Transactions of the American Fisheries Society 137:1378–1388, 2008

DOI: 10.1577/T07-222.1

[Note]

Differential Gene Expression between Fall- and Spring-Run Chinook Salmon Assessed by Long Serial Analysis of Gene Expression

JEREMIAH C. BERNIER*[1]

Coastal Oregon Marine Experiment Station, Hatfield Marine Science Center, Oregon State University, Newport, Oregon 97365, USA

SHANDA R. BIRKELAND, MICHAEL J. CIPRIANO, AND ANDREW G. MCARTHUR

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

MICHAEL A. BANKS

Coastal Oregon Marine Experiment Station, Hatfield Marine Science Center, Oregon State University, Newport, Oregon 97365, USA

Abstract.—Of all Pacific salmonids, Chinook salmon *Oncorhynchus tshawytscha* display the greatest variability in return times to freshwater. The molecular mechanisms of these differential return times have not been well described. Current methods, such as long serial analysis of gene expression (LongSAGE) and microarrays, allow gene expression to be analyzed for thousands of genes simultaneously. To investigate whether differential gene expression is observed between fall- and spring-run Chinook salmon from California's Central Valley, LongSAGE libraries were constructed. Three libraries containing between 25,512 and 29,372 sequenced tags (21 base pairs/tag) were generated using messenger RNA from the brains of adult Chinook salmon returning in fall and spring and from one ocean-caught Chinook salmon. Tags were annotated to genes using complementary DNA libraries from Atlantic salmon *Salmo salar* and rainbow trout *O. mykiss*. Differentially expressed genes, as estimated by differences in the number of sequence tags, were found in all pairwise comparisons of libraries (freshwater versus saltwater = 40 genes; fall versus spring = 11 genes; and spawning versus nonspawning = 51 genes). The gene for ependymin, an extracellular glycoprotein involved in behavioral plasticity in fish, exhibited the most differential expression among the three groupings. Reverse transcription polymerase chain reaction analysis verified the differential expression of ependymin between the fall- and spring-run samples. These LongSAGE libraries, the first reported for Chinook salmon, provide a window of the transcriptional changes during Chinook salmon return migration to freshwater and spawning and increase the amount of expressed sequence data.

Chinook salmon *Oncorhynchus tshawytscha* are the largest and farthest migrating of all Pacific salmon, and they have the most diverse set of life history strategies in terms of spawning migration timing. The diversity of this extraordinary anadromous salmonid is displayed by numerous populations, or runs, with various return times to freshwater for spawning (Groot and Margolis 1991). In the Sacramento River watershed of California's Central Valley, there are at least four distinct runs that return from the Pacific Ocean to freshwater in winter, spring–summer, fall, and late fall (Banks et al. 2000). We will focus on the two most abundant runs, fall and spring (Waples et al. 2004). Spring-run fish migrate to freshwater during April–May, whereas fall-run fish migrate in September–November. Both runs spawn in the fall, but by migrating earlier in the year during the snowmelt, the spring-run fish are able to access reaches that are not available to the fall run. These seasonal differences in return times can evolve in relatively few generations because of selective pressures from environmental effects (mainly water temperature) on the progeny of fish from different runs (Quinn et al. 2000). The differences are thought to have arisen independently in each watershed due to parallel evolution (Waples et al. 2004).

Although neither run feeds while in freshwater, fish from each run exhibit striking differences in reproductive physiology and behavior. Fall-run fish are already sexually mature when they enter freshwater; they migrate relatively short distances, and spawn in lower stretches of watersheds. In contrast, spring-run fish are not yet sexually mature upon entry into freshwater, and they usually migrate much farther upstream than fall-run fish. Spring-run fish remain in freshwater over the summer months before spawning in the fall. Spring-run fish maintain an ocean-like silver ("bright") scale color and torpedo-like shape until the fall, whereas fall-run fish develop a brown or red skin color, humped back,

* Corresponding author: jeremiah.bernier-latmani@chuv.ch
[1] Present address: Institute of Pathology, Centre Hospitalier Universitaire Vaudois (CHUV), Rue du Bugnon 25, Lausanne 1011, Switzerland.

Received October 25, 2007; accepted February 9, 2008
Published online September 15, 2008

and elongated snout while migrating (Groot and Margolis 1991).

Although there are extensive historical records of differences in return time to freshwater for these different populations (Yoshiyama et al. 1998), little is known about the molecular mechanisms resulting in differing physiology and behavior of the separate runs. Gonadotropin-releasing hormone (GnRH) has been studied extensively for its role in migration timing of chum salmon *O. keta* (Kudo et al. 1996), masu salmon *O. masou* (Bhandari et al. 2003), sockeye salmon *O. nerka* (Amano et al. 1998; Taniyama et al. 2000), and coho salmon *O. kisutch* (Dickey and Swanson 2000). The expression level of GnRH stimulates the secretion of sex steroid hormones, which in turn are differentially expressed according to developmental and environmental stimuli (Ando and Urano 2005).

New technologies allow screens for genes of interest without a priori knowledge of biological mechanisms and may help to explore questions in fish biology. Specifically, techniques such as analysis of DNA microarrays (Schena et al. 1995) and serial analysis of gene expression (SAGE; Velculescu et al. 1995) allow detection and quantification of thousands of gene transcripts simultaneously from a given tissue. For example, a DNA microarray that was designed for Atlantic salmon *Salmo salar* by use of complementary DNA (cDNA) oligonucleotides (Rise et al. 2004) has been used to study differences in brain gene expression between fish displaying distinct reproductive strategies (Aubin-Horth et al. 2005a, 2005b).

Long SAGE (LongSAGE; Saha et al. 2002), a variant of the original SAGE protocol, uses a 21-base-pair (bp) sequence tag isolated from a messenger RNA (mRNA) molecule to uniquely identify the source gene from within the genome. Short sequence tags sampled from all mRNAs containing *Nla* III restriction enzyme recognition sequences are ligated together to form long concatenated molecules that are cloned and sequenced. Quantification of all tags provides a relative measure of gene expression (i.e., mRNA abundance). LongSAGE thus provides both the identity of expressed genes and levels of their expression.

In the present study, we used the sampling power of the LongSAGE technique to examine the expression levels of thousands of genes in the brains of fall- and spring-run returning adult Chinook salmon. By creating three LongSAGE libraries (from fall-run, spring-run, and ocean samples), we were able to investigate whether brain gene expression differs between Chinook salmon returning to freshwater in spring versus fall and which genes are differentially regulated.

