



2016-03-01

Phylogenetic Relationships of Cottids (Pisces: *Cottidae*) in Upper Snake River Basin of Western North America

Sun Yeong Oh
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Biology Commons](#)

BYU ScholarsArchive Citation

Oh, Sun Yeong, "Phylogenetic Relationships of Cottids (Pisces: *Cottidae*) in Upper Snake River Basin of Western North America" (2016). *All Theses and Dissertations*. 6228.
<https://scholarsarchive.byu.edu/etd/6228>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Phylogenetic Relationships of Cottids (Pisces: *Cottidae*) in Upper Snake River Basin of
Western North America

Sun Yeong Oh

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Science

Dennis K. Shiozawa, Chair
R. Paul Evans
John S. K. Kauwe

Department of Biology
Brigham Young University

March 2016

Copyright © 2016 Sun Yeong Oh

All Rights Reserved

ABSTRACT

Phylogenetic Relationships of Cottids (Pisces: *Cottidae*) in Upper Snake River Basin of Western North America

Sun Yeong Oh
Department of Biology, BYU
Master of Science

Freshwater sculpins (*Cottus*) are common throughout temperate regions of the Northern Hemisphere. Their broad distribution in the Western North America makes them a good model for understanding phylogeographic relationships among western fishes. Within much of the interior west three lineages, *C. bairdii*, *C. confusus*, and the *C. beldingii* complex, are most prevalent. The distribution of these three overlap in the Snake River Basin. All occur below Shoshone Falls on the Snake River. However, only two currently reside in the Upper Snake River above the falls. An exception are the Lost River streams of central Idaho. While these streams are technically part of the Upper Snake River Basin, they do not directly connect with the Snake River. Preliminary studies with a single mitochondrial DNA (mtDNA) gene suggested multiple pathways for *Cottus* introduction into the Lost River stream complex. Here, three mitochondrial and five nuclear genes were examined to investigate the phylogenetic relationships of these three lineages. Sequences were obtained from 71 different populations in the Lost River streams and surrounding basins. Maximum Likelihood (ML) phylogenies were constructed using these data. Our data indicate that relationships among populations within these species are complex and that no single invasion into the Lost River streams and surrounding regions can account for the phylogenetic signals detected. Instead, it appears that multiple invasions in an evolving landscape played a significant role in the modern distribution of species in this region.

Keywords: *Cottus*, phylogenetic, transcriptome, nuclear markers, mitochondrial markers, phylogeography

ACKNOWLEDGMENTS

I would like to thank my advisor, Dennis K. Shiozawa, for spending countless hours on reviewing proposals, discussing experimental designs, and advising in the interpretation of results. Also, I would like to thank my other committee members, R. Paul Evans and John Kauwe, for their helpful suggestions and encouragement. I would like to thank Nina Laitinen and R. Paul Evans for lab technique training. I would also like to thank Perry Ridge for developing the orthology inference program and for his advice on the bioinformatics portion of the project. Additionally, I would like to thank the Montana Fish, Wildlife, and Parks (MT FWP), the Colorado Parks and Wildlife (CPW), Oregon State University (OSU), the US Environmental Protection Agency (EPA), the Wyoming Game and Fish Department (WGFD), and the US Forest Service (USFS) for help in obtaining fish specimens. Lastly, I would like to thank my friends and family for believing in me and always giving positive support.

TABLE OF CONTENTS

TITLE PAGE.....	i
ABSTRACT.....	ii
ACKNOWLEDGMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
INTRODUCTION.....	1
MATERIALS AND METHODS.....	5
<i>Specimen Collections</i>	5
<i>RNA Sequencing</i>	5
<i>Primer Construction</i>	6
<i>PCR Amplification and Sequencing</i>	6
<i>Data Analysis</i>	7
RESULTS.....	9
<i>DNA Sequencing and Alignment</i>	9
<i>Phylogenetic Analysis</i>	9
<i>C. beldingii Group</i>	10
<i>C. confusus</i>	10
<i>C. bairdii</i>	11
<i>Combined Tree</i>	11
DISCUSSION.....	12
<i>C. beldingii Group</i>	12

<i>C. confusus</i>	14
<i>C. bairdii</i>	15
REFERENCES.....	18

LIST OF TABLES

Table 1: List of Populations.....	22
Table 2: List of Primers.....	29
Table 3: List of Genes.....	30
Table 4: Code for Figure 7.....	31
Table 5: Code for Figure 8.....	36
Table 6: Code for Figure 9.....	38
Table 7: Code for Figure 11.....	39

LIST OF FIGURES

Figure 1: Yellowstone Hotspots Movement.....	41
Figure 2: Map of Basins in Western North America.....	42
Figure 3: Distribution of <i>C. beldingii</i> in Western North America.....	43
Figure 4: Distribution of <i>C. confusus</i> in Western North America.....	44
Figure 5: Distribution of <i>C. bairdii</i> in Western North America.....	45
Figure 6: Condensed Maximum Likelihood Tree	46
Figure 7: <i>C. bairdii</i> Phylogeny 1.....	47
Figure 8: <i>C. bairdii</i> Phylogeny 2.....	48
Figure 9: <i>C. confusus</i> Phylogeny 1.....	49
Figure 10: <i>C. confusus</i> Phylogeny 2.....	50
Figure 11: <i>C. beldingii</i> Phylogeny.....	51
Figure 12: Serine/Arginine Phylogeny.....	52
Figure 13: 26S Protease Regulatory Subunit 6A Phylogeny.....	53
Figure 14: Ubiquitin carboxyl-terminal hydrolase 10 Phylogeny.....	54
Figure 15: Tensin Phylogeny.....	55
Figure 16: Collection Sites of <i>C. bairdii</i>	56
Figure 17: Collection Sites of <i>C. confusus</i>	57
Figure 18: Collection Sites of <i>C. beldingii</i> Groups.....	58

INTRODUCTION

The distribution of Western North American (N.A.) fish fauna have been shaped by volcanism, climatic cycles, and tectonic processes. In the last 10 million years, a series of significant hydrological events resulted from the cooling and collapsing of the Snake River Plain as the Yellowstone Hotspot effectively moved East-Northeast (as the North American Plate moved West-Southwest) across Idaho to its current location at Yellowstone National Park (Fig.1)[1,2]. The continental divide shifted progressively eastward with the hotspot movement, thus drainage basins were disrupted in this part of Western N.A.

From approximately 10 million years ago to 2.5 million years ago (mya), the Upper Snake River entered a large rift lake, Lake Idaho, in the western Snake River Plain. The outflow of Lake Idaho was westward, through an unknown route into Oregon, and ultimately entered the Pit-Sacramento River system [2,3]. Between 3.5 and 2.5 mya, a tributary to the Salmon River of Idaho back-cut into the northern sill of Lake Idaho, capturing the lake outflow into the Salmon River Basin, which was part of the Columbia River Basin. Lake Idaho drained catastrophically, cutting what is now known as Hells Canyon on the Snake River. The main tributary to Lake Idaho, the Snake River, thus also became part of the Columbia River Basin [2,4] and the Pit-Sacramento River connection was severed. Throughout this time period, as the Yellowstone Hotspot moved progressively northeastward, the volcanic track behind the hotspot cooled and collapsed, forming the Snake River Plain. This resulted in the capture of streams from surrounding basins into the plain. In particular, the Lost River streams, on the northern border of the Snake River Plain, were captured from the Salmon River and Upper Missouri River basins [5].

The two main river systems of this area are the Big Lost River on the east side of the Pioneer Mountains and the Big Wood River on the west. In the late Miocene and Pliocene, both river systems flowed to the west once they entered the Snake River Plain. The Big Lost River extends from the Salmon River divide to the Snake River Plain [6-8]. In the early Pleistocene (2.3 Ma), volcanic flows on the Snake River Plain formed the Big Lost Trough and diverted the paleo-Big Lost River eastward [8]. This volcanically dammed basin, the Big Lost River Trough subbasin [9], includes inflow from the Big Lost River, Little Lost River, and Birch Creek. The Lost River streams further east, Crooked Creek, Medicine Lodge Creek, Beaver Creek, and Camas Creek drain into the Mud Lake subbasin. Currently, none of the Lost River streams directly connect to the Snake River, instead they sink or are lost in the lava flows that characterize the plain [8]. However, during glacial maxima, shallow lakes in these two basins may have expanded, forming a combined Lake Terreton, which is also hypothesized to have had active surface connections to the Snake River [9]. If so, this could have allowed the transfer of fish among all of the Lost River streams and between Lake Terreton and the Snake River. Whitefish in the Big Lost River and the Snake River are closely related [10], suggesting a connection between the two systems, and potentially a Pleistocene overflow of Lake Terreton into the Snake River could have generated the connection required to facilitate such movement. Alternatively, the whitefish movement into the Lost River system could be associated with earlier connections of the Lost River with the Snake River, prior to diversion of the Big Lost River into the Big Lost River Trough. Data for a high stand of the lakes in the Big Lost River Trough subbasin and the Mud Lake subbasin between 80 to 100 ka are equivocal relative to a connection at that time [9].

The other major river in the Snake River Plain, the Big Wood River, drained southward to the Glens Ferry Formation on the eastern shore of Lake Idaho in the late Pliocene (4-3 Ma) [8]. Then, in the early Pleistocene, lava flows diverted the Big Wood River from its original channel to its present ephemeral connection with the Snake River [8,11]. Whitefish also occur in the Big Wood River and they are more distantly related to the Big Lost River and Upper Snake River whitefish [10], suggesting that the connection of the Lost River postdated the Big Wood River connection. These and other hydrological events provided multiple opportunities for the transfer of fishes and aquatic invertebrates between basins.

Cottids (*Cottus*: *Cottidae*: *Scorpaeniformes*) are a widespread group of fishes in this region. Of 23 recognized cottid species in Western N.A. [12-15], three dominant lineages occur in the Lost River streams and surrounding streams: the *C. beldingii* group (consisting of *C. greenei*, *C. beldingii*, and *C. leiopomus*), *C. confusus*, and the Western form of *C. bairdii*.

Members of the *C. beldingii* group occur in the Columbia River Basin, the Lahontan Basin, the Eastern Snake River Plain, the Bonneville Basin, and the Upper Colorado River Basin (Fig. 2,3). In Idaho, *C. beldingii* occurs in the Snake River System, the Bear River system, and the Big Lost River drainage. *C. greenei*, the Shoshone sculpin, is found in the springs and spring creeks within the Thousand Springs reach of the Snake River, upstream of Bliss, Idaho. *C. leiopomus*, the Wood River sculpin, occurs in the Wood River basin of south central Idaho [7].

The distribution of *C. confusus* is centered in the Columbia River Basin (Fig. 2,4). In Idaho, it is found in the Snake River tributaries below Shoshone Falls, and the Lost River drainages in the Upper Snake River Basin. It extends upstream in the Clarks Fork and Spokane River systems into Montana [7,16]. *C. confusus* does not occur upstream of Shoshone Falls on the Snake River, but is in the Salmon River [7,16].

The Western form of *C. bairdii* occurs in the Columbia River Basin, the Eastern Snake River Plain, the Bonneville Basin, the Upper Colorado River Basin and the Upper Missouri River Basin (Fig. 2,5). In Idaho, they have been collected from the Snake River and tributaries below and above Shoshone Falls, but not in Lost River streams. The systematics of these species is not well understood. *C. bairdii* is typically assumed to have two separate forms, the Eastern form and the Western form. However, recent studies have shown that to be a simplistic view of the relationships in this group (Neeley *et al.* in prep). Instead, it is a complex set of multiple undescribed species, and it is possible that more than one species occurs in Idaho [7].

The distribution of *C. bairdii* and *C. beldingii* overlap considerably in the Snake River, Columbia, Bonneville, and Colorado River basins, but they are not sympatric in two major drainage basins: the Upper Missouri River Basin and the Lahontan Basin. In the Upper Missouri River Basin, *C. bairdii* is present but *C. beldingii* is absent. In the Lahontan Basin, *C. bairdii* is absent but *C. beldingii* is present. *C. bairdii* and *C. beldingii* both occur in the Middle Snake River (below Shoshone Falls) and Lower Snake River, where they are sympatric with *C. confusus*.

Differences in distributions of these three cottid lineages suggest unique independent phylogeographic histories. The objective of this study was to examine the phylogenetic relationships of the cottids species in this region and determine if they entered the Wood River and the Lost River streams using the same pathways. Our null hypothesis is that all cottid species entered these streams through the same pathway.

In order to examine these relationships, population-level variation of these three *Cottus* lineages were examined using mtDNA and five nuclear markers. MtDNA is widely used and can

be easily amplified from a variety of taxa. It has a high evolutionary rate and a low probability of recombination, which allows the recovery of relatively recent historical events [17].

We also examined nuclear loci for independent estimates of phylogenetic relationships [18]. The nuclear genome is significantly larger than the mitochondrial genome and it has biparental inheritance, resulting in many more potential markers [19]. Protein coding nuclear genes evolve slowly, are less prone to base-composition bias, and present fewer alignment issues than mtDNA genes [20-22]. We used RNA transcripts [23] to develop primers for five protein-coding genes of *Cottus*.

MATERIALS AND METHODS

Specimen Collections

DNA for PCR amplification and sequencing was obtained from specimens archived in the Monte L. Bean Life Science Museum. Whole fish, stored in 80% EtOH, and fin clips stored in 95% or 100% EtOH. Archived extracted DNA, stored at -80 °C, was used when available. Transcriptomes were extracted from three *Cottus* specimens: *C. bairdii* (BYU# 259560; Soldier Creek, Bonneville Basin), *C. beldingii* (BYU# 259554; Mary's River, Lahontan Basin), and *C. beldingii* (BYU# 259658; Weber River, Bonneville). The specimens were field collected and tissues extracted and placed in RNA later. The tissues were kept on ice and then stored in a -80 °C freezer in the laboratory until processed.

RNA Sequencing

RNA extractions were conducted using the Qiagen Rneasy Animal Mini Kit (Qiagen Group, Valencia, CA) following the "Purification of Total RNA from Animal Tissues" protocol.

Quality of the RNA extractions were verified on the NanoDrop. Following this, cDNA libraries were constructed using Illumina TruSeq RNA Sample Prep Kit V2 (Illumina Inc., San Diego, CA) with the “Low Throughput” (LT) protocol. cDNA quality was verified using the Agilent DNA 7500 kit (Agilent Technologies Inc., Santa Clara, CA). The cDNA libraries were submitted to the Huntsman Cancer Institute for sequencing on an Illumina HiSeq 2000 (Illumina, San Diego, CA). The three *Cottus* libraries were sequenced on a single lane of a flow cell with 50 base pair (bp) single end reads. Reads were then filtered using the Sickle (najoshi, GitHub.com) quality trimmer software and assembled de novo into contigs, using Trinity assembly software [24].

