AN ABSTRACT OF THE DISSERTATION OF

M. Renee Bellinger for the degree of Doctor of Philosophy in Fisheries Science presented on December 2, 2014.

Title: Genomics of migration: from marine distributions of salmonids to mechanisms of olfactory and magnetic cue perception for natal homing

Abstract approved: _____

Michael A. Banks

Substantial scientific investment has been directed towards understanding factors that influence distribution patterns and animals' remarkable ability for precise orientation and navigation, yet fundamental gaps in our knowledge remain. In my dissertation, I applied emerging genetic technologies to conduct a top-down and bottom-up investigation of animal movement and cue perception. First, in partnership with Project CROOS and the California Salmon Genetic Stock Identification project, stock-specific, marine migratory distributions of Chinook salmon (*Oncorhynchus tshawytscha*) were characterized for five consecutive months (2010) over 1000 km of coastline. A statistical model was developed to provide measures of relative stock-specific abundance, insights into broad factors that influence migratory distribution, and for fisheries management applications. For the second component of my dissertation, I studied specialized olfactory cells of salmonids that are proposed to contain nanometer-sized magnetite crystals that interact with earth strength magnetic fields to transduce them into neural signals. The transcriptome profiles of candidate

magnetoreceptor and non-magnetic cells isolated from olfactory rosette tissue, whole olfactory rosettes, and blood and muscle tissue were characterized from ~661 million Illumina RNA-seq reads. A total of 1,006 differentially expressed genes were identified in the magnetic cell sample type. Results, consistent with having identified genes involved in magnetite crystal formation in fish, were used to develop a genetic model of magnetic sensory perception. Finally, to provide insights into olfaction-based homing that takes place in freshwater, the olfactory repertoire of salmonids was inferred from the Rainbow trout (*O. mykiss*) genome and compared to that of 15 other teleosts and the jawless fish, sea lamprey (*Petromyzon marinus*), an ancient species. The abundance and diversity of trace amine-associated and V2R-like genes suggests that these classes of chemoreceptors have biological importance. These findings have relevance for resolving if salmon and other marine organisms imprint on magnetic fields, and contribute to our understanding of how magnetic sense, olfaction, and genetic programming are involved in migratory distributions.

Genomics of migration: from marine distributions of salmonids to mechanisms of olfactory and magnetic cue perception for natal homing

by
M. Renee Bellinger

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented December 2, 2014
Commencement June 2015

Doctor of Philosophy dissertation of M. Renee Bellinger presented on December 2, 2014

APPROVED:

Major Professor, representing Fisheries Science

Head of the Department of Fisheries and Wildlife

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

M. Renee Bellinger, Author

ACKNOWLEDGEMENTS

With deep gratitude I thank Michael Banks, my major advisor, for his mentorship and support throughout the years I spent as a member of his laboratory team. He steadfast encouraged creative thought, growth, and facilitated and supported diverse and numerous opportunities. Gil Sylvia and Peter Lawson have been instrumental to my education, challenging me to expand how I think about the world and providing guidance and advice. Special thanks to Dave Jacobson for his seemingly endless knowledge, being the "go-to" for all questions no matter how obscure, and for being a great work colleague. I thank my committee members, Drs. Philip Brownell, Clint Epps, Ken Lohmann, and Todd Mockler, for providing encouragement, perspective, and guidance throughout my PhD. The Department of Fisheries and Wildlife provided outstanding support in the form of scholarships and opportunities. Words fail to capture the depth of impact that being a part of the Banks' lab and the Hatfield Marine Science Center community has had on my life.

I am grateful to the following for providing research funding and support: Mamie Markham Research Award (2009, 2013), Dr. Hari S. & Dr. Renuka R. Sethi Graduate Scholarship, Bill Wick Marine Fisheries Award, Oregon Lottery Scholarship, H. Richard Carlson Scholarship, Thomas G. Scott Grant Scholarship for the Outstanding Ph.D. student in Fisheries and Wildlife, and the Neil Armantrout Graduate Fellowship. The receipt of private scholarship donations is an uplifting and humbling experience. Special thanks to Dr. Renuka Sethi for her kindness and support of students, and for making time to meet students during her visit back to OSU. For travel awards I thank the OSU Graduate School & Department of Fisheries and Wildlife

matching funds, Dr. Kenneth Munson, and the Western Association of Marine Laboratories.

The following individuals graciously provided fish or facilities for experiments: Mary Arkoosh, Stacy Strickland, Joe Dietrich, Greg Hutchinson, Alex Krupkin, National Marine Fisheries Service; Rob Chitwood, Smith Farm Fish Genetics and Performance Laboratory; David Noakes, Joseph O'Neil, and Ryan Couture, Oregon Hatchery Research Center; Trask River Hatchery staff, Oregon Department of Fish and Wildlife.

I thank Kathleen O'Malley, Shawn O'Neil, Nick Sard, Jonathan Minch, Dalton Hance, Melissa Evans, Alana Alexander, and Chante Davis for helpful conversations, programming tips, scripts, and/or review of work products. Mark Dasenko, OSU Center for Genome Research and Biocomputing provided Illumina RNA-Sequencing services. I appreciate contributions from Brett Tyler, Stewart Chair, Botany and Plant Pathology Director, Center for Genome Research and Biocomputing, for my genomics education and guidance during the Keck proposal process.

Many thanks to the Hatfield Marine Science Center community and friends at large for their support during the trials and tribulations of being a graduate student and all the extra surprises that life brought my way. Special thanks to Sana Banks, John and Amy Chapman, Jennifer and Justin Wimpress, Mary Arkoosh, Kristina McCann-Grosvenor and Hovey Grosvenor, Jessica Waddell, Janet Webster, Jeany Volk, and Melody Pfister. I am grateful to the anonymous group that gifted me with yoga credits, that incredibly kind and thoughtful gift was perfect for the circumstances.