Methods

Tissue collection.—All Chinook salmon individuals were collected on the Feather River near Oroville, California, at the Thermolito Afterbay outflow, except for a single fish of unknown origin, which was caught in the Pacific Ocean 16 km west of Eureka, California, in August 2005. All samples were collected by hook-and-line methods. Spring-run samples were collected between May 13 and June 26, 2005. Fall-run samples were collected between September 15 and 17, 2004, and between November 15 and 28, 2005. Fall-run samples were harvested only if they displayed characteristics associated with sexual maturity (dark skin color, pronounced snout, humped back); spring-run samples displayed a bright appearance similar to that of the ocean phase. All fall- and spring-run fish were captured between dawn and 1000 hours to reduce among-sample variability due to light effects and water temperature differences. Fish were killed immediately, and whole brains were dissected from the cranium as quickly as possible. Brain tissue was preserved in RNAlater (Ambion, Austin, Texas) and was stored at −20°C. Brains were ground with tissue grinders and homogenized in Trizol Reagent (Invitrogen, Carlsbad, California) with glass–Teflon homogenizers according to manufacturer's instructions. The RNA samples were quantified and purity was checked using a ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware); presence or absence of RNA degradation was examined by electrophoresis on a 2% agarose gel.

Long serial analysis of gene expression.—Three LongSAGE libraries were constructed using 10 μg of total RNA from brains of one spring-run female, one fall-run female, and one ocean-caught female. The fall-run sample was collected in 2004; the spring-run and ocean samples were collected in 2005. Because pooling of samples increases the number of highly expressed mRNAs binding to the oligo(dT) beads and therefore can exclude weakly expressed genes (Morris et al. 2003), libraries were constructed with individual samples to increase the probability of observing novel gene sequences. LongSAGE library construction and subsequent analysis used only RNA from the brains of female Chinook salmon.

LongSAGE libraries were constructed using the I-SAGE Long Kit protocol (Invitrogen) with the following modifications. LongSAGE concatemers were partially digested with *Nla* III (60°C for 1 min) to decrease the prevalence of concatemer circularization after ditag ligation (Gowda et al. 2004). Additionally, we used pGEM-3Z cloning vector (Promega, Madison, Wisconsin) with blue–white

screening instead of pZErO-1 (Invitrogen). Together, these changes improved LongSAGE tag sequencing efficiency by increasing the number of tags per clone while decreasing the amount of empty vectors. Plasmid preparations were performed on GeneMachines RevPrep Orbit (Genomic Solutions, Ann Arbor, Michigan) and RoboPrep 2500 (MWG Biotech, High Point, North Carolina) platforms, and sequencing was conducted with a 3730xl DNA Sequencer (Applied Biosystems, Inc. [ABI], Foster City, California) using Big Dye Terminator chemistry (ABI).

Sequences were analyzed with software created by authors Cipriano and McArthur for LongSAGE tag extraction and analysis. The SAGE software parses concatemer sequence from vector sequence and extracts and orients LongSAGE tags from concatemers by recognition of the 5′-CATG anchoring site. The tags are recognized by the unique 21-bp sequence and enumerated. The pipeline also excludes tags with sequence ambiguity and putative sequencing error based on sequence quality scores. Putative sequencing error tags were those observed only once among the libraries. Some of these were saved from exclusion by matches to the expressed sequence tag (EST) database. Tag sequences were matched to available EST databases for annotation. Because only 2,000 Chinook salmon singleton ESTs are currently available (The Institute for Genomic Research [TIGR]), we additionally used 88,362 ESTs from the closely related rainbow trout *O. mykiss* (TIGR) and Atlantic salmon (Genomic Research on Atlantic Salmon Project; Davey et al. 2001) for tag-to-gene annotation. The EST contigs (sets of overlapping DNA segments) were annotated using BLASTX software to the National Center for Biotechnology Information (NCBI) nonredundant protein database. Open reading frames (ORFs) were called using Glimmer software (TIGR) on each contig, and each ORF was annotated using BLASTP software (Altschul et al. 1990). LongSAGE tags were assigned to contigs only by perfect 21-bp matches in the correct orientation, and LongSAGE tags with one-nucleotide mismatches were assumed to be real tags only if they mapped to a site where there was an existing LongSAGE tag. These cases were manually reviewed and added to the list of annotated LongSAGE tags if there was significant homology to known genes in other fish species. Those tags with no exact match in the EST database were then compared with all available genomic sequence data by BLASTN (Altschul et al. 1990). In some cases, multiple tags were mapped to the same gene. This phenomenon is possible due to inefficient digestion by *Nla* III, which results in a LongSAGE tag sequence that is not necessarily associated with the most 3′ CATG site. With efficient digestion, *Nla* III cuts the most 3′ CATG site of an mRNA molecule, resulting in a primary sense tag. The LongSAGE tags resulting from inefficient digestion by *Nla* III are designated as alternate sense tags but are annotated to the same gene in the EST reference database. Those tags without a match to an ORF, cDNA contig, or NCBI DNA database were considered unknown.

To assess the likelihood of putative differences in gene expression between libraries, we used the log-likelihood ratio statistic, *R* (Stekel et al. 2000), which scores tags by departures from the null hypothesis of equal counts in each library given the total number of tags sampled from each library (Stekel et al. 2000). Higher *R*-values indicate a greater probability of differential expression, whereas *R*-values near zero represent constitutive expression. An analysis described by Stekel et al. (2000) was conducted to establish a threshold for differential expression and highlight only putatively differentially expressed genes; only those tags with an *R*-value of 4 or greater were used for LongSAGE tag and reverse transcription polymerase chain reaction (RT-PCR) analysis.