Primer Construction

For nuclear markers, an orthology inference program (Perry Ridge, in review) was used to find potential orthologs (90% or more homology). GeneiousPro version 6.0.5 (Biomatters; Auckland, NZ) was used to assemble consensus sequences with the stickleback reference genome (GenBank: AANH000000000.1). Consensus sequences that overlapped with the reference genome were considered as possible primer sites. Some consensus sequences that were not overlapped were also considered. Primers were constructed using GeneiousPro or Primer3Plus. Primer sequences for mitochondrial genes were provided by Peter Unmack (Neeley *et al.*, in prep).

PCR Amplification and Sequencing

We targeted three mitochondrial genes and portions of five nuclear genes for amplification: ATPase (c. 844 bp), ND1 (c. 978 bp), ND2 (c. 1052 bp), Serine/Arginine (c. 599

bp), Tensin (c. 567 bp), Tob1 (c. 196 bp), Ubiquitin carboxyl-terminal hydrolase 10 (c. 429 bp), and 26S protease regulatory subunit 6A (c. 1082 bp) (Table 1; Fig. 16-18). For *C. confusus*, the ND4 gene was added for better phylogenetic signal. PCR reactions were run using the following reaction mixture: 2.25 μ L nuclease-free water, 0.5 μ L forward primer, 0.5 μ L reverse primer, 6.25 μ L Taq polymerase, and 3 μ L DNA for a total reaction volume of 12.5 μ L. The cyclic PCR reactions' thermal profile consisted of 2 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s annealing at 48°C (Table 2), and 1 min 30 s extension at 72°C. Amplification success was verified by ultraviolet visualization following gel electrophoresis on 1% agarose gels. We purified the PCR product using PrepEase Purification 96 well plates.

Purified PCR products, for both forward and reverse directions, were cycle-sequenced with Big Dye (Applied Biosystems, Inc. Foster City, CA) chemistry. This was accomplished with the following reaction mixture: 2.75 μ L nuclease-free water, 1.75 μ L 5x buffer, 0.5 μ L Big Dye, 0.5 μ L primer (~10 pmoles), and 5.0 μ L of purified PCR product for a total reaction volume of 10.5 μ L. Products of cycle sequencing were purified with Sephadex spin columns (G-50; Sigma-Aldrich Co., St. Louis, MO). Dry samples were submitted to the Brigham Young University DNA Sequencing Center to be sequenced using an ABI 3730xl automated sequencer (Applied Biosystems, Inc. Foster City, CA). Sequences were imported into Sequencher 4.8 (Gene Codes Corporation, Ann Arbor, MI) and edited and aligned using both forward and reverse sequences.

Data Analysis

Six analyses were run to build phylogenetic trees of *Cottus*: mtDNA (concatenated ATPase, ND1, and ND2), Serine/Arginine, Tensin, Tob1, Ubiquitin carboxyl-terminal hydrolase

10, and 26S protease regulatory subunit 6A. Nuclear genes were blasted against proteins to trim down the introns. Aligned sequences were run on TCS v1.21 [25] to identify unique haplotypes (Table 3). Different haplotypes sequences were run on Mega 6 [26] to select the best model of evolution for each tree. Once the best model was selected, we ran 100 bootstrap replications with a very strong branch swap filter for maximum likelihood. All gaps were treated as missing data and removed using the Complete-Deletion option in MEGA 6. Pacific Staghorn sculpin (*Leptocottus armatus*) and Buffalo Sculpin (*Enophrys bison*) were used as outgroups to root the tree. The mtDNA tree was analyzed using the Tamura Nei with gamma (TN93+G) model. The Serine/Arginine tree was analyzed using the Jukes Cantor (JK) model. The Tensin tree was analyzed using the Hasegawa-Kishino-Yano (HKY) model. The Tob1 tree was analyzed using the Jukes Cantor (JK) model. The Ubiquitin carboxyl-terminal hydrolase 10 tree was analyzed using the Kimura 2-parameter (K2) model. Lastly, the 26S protease regulatory subunit 6A tree was analyzed using the Tamura-Nei with gamma (TN93+G) model.

Two additional analyses were also run for combined mtDNA sequences: Maximum Parsimony and Bayesian. Maximum Parsimony- Aligned sequences of mtDNA were run on Mega6. It used subtree-pruning-regrafting (SPR) method. Bayesian- Aligned sequences of mtDNA were run on MrBayes program. It used HKY+G model, since the best model chosen was not available for MrBayes. We used the next best model available in the program. This data ran until the average standard deviation of split frequency was close to zero. Both trees show similar trend as the Maximum Likelihood tree.

RESULTS

DNA Sequencing and Alignment

We sequenced three mtDNA genes and the exons from five nuclear genes. The three mtDNA genes are ATPase, ND1, and ND2 which collectively contain 2874 base-pairs. The five nDNA exons are Serine/Arginine (583 characters), Tensin (567 characters), Tob1 (196 characters), Ubiquitin carboxyl-terminal hydrolase 10 (429 characters), and 26S protease regulatory subunit 6A (1082 characters). Preliminary comparisons of mtDNA and the nuclear markers indicated much lower resolution in the nDNA sequence data. Therefore, we ran fewer individuals for each population with the nDNA genes. Combined mtDNA sequences, from 330 individuals, resulted in 153 haplotypes and 1078 polymorphic sites. Serine/Arginine sequences, from 58 individuals, resulted in 21 haplotypes and 112 polymorphic sites. Tensin sequences, from 64 individuals, resulted in 26 haplotypes and 89 polymorphic sites. Tob1 sequences, from 60 individuals, resulted in 11 haplotypes with 30 polymorphic sites. Ubiquitin carboxyl-terminal hydrolase 10 sequences, from 56 individuals, resulted in 18 haplotypes and 44 polymorphic sites. Lastly, 26S protease regulatory subunit 6A, from 43 individuals, resulted in 25 haplotypes and 111 polymorphic sites (Table 3).

Phylogenetic Analysis

The concatenated mtDNA ML tree (Fig.6), using the TN93+G model of sequence evolution, was rooted with *L. armatus* and *E. bison* as outgroups. *C. greenei*, which was expected to be in the *C. beldingii* clade [27], form polytomy with the rest of the cottid groups. The bootstrap value (bv) that supports *C. greenei* is low (bv=51), which will collapse into five lineage polytomies. Even with the collapsing, *C. greenei* still retains its independence from the

other *C. beldingii* species groups. *C. confusus*, and the Columbia River Basin *C. bairdii* also form independent lineages, separate from other populations of western *C. bairdii* and *C. rhotheus*. *C. rhotheus* is ancestral to the Snake River, Washington, Eastern Montana, and Intermountain *C. bairdii*, but not to the Western Montana and Wisconsin *C. bairdii*. None of the nuclear genes provided significant resolution.

C. beldingii Group

As noted above, *C. greenei* does not form a monophyletic group with the rest of the *C. beldingii* clade (Fig.6). The *C. beldingii* clade is composed of four subgroups: Lower Columbia River, Lahontan Basin, *C. leiopomus*, and Upper Snake River/Bonneville/Colorado River (Fig. 11). The Lower Columbia River Basin *C. beldingii* is basal to the rest of the *C. beldingii* group. Lahontan Basin *C. beldingii* and *C. leiopomus* are sister clades and are basal to the Upper Snake River, Bonneville, and Colorado *C. beldingii*. Colorado Basin and Lost River *C. beldingii* form sister clades are ancestral to the rest of the Upper Snake River and Bonneville *C. beldingii*. Unlike mtDNA, Serine/Arginine places *C. greenei* as monophyletic with other *C. beldingii* (Fig. 12).

C. confusus

C. confusus' phylogenetic relationships with the other lineages were not very clear with the combined markers utilized in the overall analysis (Fig.9). *C. confusus* is a polytomy with both *C. bairdii* and *C. beldingii*. Within the *C. confusus* clade, bootstrap values were also low so the relationships among *C. confusus* in the region of interest were also not well resolved. Therefore, for these species, we added the mitochondrial ND4 gene. Adding the ND4 gene

significantly increased the bootstrap values (Fig. 10). Populations of *C. confusus* from the Lost River and the Salmon River systems were sister to each other. *C. confusus* in the Boise/Payette area are basal to those two drainage systems. The 26S protease regulatory subunit 6A shows similar relationship (Fig.13), although Billingsley Creek (*C. greenei*) roots with *C. confusus*.

C. bairdii

The *C. bairdii* clade is composed of six subclades: Snake River, Washington, Intermountain, Eastern Montana, Western Montana, and *C. rhotheus* from Western Montana. But the Columbia River Basin *C. bairdii* are part of the polytomy with *C. confusus* and the *C. beldingii* groups (Fig.6). The Western Montana *C. bairdii* are sister clades to *C. bairdii* in Wisconsin (Fig.7,8). These two clades are weakly basal to the remainder of the Western *C. bairdii* and the Western Montana *C. rhotheus* samples. The Eastern Montana *C. bairdii* (Fig.7,8) are basal to the Intermountain, Snake River, and Washington *C. bairdii*. However, the relationships among the Intermountain *C. bairdii* populations, the Snake River, and the Washington *C. bairdii* isn't clear due to low bootstrap values. Three nuclear markers, Serine/Arginine, Ubiquitin carboxyl-terminal hydrolase 10 and Tensin, show that *C. bairdii* in the Columbia River Basin are monophyletic from the other *C. bairdii* populations examined (Fig. 12,14-15).

Combined tree

When all mitochondrial and nuclear markers are combined, the three cottids lineages retain similar relationships in the combined mtDNA tree.

DISCUSSION

Lost River streams consist of Little Lost River, Big Lost River, Birch Creek, Medicine Lodge, Crooked Creek, Camas Creek, and Beaver Creek. Currently, five recognized species surround the Lost River streams: *C. beldingii* group (*C. beldingii*, *C. greenei*, and *C. leiopomus*), *C. confusus*, and *C. bairdii*. Of these, three occur in the Lost River streams, *C. beldingii*, *C. confusus*, and *C. bairdii*. Our data indicate that phylogeographic relationships among these species are complex and that no single invasion route into the Lost River streams and surrounding regions can account for the phylogenetic signals detected within species. Instead, it appears that the timing of invasions into an evolving landscape may have played a significant role in the modern distribution of species in this region. For example, the Snake River Plain is characterized by subsidence behind the active Yellowstone Hotspot. Species that entered the Snake River Plain earlier would be confronted with different drainage patterns than those that entered the drainage significantly later in time.

C. beldingii Group

In our analysis, *C. greenei*, was expected to be in the *C. beldingii* clade [27], but instead our ML analysis generated weak bootstrap support and essentially *C. greenei* forms a polytomy with the *C. beldingii* lineage, *C. confusus* and two *C. bairdii* lineages (Fig.6). However, Kinzinger *et al.* (2005) is likely correct. In a separate study, Neeley *et al.* (manuscript in prep), using over 8,000 bp of mitochondrial sequence, shows that *C. greenei* is a strongly supported sister taxon to the other *C. beldingii* lineage cottids, including those from the lower Columbia-Willamette River region. This supports an origin of the *C. beldingii* lineage in the Pliocene Snake-Lahontan-Sacramento system. *C. greenei* is only found in springs and spring creeks within the Thousand

Springs reach of the Snake River [7]. This could mean that *C. greenei* is a remnant from Pliocene Lake Idaho. The data that we have do not resolve this but springs and spring creeks provide a much more lentic-like habitat than in the Snake River.

Cottids ancestral to the Lower Columbia-Willamette form of *C. beldingii* could have dispersed downstream with the initial capture of Lake Idaho, or through an earlier connection between Lake Idaho and the Columbia River Basin [2-4]. The remaining three *C. beldingii* lineages occur in drainages associated with Pliocene Lake Idaho. *C. beldingii* in the Lahontan Basin, *C. beldingii* in the Upper Snake River and *C. leiopomus* in the Wood River Basin. These three form a polytomy in our analysis, but given the recognition of *C. leiopomus* as a separate species, it suggests that at this node at least one additional species of cottid, the Upper Snake River group, should be recognized. Another location of *C. beldingii* is in the Lahontan Basin [28] so the Lahontan form would retain that name. Because our discussion focus on the Lost River region of Idaho, the Lahontan Basin will not be discussed further.

C. beldingii could have entered the ancestral Wood River, generating the lineage that gives rise to *C. leiopomus*, during the time the Wood River was depositing the late Pliocene fluvial sediments in the Glens Ferry Formation in eastern Lake Idaho [8]. Alternatively, the cottids could have entered the Wood River system in the early Pleistocene, after the draining of Lake Idaho [2,4] and lava flows diverted the Big Wood River to its present ephemeral connection with the Snake River [8,11]. However, descriptions of the present-day condition of the ephemeral connection of the Big Wood River suggests that it could be a barrier to the movement of cottids into the Big Wood River Basin [10, 29].

The remaining *C. beldingii* clade are those in the Upper Snake River, Colorado River, and Bonneville basins. One line from this clade occurs in the western Lost River streams. Since

these streams have no current active connections to the Snake River, their origin is not known. Two possible routes exist for the *C. beldingii* invasion into the Lost River system. First, *C. leiopomus* could have entered the Lost River streams from a previous connection to the Upper Snake River. Second, *C. leiopomus* could have entered the Lost River streams through a headwater transfer from the Salmon River. *C. beldingii* are not known from the Salmon River [7,28], which suggests that they did not enter the Lost Rivers from the Salmon. However, their absence in that drainage could also be due to extinction or a lack of sampling.

Interestingly, the *C. beldingii* group phylogeny is similar to that of whitefish in this region [10]. In Miller's study, whitefish from the Wood River System broke off from the whitefish in the Snake River System earlier than the ones in the Lost River streams. In our study, we see the same pattern in the *C. beldingii* group. Both suggest that fish from the Big Wood River are more distantly related to the Big Lost River and Upper Snake River lineages [10]. One possible event is that drainage of Lake Idaho severed the Big Wood River and the Snake River, which was the route used for both fishes to go in those rivers. Later on, around 2.3 mya, basaltic flows from the volcanic rift zones on the Snake River Plain severed the Snake River and Big Lost River [8,10].

C. confusus

In our analysis, *C. confusus* arises from a polytomy with the *C. beldingii* lineage and two *C. bairdii* lineages (Fig. 6). Its distribution suggests a Columbia River Basin origin, but our sampling was not intensive enough to adequately address that question. Instead, we were interested in understanding how *C. confusus* entered the Lost River streams. The invasion routes for *C. confusus* into the Lost River system could be via the Snake River into the Lost River

streams concurrent with the movement of *C. beldingii*. Or, it could be via the Salmon River. If the *C. confusus* took the same route as the *C. beldingii* group, their movement could be concordant. However, *C. confusus* is not found in Lahontan Basin, the Snake River above Shoshone Falls, nor in the Wood River Basin. *C. beldingii* occurs in all of these regions. The absence of *C. confusus* may be due to regional extinctions, but the abundance of *C. confusus* in the Columbia River Basin below Shoshone Falls suggests that this species is centered in the Columbia River Basin itself. It is likely that, prior to the drainage of Lake Idaho, *C. confusus*, was restricted to the Columbia and Salmon River basins, gaining entry to the middle Snake River only after the connection with the Snake River. But *C. confusus* does occur in the Lost River system.