Last, but certainly not least, I thank my family and extended family of friends for their inspiration and support. My parents, Richard Bellinger and Joan Graves, instilled a love of knowledge and fostered my education along its circuitous path. Along with my parents, Rose Bellinger O'Neil, Alice Bellinger, Ruth Huwe and Marlys Powell, and the rest of my family, provided invaluable advice and encouragement. Eliana Bellinger-Thomas, thank you for being patient with me during the busy times. I thank Dr. Holbrook for dedicating his life to medicine, and for giving me and Jason the gift of more time together. Jason, you are my rock, none of this grand adventure would have been possible without you.

CONTRIBUTIONS OF AUTHORS

Chapter 2. Drs. Michael A. Banks, Gil Sylvia, and Peter W. Lawson initiated the development of the Oregon-based at-sea Chinook salmon research project and electronic data collection information systems under the umbrella of "Project CROOS, Collaborative Research on Oregon Ocean Salmon". Sarah J. Bates managed the California's project fisheries sampling and data. Eric D. Crandall performed genetic stock identification analysis for the California Chinook salmon samples and created Figure 2.1. John Carlos Garza supervised the development of the California SNP baseline and served as project advisor. All authors contributed to editorial review of the manuscript and provided valuable feedback for manuscript content.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF APPENDIXES

DEDICATION

"You can do it, stay determined"
for my family, especially Eliana and her cousins, our next generation

Genomics of migration: from marine distributions of salmonids to mechanisms of olfactory and magnetic cue perception for natal homing

1 – GENERAL INTRODUCTION

Long distance migration is an important strategy used by organisms to maximize fitness in seasonal environments (Alerstam et al 2003). Substantial scientific investment has been directed towards understanding factors that influence distribution patterns and animals' remarkable ability for precise orientation and navigation, yet fundamental gaps in our knowledge remain. One of the most intriguing questions is which external cues are utilized by animals to identify their geographic position, and how at a molecular level these informational cues are converted from an external to an internal stimulus. Answers to these questions have implications for conservation and management efforts directed towards species at risk for population decline from anthropogenic changes to the environment. To fully appreciate the complexity migratory strategies, research endeavors are limited by the availability of census methods capable of cataloguing population-specific movement patterns. A comprehensive understanding of migration is increasingly urgent because global climate change can cause spatial mismatch in resource availability and animals' ability to adapt to environmental fluctuations is potentially at risk (Both and Visser 2001; Schweiger et al 2008).

The evolution of a migration is driven by complex interaction among geographical, historical, social, and ecological factors. The substantial diversity of migratory strategies is indicated by the wide spectrum of species, population, and individual specific patterns of movements. For example, species or populations breeding in discrete geographic regions may seasonally migrate and converge at a distant destination (Baker et al 1986; Lopes et al 2006), or populations or groups of individuals from a common region may segregate during the non-breeding period and migrate to

1 – GENERAL INTRODUCTION

1.1 OVERVIEW OF MIGRATION

Long distance migration is an important strategy used by organisms to maximize fitness in seasonal environments (Alerstam et al 2003). Substantial scientific investment has been directed towards understanding factors that influence distribution patterns and animals' remarkable ability for precise orientation and navigation, yet fundamental gaps in our knowledge remain. One of the most intriguing questions is which external cues are utilized by animals to identify their geographic position, and how at a molecular level these informational cues are converted from an external to an internal stimulus. Answers to these questions have implications for conservation and management efforts directed towards species at risk for population decline from anthropogenic changes to the environment. To fully appreciate the complexity migratory strategies, research endeavors are limited by the availability of census methods capable of cataloguing population-specific movement patterns. A comprehensive understanding of migration is increasingly urgent because global climate change can cause spatial mismatch in resource availability and animals' ability to adapt to environmental fluctuations is potentially at risk (Both and Visser 2001; Schweiger et al 2008).

The evolution of a migration is driven by complex interaction among geographical, historical, social, and ecological factors. The substantial diversity of migratory strategies is indicated by the wide spectrum of species, population, and individual specific patterns of movements. For example, species or populations breeding in discrete geographic regions may seasonally migrate and converge at a distant destination (Baker et al 1986; Lopes et al 2006), or populations or groups of individuals

from a common region may segregate during the non-breeding period and migrate to geographically discrete locations (Baker et al 1990; Lagerquist et al 2008; Alves et al 2012). Further, geographically proximate but genetically distinct breeding populations may undertake radically different migratory routes (Weitkamp 2010; Delmore et al 2012). Depending on the species and population, Pacific salmonids (genus *Oncorhynchus*) employ one or a mix of these types of movement patterns (Nicholas and Hankin 1988; Weitkamp and Neely 2002; Weitkamp 2010). Each of these migratory strategies carries a fitness consequence owing to energetic costs of travel and potential differences in benefits of selecting one particular habitat over another (Alves et al 2012). The degree to which stock-specific marine distributions respond to short-term prey density and oceanic conditions relative to innate guidance programs developed as an evolutionary adaptation to oceanic productivity patterns, and the interaction among these factors, is the subject of continued research and debate (Kallio-Nyberg et al 1999; Weitkamp 2010; Satterthwaite et al 2013).