Reverse transcription polymerase chain reaction.—Three samples, each from fall- and spring-run fish harvested in 2005, were used for RT-PCR. Separate samples were used for LongSAGE and RT-PCR. Total RNA was extracted as detailed above for LongSAGE and treated with DNase I (Sigma-Aldrich, St. Louis, Missouri; enzyme number 3.1.21.1, IUBMB 1992) to prevent DNA contamination. The RT-PCR analysis was performed using SuperScript III One-Step RT-PCR System (Invitrogen) with Platinum *Taq* DNA polymerase (Invitrogen; 2.7.7.7) according to the manufacturer's instructions. Primers were designed from EST sequence matched to the rainbow trout ependymin gene (NCBI accession number M93697) and the Chinook salmon glyceraldehyde-3-phosphate dehydrogenase (1.2.1.12) gene (*GAPDH**; AB177405). Both forward and reverse gene-specific primers were used at a concentration of 10 μM. For ependymin, the forward primer was 5′-CCCTCCACGAGAAGATGCAG-3′ and the reverse primer was 5′-GGCCCTCAGGAGCCTCCT-3′; for *GAPDH**, the forward primer was 5′-CGCCTACACCGCCACCCA-3′ and the reverse primer was 5′-CGCGGTGGCTGTAGCCAAA-3′. The RT-PCR reactions (total = 25 μL) contained 12.5 μL of 2X Reaction Mix (Invitrogen), 2.5 μL of total RNA (20 ng/μL), 8 μL of H_2O, 1 μL of reverse transcriptase (2.7.7.49)–*Taq* polymerase mix (Invitrogen), 1 μL of forward primer, and 1 μL of reverse primer. We performed cDNA synthesis at 55°C for 30 min. The PCR cycling conditions were (1) 1 cycle at 94°C for 2 min and (2) 30 cycles at 94°C for 15

TABLE 1.—Summary of long serial analysis of gene expression tag counts for fall-run, spring-run, and ocean-caught Chinook salmon ($n = 1$ female/library) originating from the Central Valley, California.

Library	Total number of tags	Total number of unique tag sequences
Ocean	29,372	7,018
Spring run	26,414	7,250
Fall run	25,512	6,671
Total	81,298	

s, 60°C (ependymin) or 63°C (*GAPDH**) for 30 s, and 68°C for 45 s. Aliquots were taken at cycle 12, 18, 24, and 30 and were visualized with ethidium bromide staining on a 2% agarose gel. The RT-PCR of *GAPDH** was performed to control for equal loading of PCR product among samples.

Results

Long Serial Analysis of Gene Expression

After correcting for sequencing error, 81,298 LongSAGE tags with 9,080 unique tag sequences were obtained; 25,512–29,372 LongSAGE tags were obtained from each library, and 6,671–7,250 unique tags/library were detected (Table 1). Of these tag sequences, 2,822 (31%) were observed at least once in all three libraries (data not shown). After automated and manual annotation of LongSAGE tag sequences using salmonid cDNA libraries, 6,544 (72.07%) matched solely to the NCBI DNA database, 1,224 (13.48%) were considered unknown, 606 (6.67%) matched to called ORFs in our reference EST database, 274 (3.01%) matched cDNA contigs with no ORF call, 221 (2.43%) were manually matched to publicly available sequence data, and 211 (2.32%) had multiple matches in the cDNA reference database and were considered unresolved (data not shown). All SAGE data were deposited to GenBank's Gene Expression Omnibus database (accession number GSE6009).

To organize and analyze the LongSAGE tags between libraries, we counted each tag's abundance in the two libraries being compared. By grouping the libraries according to the physical state of the fish or the season of sampling, we were able to analyze and filter the data to make more-valid biological comparisons of LongSAGE counts (Figure 1).

We first compared tag counts in the fall- and spring-run libraries to those of the ocean library. This allowed analysis of genes differentially regulated between freshwater and ocean samples (Figure 1A). We first generated lists of tags that were differentially expressed ($R > 4$) between the fall- or spring-run library (freshwater samples) and the ocean library. Of the 40 tags on this list, 7 were upregulated in the freshwater samples and 28 were downregulated (Table 2). Of the tags that were upregulated in the freshwater samples, three were annotated as the ependymin gene, which is strongly expressed in fish brain tissue (the product is an

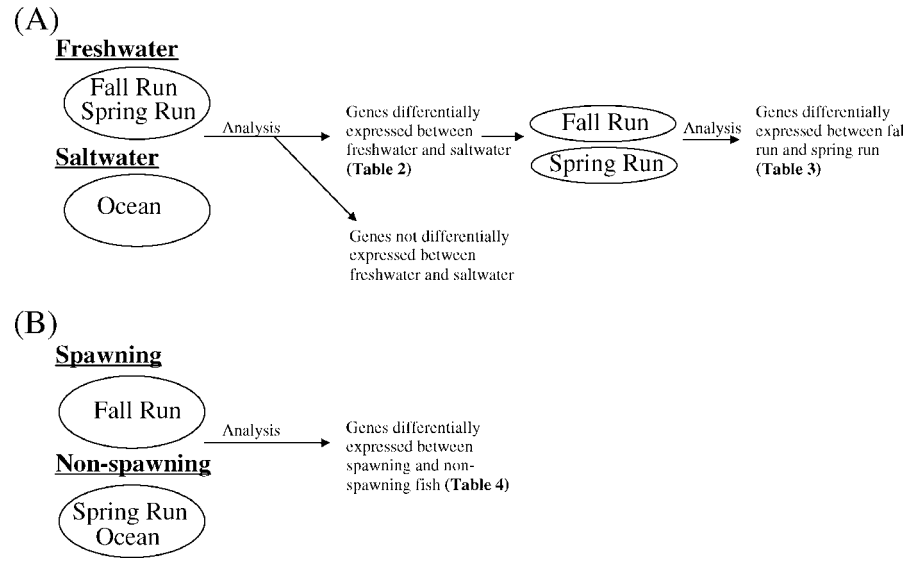

FIGURE 1.—Flow chart of long serial analysis of gene expression in fall- and spring-run Chinook salmon collected from the Feather River, California, and ocean-caught Chinook salmon: **(A)** differential gene expression in freshwater versus ocean-caught fish and spring- versus fall-run fish; and **(B)** differential gene expression in spawning (fall-run) versus nonspawning (spring-run and ocean-caught) fish.

TABLE 2.—Sequences and counts of differentially expressed tags determined by use of long serial analysis of gene expression in freshwater (fall- and spring-run) and ocean-caught samples of Chinook salmon originating from the Central Valley, California. Tags that were upregulated or downregulated in freshwater samples or that had counts in between those of fall-and spring-run libraries are indicated. Annotations include species of origin (those without species are from Chinook salmon, Atlantic salmon, or rainbow trout) and National Center for Biotechnology Information (NCBI) accession numbers (codes for accession numbers appearing more than once: * = alternate sense tag; # = alternate antisense tag; § = tag listed to the given accession number but matched to a unique expressed sequence tag).