The ML analysis places the Salmon River *C. confusus* as sister clades to the Lost River stream *C. confusus* and the Boise and Payette River populations are basal to those two clades (Fig. 10). This supports *C. confusus* using the Salmon River route to enter the Lost River streams through headwater capture from the Salmon River as the Snake River Plain subsided [5]. Further, our analysis suggests that *C. confusus* in the Lost River streams are monophyletic, which would be expected from a single transfer event into one of the Lost River streams. These *C. confusus* subsequently dispersed to other western Lost River streams.

C. bairdii

The Western form of *C. bairdii* occurs in the Columbia River Basin, the Eastern Snake River Plain, the Bonneville Basin, the Upper Colorado River Basin and the Upper Missouri River Basin (Fig. 2, 3). In Idaho, they have been collected from the Snake River and tributaries below and above Shoshone Falls, but not in western Lost River streams. It's also absent in

Lahontan Basin. *C. bairdii*'s absence in Lahontan Basin may be due to its extinction or *C. bairdii* may not have existed in the Snake River Basin until after the Late Pliocene capture of Lake Idaho into the Columbia River Basin. That capture event would have severed access to the Lahontan Basin so that later invaders could not enter the Lahontan Basin. Fossil evidence of *C. bairdii* has been found in late Pleistocene fluvial sediments in the Snake River above Shoshone Falls [30]. It has also been reported from fluvial deposits in the Glens Ferry Formation [31,32]. These deposits are thought to have been laid down as Lake Idaho drained [3], placing these fossils in the early Pleistocene. No fossil evidence exists for this species in an earlier time period.

Our phylogenetic relationships of the *C. bairdii* group indicate that *C. bairdii* in the Snake River/Bonneville/Colorado River basins are sister taxa to *C. bairdii* in Eastern Montana (Fig.6). We hypothesize that *C. bairdii* in the Snake River system likely originated from an invasion of the Eastern Montana form of *C. bairdii* into the headwaters of the Snake River in the late Pliocene or early Pleistocene, after the capture of Lake Idaho into the Columbia River Basin. That would be concordant with its absence in the Lahontan Basin and could place its presence in the Snake River after the severing of the Wood River Basin connection that facilitated the movement of *C. beldingii* into those streams.

We do not see evidence of the intermountain *C. bairdii* in the Lost River streams which enter the Lost River Trough, even though the Lost River whitefish are closely related to the Snake River whitefish. A transfer of whitefish should have allowed a transfer of cottids. However, the shallow pluvial Lake Terreton may have provided a barrier to cottid movement but not to whitefish. Alternatively, the whitefish may have entered through a Lost River connection

prior to the invasion of *C. bairdii* into the Upper Snake River from the Missouri River headwaters.

Cottids are also found in two drainages of the eastern-most Lost River streams. Wallace and Zaroban [7] indicate that *C. confusus* inhabits both the Medicine Lodge and Beaver Creek of the Camas Creek drainage, the eastern-most of the Lost River streams, and report that *C. bairdii* was introduced to a few locations on Beaver Creek. Our investigation showed that the cottids in both Medicine Lodge Creek and Beaver Creek were the Eastern Montana form of *C. bairdii*.

Two potential sources exist for these fish. One is translocation by man. A second possibility is a stream transfer, similar to what occurred with the Salmon River – Lost River transfer of *C. confusus*, except that the eastern Lost River stream systems head at the border of the Missouri River system rather than the Salmon River. Thus, the likely invader to these Lost River streams through stream capture would be the eastern Montana *C. bairdii*. While our data do not allow us to distinguish between these two sources for the cottids in the eastern Lost River streams, it should be noted that both Beaver Creek and Camas Creek flow together into Mud Lake, the eastern basin of pluvial Lake Terretton. Medicine Lodge Creek flows into Mud Lake separately. Yet Camas Creek does not have cottids. We feel that the Medicine Lodge and Beaver Creek cottids need to be more carefully investigated along with Montana populations of the Eastern Montana *C. bairdii* populations, both adjacent to the headwaters of these Idaho streams and in streams adjacent to population centers where potential translocations may have originated. If, instead of being human-mediated, these cottids represent a new wave of natural invaders into the Snake River system, they would represent a continuation of the natural invasion processes we have seen with other cottid species in the Upper Snake River and surrounding basins.

REFERENCES

1. Leeman W, Schutt D, Hughes S. Thermal structure beneath the Snake River Plain: Implications for the Yellowstone hotspot. *Journal of Volcanology and Geothermal Research* 2009; 188(1-3):57-67.
2. Billman E, Lee J, Young D, McKell M, Evans R, Shiozawa D. Phylogenetic Divergence in a Desert Fish: Differentiation of Speckled Dace within the Bonneville, Lahontan, and Upper Snake River Basins. *Western North American Naturalist*.2010;70(1):39-47.
3. Smith G, Morgan N, Gustafson E. *Fishes of the Mio-Pliocene Ringold formation, Washington*. Ann Arbor, Mich.: Museum of Paleontology, the University of Michigan; 2000.
4. Hodges M, Link P, Fanning C. The Pliocene Lost River found to west: Detrital zircon evidence of drainage disruption along a subsiding hotspot track. *Journal of Volcanology and Geothermal Research*. 2009;188(1-3):237-249.
5. Corsi C, Elle S. *Regional Fisheries Management Investigations. Region 6 Rivers and Streams Investigations- Big Lost and Little Lost Rivers, and Birch and Medicine Lodge Creek Survey*. Idaho Department of Fish and Game, Boise, Idaho. 1989;.
6. Baldwin E. Faulting in the Lost River Range area of Idaho. *American Journal of Science*. 1951;249(12):884-902.
7. Wallace R, Zaroban D. *Native fishes of Idaho*. Bethesda, Maryland, USA: American Fisheries Society; 2013.
8. Link P, Hodges M. The Neogene drainage history of south-central Idaho. *The Geological Society of America*. 2011;21.

9. Gianniny G, Thackray G, Kaufman D, Forman S, Sherbondy M, Findeisen D. Late Quaternary highlands in the Mud Lake and Big Lost Trough subbasins of Lake Terreton, Idaho. *Special Papers: Geological Society of America*. 2002;353.
10. Miller B. The Phylogeography of *Prosopium* in Western North America [Master Thesis]. Brigham Young University. 2006;.
11. Parr C, Jordan J, Erickson J, Graham R, Rainey E, Rigby J et al. Resource Inventory Upper Snake River Basin. Idaho Department of Water Resources Water Planning Bureau. 1998;.
12. Nelson JS. *Fishes of the world*. 4th ed. ed.: John Wiley, Hoboken, N.J.; 2006.
13. Froese R, Pauly D. *FishBase*. www.fishbase.org (accessed 1 Jan 2016).
14. Etnier DA. The fishes of Tennessee. Starnes WC, editor. University of Tennessee Press, Knoxville; 1993.
15. Cheng J, Sedlazeck F, Altmuller J, Nolte AW. Ectodysplasin signalling genes and phenotypic evolution in sculpins (*Cottus*). *Proceedings Biological sciences / The Royal Society*. 2015 Sep 22;282(1815). PubMed PMID: 26354934. Pubmed Central PMCID: 4614746.
16. Databasin.org. Shorthead sculpin (*Cottus confusus*) distribution and status by HUC8 | Data Basin [Internet]. 2016 [cited 16 February 2016]. Available from: <http://databasin.org/datasets/8855fe6b49f24168b58259ad87e5ffd2>
17. Hurst G, Jiggins F. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*. 2005;272(1572):1525-1534.

18. Shaw K. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*. 2002;99(25):16122-16127.
19. Cuéllar-Pinzón J, Presa P, Hawkins S, Pita A. Genetic markers in marine fisheries: Types, tasks and trends. *Fisheries Research*. 2016;173:194-205.
20. Wiegmann B, Mitter C, Regier J, Friedlander T, Wagner D, Nielsen E. Nuclear Genes Resolve Mesozoic-Aged Divergences in the Insect Order Lepidoptera. *Molecular Phylogenetics and Evolution*. 2000;15(2):242-259.
21. Lin C, Danforth B. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution*. 2004;30(3):686-702.
22. Gaines C, Hare M, Beck S, Rosenbaum H. Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proceedings of the Royal Society B: Biological Sciences*. 2005;272(1562):533-542.
23. Betancur-R. R, Broughton R, Wiley E, Carpenter K, López J, Li C et al. The Tree of Life and a New Classification of Bony Fishes. *PLoS Curr*. 2013;.
24. Grabherr M, Haas B, Yassour M, Levin J, Thompson D, Amit I et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644-652.
25. Clement M, Posada D, Crandall K. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*. 2000;9(10):1657-1659.

26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*. 2013;30(12):2725-2729.
27. Kinziger A, Wood R, Neely D. Molecular Systematics of the Genus *Cottus* (Scorpaeniformes: Cottidae). *Copeia*. 2005;2005(2):303-311.
28. Lee D. Atlas of North American freshwater fishes, 1980- et seq. [Raleigh, N.C.]: North Carolina State Museum of Natural History; 1980.
29. Hubbs C, Miller R. The zoological evidence: correlation between fish distribution and hydrographic history in the desert basins of western United States. In: *The Great Basin, with emphasis on glacial and postglacial times*. Bulletin of the University of Utah. 1948;38:17-166.
30. Spencer JE, Smith GR, Dowling TE. Middle to late Cenozoic geology, hydrography, and fish evolution in the American Southwest. The Geological Society of America 2008; Special Paper 439:279-99.
31. Smith G, Swirydezuk K, Kimmel P, Wilkinson B. Fish Biostratigraphy of Late Miocene to Pleistocene Sediments of the Western Snake River Plain, Idaho. *Cenozoic Geology of Idaho: Idaho Bureau of Mines and Geology Bulletin*. 1982;26(519-541).
32. Smith G, Dowling T, Gobalet K, Lugaski T, Shiozawa D, Evans R. Biogeography and timing of evolutionary events among Great Basin fishes. *Great Basin Aquatic Systems History: Smithsonian Contributions to Earth Sciences*. 2002;(33):175-234.

Table 1: List of Populations. List of populations included in phylogenetic analysis.

Species	Location	Population #	Atpase (BYU ID)	ND1 (BYU ID)	ND2 (BYU ID)	Serine/ Arginine (BYU ID)	Tensin (BYU ID)	Tob1 (BYU ID)	Ubiquitin carboxyl-terminal hydrolase 10 (BYU ID)	26S protease regulatory subunit 6A (BYU ID)
<i>C. bairdii</i>	Beaver River, UT	1279	56367-56371, 56374-56377	56367-56371, 56374-56377	56367-56371, 56374-56377	56367	56367	56367	56367	56367
<i>C. bairdii</i>	Little Bear River, UT	1380	57385, 57388	57385, 57388	57385, 57388	57388	57388	57388	57388	57388
<i>C. bairdii</i>	Blacksfork, UT	2608	60676-60677, 60686-60688, 60690-60693	60676-60677, 60686-60688, 60690-60693	60676-60677, 60686-60688, 60690-60693	60686	60686	60686	60686	60686
<i>C. bairdii</i>	Willow Creek, ID	722	89246-89249, 89251	89248, 89251	89246-89251		89246		89246	
<i>C. bairdii</i>	Weiser River, ID	1571	60168-60171	60168-60171	60168-60171	60168	60168	60168	60168	
<i>C. bairdii</i>	Medicine Lodge, ID	3766	142409-142413, 142415-142419	142409-142413, 142415-142419	142409-142413, 142415-142419	142409	142409	142409	142409	142409
<i>C. bairdii</i>	Goose Creek, ID	6951	229481, 229482, 229485, 229487, 229488	229481, 229482, 229485, 229487, 229488	229481, 229482, 229485, 229487, 229488		229481	229481	229481	
<i>C. bairdii</i>	Raft River, ID	6952	229474-229477	229478-229479						
<i>C. bairdii</i>	Grimes Creek, ID	6953	229458, 229464	229458	229464	229464	229464	229464	229464	
<i>C. bairdii</i>	Jordan Creek, ID	6956	229425	229425						

Table 1: (continued)

C. bairdii	Middle Creek, ID	7318	227668, 227671-227672	227668, 227671-227672	227668, 227671-227672	227668	227668	227668	227668	
C. bairdii	Sheridan Creek, ID	7319	227688-227690	227688-227690	227688-227690	227688	227688	227688	227688	
C. bairdii	Beaver Creek, ID	7321	227762	227762	227762	227762	227762	227762	227762	227762
C. bairdii	Jarbridge, ID	7336	227795-227796	227795-227796	227795-227796	227795	227795	227795		
C. bairdii	Mores Creek, ID	7340	227858-227859	227858-227859	227858-227859	227858	227858	227858	227858	227858
C. bairdii	Payette River, ID	7344	227927-227928	227927-227928	227927-227928	227928	227928	227928	227928	227928
C. bairdii	White River, CO	1296	88974-88978, 88980-88985	88974-88978, 88980-88985	88974-88978, 88980-88985	88974	88974	88974	88974	88974
C. bairdii	Florida River-Edgemont Ranch, CO	6979	227407	227407	227407	227407	227407	227407	227407	227407
C. bairdii	Florida River-Rainbow Ranch, CO	6983	227411	227411	227411	227411	227411	227411	227411	227411
C. bairdii	Dolores River, CO	6984	227415	227415	227415	227415	227415	227415	227415	227415
C. bairdii	Duck Creek, MT	3130	59199-59210	59199-59210	59199-59210	59199	59199	59199	59199	59199
C. bairdii	Flatwillow Creek, MT	3147	59417-59421, 59425-59427, 59431, 59435-59436	59417-59421, 59425-59427, 59431, 59435-59436	59417-59421, 59425-59427, 59431, 59435-59436	59419	59419			

Table 1: (continued)

C. bairdii	Spring Creek, MT	3157	59546, 59549-59550, 59552, 59555-59556	59546, 59549-59550, 59552, 59555-59556	59546, 59549-59550, 59552, 59555-59556	59546	59546	59546		59546
C. bairdii	Libby Creek, MT	3170	59666, 59668-59671, 59674-59676	59666, 59668-59671, 59674-59676	59666, 59668-59671, 59674-59676	59666	59666	59666	59666	59666
C. bairdii	Yaak River, MT	3172	59730-59731, 59734, 59736-59737, 59739-59741	59730-59731, 59734, 59736-59737, 59739-59741	59730-59731, 59734, 59736-59737, 59739-59741	59730	59730	59730	59730	
C. bairdii	Pipe Creek, MT	3173	59781, 59785-59787, 59789-59792	59781, 59785-59787, 59789-59792	59781, 59785-59787, 59789-59792	59789	59789			
C. bairdii	St. Mary, MT	3224	68251-68253	68251-68253	68251-68253	68251	68251		68251	68251
C. bairdii	Midvale Creek, MT	3227	68396-68400, 68402-68407	68396-68400, 68402-68407	68396-68400, 68402-68407	68396	68396	68396	68396	
C. bairdii	Teton, MT	3228	68312-68313, 68316, 68320-68322, 68324-68325, 68329	68312-68313, 68316, 68320-68322, 68324-68325, 68329	68312-68313, 68316, 68320-68322, 68324-68325, 68329	68312	68312	68312	68312	68312

