplotypes were detected, indicating an absence of species hybridization. The barcode region exhibited 32-fold greater divergence for congeneric species (8.22%) compared to conspecific individuals (0.26%), and all species demonstrated a barcode gap when full-length sequences were analyzed. These results indicate that DNA barcodes can reliably identify salmon and trout species in the North American commercial market. Furthermore, three mini-barcode regions were identified to have strong diagnostic power among the salmonids, enabling differentiation of all species in this study. Future research efforts may be directed toward the development of appropriate mini-barcode primers and validation of this method in heavily processed products. Work will also be undertaken to develop a species-specific multiplex PCR assay to enable the rapid identification of salmon species in commercial food products. On a larger scale, the development of a COI barcode oligonucleotide microarray for high-throughput identification of commercial fish species is another potential area of research in this field.

ABBREVIATIONS USED

COI, cytochrome *c* oxidase subunit I; DNA, deoxyribonucleic acid; FDA, Food and Drug Administration; FISH-BOL, Fish Barcode of Life; K2P, Kimura two-parameter.

ACKNOWLEDGMENT

We thank Haile Yancy at the U.S. FDA Center of Veterinary Medicine and staff at the Canadian Center for DNA Barcoding, especially Natalia Ivanova, Janet Topan, Claudia Bertrand, Rick Turner, Alex Borisenko, Brianne St. Jacques, and Megan Milton for their advice. Thanks also to Greta Klungness for assistance with ArcMap. Finally, we thank the sample donors, which enabled this project: Alaska Department of Fish and Game Gene Conservation Laboratory, American Gold Seafoods, Casitas Municipal Water District, Clear Springs Foods, Creative Salmon, Marine Harvest Canada, National Marine Fisheries Southwest Fisheries Science Center, Oregon Department of Fish and Wildlife, Dr. Michael Banks and Renee Bellinger from the Marine Fisheries Genetics Laboratory at Hatfield Marine Science Center (Oregon State University Hatfield), Salmon of the Americas, Washington Department of Fish and Wildlife Molecular Genetics Lab, Pacific Salmon Treaty, and the Washington State General Fund.

Supporting Information Available: Sample and sequence details for all DNA barcodes obtained in this study, including species names, GenBank accession numbers, sequence lengths, and geographic sampling regions (Appendix S1). Neighbor-joining tree displaying the results of this project combined with the results of the "Barcoding of Canadian Freshwater Fishes" project (Appendix S2). The tree was generated on BOLD based on the Kimura two-parameter distance method, including barcode sequences from a total of 979 individuals representing 9 species. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Johnson, H. M. *2006/2007 Annual Report on the United States Seafood Industry*, 14th ed.; H. M. Johnson and Associates: Jacksonville, OR, 2007; p 101.

(2) Fishhawk Fisheries, Inc. Astoria, OR. Personal communication, 2009.

(3) Yancy, H. F.; Zemlak, T. S.; Mason, J. A.; Washington, J. D.; Tenge, B. J.; Nguyen, N. T.; Barnett, J. D.; Savary, W. E.; Hill, W. E.; Moore, M. M.; Fry, F. S.; Randolph, S. C.; Rogers, P. L.; Hebert, P. D. N. Potential use of DNA barcodes in regulatory science: Applications of the *Regulatory Fish Encyclopedia*. *J. Food Prot.* **2008**, *71*, 456–458.

(4) Hebert, P. D. N.; Cywinska, A.; Ball, S. L.; deWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. London, Ser B* **2003**, *270*, 313–319.

(5) Ward, R. D.; Hanner, R.; Hebert, P. D. N. The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* **2009**, *74*, 329–356.

(6) Waples, R. S.; Gustafson, R. G.; Weitkamp, L. A.; Myers, J. M.; Johnson, O. W.; Busby, P. J.; Hard, J. J.; Bryant, G. J.; Waknitz, F. W.;