Tag annotation	NCBI accession number	Tag sequence	Library		
			Fall run	Spring run	Ocean
Upregulated in freshwater samples					
Ependymin					
Ependymin I	M93697	CATGTTAATGTAATCTCTACG	745	384	18
Ependymin I	M93697*	CATGACAACTTAACATTAACT	70	36	2
Ependymin II	M93698	CATGACAACTTGAAATTAACT	59	33	0
Others					
Possible C-type natriuretic peptide	AB76602	CATGAATAAATGTACACTGAA	25	39	3
Fatty acid binding protein H6-isoform (long-fingered icefish *Cryodraco antarcticus*)	U92443	CATGAAGCTTTCAATAAAGTC	21	18	1
Unknown					
	-	CATGCACTCTGTTATCGGGGC	24	17	0
	-	CATGCAAAGACAGTAGTAGTA	250	70	0
Downregulated in freshwater samples					
Mitochondrial					
Cytochrome-c oxidase[a] subunit I (*COX1**)	NP_148940	CATGTGCTGGTTTCAAGCCAA	175	475	1,181
Cytochrome-c oxidase subunit III (*COX3**)	NP_148944	CATGGCCTACACGTGATTATT	240	342	565
NADH dehydrogenase[b] subunit 1 (*NDI**)	NP_148938	CATGCTTAAGGCACGCCCCAC	200	209	372
Probable mitochondrial gene	AF392054	CATGCAAGTCTCCGCATTCCT	88	114	266
Probable *COX1**	AAK70887	CATGGGGGTTCAATTCCTCCC	15	49	124
*COX3**	NP_148944#	CATGGAATCCGGTGGCGACAA	4	5	33
Na^+/K^+ ATPase[c] alpha subunit isoform 1c	AY319389	CATGTGTGCTTGCACGTTTCC	12	14	55
*COX3**	NP_148944#	CATGTAACGGTGACACCAGAT	4	3	37
Myelin					
Myelin basic protein (zebrafish *Danio rerio*)	AAW52552	CATGTAGTTTTGCAAATTCTC	0	23	71
Myelin basic protein (zebrafish)	AAW52552§	CATGTAATTTTGCAAATTCTC	23	15	70
Myelin proteolipid protein	AAB39006	CATGATCTAGCTGCTTTTGCT	2	10	52
Others					
Alpha tubulin (Chum salmon)	X66973	CATGTTGTGTGTATTCTCAAT	42	36	119
Short-chain dehydrogenase–reductase[d] (zebrafish)	NP_987120	CATGTATTTTGTGATATCATC	26	41	118
Unnamed, hypothetical					
Unnamed protein product (spotted puffer *Tetraodon nigroviridis*)	CAF99308	CATGTGAGGAGGCAGCACTTG	57	89	285
Hypothetical protein LOC553718 (zebrafish)	NP_1018525	CATGAATGATTTCCCAGCAGC	8	5	108
Unknown					
	-	CATGGTTCATTGAAGCCAATA	15	55	129
	-	CATGTTGTGTCAGGGTTCTCA	42	51	116
	-	CATGTGAGGAGGCGGAGCCTG	11	10	58
	-	CATGGGACTTAAACTCCAAAA	4	2	30
	-	CATGTCTTAAAGTGTGCGTGC	3	34	120
	-	CATGCCTAATAAAACAACAAC	6	19	65
	-	CATGCTGTGATATTCCTCCCA	2	2	27
	-	CATGTATAAACTGCTAAGAAT	1	3	35
	-	CATGTTGCCTTATCCAGCACT	0	2	24
	-	CATGATCAAAACAATAAAAAT	1	1	20
	-	CATGGGATAGATTTCGCTTGT	2	0	28
	-	CATGTCTTAATGCGCACGTGT	0	0	49
	-	CATGGAGAGTAGGGAGCCGTT	0	0	16
Medium level in ocean					
Globins					
Beta-globin	CAA65953	CATGGGCAGTCGGTACTTCTA	578	28	234
Beta-globin	CAA65948	CATGGGTAACCCCGCCGTGGC	498	85	176
Alpha-globin I	BAA13533	CATGGACGATCTCTTTGGTTT	246	23	76
Alpha-globin IV	BAA13534	CATGAAAGTCCATCATTGGAC	456	23	104
Alpha-globin IV	BAA13534§	CATGAAAGTCCAAACTTGGAC	387	14	61

[a] Enzyme number 1.9.3.1 (IUBMB 1992).
[b] 1.6.99.3.
[c] 3.6.3.9.
[d] 1.3.99.2.

TABLE 3.—Sequences and counts of genes identified as differentially expressed between fall- and spring-run Chinook salmon originating from the Central Valley, California, based on long serial analysis of gene expression (LongSAGE; this list is corrected for differential expression between freshwater and ocean-caught samples).

Annotation	Tag sequence	Library	
		Fall run	Spring run
Upregulated in spring-run library			
Cytochrome c oxidase subunit I (*COX1**)	CATGTGCTGGTTTCAAGCCAA	175	475
Myelin basic protein	CATGTAGTTTTGCAAATTCTC	0	23
Unknown	CATGTCTTAAAGTGTGCGTGC	3	34
Unknown	CATGGTTCATTGAAGCCAATA	15	55
Upregulated in fall-run library			
Ependymin I	CATGTTAATGTAATCTCTACG	745	384
Beta-globin	CATGGGTAACCCCGCCGTGGC	498	85
Beta-globin	CATGGGCAGTCGGTACTTCTA	578	28
Alpha-globin IV	CATGAAAGTCCAAACTTGGAC	387	14
Alpha-globin I	CATGGACGATCTCTTTGGTTT	246	23
Alpha-globin IV	CATGAAAGTCCATCATTGGAC	456	23
Unknown	CATGCAAAGACAGTAGTAGTA	250	70

extracellular glycoprotein involved in fish behavioral plasticity; for review, see Shashoua [1977, 1991]). Tags annotated as the C-type natriuretic peptide gene or the fatty acid binding protein H6-isoform gene were all upregulated in the freshwater samples. The C-type natriuretic peptide is a regulator of salt concentration in the blood (Tervonen et al. 1998), and fatty acid binding protein H6 is believed to be involved in the uptake of long-chain fatty acids (Stewart 2000). Of the tags that were downregulated in the freshwater samples, eight were annotated to mitochondrial genes, including the cytochrome-c oxidase subunit 1 (*COX1**), cytochrome-c oxidase subunit 3, and NADH dehydrogenase genes (Table 2). Additional genes that were downregulated in the freshwater samples included those for α-tubulin (a major component of microtubules), myelin (part of a sheath that insulates neurons), and short-chain dehydrogenase–reductase (a member of a large family of proteins that reduce many different substrates; Jornvall et al. 1995).