Table 1: (continued)

C. bairdii	Big Hole River-Pool Habitat, MT	3229	68344-68350, 68352, 68354-68356	68344-68350, 68352, 68354-68356	68344-68350, 68352, 68354-68356	68344	68344	68344	68344	68344
C. bairdii	Milk River, MT	3231	68365-68376	68365-68376	68365-68376	68365	68365	68365	68365	68365
C. bairdii	Yellowstone River, MT	3266	68720	68720	68720					
C. bairdii	Okanogan, WA	5302	116809-116811, 116813, 116815	116809-116811, 116813, 116815	116809-116811, 116813, 116815		116810	116810	116810	116810
C. bairdii	Willamette, OR	5310	116831-116832	116831-116832	116831-116832	116831	116831	116831		
C. bairdii	Sprague, OR	5965	112378, 112383-112384, 112388	112378, 112383-112384, 112388	112378, 112383-112384, 112388	112378			112378	
C. bairdii	Willow Creek, WI	3120	59089	59089	59089			59089		
C. bairdii	Hoback River, WY	836	99361, 99364-99365	99361, 99364-99365	99361, 99364-99365				99361	
C. bairdii	Butterfield, NV	5217	116722, 116728	116722, 116728	116722, 116728	116728	116728	116728	116728	
C. bairdii	Rock Creek, MT	3119	59046, 59048, 59051-59054, 59059-59060, 59065-59067	59046, 59048, 59051-59054, 59059-59060, 59065-59067	59046, 59048, 59051-59054, 59059-59060, 59065-59067	59046	59046	59046	59046	
C. beldingii	Mary's River, NV	2798	58058	58058	58058	58058	58058	58058	58058	58058

Table 1: (continued)

C. beldingii	Humboldt River, NV	2802	58135, 58142, 58144, 58148-58154	58135, 58142, 58144, 58148-58154	58135, 58142, 58144, 58148-58154	58135	58135	58135	58135	58135
C. beldingii	Desert Creek, NV	3047	59016-59017	59016-59017	59016-59017	59017	59017	59017	59017	59017
C. beldingii	Stump Creek, ID	905	99508, 9511-99513	99508, 9511-99513	99508, 9511-99513	99508	99508	99508	99508	99508
C. beldingii	Cherry Creek, ID	1681	60301-60302	60301-60302	60301-60302	60299	60299	60299	60299	60299
C. beldingii	Pass Creek, ID	1683	60320-60325	60320-60325	60320-60325	60325	60325	60325	60325	
C. beldingii	Weber, UT	1737	60530-60531	60530-60531	60530-60531	60531	60531	60531	60531	60531
C. beldingii	Blacksmith Fork, UT	3253	57236, 57240	57236, 57240			57236	57236	57236	57236
C. beldingii	Squirrel Creek, ID	6950	229498-229500	229500	229498-229500					
C. beldingii	Toponce Creek, ID	6955	229448-229451	229448-229449	229449-229450		229448			
C. beldingii	Provo River, UT	7303	227602-227603	227602-227603	227602	227602	227602	227602	227602	227602
C. beldingii	Eagle River, CO	7041	227424	227424	227424	227424	227424	227424	227424	227424
C. beldingii	Willow Creek, CO	8528	260163	260163	260163	260163	260163	260163	260163	260163
C. beldingii	Yakima River, WA	3268	68823-68824	68823-68824	68823-68824	68823	68823	68823	68823	68823
C. leiopomus	Big Wood River, ID	2924	62162-62165	62162-62165	62162-62165		62162	62162	62162	
C. greenei	Billingsley Creek, ID	7955	250918, 250920, 250922, 250931, 250933	250918, 250920, 250922, 250931, 250933	250918, 250920, 250922, 250931, 250933	250918	250918	250918	250918	250918

Table 1: (continued)

C. confusus	Birch Creek, ID	1680	60289-60290	60289-60290	60289-60290	60290	60290	60290	60290	60290
C. confusus	Big Gulch Creek, ID	1682	60309-60312, 60314-60316, 60318	60309-60312, 60314-60316, 60318	60309-60312, 60314-60316, 60318	60309	60309	60309	60309	60309
C. confusus	18 mile, ID	1684	60329, 60332	60329, 60332	60329, 60332	60329	60329	60329	60329	60329
C. confusus	Wildhorse Creek, ID	1685	60341-60343, 60345-60348	60341-60343, 60345-60348	60341-60343, 60345-60348	60341	60341	60341	60341	60341
C. confusus	Big Lost River, ID	1688	60371-60372, 60375, 60377	60371-60372, 60375, 60377	60371-60372, 60375, 60377	60371	60371	60371	60371	
C. confusus	Boise River, ID	2804	58074, 58077	58074, 58077	58074	58074	58074	58074	58074	58074
C. confusus	Salmon River, ID	4870	69174	69174	69174	69174	69174	69174	69174	69174
C. confusus	Clear Creek, ID	7341	227879-227880	227879-227880	227879	227880	227880	227880	227880	227880
C. confusus	Crooked Creek, ID	7342	227906-227907	227906-227907	227906	27907	27907	27907		27907
C. confusus	Big Gulch Creek, ID	8735	261463, 261465-261470, 261472	261463, 261465-261470, 261472	261463, 261465-261470, 261472	261463	261463	261463	261463	
E. bison	Padilla, WA	2843	61901	61901	61901	61901	61901	61901	61901	61901
C. aleuticus	Doris, Canada	5628	114350	114350	114350	114350	114350	114350	114350	114350
L. armatus	Snohomish, WA	3043	58155	58155	58155	58155	58155	58155	58155	58155

Table 1: (continued)

C. cognatus	Skalkaho Creek, MT	2914	62473- 62475, 62477- 62481, 62483- 62485	62473- 62475, 62477- 62481, 62483- 62485	62473- 62475, 62477- 62481, 62483- 62485		62473	62473	62473	62473
C. rhotheus	Libby Creek, MT	3171	59696, 59698- 59699, 59701- 58706	59696, 59698- 59699, 59701- 58706	59696, 59698- 59699, 59701- 58706	59696	59696	59696		
C. rhotheus	Pleasant Valley, MT	3174	59815, 59817, 59819, 59820, 59822- 59824, 59832, 59853	59815, 59817, 59819, 59820, 59822- 59824, 59832, 59853	59815, 59817, 59819, 59820, 59822- 59824, 59832, 59853	59853	59853	59853		

Table 2: List of Primers. List of primers sequences used in this study and annealing temperatures.

Gene	Primers	Sequences	Annealing temperatures (°C)
ND1	16ScottF	Peter Unmack in prep	48 °C
	Ille33R	Peter Unmack in prep	48 °C
ND2	Gln56F	Peter Unmack in prep	48 °C
	Trp62R	Peter Unmack in prep	48 °C
ATPase	Lys.22F	Peter Unmack in prep	48 °C
	CO3.23R	Peter Unmack in prep	48 °C
Serine/Arginine	ENSG.F	CAGTGATTGGGAGGACTGGG	62 °C
	ENSG.R	CAAAAGCCGAAGAGATGCC	62 °C
Tensin	TNS1-F	ACAGACCTSTCTTGCAWCGC	58 °C
	TNS1-R	TCCGCAAGAAAGACTCCCTG	58 °C
Tob1	TOB1-F	CAAYGCCAAGGAGTTTGTG	52 °C
	TOB1-R	TGCAGCCATCACAGATCAC	52 °C
Ubiquitin carboxyl-terminal hydrolase 10	USP10-F	CRGCAGCCAAGTTTKCCCC	55 °C
	USP10-R	CAGAGGCTTGGAGTTGTGGAAG	55 °C
26S protease regulatory subunit 6A	PSMC3-F	AGGCGTAGTACTGAAGGTTG	55 °C
	PSMC3-R	ACGAAGTGGAGGATGGAATC	55 °C

Table 3: List of Genes. Genes used in the study and their base pairs, variance, haplotypes, ratio of haplotypes to base pairs, and ratio of variance to base pairs.

Gene	# Base Pairs	# Variance	# Haplotypes	# Haplotypes/ # Base Pairs	# Variance/ # Base Pairs
mtDNA (ATPase, ND1, ND2)	2874	1087	153	0.053	0.38
Serine/ Arginine	583	112	21	0.036	0.19
Tensin	567	89	26	0.046	0.16
TOB1	196	30	11	0.056	0.15
Ubiquitin carboxyl-terminal hydrolase 10	429	44	18	0.042	0.10
26S protease regulatory subunit 6A	1082	111	25	0.023	0.10

Table 4: Code for Figure 7.

Code	Location (Population #)	Basin	BYU ID
BE	Beaver Creek, ID (7321)	Columbia River Basin	227762
BEV	Beaver River, UT (1279)	Bonneville Basin	56367-56371 , 56374-56377
BI.1	Big Hole River, MT (3229)	Upper Missouri River Basin	68356
BI.2	Big Hole River, MT (3229)	Upper Missouri River Basin	68354
BI.3	Big Hole River, MT (3229)	Upper Missouri River Basin	68348
BI.4	Big Hole River, MT (3229)	Upper Missouri River Basin	68355
BI.5	Big Hole River, MT (3229)	Upper Missouri River Basin	68345
BI.6	Big Hole River, MT (3229)	Upper Missouri River Basin	68346
BI.7	Big Hole River, MT (3229)	Upper Missouri River Basin	68347
BI.8	Big Hole River, MT (3229)	Upper Missouri River Basin	68349, 68350, 68352
BI.9	Big Hole River, MT (3229)	Upper Missouri River Basin	68344

Table 4: (continued)

BL.1	Blacks Fork, UT (2608)	Upper Colorado River Basin	60676
BL.2	Blacks Fork, UT (2608)	Upper Colorado River Basin	60677, 60687, 60690- 60692
BL.3	Blacks Fork, UT (2608)	Upper Colorado River Basin	60686, 60688
BL.4	Blacks Fork, UT (2608)	Upper Colorado River Basin	60693
BU	Butterfield, NV (5217)	White River	116722, 116728
DO	Dolores River, CO (6984)	Upper Colorado River Basin	227415
DU.1	Duck Creek, MT (3130)	Upper Missouri River Basin	59202
DU.2	Duck Creek, MT (3130)	Upper Missouri River Basin	59199, 59200, 59201, 59203-59210
FL	Florida River-Edgemont Ranch, CO (6979)	Upper Colorado River Basin	227407
FLA	Flatwillow Creek, MT (3147)	Upper Missouri River Basin	59417-59421, 59425-59427, 59431, 59435- 59436
GO	Goose Creek, ID (6951)	Columbia River Basin	229481, 229482, 229485, 229487, 229488
GR	Grimes Creek, ID (6953)	Columbia River Basin	229458, 229464

Table 4: (continued)

HO	Hoback River, WY (836) Raft River, ID (6952) Willow Creek, ID (722)	Columbia River Basin	99361, 99364, 99365 229478, 229479 89249, 89251
JA	Jarbridge, ID (7336)	Columbia River Basin	227795, 227796
JO	Jordan Creek, ID (6956)	Columbia River Basin	229425
LI	Little Bear River, UT (1380)	Bonneville Basin	57385, 57388
MD.1	Midvale Creek, MT (3227) Milk River, MT (3231) Teton River, MT (3228)	Upper Missouri River Basin	68396-68397, 68399- 68400, 68403-68407, 68365- 68376, 68312, 68313, 68316, 68320- 68322, 68324-68325, 68329
MD.2	Midvale Creek, MT (3227)	Upper Missouri River Basin	68398
MD.3	Midvale Creek, MT (3227)	Upper Missouri River Basin	68402
ME.1	Medicine Lodge, ID (3766)	Columbia River Basin (Lost River streams)	142417, 142418
ME.2	Medicine Lodge, ID (3766)	Columbia River Basin (Lost River streams)	142409- 142413, 142415, 142416
ME.3	Medicine Lodge, ID (3766)	Columbia River Basin (Lost River streams)	142419
MID	Middle Creek, ID (7318)	Columbia River Basin	227668, 227671, 227672

Table 4: (continued)

MR	Mores Creek, ID (7340) Raft River, ID (6952)	Columbia River Basin	227858, 227859 229474- 229477
OK	Okanogan, WA (5302) Yakima River, WA (3268)	Columbia River Basin	116809-116811, 116813,116815 68824
PA	Payette River, ID (7344)	Columbia River Basin	227927, 227928
RA	Florida River- Rainbow Ranch, CO (6983)	Upper Colorado River Basin	227411
RO	Rock Creek, MT (3119) Spring Creek, MT (3157)	Upper Missouri River Basin	59046, 59048, 59051- 59054, 59059, 59060, 59065- 59067, 59546, 59549- 59550, 59552
SH	Sheridan Creek, ID (7319)	Columbia River Basin	227688, 227689, 227690
SP	Spring Creek, MT (3157)	Upper Missouri River Basin	59555, 59556
ST	St. Mary River, MT (3224)	Upper Missouri River Basin	68251- 68253
WE.1	Weiser River, ID (1571)	Columbia River Basin	60170
WE.2	Weiser River, ID (1571)	Columbia River Basin	60171
WE.3	Weiser River, ID (1571)	Columbia River Basin	60169
WE.4	Weiser River, ID (1571)	Columbia River Basin	60168

Table 4: (continued)

WH.1	White River, CO (1296)	Upper Colorado River Basin	88975
WH.2	White River, CO (1296)	Upper Colorado River Basin	88978, 88982
WH.3	White River, CO (1296)	Upper Colorado River Basin	88980
WH.4	White River, CO (1296)	Upper Colorado River Basin	88983, 88985
WH.5	White River, CO (1296)	Upper Colorado River Basin	88981, 88984
WH.6	White River, CO (1296)	Upper Colorado River Basin	88976
WH.7	White River, CO (1296)	Upper Colorado River Basin	88974, 88977
WS	Willow Creek, ID (722) Squirrel Creek, ID (6950)	Columbia River Basin	89246- 89248 229498, 229499
YE	Yellowstone River, MT (3266)	Upper Missouri River Basin	68720

Table 5: Code for Figure 8.