Salmon stock-specific migration patterns are genetically encoded and passed to offspring through some mechanism that is at present unknown. A recent meta-analysis of 30 years of coded-wire-tag (CWT, small individually marked tags placed in snouts of juvenile fish) recovery data from Chinook salmon indicates that populations that originate from common freshwater regions exhibit similar distributions, yet these distributions are distinct from those of fish that originate from adjacent freshwater populations (Weitkamp 2010, see also early work by Nicholas and Hankin 1988). In some cases, patterns between neighboring regions are acutely different. One such case is a migratory divide that occurs along the mid Oregon coast. Stocks originating from

Northern Oregon coastal regions are recovered primarily in northern British Columbia and Alaska, in contrast to mid and southern Oregon coastal stocks that tend to remain local or migrate southward. Similar consistency of recovery patterns was also observed for Coho salmon (Weitkamp and Neely 2002). These broad-scale stock-specific patterns inferred through CWT recovery data occurred across multiple years and in spite of highly variable inter-annual ocean conditions. Evidence for the genetic basis of migratory distribution is further supported by experimental research. In one study, two Chinook salmon stocks with differing migratory pathways were hybridized and offspring reared, released and tagged from a single site (Quinn et al 2011). Coded-wire tag recovery data indicated that the migratory route of offspring was intermediate to that of the parental stocks (see also (Kallio-Nyberg et al 2000) for a similar finding in Atlantic salmon). Migratory distribution patterns thus appear to be inherited, but how this trait is transmitted from parent to offspring and which external cues are involved remains a mystery.

Salmon stocks exhibit consistency in space and time distribution patterns, but the exact mechanism(s) utilized for long-distance homing from distant migratory locales are not known (Dittman and Quinn 1996). Coastal migrant Chinook salmon stocks display latitudinal trends in total distance traveled from natal regions (Nicholas and Hankin 1988; Weitkamp 2010) and appear capable of concerted, well-timed returns to rivers. This suggests knowledge of position relative to their river goal (Neave 1964). Little is known about offshore longitudinal patterns, except that some stocks with stream-type life history move offshore rapidly in contrast to ocean-type which appears to migrate along the coastal shelf (Healey 1983; Weitkamp 2010). Environmental variables that change from

north to south and are thus possible components of a guidance system include length of day, altitude, and inclination of the sun's arc (rising by 1° per degree of latitude (Hasler and Schwassman 1960; Braemer 1960), magnetic inclination and intensity (reviewed by (Freake et al 2006), and odors (Hasler and Wisby 1951). Although we do not know for sure exactly which cues salmon use for position determination during long distant migration, mounting evidence suggests that magnetic field elements are likely important navigational aids (Quinn and Brannon 1982; Putman et al 2013; Putman et al 2014). In freshwater, strong evidence supports olfaction as a primary mechanism for guidance back to natal regions (Dittman and Quinn 1996), yet which chemical cues are involved is at present unknown.

Collecting detailed information on population-specific patterns of movements to elucidate factors that influence migration is difficult because in typical cases stock origin cannot be identified through morphological characteristics. For aquatic organisms, visualization of subjects is occluded by water. Thus, methods to census fish populations require overcoming substantial logistic challenges.

To improve our understanding of migration, I applied emerging genetic technologies to conduct a top-down and bottom-up investigation of animal movement and cue perception. In chapter two, stock-specific migratory distributions of Chinook salmon during their marine migration were characterized over 1000 km of coastline. A statistical model developed for fisheries management applications yields informative measures of stock-specific relative abundance and broad insights into factors that influence migratory distribution. The bottom-up approach encompassed characterization of tissue-specific transcriptomes of Chinook salmon to identify (a, Chapter Three) the

genetic basis of magnetite formation in vertebrates and (b, Chapter Four) olfactory gene expression patterns relative to the full olfactory repertoire of salmonids as deduced from the genome of Rainbow trout (Berthelot et al 2014). The olfactory repertoire of salmonids was compared to that of other teleosts, using data drawn from public genetic repositories, and the jawless fish, sea lamprey (*Petromyzon marinus*), using genes predicted from the recently published genome (Smith et al 2013). Chapter five provides conclusions for the study as a whole and recommendations for future studies with emphasis on validation of the magnetite hypothesis of magnetic signals transduction. In sum, this research contributes a novel framework for categorizing space-time movements in fish and provides fundamental insights into the molecular basis of sensory mechanisms that aid navigation.

1.2 REFERENCES

Alerstam T, Hedenstrom A, Akesson S (2003) Long-distance migration: evolution and determinants. Oikos 103:247–260. doi: 10.1034/j.1600-0706.2003.12559.x

Alves JA, Gunnarsson TG, Hayhow DB, et al (2012) Costs, benefits and fitness consequences of different migratory strategies. Ecology 94:11–17. doi: 10.1890/12-0737.1

Baker CS, Herman LM, Perry A, et al (1986) Migratory movement and population structure of humpback whales (Megaptera novaeangliae) in the central and eastern North Pacific. Mar Ecol Prog Ser 31:105–119.

Baker CS, Palumbi SR, Lambertsen RH, et al (1990) Influence of seasonal migration on geographic distribution of mitochondrial DNA haplotypes in humpback whales. Nature 344:238–240. doi: 10.1038/344238a0

Berthelot C, Brunet F, Chalopin D, et al (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun. doi: 10.1038/ncomms4657

Both C, Visser ME (2001) Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. Nature 411:296–298. doi: 10.1038/35077063

Braemer W (1960) A Critical Review of the Sun-Azimuth Hypothesis. Cold Spring Harb Symp Quant Biol 25:413–427. doi: 10.1101/SQB.1960.025.01.044

Delmore KE, Fox JW, Irwin DE (2012) Dramatic intraspecific differences in migratory routes, stopover sites and wintering areas, revealed using light-level geolocators. Proc R Soc B Biol Sci 279:4582–4589. doi: 10.1098/rspb.2012.1229

Dittman AH, Quinn TP (1996) Homing in Pacific salmon: mechanisms and ecological basis. J Exp Biol 199:83–91.