We were also interested in identifying potential genes influencing migration timing between the fall and spring runs on the Feather River. That is, we wanted to investigate whether expression of a particular gene influences a Chinook salmon to migrate in spring rather than fall. We only considered for analysis those genes that were differentially regulated between freshwater and ocean samples (Table 2). We then generated a list of tags that were differentially regulated ($R > 4$) between the fall- and spring-run libraries and cross-checked the list with that given in Table 2 (Figure 1A. Only tags that were present on both lists were designated as differentially regulated after correcting for freshwater effects (Table 3). Tables 2 and 3 both include tags for *COX1** and the ependymin gene, because these tags were the two most differentially regulated among the three libraries. Eleven tags were differentially expressed between the fall- and spring-run libraries. The four tags that were upregulated in the spring-run library included annotations to *COX1** and myelin basic protein, as well as two tags that could not be annotated (Table 3). Of the seven tags that were upregulated in the fall run, only two (annotated to ependymin and an unknown gene) were not annotated to a globin gene (Table 3).

The spawning condition of fall-run samples also allowed analysis of genes that were differentially expressed in fish preparing to spawn. The analysis of LongSAGE tags was conducted as described above with the following modifications. We first generated a list of differentially expressed genes ($R > 4$) between the fall-run and ocean-sample libraries and between the fall- and spring-run libraries. From the two primary lists under comparison, we created a subset consisting of only those tags that matched between primary lists (Figure 1B). This subset contained 51 tags that were differentially expressed in the fall-run library relative to the spring-run or ocean library. Of the tags in this subset, 39 tags were upregulated and 12 tags were downregulated in the spawner sample. Of the upregulated tags, 16 were annotated as ribosomal protein or ribosomal RNA genes; five tags annotated to globin genes were all upregulated in the spawner library, and five tags annotated to egg protein genes (vitellogenin, vitelline envelope protein, and zona radiata) were all found uniquely in the fall-run library (Table 4). Annotation of the six remaining tags that were upregulated in the spawner library included antifreeze

TABLE 4.—Sequences and counts of genes identified as differentially expressed between Chinook salmon spawners (fall-run samples) and nonspawners (spring-run and ocean-caught samples) originating from the Central Valley, California, based on long serial analysis of gene expression (LongSAGE; rRNA = ribosomal RNA). Tags that were upregulated in spawner or nonspawner samples are indicated. Annotations include species of origin (those without species are from Chinook salmon, Atlantic salmon, or rainbow trout) and National Center for Biotechnology Information (NCBI) accession numbers (codes for accession numbers appearing more than once: * = alternate sense tag; # = alternate antisense tag; § = tag listed to the given accession number but matched to a unique expressed sequence tag).

Tag annotation	NCBI accession number	Tag sequence	Library		
			Fall run	Spring run	Ocean
Upregulated in spawner sample					
Ribosomal					
40S ribosomal protein S21 (Formosan landlocked salmon [FLS] *Oncorhynchus masou formosa*)	ABY28376	CATGCAGCTCTGGGATCTCTT	162	59	66
Ribosomal protein L21 (Senegalese sole)	BAF98669	CATGGCATAACTATGCTGACC	108	46	42
Ribosomal protein L21 (Senegalese sole)	BAF98669§	CATGGCATAACTGTGCTGACC	89	36	22
40S ribosomal protein S29 (channel catfish *Ictalurus punctatus*)	AAQ63317	CATGGGATGTTGCATTATTCC	157	56	42
40S ribosomal protein S3 (FLS)	ABY28373	CATGGATGACAGCAACCGGTC	64	16	12
Ubiquitin and ribosomal protein S27a (zebrafish)	NP_956796	CATGGCCAGCCACTTTGATAG	46	12	7
40S ribosomal protein S14 (FLS)	ABY28375	CATGAAGATTGGACGCATCGA	41	7	13
Probable 18S rRNA (bambooleaf wrasse *Pseudolabrus sieboldi*)	AB900393	CATGGTGACCACGGGTAACGG	179	3	13
Putative ribosomal protein L14 (Senegalese sole)	BAF98662	CATGATCTAAAATAAATCCTC	42	11	12
28S rRNA	U4341	CATGTTAGAACAATGTATGTA	20	1	2
18S ribosomal gene (masu salmon)	AY856868	CATGCTAACTAGTTATGCGGC	23	0	2
Ribosomal protein L30 (channel catfish)	AK95157	CATGTTGGCCAAGACTGGTGT	46	11	12
Ribosomal protein L10 (FLS)	ABY28368	CATGGAGGAGGAACAACTAGC	18	0	0
18S rRNA gene	AF308735	CATGCGGGCCAATCTCGGTTG	17	0	1
18S rRNA gene (European seabass *Dicentrarchus labrax*)	AM19038	CATGACCCGCCGGGCAGCGTC	33	0	2
18S rRNA gene (European hake *Merluccius merluccius*)	EU22565	CATGATTAAGAGGGACGGCCG	31	0	1
Globin					
Beta-globin	CAA65948	CATGGGTAACCCCGCCGTGGC	498	85	176
Beta-globin	CAA65953	CATGGGCAGTCGGTACTTCTA	578	28	234
Alpha-globin IV	BAA13534	CATGAAAGTCCAAACTTGGAC	387	14	61
Alpha-globin I	BAA13533	CATGGACGATCTCTTTGGTTT	246	23	76
Alpha-globin IV	BAA13534§	CATGAAAGTCCATCATTGGAC	456	23	104
Egg proteins					
Zona radiata structural protein	AAK97529	CATGTGCAAATAAATCTGACT	163	0	0
Vitellogenin	Q92093	CATGTAGTGCTGTGGTGGAGC	199	0	0
Vitellogenin	Q92093§	CATGTGGCCTGTCGCTGCTCT	75	0	0
Vitelline envelope protein gamma	AAF71260	CATGCCCATTGAAATAAACAA	71	0	0
Vitelline envelope protein alpha	AAF71258	CATGATTGACCCACGTTTTTA	21	0	0
Others					
Ependymin I	M93697	CATGTTAATGTAATCTCTACG	745	384	18
Antifreeze protein type IV (Atlantic cod *Gadus morhua*)	Q56TUO	CATGTCAACAACATTGGTCTT	37	7	0
Melanin-concentrating hormone 2 (chum salmon)	P69155	CATGGTGGGAAGGGTGTACCG	28	0	0
Precerebellin-like protein	AAF4305	CATGACTTAGCAAGAAATGGA	18	0	0
Unnamed protein product (spotted puffer)	CAG4449	CATGGTCTGTGTGAATCTGCT	17	0	0
14-kilodalton apolipoprotein (yellow perch *Perca flavescens*)	ABW6868	CATGTATATAACTATGTGTTT	21	0	0
Unknown					
	-	CATGCAAAGACAGTAGTAGTA	250	70	0
	-	CATGAGAGGTGTAGAATAAGT	53	6	8
	-	CATGAAAATTTAAAATTTAAA	53	3	0
	-	CATGGGCTTATAGGGCGGTGC	35	1	0
	-	CATGACCCGCCGAAACCAAGC	35	1	1
	-	CATGAGTTTCAAATGTCAATG	38	0	0
	-	CATGAATGGATGAACGAGATT	26	0	2
Upregulated in nonspawner samples					
Mitochondrial					
Cytochrome-c oxidase subunit I (*COX1**)	NP_148940	CATGTGCTGGTTTCAAGCCAA	175	475	1181
Cytochrome b *CYTB**)	NP_148950	CATGCCCGTAGAACACCCATT	128	226	307
ATP synthase[a] F0 subunit 6 (*ATP6**)	NP_148943	CATGGGTTTAGCGGTCCCATT	16	100	157
NADH dehydrogenase subunit 2 (*ND2**)	NP_148939	CATGTTAGCTTTGTAGTAAAA	22	77	130