Code	Location (Population #)	Basin	BYU ID
LI.1	Libby Creek, MT (3171) Pleasant Valley, MT (3174)	Columbia River Basin	59696, 59699, 59701, 59702, 59704, 59706, 59820
LI.2	Libby Creek, MT (3171)	Columbia River Basin	59698
LI.3	Libby Creek, MT (3171)	Columbia River Basin	59705
LI.4	Libby Creek, MT (3171)	Columbia River Basin	59703
LI.5	Libby Creek, MT (3170)	Columbia River Basin	59666, 59669- 59671, 59674- 59676
LI.6	Libby Creek, MT (3170)	Columbia River Basin	59668
PI.1	Pipe Creek, MT (3173)	Columbia River Basin	59791
PI.2	Pipe Creek, MT (3173)	Columbia River Basin	59781, 59785- 59787, 59789, 59790, 59792
PL.1	Pleasant Valley, MT (3174)	Columbia River Basin	59823, 59824
PL.2	Pleasant Valley, MT (3174)	Columbia River Basin	59832

Table 5: (continue)

PL.3	Pleasant Valley, MT (3174)	Columbia River Basin	59815, 59817, 59819, 59822, 59853
SK.1	Skalkaho, MT (2914)	Columbia River Basin	62473, 62477, 62481, 62483, 62485
SK.2	Skalkaho, MT (2914) Yaak River, MT (3172)	Columbia River Basin	62474, 62475, 62478, 62484, 59731, 59734, 59736, 59737, 59739, 59741
SK.3	Skalkaho, MT (2914)	Columbia River Basin	62479, 62480
WI	Willow Creek, WI (3120)	Mississippi River	59089
YA.1	Yaak River, MT (3172)	Columbia River Basin	59740
YA.2	Yaak River, MT (3172)	Columbia River Basin	59730

Table 6: Code for Figure 9

Code	Location (Population #)	Basin	BYU ID
BO	Boise River, ID (2804)	Columbia River Basin	58074, 58077
BR	Birch Creek, ID (1680)	Columbia River Basin (Lost River streams)	60289, 60290
BW	Big Lost River, ID (1688) Wildhorse Creek, ID (1685)	Columbia River Basin (Lost River streams)	60371, 60372, 60375, 60377 60341- 60343, 60345- 60348
CL.1	Clear Creek, ID (7341)	Columbia River Basin	227879
CL.2	Clear Creek, ID (7341)	Columbia River Basin	227880
CR	Crooked Creek, ID (7342)	Columbia River Basin (Lost River streams)	227906, 227907
SB	Salmon River, ID (4870) Big Gulch Creek, ID (8735) Big Gulch Creek, ID (1682) 18 mile, ID (1684)	Columbia River Basin	69174 261463, 261465- 261470 60309- 60312, 60314- 60316, 60318 60329, 60332

Table 7: Code for Figure 11.

Code	Location (Population #)	Basin	BYU ID
BG	Big Wood River, ID (2924)	Columbia River Basin	62162- 62165
BP	Cherry Creek, ID (1681) Pass Creek, ID (1683)	Columbia River Basin	60301-60302 60320-60323
DE	Desert Creek, NV (3047)	Lahontan Basin	59016, 59017
EW	Eagle River, CO (7041) Willow Creek, CO (8528)	Upper Colorado River Basin	227424 260163
HU	Humboldt River, NV (2802)	Lahontan Basin	58135, 58142, 58144, 58148- 58154
MA	Mary's River, NV (2798)	Lahontan Basin	58058
SB	Stump Creek, ID (905) Blacksmith Fork, UT (3253) Toponce Creek, ID (6955)	Columbia River Basin Bonneville Basin Columbia River Basin	99508, 99511-99513 57236, 57240 229448, 229449- 229451
SQ	Squirrel Creek, ID (6950)	Columbia River Basin	229500

Table 7: (continued)

WP	Weber, UT (1737) Provo River, UT (7303)	Bonneville Basin	60530, 60531 227602, 227603
WY	Willamette, OR (5310) Yakima River, WA (3268)	Upper Willamette Columbia River Basin	116832 68823

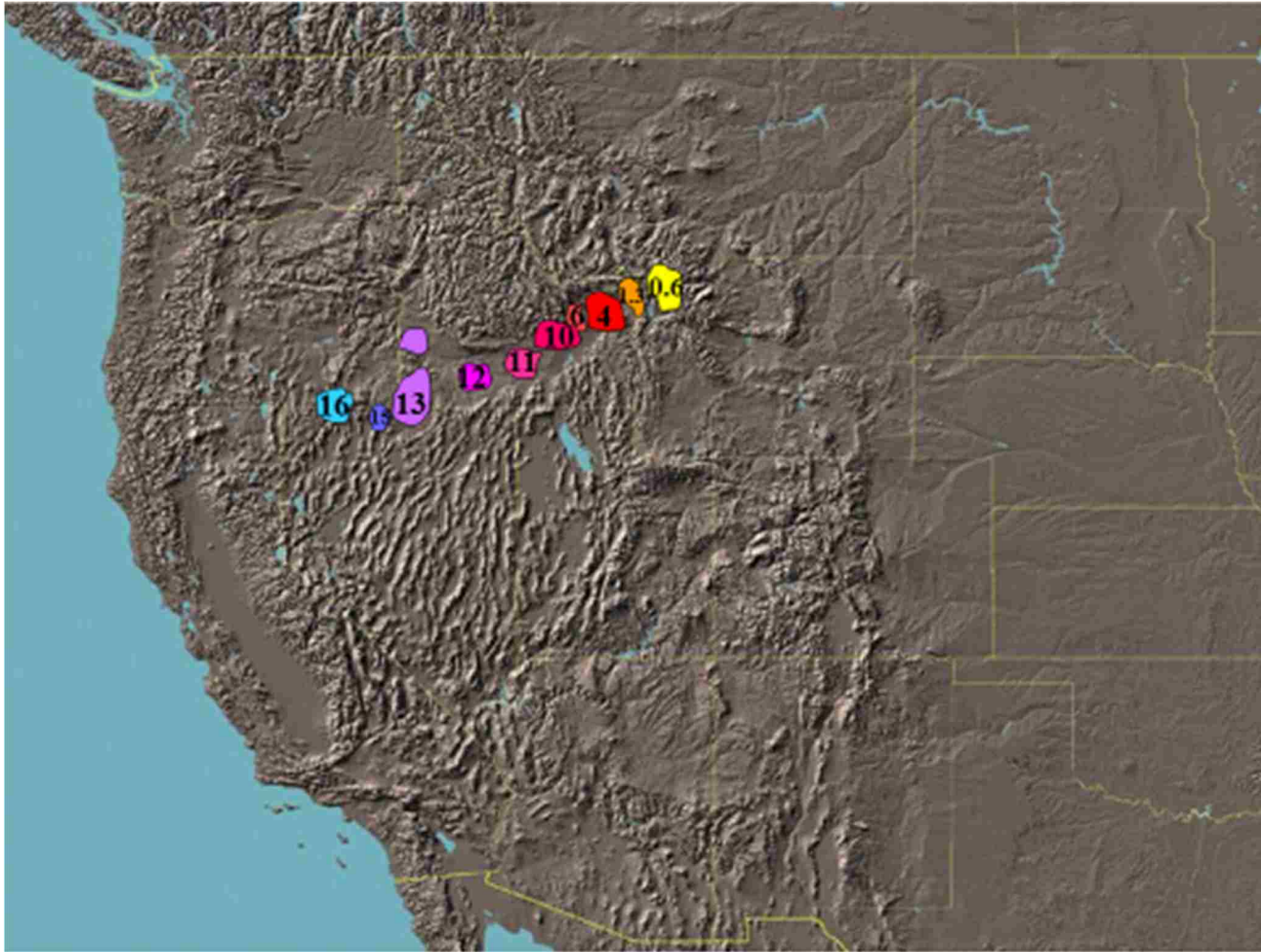


Figure 1: Yellowstone Hotspots Movement. Map of Western North America with Yellowstone hotspots movement with their respective years (mya).

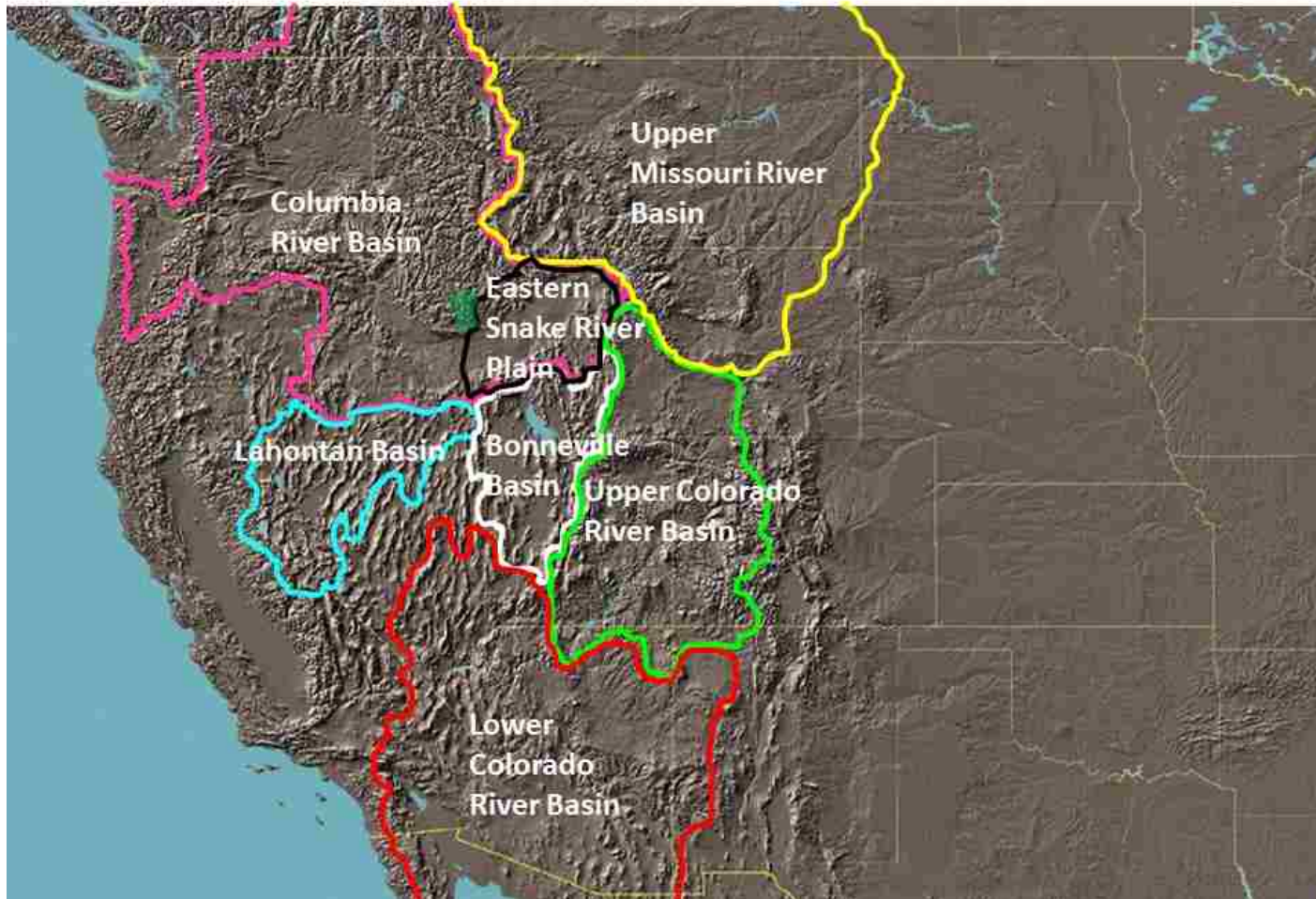


Figure 2: Map of Basins in Western North America. Map of Western North America with Columbia River Basin, Upper Missouri River Basin, Eastern Snake River Plain (Upper Snake River Basin), Lahontan Basin, Bonneville Basin, Upper Colorado River Basin, and Lower Colorado River Basin.

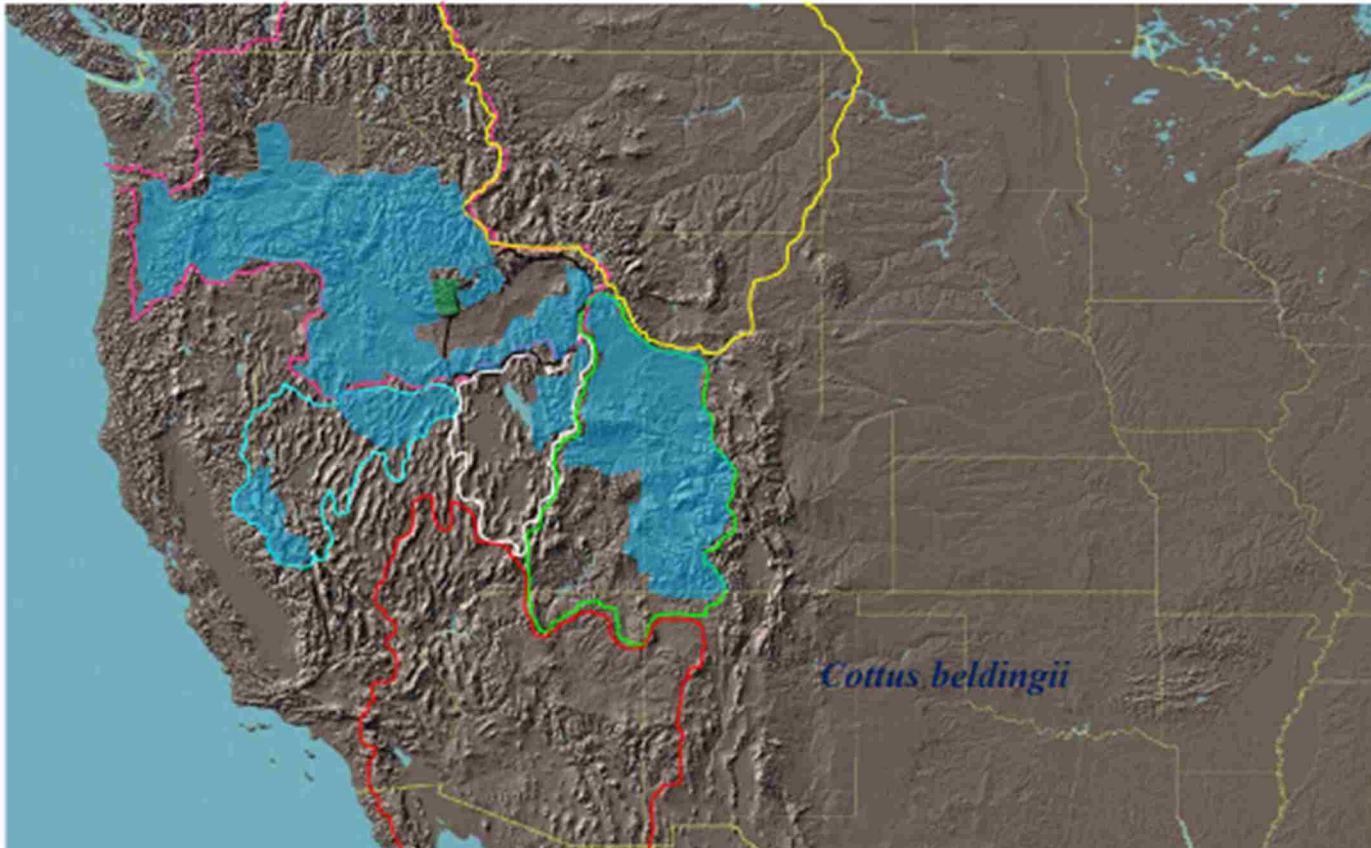


Figure 3: Distribution of *C. beldingii* in Western North America. Distribution of *C. beldingii* is highlighted in blue.



Figure 4: Distribution of *C. confusus* in Western North America. Distribution of *C. confusus* is highlighted in green.