Freake MJ, Muheim R, Phillips JB (2006) Magnetic maps in animals: a theory comes of age? Q Rev Biol 81:327–347.

Hasler AD, Schwassman HO (1960) Sun orientation of fish at different latitudes. Cold Spring Harb Symp Quant Biol 25:429–441.

Hasler AD, Wisby WJ (1951) Discrimination of Stream Odors by Fishes and Its Relation to Parent Stream Behavior. Am Nat 85:223–238. doi: 10.1086/281672

Healey MC (1983) Coastwide distribution and ocean migration patterns of stream- and ocean-type Chinook salmon, Oncorhynchus tshawytscha. Can Field Nat 97:427–433.

Kallio-Nyberg I, Koljonen M-L, Saloniemi I (2000) Effect of maternal and paternal line on spatial and temporal marine distribution in Atlantic salmon. Anim Behav 60:377–384. doi: 10.1006/anbe.2000.1465

Kallio-Nyberg I, Peltonen H, Rita H (1999) Effects of stock-specific and environmental factors on the feeding migration of Atlantic salmon (*Salmo salar*) in the Baltic Sea. Can J Fish Aquat Sci 56:853–861. doi: 10.1139/cjfas-56-5-853

Lagerquist BA, Mate BR, Ortega-Ortiz JG, et al (2008) Migratory movements and surfacing rates of humpback whales (Megaptera novaeangliae) satellite tagged at Socorro Island, Mexico. Mar Mammal Sci 815–830. doi: 10.1111/j.1748-7692.2008.00217.x

Lopes RJ, Marques JC, Wennerberg L (2006) Migratory connectivity and temporal segregation of dunlin (Calidris alpina) in Portugal: evidence from morphology, ringing recoveries and mtDNA. J Ornithol 147:385–394. doi: 10.1007/s10336-005-0048-y

Neave F (1964) Ocean migration of Pacific salmon. J Fish Res Board Can 21:1227–1244.

Nicholas JW, Hankin DG (1988) Chinook salmon populations in Oregon coastal river basins: description of life histories and assessment of recent trends in run strengths. Oregon Department of Fish and Wildlife, Fish Division Information Reports 88-1, Corvallis.

Putman NF, Lohmann KJ, Putman EM, et al (2013) Evidence for Geomagnetic Imprinting as a Homing Mechanism in Pacific Salmon. Curr Biol 23:312–316. doi: 10.1016/j.cub.2012.12.041

Putman NF, Scanlan MM, Billman EJ, et al (2014) An Inherited Magnetic Map Guides Ocean Navigation in Juvenile Pacific Salmon. Curr Biol 24:446–450. doi: 10.1016/j.cub.2014.01.017

Quinn TP, Brannon EL (1982) The use of celestial and magnetic cues by orienting sockeye salmon smolts. J Comp Physiol A 147:547–552. doi: 10.1007/BF00612020

Quinn TP, Chamberlin J, Brannon EL (2011) Experimental evidence of population-specific marine spatial distributions of Chinook salmon, Oncorhynchus tshawytscha. Environ Biol Fishes 92:313–322. doi: 10.1007/s10641-011-9841-z

Satterthwaite WH, Mohr MS, O'Farrell MR, Wells BK (2013) A comparison of temporal patterns in the ocean spatial distribution of California's Central Valley Chinook salmon runs. Can J Fish Aquat Sci 70:574–584. doi: 10.1139/cjfas-2012-0395

Schweiger O, Settele J, Kudrna O, et al (2008) Climate change can cause spatial mismatch of trophically interacting species. Ecology 89:3472–3479. doi: 10.1890/07-1748.1

Smith JJ, Kuraku S, Holt C, et al (2013) Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nat Genet 45:415–421. doi: 10.1038/ng.2568

Weitkamp LA (2010) Marine Distributions of Chinook Salmon from the West Coast of North America Determined by Coded Wire Tag Recoveries. Trans Am Fish Soc 139:147–170. doi: 10.1577/T08-225.1

Weitkamp L, Neely K (2002) Coho salmon (Oncorhynchus kisutch) ocean migration patterns: insight from marine coded-wire tag recoveries. Can J Fish Aquat Sci 59:1100–1115. doi: 10.1139/f02-075

2 - Geo-referenced, abundance calibrated ocean distribution of Chinook salmon (*Oncorhynchus tshawytscha*) stocks across the west coast of North America

M. Renee Bellinger, Michael A. Banks, Sarah J. Bates, Eric D. Crandall, John Carlos Garza, Gil Sylvia, and Peter W. Lawson

In review

2.1 ABSTRACT

Pacific salmon (genus *Oncorhynchus*) forecasting models are used to estimate stock composition, abundance, and distribution for assessment of proposed fisheries impacts, but most models fail to account for variability in survival and the influence of biophysical factors on migratory distributions. In this study, ocean distribution and relative abundance of Chinook salmon (*O. tshawytscha*) stocks encountered in the California Current large marine ecosystem were inferred using at-sea catch per unit effort (CPUE) fisheries and genetic stock identification data. In contrast to stock distributions estimated through coded-wire-tag recoveries (typically limited to hatchery salmon), stock-specific CPUE provides information for both wild and hatchery fish. Furthermore, this metric is independent of other stocks and is easily interpreted over multiple temporal or spatial scales, in contrast to stock composition results. Using empirical data, the stock-specific CPUE and stock composition estimates were compared to identify conditions under which these two measures were maximally different. Samples and data used here were collected at-sea using a combination of retention and non-retention fishery sampling protocols. Understanding the effects of fishery sampling method on catch rates is important when using fishery-dependent data to infer relative stock abundances. A weak effect of fishery sampling method on catch rates was observed in some, but not all analyzed, cases. Novel visualizations of stock-specific ocean distribution patterns facilitate consideration of how highly refined, spatial and genetic information could be incorporated in ocean fisheries management systems and used to investigate

oceanographic and biogeographic factors that influence migratory distributions of fish in the coastal ocean.