TABLE 4.—Continued.

Tag annotation	NCBI accession number	Tag sequence	Library		
			Fall run	Spring run	Ocean
NADH dehydrogenase subunit 4 (*ND4**) (cherry salmon *O. masou masou*)	YP_961378	CATGGGACTAGTCGCAGGGGG	10	60	118
Others					
Hypothetical protein LOC93203 (zebrafish)	NP_956528	CATGTGTTCTGACATCAAAAA	97	226	291
Beta tubulin 2 (zebrafish)	AAH56533	CATGTCAATAAAATTTCTTTT	43	107	135
Glyceraldehyde-3-phosphate dehydrogenase	BAD16620	CATGTAATATTGGTGTTTAAA	17	62	64
Myelin basic protein (zebrafish)	AAW52552	CATGTAGTTTTGCAAATTCTC	0	23	71
Unknown					
	-	CATGTTCCACAACAAAACAAA	30	79	83
	-	CATGGTTCATTGAAGCCAATA	15	55	129
	-	CATGTCTTAAAGTGTGCGTGC	3	34	120

[a] Enzyme number 3.6.3.14 (IUBMB 1992).

protein, which depresses the serum freezing point in teleost fish (Ewart et al. 1999); melanin concentrating hormone 2, a regulator of pigment change (Baker 1991); precerebellin-like protein, which is of unknown function but may be part of immune response (Bayne and Gerwick 2001); and 14-kilodalton apolipoprotein, a lipid-binding protein. Of the tags that were downregulated in the spawner library, five were annotated to mitochondrial genes (Table 3). The four remaining downregulated tags were annotated to *GAPDH** (an enzyme involved in glycolysis), β-tubulin (a component of microtubules), myelin basic protein, and a hypothetical protein product from zebrafish (Table 4).

Reverse Transcription Polymerase Chain Reaction

A tag annotated to ependymin was one of the most variably expressed tags among all three LongSAGE libraries (Tables 2–4). To verify the relative expression of ependymin between the fall- and spring-run libraries, we conducted RT-PCR on three samples from each run. Unfortunately, lack of additional ocean samples prevented RT-PCR for this group. Ependymin was more strongly expressed in all three fall-run samples than in spring-run samples, whereas the constitutively expressed *GAPDH** showed equal expression in all samples (Figure 2).

Discussion

The molecular mechanisms controlling salmon spawning and migration times are largely unknown. In this work, we have constructed the first LongSAGE libraries for Chinook salmon and compared differences in brain gene expression between fish of differing physiological condition that were sampled during different freshwater migration seasons and at different salinity levels. The LongSAGE libraries yielded novel expressed sequences from Chinook salmon, which may be useful in future studies of salmonid migration, behavior, and physiological changes during the transition from ocean to freshwater. Furthermore, we have demonstrated that LongSAGE can identify genes that are differentially expressed in Chinook salmon, and this differential expression can be subsequently verified by RT-PCR.

That ependymin exhibits stronger expression in the fall run than the spring run is an interesting observation. Ependymin has long been thought to be an effector of long-term memory potentiation in fish (for review, see Shashoua [1977, 1991]). Injection of ependymin antibodies into the brains of trained zebrafish caused significant loss of the retention of a learned behavior (Pradel et al. 1999). In our study, the relative level of ependymin expression in the brains of Chinook salmon was lowest in the ocean sample, intermediate in the spring run, and highest in the fall run (Table 2). All salmon display spawning site fidelity and tend to

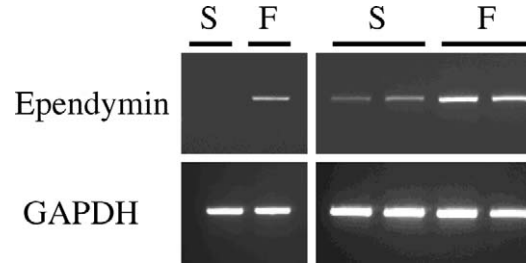

FIGURE 2.—Comparison of reverse transcription (RT) polymerase chain reaction (PCR) analysis results for the ependymin gene of fall- (F) and spring-run (S) Chinook salmon ($n = 3$ samples/run) collected from the Feather River, California. The RNA samples were prepared by the same methods used for long serial analysis of gene expression. After RT and 30 PCR cycles, amplified DNA was run on a 2% agarose gel. For each sample, control reactions were performed using primers for the glyceraldehyde-3-phosphate dehydrogenase gene (*GAPDH**).

spawn in natal streams, and olfactory cues are thought to lead Pacific salmon back to the stream of origin (Quinn 1993). Although fish from both the spring and fall runs were migrating, the fall-run samples were ripe (ready to spawn immediately), while the spring-run samples were not yet sexually mature. If ependymin does play a role in Chinook salmon homing, it is either (1) more strongly expressed in fall run fish or (2) more highly expressed as all Chinook salmon approach spawning. Because ependymin has been shown to be involved in memory formation, it may have a role in the homing and spawning of Chinook salmon.