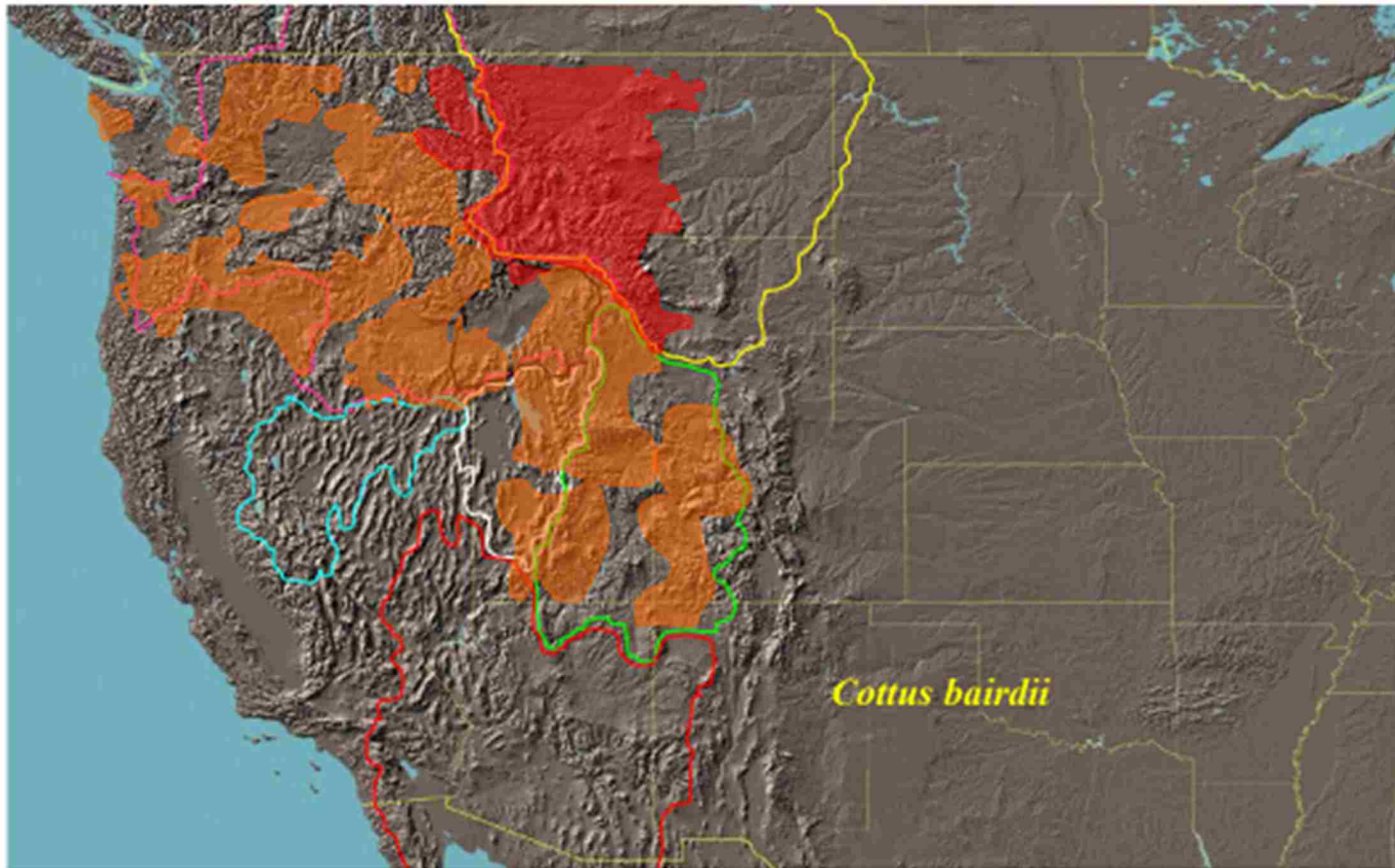


Figure 5: Distribution of *C. bairdii* in Western North America. Distribution of western *C. bairdii* is highlighted in orange and Montana *C. bairdii* is highlighted in red.

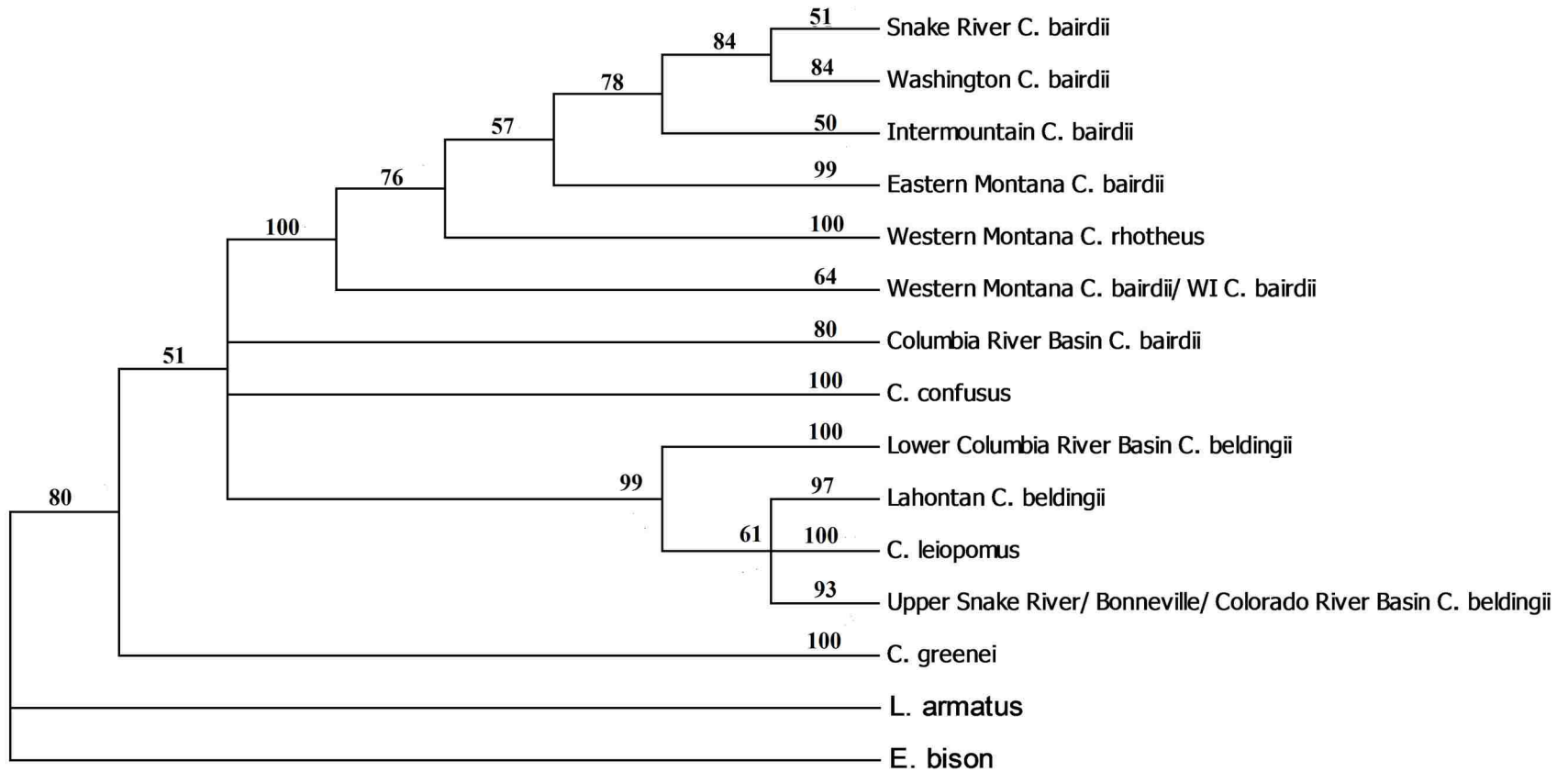


Figure 6: Condensed Maximum Likelihood Tree. Maximum Likelihood tree of cottids, using concatenated mtDNA (ATPase, ND1, and ND2). Bootstrap values lower than 50 were collapsed.

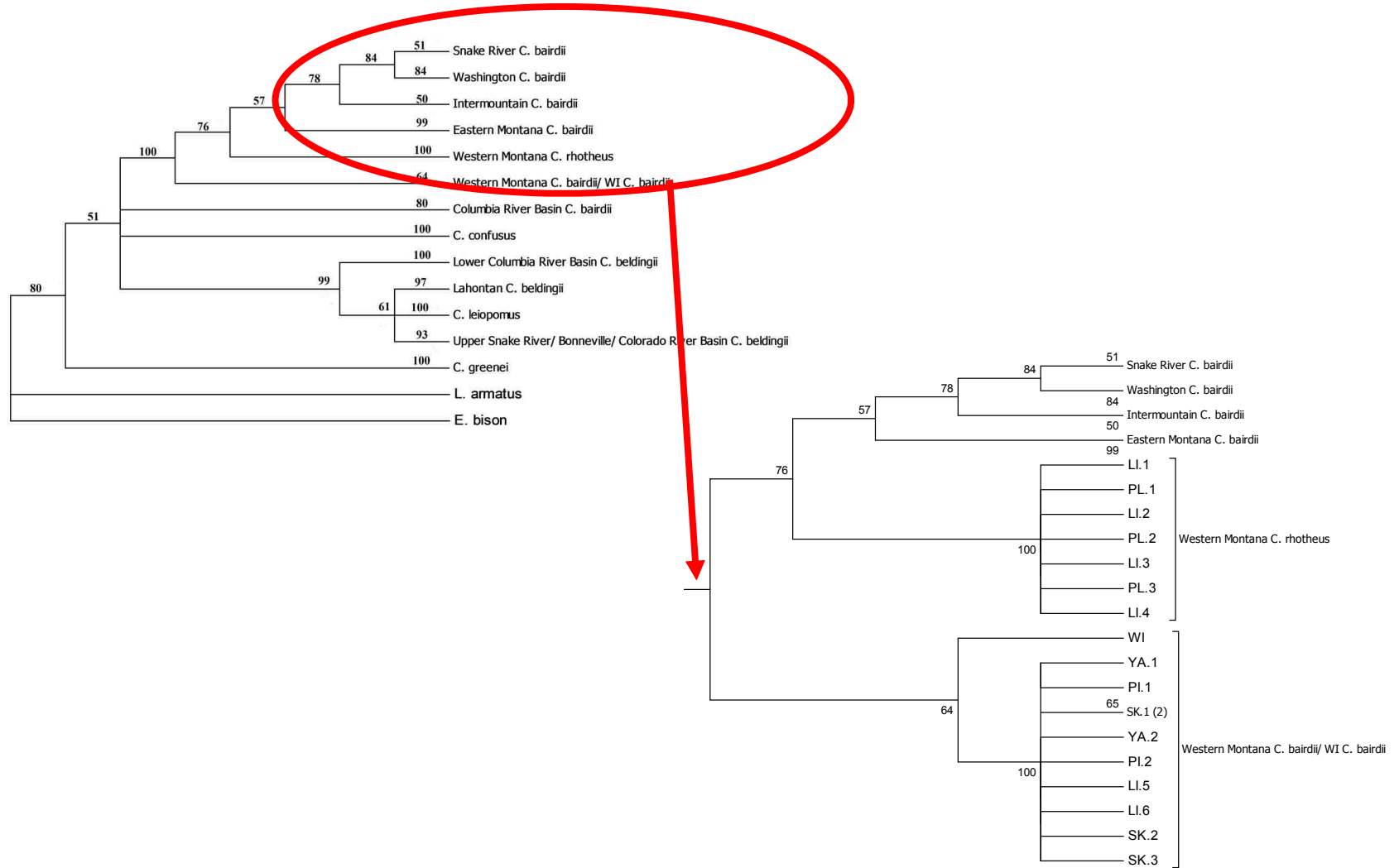


Figure 8: *C. bairdii* Phylogeny 2. Western North America *C. bairdii* subtree 2, using mtDNA (ATPase, ND1, and ND2). Bootstrap values lower than 50 were collapsed. Table for the codes are in Table 5.

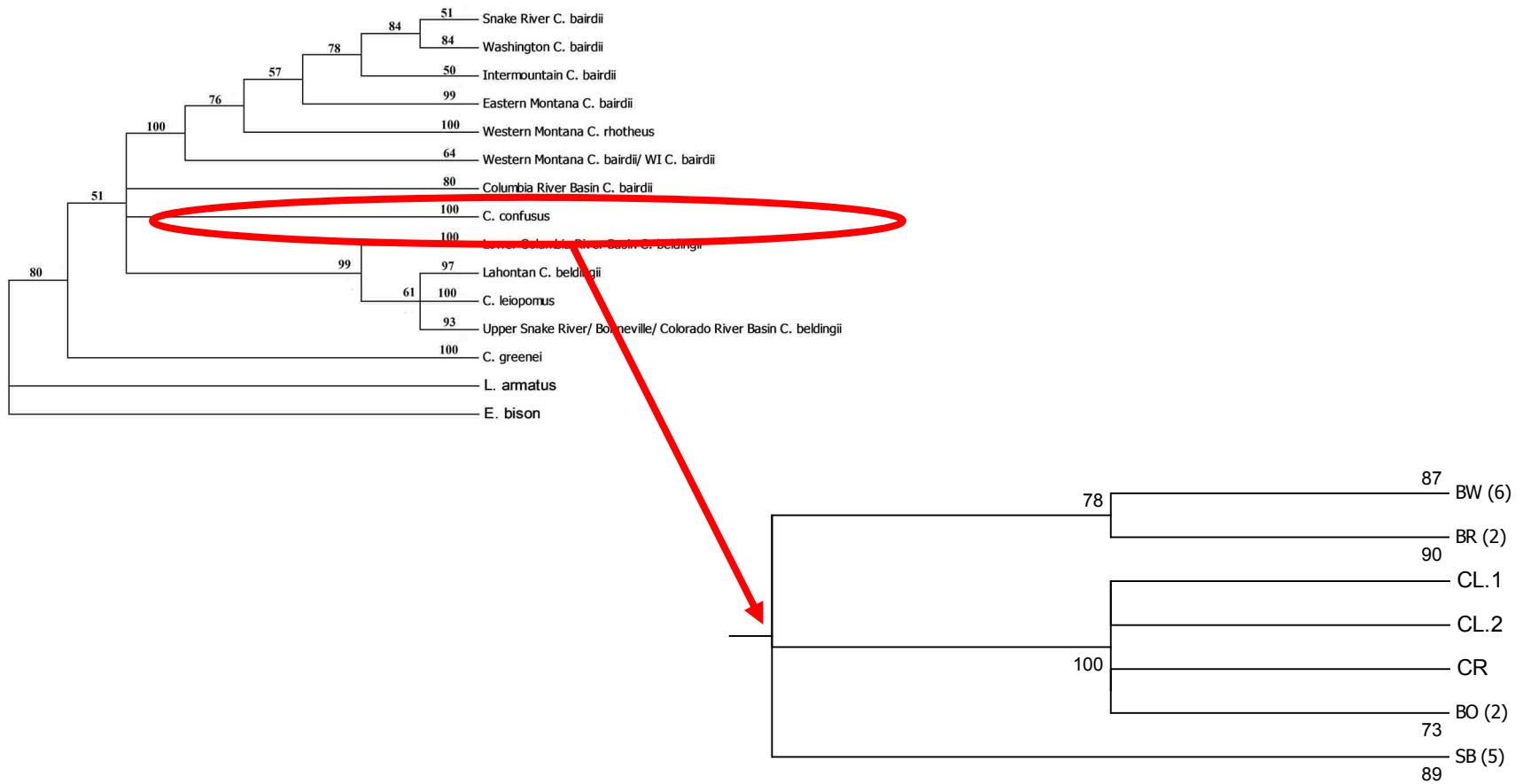


Figure 9: *C. confusus* Phylogeny 1. Western North America *C. confusus* subtree, using mtDNA (ATPase, ND1, and ND2). Bootstrap values lower than 50 were collapsed. Table for the codes are in Table 6.