2.2 INTRODUCTION

Understanding the abundance of individual fish stocks as well as the timing and direction of their migratory movements is critical for fisheries management to meet the dual objectives of conservation and harvest (McDonald 1981). Because anadromous Pacific salmon (*Oncorhynchus*) generally return to their natal streams to breed most stocks are genetically distinct (Banks et al 2000; Waples et al 2004; Moran et al 2013). Thus, on an annual basis during the fisheries harvest season setting process, fisheries managers forecast for each stock and age-class abundance along with proposed harvest impacts. When forecasted abundance of a particular stock is projected to fall below conservation targets, large time-area closures may be implemented as precautionary measures to protect these "weak" stocks. If actual stock abundance is higher than forecasted abundance, a fishery closure can result in lost economic opportunity associated with foregone harvest of the stocks that could have sustained harvest pressures. Alternatively, if stock forecasts are too high, over-harvest of stocks may occur and result in failure to meet conservation goals. Most Pacific salmon forecasting models fail to account for annual or spatial variability in survival (Burke et al 2013) and the influence of biophysical factors on migratory distributions (Kallio-Nyberg et al 1999; Satterthwaite et al 2013). Moreover, abundance has been highly variable over the last few decades (Heard et al 2007; Shaul et al 2007). Shortfalls of fishery forecast models have motivated new

approaches to fisheries management, with genetic–based stock-identification techniques holding promise to provide fine-scale stock-specific abundance and migratory distribution data to inform fisheries science and management at spatial and temporal scales that surpass contemporary systems.

The Pacific Fishery Management Council (PFMC) develops annual management measures for Chinook salmon (*O. tshawytscha*) ocean fisheries that occur in the southern portion of the California Current large marine ecosystem off the coast of the contiguous United States (US) (Pacific Fishery Management Council 2010). A single-season modeling tool called the "Fishery Regulation Assessment Model" (FRAM) is used by the PFMC to predict cohort-based stock abundance, time and area stock compositions, and to assess impacts to stocks under different fishery harvest scenarios (Pacific Fishery Management Council 2007). This model relies heavily on data from fish (mostly from hatcheries) that have been hypodermically implanted with a coded-wire-tag (CWT) that indicates source stock and cohort year (Jefferts et al 1963; Johnson 2004). The CWT mark/recapture data are "expanded" using sampling and marking rate information to represent all tagged and untagged stocks (unmarked hatchery and wild fish) in the modeled fishery (Johnson 2004). The FRAM model assumes that CWT fish accurately represent the modeled stock and that stock distribution and migration is constant from year to year. However, stock distribution and survival are influenced by marine environmental conditions (Kallio-Nyberg et al 1999; Hobday and Boehlert 2001; Wells et al 2008) that fluctuate both spatially and on inter-annual and intra-annual bases, and survival of CWT-tagged hatchery fish may differ from natural stock counterparts (Peyronnet et al 2008;

Thériault et al 2011). Accuracy of fisheries forecasts would be improved with a better understanding of annual variation in migratory distribution by improving estimates of individual stock proportions for a given time and ocean area.

Both genetic stock identification (GSI) (Grant et al 1980; Milner et al 1985; Hedgecock et al 2001) and CWTs can be used to estimate the proportions of a mixed sample that come from different genetically differentiated populations and to identify individual fish provenance. But, because all salmon carry genetic "tags" that can help identify their stock of origin, GSI estimates are not biased by expansion factors inherent to marking or sampling rates, as are those from CWT recoveries. Tissue samples for GSI can be obtained non-lethally and processed rapidly, permitting pre- or in-season assessments of stock composition. In contrast, CWT recovery requires lethal sampling of fish. Although GSI holds great promise for improving salmon fisheries management (Shaklee et al 1999; Waples et al 2008), incorporation into marine harvest management in mainland US waters has been limited. In Canada (Beacham et al 2004; Beacham et al 2008) and Alaska (Dann et al 2013), stock-specific exploitation targets, guided by in-season GSI sampling or test fisheries, has resulted in greater fishing opportunity and strengthened conservation for stocks of concern.

Chinook and coho salmon (O. *kisutch*) are the two predominant salmon species encountered in salmon fisheries of the California Current, due to their southern spawning distributions (Waples 2001) and use of cool, upwelled water of the coastal shelf as a migratory corridor and feeding ground (Pearcy 1992; Hinke et al 2005). Over the past few decades, harvest of coho salmon off the coast of the U.S.

state of Oregon (OR) has been severely restricted and off the coast of California (CA) completely closed due to conservation concerns (Shaul et al 2007). The Chinook salmon fishery has persisted, but in recent years the failure of some stocks to meet conservation targets resulted in large closures, highlighting the need for development of new techniques to increase accuracy and spatial resolution of fisheries stock forecasts and provide for better control of harvest impacts. Starting in 2006, commercial salmon fishermen, fisheries managers, and scientists in OR, CA, and Washington state developed a partnership to elucidate fine-scale Chinook salmon stock distribution in the California Current ecosystem, utilizing newly developed genetic resources (Seeb et al 2007; Clemento et al 2014), and geo-referenced catch and fishing effort data. This collaborative effort has used a combination of at sea sampling during the course of normal (retention) fisheries and in test (non-retention) fisheries with catch and release protocols in times when harvest is not permitted.