Our comparison of LongSAGE tag counts between freshwater and ocean samples suggests a change in brain metabolism during the ocean-to-freshwater transition. Salmon feed in the ocean and abruptly stop feeding upon entry into freshwater, and they use stored fats and muscle as energy during the freshwater spawning migration (Bardonnet and Baglinière 2000). It is thought that salmon use glycolysis for energy production in the ocean phase and for the metabolism of fat and protein during migration and spawning. In our comparison of freshwater and ocean samples, mitochondrial gene expression was clearly downregulated in the freshwater samples, whereas the fatty acid binding protein H6-isoform gene was upregulated (Table 2). These results reflect the changes in gene expression necessary for migration from salt water to freshwater.

The primary advantage of using LongSAGE is the discovery of novel and rare transcripts (Chen et al. 2002b; Sun et al. 2004). Microarrays, on the other hand, only quantify transcripts at genes that are physically placed on the array; therefore, microarrays may miss unidentified transcripts that are not printed on the array (Lu et al. 2004). LongSAGE simply captures RNA directly from samples, without construction of cDNA libraries or spotting of microarray chips. Although genome information is useful in annotation, one can build LongSAGE libraries without prior knowledge of the genome of interest (Velculescu et al. 1995). LongSAGE samples most of the transcriptome (polyadenylated RNA with an *Nla* III restriction site), but lack of genomic sequence data can result in many LongSAGE tags of unknown origin (Pleasance et al. 2003). In the present study, 1,224 tag sequences could not be annotated to known genes, and 23 of these sequences were differentially expressed. This finding is unsurprising because, in contrast to model organisms, there is a relative lack of genomic and expressed sequence data available for Chinook salmon.

The sensitivity of LongSAGE can be seen in the effect of small sample size on tag counts. We found that *GAPDH**, which is commonly used as control for equal DNA and protein loading, was differentially expressed between fall- and spring-run fish after LongSAGE. However, RT-PCR of more samples showed equal expression of *GAPDH** between runs, as expected. This highlights the need for downstream verification of high-throughput gene expression analysis, especially when using libraries made from a single individual. We demonstrated with RT-PCR analysis that the expression pattern of ependymin is consistent between LongSAGE and RT-PCR. Although sampling of individual fish with LongSAGE may allow the discovery of rare transcripts, more-downstream verification of differential gene expression is needed. Alternatively, constructing libraries of many pooled samples may reduce random sampling error, but the risk is that only highly transcribed genes may be observed.

Overall, the LongSAGE in our study provides a window on Chinook salmon gene expression changes in response to migration and spawning. These 9,080 novel LongSAGE tags significantly increase the pool of Chinook salmon cDNA that is already sequenced. If the Chinook salmon genome or a broader sampling of ESTs is sequenced, a more-accurate annotation of these LongSAGE tags should be possible. Techniques such as 5′ and 3′ rapid amplification of cDNA ends (Chen et al. 2002a; Hwang et al. 2004) can use LongSAGE tags as primers to amplify full-length cDNA, allowing annotation of LongSAGE tags with interesting expression profiles and exploration of the functional importance of their source genes in Chinook salmon migration and spawning timing.

Acknowledgments

This work was supported with a grant from the California Department of Water Resources awarded to M.A.B.; J.C.B. received additional funding from the North Umpqua Foundation, Roseburg, Oregon. We thank Marc and Harry Johnson, Jason Kindopp, and others at the California Department of Water Resources, Oroville, for assistance in sample collection. Partial support was provided by the Marine Biological Laboratory's Program in Global Infectious Diseases, which is funded by the Ellison Medical Foundation. Computational resources were provided by the Marine Biology Laboratory's Josephine Bay Paul Center for Comparative Molecular Biology and Evolution through funds provided by the W. M. Keck Foundation and the G. Unger Vetlesen Foundation. We would also like to thank Phillip Shaw, Fred Utter, and three anonymous reviewers whose comments greatly improved the manuscript.

References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Amano, M., M. Ashihara, Y. Yoshiura, S. Kitamura, K. Ikuta, and K. Aida. 1998. Two differing salmon GnRH precursor mRNAs are co-expressed in the brain of sockeye salmon. Cell and Tissue Research 292:267–273.

Ando, H., and A. Urano. 2005. Molecular regulation of gonadotropin secretion by gonadotropin-releasing hormone in salmonid fishes. Zoological Science 22:379–389.

Aubin-Horth, N., C. R. Landry, B. H. Letcher, and H. A. Hofmann. 2005a. Alternative life histories shape brain gene expression profiles in males of the same population. Proceedings of the Royal Society Biological Sciences Series B 272:1655–1662.

Aubin-Horth, N., B. H. Letcher, and H. A. Hofmann. 2005b. Interaction of rearing environment and reproductive tactic on gene expression profiles in Atlantic salmon. Journal of Heredity 96:261–278.

Baker, B. I. 1991. Melanin-concentrating hormone—a general vertebrate neuropeptide. International Review of Cytology 126:1–47.

Banks, M. A., V. K. Rashbrook, M. J. Calavetta, C. A. Dean, and D. Hedgecock. 2000. Analysis of microsatellite DNA resolves genetic structure and diversity of Chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Canadian Journal of Fisheries and Aquatic Sciences 57:915–927.

Bardonnet, A., and J.-L. Baglinière. 2000. Freshwater habitat of Atlantic salmon (*Salmo salar*). Canadian Journal of Fisheries and Aquatic Sciences 57:497–506.

Bayne, C. J., and L. Gerwick. 2001. The acute phase response and innate immunity of fish. Developmental and Comparative Immunology 25:725–743.

Bhandari, R. K., S. Taniyama, T. Kitahashi, H. Ando, K. Yamauchi, Y. Zohar, H. Ueda, and A. Urano. 2003. Seasonal changes of responses to gonadotropin-releasing hormone analog in expression of growth hormone/prolactin/somatolactin genes in the pituitary of masu salmon. General and Comparative Endocrinology 130:55–63.