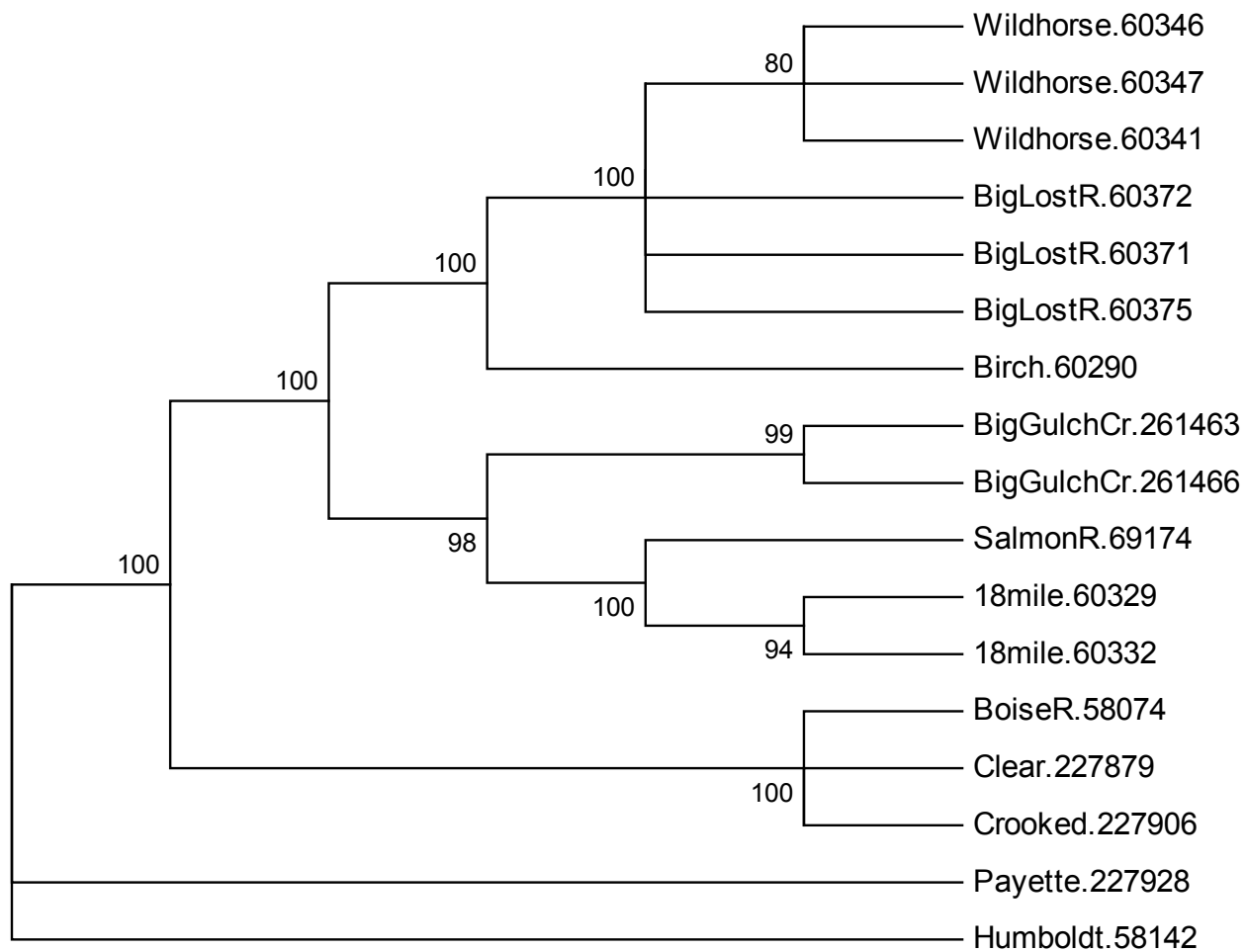


Figure 10: *C. confusus* Phylogeny 2. Western North America *C. confusus* subtree with additional gene ND4 (ATPase, ND1, ND2, and ND4). Bootstrap values lower than 50 were collapsed.

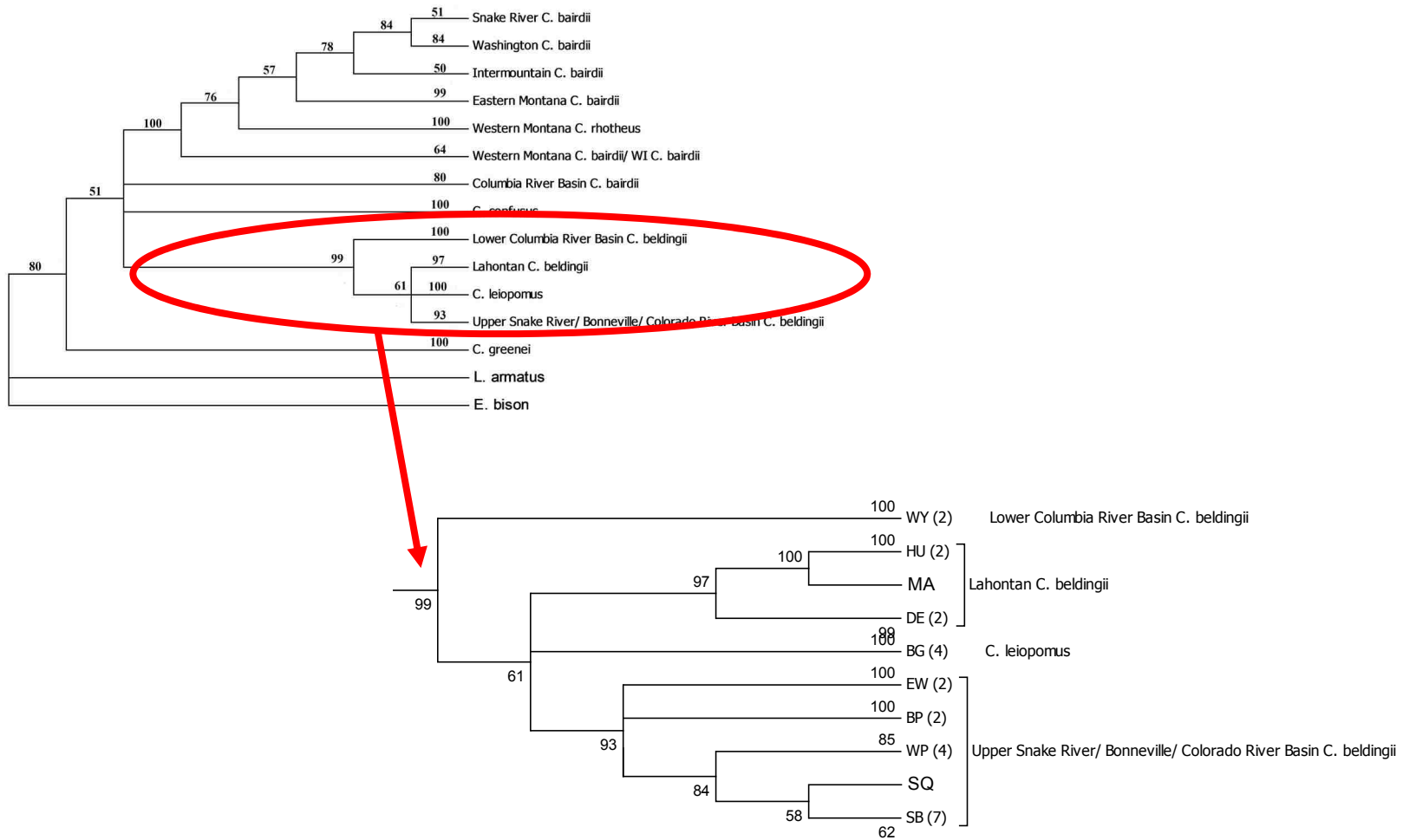


Figure 11: *C. beldingii* Phylogeny. Western North America *C. beldingii* subtree, using mtDNA (ATPase, ND1, and ND2). Bootstrap values lower than 50 were collapsed. Table for the codes are in Table 7.

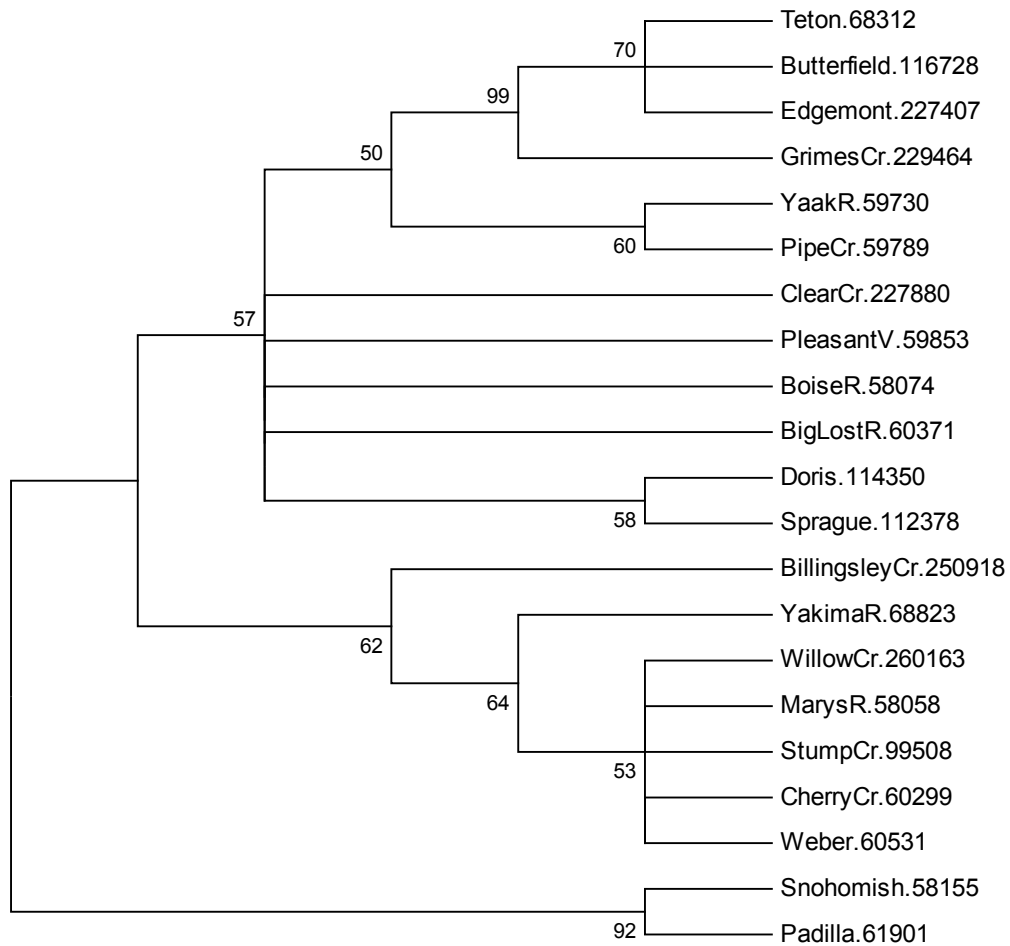


Figure 12: Serine/Arginine Phylogeny. Bootstrap values lower than 50 were collapsed.

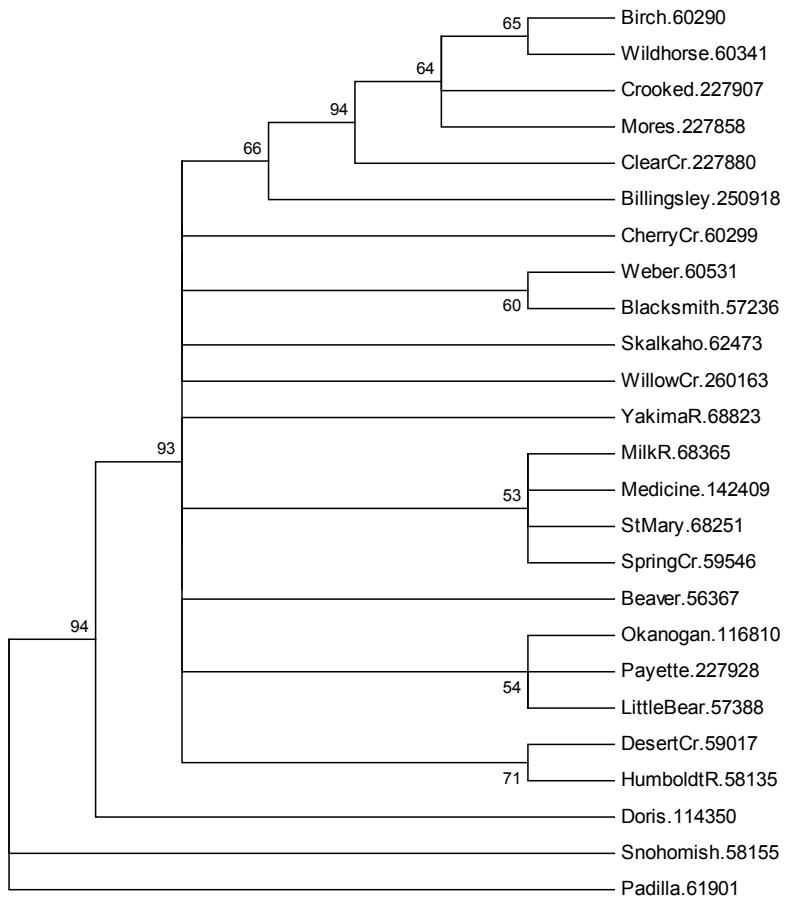


Figure 13: 26S protease regulatory subunit 6A Phylogeny. Maximum Likelihood tree of 26S protease regulatory subunit 6A. Bootstrap values lower than 50 were collapsed.