In this study, we present proof of concept and results for a large-scale GSI sampling and analysis program for the year 2010. Patterns of stock diversity, relative distribution and abundance were inferred from a metric we term Stock-Specific Catch Per Unit Effort (SSCPUE), which combines fishing effort, catch locations and GSI results. In contrast to stock composition results, which are reliant on the relative abundance of all stocks in the sample, SSCPUE is easily interpreted over multiple temporal and spatial scales and is mostly independent of other stocks. The correlation between SSCPUE and stock composition results was evaluated to identify conditions under which these two measures are maximally different. Implementation of non-retention and retention fishery sampling protocols within single areas over short time-

periods permitted evaluating the impact of sampling methodology on CPUE. Novel visualization of patterns of stock-specific ocean distribution facilitates the consideration of how highly refined spatial information might be incorporated in ocean salmon fishery management regimes and used to investigate factors that influence migratory distributions of these ecologically, culturally, and economically important fish.

2.3 METHODS

2.3.1. At-sea data collection and sampling

At-sea data collection and biological sampling were conducted by commercial troll salmon fishermen in coastal waters of the California Current large marine ecosystem from Cape Falcon, OR (lat 45.767º North (N)) southward to near the CA Channel Islands (lat 32.53º N) as part of both regular commercial fishing operations and non-retention sampling in times and areas closed to harvest. A stratified sampling plan was implemented with the objective of obtaining samples from 200 legal-sized Chinook salmon (typically three years of age or older) per week (~ 800 per month) from each of the seven fisheries management zones managed by the PFMC (Figure 2. 1): North OR Coast (NO), Central OR Coast (CO), Klamath Zone OR (KO), Klamath Zone CA (KC-n, with no sampling permitted in the KC-s), Fort Bragg (FB), San Francisco (SF), and Monterey (MO). Although both the San Francisco and Monterey zones are managed as a single unit, due to the latitudinal breadth of these regions and the prospect for these geographic features to delineate areas with different biogeographic patterns, the SF area was divided at Point Reyes

into north (SF-n) and south (SF-s) areas for sampling with sample size objectives as above, and the MO area was divided into north (MO-n) and south (MO-s) at Point Sur for analysis purposes. Accordingly, results and data are presented for a total of nine different area strata. In OR, the commercial fishery was open in all areas from May through August, except for the KO during the month of June. In CA, the commercial season was severely restricted due to forecasted low abundance for Central Valley fall run (Pacific Fishery Management Council 2010). The area KC-n was closed for the entire season, and areas SF and MO were only open for eight days, July 1 – 4 and 8 - 11. The area FB was open during those days, plus July 15 – 29 and all of August. All areas were closed from Cape Falcon, OR to the US/Mexico international boundary during September.

Fishermen used individual knowledge, and social cues when available, to catch fish with maximal efficiency. In areas open to commercial fishing, samples were collected from fish retained as part of normal fishing activities. In closed areas, fishermen used a non-retention protocol designed to minimize impacts to fish (adapted from the Natural Resources Consultants (Natural Resources Consultants, Inc. 1998)). The closed-area GSI sampling was regulated by the Pacific Fishery Management Council (Pacific Fishery Management Council 2010) and conducted under a National Marine Fisheries Service Scientific Research Permit, Scientific Collecting Permits issued by the OR Department of Fish and Wildlife and the CA Department of Fish and Wildlife, and a letter from the International Pacific Halibut Commission. Fishing effort was recorded using hand-held Global Positioning System (GPS) units programmed to log vessel latitude (lat), longitude, and time in five-

minute intervals. During retention fisheries, fishermen sampled in whichever area they believed would maximize harvest and there was no limit to the number of commercial fishing vessels allowed per area (both sampling and non-participating fishing vessels). In contrast, closed-area sampling constrained fishermen to designated sampling areas, and allowed only a maximum of five vessels to expend up to 15 vessel-days of effort to collect samples from up to 200 legal-sized Chinook salmon per week-area stratum, with no harvest allowed and all fish released after sampling. Shore-based fleet managers in OR and CA coordinated fishing effort on a daily or weekly basis to actively manage progress towards sampling goals.

Fishermen collected a small tissue clip from the caudal fin (for genetic analysis) and scale samples (for aging) from legal-sized Chinook salmon landed during commercial fisheries or from legal and sub-legal-sized fish encountered during non-retention fisheries. Retention size limits differed slightly between OR and CA (28 and 27 inches total length, respectively). Fish length was recorded for most fish, but those with missing length data and sampled during non-retention fisheries were estimated to be legal or sub-legal sized based on the proportion of legal to total fish sampled in the same month-area stratum. Catch locations and times were electronically logged by manually marking waypoints on the GPS unit when fish were landed on the vessel deck.

To gauge the relevance of study findings with regard to the commercial fishery, the level of commercial fisheries sampling in this project relative to the total 2010 OR and CA commercial fishery was evaluated on the basis of catch, vessel-days of effort, and numbers of participating fishermen. Zero catch days were excluded

from study effort data to match agency practice of recording vessel effort only for trips having at least one fish landed. Commercial data were obtained from the PFMC's Salmon Document Library: Historical Data of Ocean Salmon Fisheries "Blue Book" Appendix A, Ocean Salmon Fishery Effort and Landing, and Appendix D, Economic Data (available from http://www.pcouncil.org/salmon/background/document-library/historical-data-of-ocean-salmon-fisheries/). A simplifying assumption was that all PFMC data reported for ports Astoria and Tillamook were from fisheries conducted north or south of Cape Falcon, respectively. The OR September and October terminal fisheries catch and effort data were excluded from analysis.