Chen, J. J., S. G. Lee, G. L. Zhou, and S. M. Wang. 2002a. High-throughput GLGI procedure for converting a large number of serial analysis of gene expression tag sequences into 3′ complementary DNAs. Genes Chromosomes and Cancer 33:252–261.

Chen, J. J., M. Sun, S. G. Lee, G. L. Zhou, J. D. Rowley, and S. Wang. 2002b. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. Proceedings of the National Academy of Sciences of the USA 99:12257–12262.

Davey, G. C., N. C. Caplice, S. A. Martin, and R. Powell. 2001. A survey of genes in the Atlantic salmon (*Salmo salar*) as identified by expressed sequence tags. Gene 263:121–130.

Dickey, J. T., and P. Swanson. 2000. Effects of salmon gonadotropin-releasing hormone on follicle stimulating hormone secretion and subunit gene expression in coho salmon (*Oncorhynchus kisutch*). General and Comparative Endocrinology 118:436–449.

Ewart, K. V., Q. Lin, and C. L. Hew. 1999. Structure, function and evolution of antifreeze proteins. Cellular and Molecular Life Sciences 55:271–283.

Gowda, M., C. Jantasuriyarat, R. A. Dean, and G. L. Wang. 2004. Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. Plant Physiology 134:890–897.

Groot, C., and L. Margolis. 1991. Pacific salmon life histories. University of British Columbia Press, Vancouver.

Hwang, B. J., H. M. Muller, and P. W. Sternberg. 2004. Genome annotation by high-throughput 5′ RNA end determination. Proceedings of the National Academy of Sciences of the USA 101:1650–1655.

IUBMB (International Union of Biochemistry and Molecular Biology). 1992. Enzyme nomenclature 1992. Academic Press, San Diego, California.

Jornvall, H., B. Persson, M. Krook, S. Atrian, R. Gonzalez-duarte, J. Jeffery, and D. Ghosh. 1995. Short-chain dehydrogenases reductases (SDR). Biochemistry 34:6003–6013.

Kudo, H., S. Hyodo, H. Ueda, O. Hiroi, K. Aida, A. Urano, and K. Yamauchi. 1996. Cytophysiology of gonadotropin-releasing-hormone neurons in chum salmon (*Oncorhynchus keta*) forebrain before and after upstream migration. Cell and Tissue Research 284:261–267.

Lu, J., A. Lal, B. Merriman, S. Nelson, and G. Riggins. 2004. A comparison of gene expression profiles produced by SAGE, long SAGE, and oligonucleotide chips. Genomics 84:631–636.

Morris, J. S., K. A. Baggerly, and K. R. Coombes. 2003. Bayesian shrinkage estimation of the relative abundance of mRNA transcripts using SAGE. Biometrics 59:476–486.

Pleasance, E. D., M. A. Marra, and S. J. M. Jones. 2003. Assessment of SAGE in transcript identification. Genome Research 13:1203–1215.

Pradel, G., M. Schachner, and R. Schmidt. 1999. Inhibition of memory consolidation by antibodies against cell adhesion molecules after active avoidance conditioning in zebrafish. Journal of Neurobiology 39:197–206.

Quinn, T. P. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fisheries Research 18:29–44.

Quinn, T. P., M. J. Unwin, and M. T. Kinnison. 2000. Evolution of temporal isolation in the wild: genetic divergence in timing of migration and breeding by introduced Chinook salmon populations. Evolution 54:1372–1385.

Rise, M. L., K. R. von Schalburg, G. D. Brown, M. A. Mawer, R. H. Devlin, N. Kuipers, M. Busby, M. Beetz-Sargent, R. Alberto, A. R. Gibbs, P. Hunt, R. Shukin, J. A. Zeznik, C. Nelson, S. R. M. Jones, D. E. Smailus, S. J. M. Jones, J. E. Schein, M. A. Marra, Y. S. N. Butterfield, J. M. Stott, S. H. S. Ng, W. S. Davidson, and B. F. Koop. 2004. Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. Genome Research 14:478–490.

Saha, S., A. B. Sparks, C. Rago, V. Sakmaev, C. J. Wang, B. Vogelstein, K. W. Kinzler, and V. E. Velculescu. 2002. Using the transcriptome to annotate the genome. Nature Biotechnology 20:508–512.

Schena, M., D. Shalon, R. W. Davis, and P. O. Brown. 1995. Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. Science 270(5235):467–470.

Shashoua, V. E. 1977. Brain protein metabolism and the acquisition of new patterns of behavior. Proceedings of the National Academy of Sciences of the USA 74:1743–1744.

Shashoua, V. E. 1991. Ependymin, a brain extracellular glycoprotein, and CNS plasticity. Annals of the New York Academy of Sciences 627:94–114.

Stekel, D. J., Y. Git, and F. Falciani. 2000. The comparison of gene expression from multiple cDNA libraries. Genome Research 10:2055–2061.

Stewart, J. M. 2000. The cytoplasmic fatty-acid-binding proteins: thirty years and counting. Cellular and Molecular Life Sciences 57:1345–1359.

Sun, M., G. L. Zhou, S. Lee, J. J. Chen, R. Z. Shi, and S. M. Wang. 2004. SAGE is far more sensitive than EST for detecting low-abundance transcripts. BMC Genomics 5:1–5.

Taniyama, S., T. Kitahashi, H. Ando, M. Kaeriyama, Y. Zohar, H. Ueda, and A. Urano. 2000. Effects of gonadotropin-releasing hormone analog on expression of genes encoding the growth hormone/prolactin/somatolactin family and a pituitary-specific transcription factor in the pituitaries of prespawning sockeye salmon. General and Comparative Endocrinology 118:418–424.

Tervonen, V., O. Arjamaa, K. Kokkonen, H. Ruskoaho, and O. Vuolteenaho. 1998. A novel cardiac hormone related to A-, B- and C-type natriuretic peptides. Endocrinology 139:4021–4025.

Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. Science 270(5235): 484–487.

Waples, R. S., D. J. Teel, J. M. Myers, and A. R. Marshall. 2004. Life-history divergence in Chinook salmon: historic contingency and parallel evolution. Evolution 58:386–403.

Yoshiyama, R. M., F. W. Fisher, and P. B. Boyle. 1998. Historical abundance and decline of Chinook salmon in the Central Valley region of California. North American Journal of Fisheries Management 18:487–521.