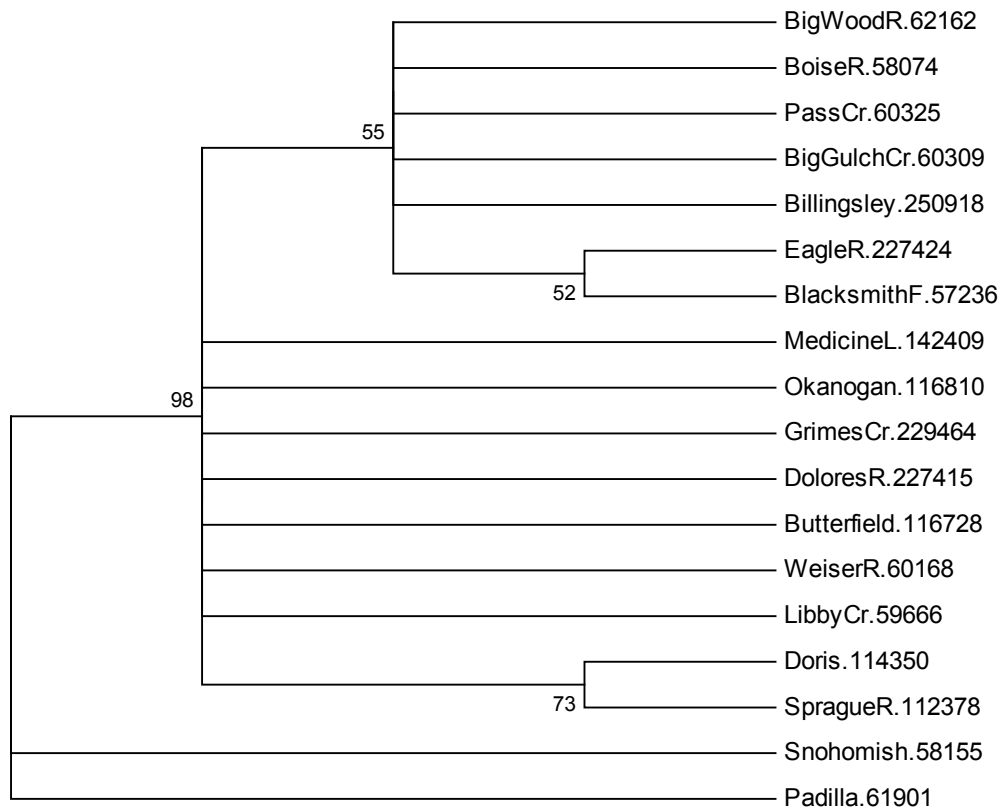


Figure 14: Ubiquitin carboxyl-terminal hydrolase 10 phylogeny. Bootstrap values lower than 50 were collapsed.

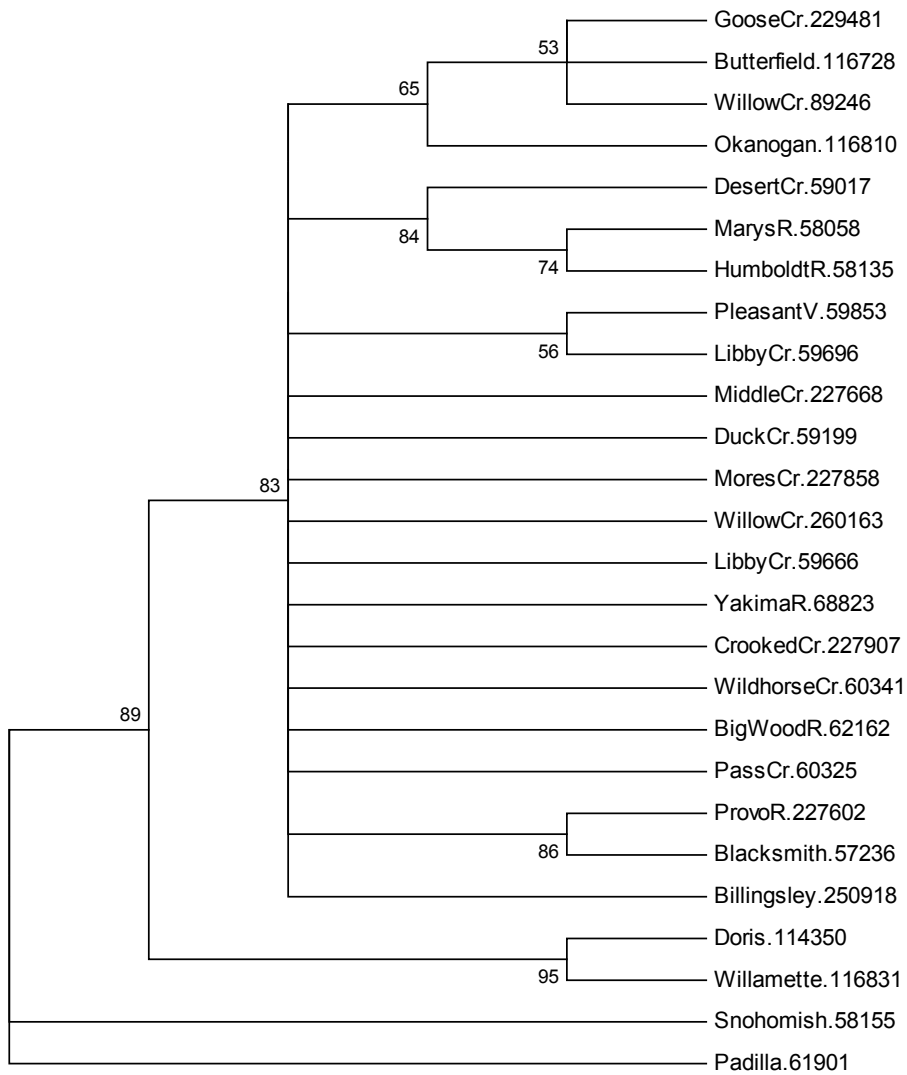


Figure 15: Tensin Phylogeny. Bootstrap values lower than 50 were collapsed.

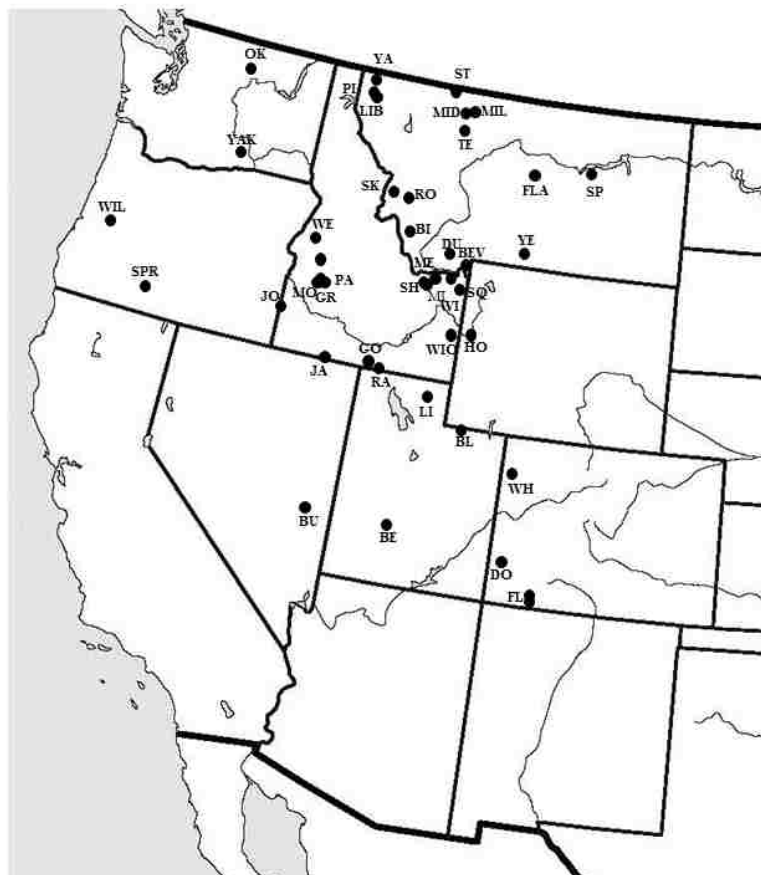


Figure 16: Collection Sites of *C. bairdii*. BE= Beaver River, UT; BEV= Beaver Creek, ID; BI= Big Hole River, MT; BL= Blacksfork, UT; BU= Butterfield, NV; DO= Dolores River, CO; DU= Duck Creek, MT; FL= Florida River, CO; FLA= Flatwillow Creek, MT; GO= Goose Creek, ID; GR= Grimes Creek, ID; HO= Hoback River, WY; JA= Jarbridge, ID; JO= Jordan Creek, ID; LI= Little Bear River, UT; LIB= Libby Creek, MT; ME= Medicine Lodge, ID; MI= Middle Creek, ID; MID= Midvale Creek, MT; MIL= Milk River, MT; MO= Mores Creek, ID; OK= Okanogan, WA; PA= Payette River, ID; PI= Pipe Creek, MT; RA= Raft River, ID; RO= Rock Creek, MT; SH= Sheridan Creek, ID; SK= Skalkaho, MT; SP= Spring Creek, MT; SPR = Sprague, OR; SQ= Squirrel, ID; ST= St. Mary, MT; TE= Teton, MT; WE= Weiser River, ID; WH= White River, CO; WI= Willow Creek, ID; WIL= Willamette, OR; WIO= Willow Creek, WI; YA= Yaak River, MT; YAK= Yakima River, WA; YE= Yellowstone River, MT

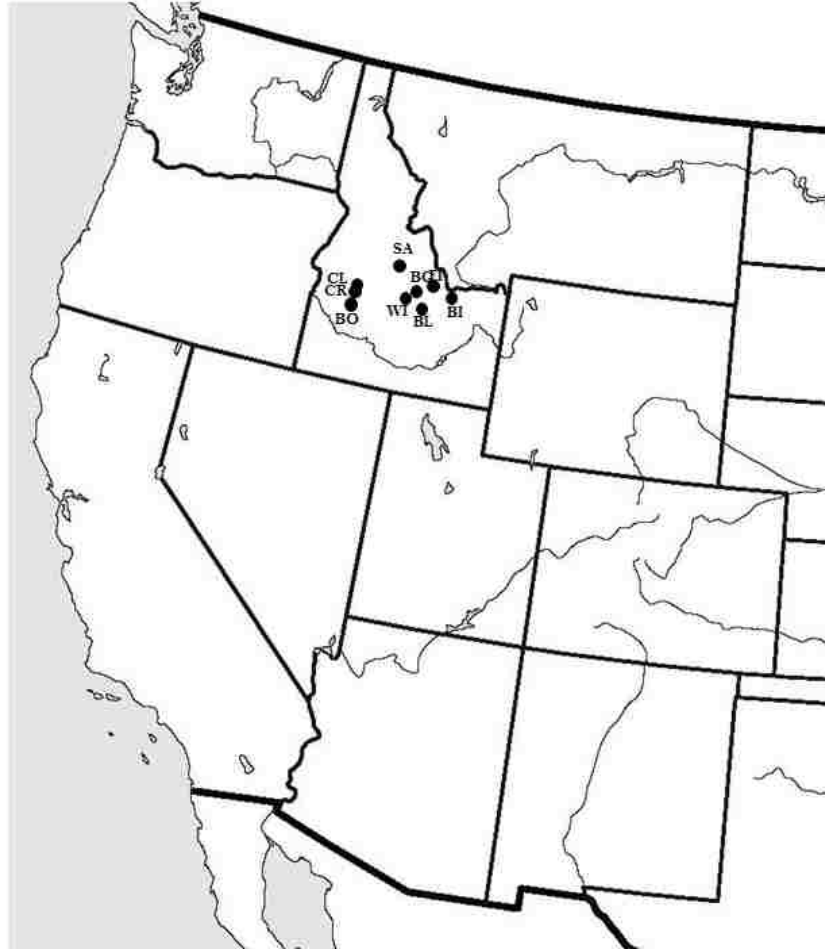


Figure 17: Collection Sites of *C. confusus*. BG= Big Gulch Creek, ID; BI= Birch Creek, ID; BL= Big Lost River, ID; BO= Boise River, ID; CL= Clear Creek, ID; CR= Crooked Creek, ID; EI= 18 mile, ID; SA= Salmon River, ID; WI= Wildhorse Creek, ID;

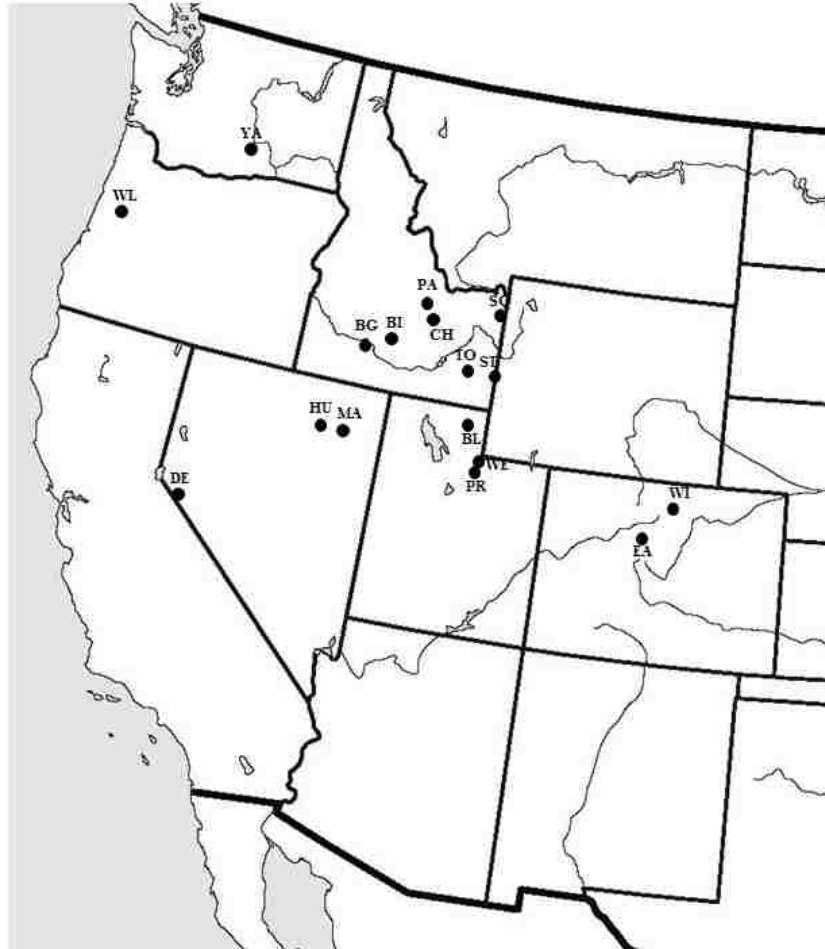


Figure 18: Collection Sites of *C. beldingii* Groups. BG= Billingsley Creek, ID; BI= Big Wood River, ID; BL= Blacksmith Fork, UT; CH= Cherry Creek, ID; DE= Desert Creek, NV; EA= Eagle River, CO; HU= Humboldt River, NV; MA= Mary's River, NV; PA= Pass Creek, ID; PR= Provo River, UT; SQ= Squirrel Creek, ID; ST= Stump Creek, ID; TO= Toponce Creek, ID; WE= Weber, UT; WI= Willow Creek, CO; WL= Willamette, OR; YA= Yakima River, WA