2.3.2 Genetic stock identification

2.3.2.1. Oregon - microsatellites

Genomic DNA was extracted from fin-clips using silica-fiber Pall-plates (Ivanova et al 2006) and arrayed into 384 well plates for genotyping. Polymerase chain reaction (PCR) was used to amplify 13 microsatellite loci standardized as part of an international baseline for Chinook salmon v3.0 (Seeb et al 2007; Moran et al 2013). This baseline contains over 30,000 Chinook salmon genotypes from 233 populations distributed from CA to Alaska (Appendix 2.1. Forward primers were fluorescently labeled and PCR products visualized using an Applied Biosystems model 3730xl Genetic Analyzer. GeneMapper software was used to assign standardized allele calls. Fish with identical or nearly identical genotypes (> 90% similarity) were identified using Microsatellite toolkit (Park 2001) and excluded from

analysis. Only fish that provided useful data at 7 or more loci were included in the final genetic data set used for stock composition and SSCPUE estimations.

2.3.2.2. California – single nucleotide polymorphisms

Samples collected off the coast of CA were genotyped using a panel of 96 single nucleotide polymorphism (SNP) markers (Clemento et al 2011) and the associated genetic baseline designed specifically for use in estimating stock composition in PFMC-managed fisheries (Clemento et al 2014). SNP markers are both cheaper and faster to assay than microsatellites and have lower genotyping error and missing data rates. Genomic DNA was extracted from fin clips using DNEasy 96 filter kits on a BioRobot 3000 (QIAGEN Inc.) after digestion in proteinase K, and then a preliminary PCR performed with primers for all 96 SNP loci. The individual locus PCRs were then performed on 96.96 Dynamic Genotyping Arrays (Fluidigm Corporation) and visualized using the EP1 instrumentation (Fluidigm) according to manufacturer's protocols. Genotypes were scored with Fluidigm SNP Genotyping analysis software and identical or nearly identical genotypes were identified and filtered as detailed above.

The SNP baseline database includes 68 populations, with denser sampling of CA Chinook salmon populations than the microsatellite baseline, and contains stock representatives for > 99% of all fish reporting groups found in ocean fisheries off CA and OR (Clemento et al 2011). Chinook and coho salmon are sister species that are occasionally misidentified in the field. Genotypes of coho salmon at these 96 markers were added to the baseline to identify misidentified fish in the field. These

coho salmon, together with fish with genotypes that were missing more than 20 loci or which had individual heterozygosities greater than 0.16 or less than 0.56 were removed from the final dataset to correct for allelic dropout and contamination respectively.

2.3.3. Mixed stock fishery analysis

Genetic identifications were aggregated into 'reporting groups' consistent with Seeb et al. (Seeb et al 2007), with the exception here of placing Central Valley spring run from the Feather River into the Central Valley fall run reporting group, because of known hybridization between these stocks (Hedgecock et al 2001). After mixed stock fishery analysis was performed, higher-level regional groupings of Alaska, British Columbia, Canada (two groups: Vancouver Island / mainland and Fraser River basin) and Puget Sound stocks were used to reduce the total number of reporting groups. At the regional grouping levels used in this study, almost all reporting units are easily resolved with both baselines (Seeb et al 2007; Clemento et al 2014). Known exceptions for the microsatellite baseline are low power to correctly assign fish to Deschutes fall (Seeb et al 2007; Hess et al 2011) and some Columbia River (e.g., Snake River fall, Lower Columbia River spring (Hess et al 2011)) runs.

The accuracy of individual assignment with the microsatellite baseline was empirically tested by comparing GSI results to stock identifications for fish with CWTs recovered during commercial fishery dockside sampling in OR. For that fishery, GSI sampled fish were labeled with barcodes, which enabled cross-referencing. Low confidence assignments (individual posterior probabilities of

assignment < 0.90) were excluded from CWT-GSI stock comparison. Fish from stocks reared or released out-of-basin were excluded from analysis. Using the SNP baseline, a similar comparison to CWT data in California was performed by Clemento et al. (Clemento et al 2014), but on a separate set of fishery samples.

The program gsi_sim, which uses both genotype frequencies and mixture proportions when estimating the origin of individuals (available at http: http://swfsc.noaa.gov/textblock.aspx?Division=FED&ParentMenuId=54&id=12964) (Kalinowsi et al 2007; Anderson et al 2008), was used for mixed stock analysis and individual assignments. A sliding-window approach was used as follows to accurately represent the proportion of each reporting unit in the Bayesian prior. Genotypes were partitioned into weekly strata for each fishery management zone. Each week's data were analyzed in the context of genotype frequencies observed in the weeks immediately before and after the focal week. Individual assignments were then collated into monthly stock proportions. Since GSI techniques generally assign all fish of unknown origin to a baseline stock, a novel maximum likelihood method implemented in gsi_sim and described by Clemento et al. (Clemento et al 2014) was used to evaluate whether fish may actually have originated from a stock/reporting group not represented in the baseline.