Neely, K.; Teel, D.; Grant, W. S.; Winans, G. A.; Phelps, S.; Marshall, A.; Baker, B. M. Characterizing diversity in salmon from the Pacific Northwest. *J. Fish Biol.* **2001**, *59*, 1–41.
(7) Smith, G. R. Introgression in fishes: Significance for paleontology, cladistics, and evolutionary rates. *Syst. Biol.* **1992**, *41*, 41–57.
(8) Schlei, O. L.; Crete-Lafreniere, A.; Whiteley, A. R.; Brown, R. J.; Olsen, J. B.; Bernatchez, L.; Wenburg, J. K. DNA barcoding of eight North American coregonine species. *Mol. Ecol. Resour.* **2008**, *8*, 1212–1218.
(9) Hubert, N.; Hanner, R.; Holm, E.; Mandrak, N. E.; Taylor, E.; Burridge, M.; Watkinson, D.; Dumont, P.; Curry, A.; Bentzen, P.; Zhang, J.; April, J.; Bernatchez, L. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One* **2008**, *3*, e2490.
(10) Behnke, R. J. *Native Trout of Western North America*. American Fisheries Society, monograph 6. American Fisheries Society: Bethesda, MD, 1992.
(11) Ivanova, N. V.; Dewaard, J. R.; Hebert, P. D. N. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol. Ecol. Notes* **2006**, *6*, 998–1002.
(12) Ivanova, N. V.; Fazekas, A. J.; Hebert, P. D. N. Semi-automated, membrane-based protocol for DNA isolation from plants. *Plant Mol. Biol. Rep.* **2008**, *26*, 186–198.
(13) Ivanova, N. V.; Zemlak, T. S.; Hanner, R. H.; Hebert, P. D. N. Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Notes* **2007**, *7*, 544–548.
(14) Kumar, S.; Dudley, J.; Nei, M.; Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings Bioinf.* **2008**, *9*, 299–306.
(15) Ratnasingham, S.; Hebert, P. D. N. BOLD: The barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* **2007**, *7*, 355–364.
(16) Hajibabaei, M.; Smith, M. A.; Janzen, D. H.; Rodriguez, J. J.; Whitfield, J. B.; Hebert, P. D. N. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* **2006**, *6*, 959–964.
(17) Meusnier, I.; Singer, G. A. C.; Landry, J. F.; Hickey, D. A.; Hebert, P. D. N.; Hajibabaei, M. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **2008**, *9*, 214.
(18) Kimura, M. A simple method of estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120.
(19) Hall, T. A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symp. Ser.* **1999**, *41*, 95–98.
(20) Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.
(21) Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596–1599.
(22) Park, L. K.; Brainard, M. A.; Dightman, D. A.; Winans, G. A. Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorhynchus keta*). *Mol. Mar. Biol. Biotechnol.* **1993**, *2*, 362–370.
(23) Gharrett, A. J.; Gray, A. K.; Brykov, V. Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. *Fish. Bull.* **2001**, *99*, 528–544.
(24) Martin, M. A.; Shiozawa, D. K.; Loudenslager, E. J.; Jensen, J. N. Electrophoretic study of cutthroat trout populations in Utah. *Great Basin Nat.* **1985**, *45*, 677–687.
(25) Campbell, M.; Kozfkay, C. C.; Boone, A.; Teuscher, D. FY2006 Final Report: Genetic investigations of Bonneville cutthroat trout in the bear river drainage, Idaho: Distribution of mitochondrial DNA diversity and rainbow trout hybridization and introgression. IDFG Report Number 07-36, Idaho Department of Fish and Game, **2007**; pp 1−28.
(26) Ward, R. D.; Zemlak, T. S.; Innes, B. H.; Last, P. R.; Hebert, P. D. N. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc., B* **2005**, *360*, 1847–1857.
(27) Ward, R. D.; Holmes, B. H. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome *c* oxidase I (*cox1*) in fishes. *Mol. Ecol. Notes* **2007**, *7*, 899–907.
(28) Park, J. Y.; Lee, H. J.; Kim, W. J.; Lee, J. H.; Min, K. S. Mitochondrial cytochrome *b* sequence variation in Korean salmonids. *J. Fish Biol.* **2000**, *56*, 1145–1154.
(29) Wilson, G. M.; Thomas, W. K.; Beckenbach, A. T. Mitochondrial DNA analysis of Pacific Northwest populations of *Oncorhynchus tshawytscha*. *Can. J. Fish. Aquat. Sci.* **1987**, *44*, 1301–1305.
(30) Wilson, G. M.; Thomas, W. K.; Beckenbach, A. T. Intra- and inter-specific mitochondrial DNA sequence divergence in *Salmo*: Rainbow, steelhead, and cutthroat trouts. *Can. J. Zool.* **1985**, *63*, 2088–2094.
(31) Cronin, M. A.; Spearman, W. J.; Wilmot, R. I.; Patton, J. C.; Bickham, J. W. Mitochondrial DNA variation in chinook and chum salmon detected by restriction enzyme analysis of PCR products. *Can. J. Fish. Aquat. Sci.* **1993**, *50*, 708–715.
(32) Thomas, W. K.; Beckenbach, A. T. Variation in salmonid mitochondrial DNA: Evolutionary constraints and mechanisms of substitution. *J. Mol. Evol.* **1989**, *29*, 233–245.
(33) McVeigh, H. P.; Davidson, W. S. A salmonid phylogeny inferred from mitochondrial cytochrome *b* gene sequences. *J. Fish Biol.* **1991**, *39* (supplement A), 277–282.
(34) Thomas, W. K.; Withler, R. E.; Beckenbach, A. T. Mitochondrial DNA analysis of Pacific salmonid evolution. *Can. J. Fish. Aquat. Sci.* **1986**, *64*, 1058–1064.
(35) Augerot, X.; Steinback, C.; Fobes, N. *Atlas of Pacific Salmon*; University of California Press: Berkeley, CA, 2005; p 150.
(36) Wong, E. H. K.; Hanner, R. H. DNA barcoding detects market substitution in North American seafood. *Food Res. Int.* **2008**, *41*, 828–837.
(37) Smith, P. J.; McVeagh, S. M.; Steinke, D. DNA barcoding for the identification of smoked fish products. *J. Fish Biol.* **2008**, *72*, 464–471.
(38) Rasmussen, R. S.; Morrissey, M. T. Application of DNA-based methods to identify fish and seafood substitution on the commercial market. *Compr. Rev. Food Sci. Food Saf.* **2009**, *8*, 118–154.

Received May 14, 2009. Revised manuscript received July 21, 2009. Accepted August 1, 2009. This study was funded by the Oregon Innovation Council through the Oregon Economic Development Department. Sequence analysis was supported through grants to Paul D. N. Hebert from Genome Canada through the Ontario Genomics Institute and from the Natural Sciences and Engineering Research Council of Canada.