2.3.4. Stock richness, distribution and CPUE-based abundance patterns

Stock richness, distribution, and abundance for month-area strata were inferred using SSCPUE. Results for all stocks and strata are graphically presented with bar graphs in a "small multiples" (*sensu* Tufte 1986) format, with each element

combining overall effort and log-transformed SSCPUEs for the month-area stratum. SSCPUEs were calculated for each stratum by multiplying stock composition estimates from mixture analysis by observed CPUE, defined as the sum of legal-sized fish encounters (sampled and unsampled fish) divided by the sum of days fished within the area during that month. Thus, SSCPUE for stock i in stratum j would be calculated as:

$SSCPUE_{\text{stock i, stratum j}}$ = stock composition value $_{\text{stock i, stratum j}}$ x (n legal-sized fish encounters / n vessel-days effort) $_{\text{stratum j}}$. The resultant values represent the number (usually a fraction) of fish from each stock that fishermen would, on average, encounter per vessel-day of fishing effort in a given stratum. This method accounts for unsampled fish and those that did not meet genotyping or GSI assignment criteria. If a vessel crossed over an area boundary during a single day, the effort was assigned in proportion to time spent in each area, as calculated from GPS tracklogs, and catch was allocated to the area where the fish was landed. Reporting groups that contributed to three or fewer strata were excluded to minimize the numbers of stocks in figures. Confidence intervals for individual SSCPUE values were not included, but overall sampling error can be inferred from CPUE modeling results (see next section). Simplifying assumptions were that GSI stock composition estimates were accurate, CPUE was unaffected by fishery sampling technique, and that CPUE was proportional to abundance. General patterns in SSCPUE results were further explored in a month by area format using filled contour plots of log-CPUE generated in SigmaPlot v11. Breaks in sampling coverage were not incorporated into the figure by masking, because imperfect sampling coverage results in numerous breaks,

depending on the time-area scale, and choosing which sections to mask is subjective. The sampling restriction in the KC-s area resulted in a definite break in data coverage, but CPUEs were likely intermediate to measurements made in adjacent sampling areas. The time-frame "month" was selected for SSCPUE analyses because it is the interval used in PFMC fisheries management.

2.3.5. Catch per unit effort and statistical modeling

Generalized Linear Models (GLMs) were used to assess time-area variability in CPUE with associated error and how CPUE estimates were effected by the exogenous variables sample time-period (at week or month intervals), sample area strata, sampling technique (retention or non-retention), and fishermen effect (a measure of individual fisherman power). The confidence intervals were assessed to determine the effects of sample effort and catch patterns on CPUE error, which aids in the evaluation of study design. The modeled CPUE results were compared to observed CPUEs calculated as described in the SSCPUE section. To identify the most appropriate GLM, histograms of catch per vessel-day and modeled Poisson and negative binomial distributions were visually assessed for fit, followed by running a model without terms (by setting $\beta = 1$) and using a likelihood ratio test to evaluate overdispersion. Finally, the adequacy of Poisson and log-linear negative binomial models was statistically assessed using a Vuong non-nested hypothesis test (Vuong 1989), selecting the better supported model type for subsequent GLM analyses. The fit of models to the data was assessed by calculating delta Akaike Information Criteria (ΔAIC) between a model with no terms and alternative models with terms.

An ANOVA was used to test terms for significant effects, using a cutoff of $p < 0.05$ for significance. The strength of each term's effect was evaluated by the amount that residual deviance decreased from the term's inclusion. Consistent with SSPCPUE analyses, CPUE was defined as legal-sized catch per unit vessel effort, but only vessel-days (and catch) having GPS track log records and at least 85% of fishing effort expended in a single sampling area during a single day were included in CPUE modeling. The CA July retention and non-retention data were combined for the month-area model. Week-area combinations having zero catch for all sample days (n = 14 days representing 9 of 168 week-area combinations) were excluded because that pattern of data results in null values in the maximum likelihood estimator (quasi-complete separation problem (Allison 2008)). For similar reasons, the terms time and area could not be modeled with the term fishermen effect. Analyses were performed in R version 2.15.2 with packages foreign (v 0.8-54), mass (v 7.3-23), car (v 2.0-18), lattice (0.20-23) and pscl (1.04.4).

Individual fisherman skill and vessel efficiency is expected to vary across the fishing fleet, but measuring these effects on CPUE is confounded by inherent limitations to the study design. Most fishermen sampled for short periods of time and on intermittent bases, and typically fished out of a single port (mostly remaining in a single sampling area) for the season. Overall variability in spatial and temporal abundance of fish confounds attributing individual fishermen CPUE to their personal fishing power. In spite of these limitations, the effect of individual fisherman performance on CPUE was evaluated using a GLM with the independent variable

"fisherman effect". Results provide a rough idea of between-fisherman catch success, regardless of the cause, and findings could help guide future study design.

2.3.6. Evaluation of fisheries sampling methodology on fish catchability

Non-retention fishery sampling, combined with GSI, is a potential tool for pre-season evaluation of stock composition and abundance forecasts. However, accounting for fleet dynamics that affect catchability of fish is essential when using fishery-dependent CPUE data to draw inferences about fish population parameters (Maunder et al 2006). Catch rates in retention fisheries may be higher than in non-retention fisheries because incentives to catch fish are higher for boats that can keep and sell their catch. Moreover, open fisheries typically have a larger number of fishermen on the water searching for fish aggregations which increases the probability of success in finding them, and fish school location information is shared through fishermen cooperation. The potential impact of sampling methodology on fish catchability was investigated using two approaches. First, GLMs were used to statistically test for differences in estimated mean CPUE between retention and non-retention fisheries consecutively sampled in single areas. Five separate GLMs were run, each using data from single areas with both types of sampling. The area data were not pooled because spatial differences in fish abundance coupled with unequal sample effort across area strata would likely confound interpretation of CPUE results. In four of the analyses, retention fisheries data collected July 1 – 4 and 8 – 11 were compared to non-retention fisheries data collected July 13 – 28 (areas SF-n, SF-s, MO-n, MO-s). Ideally, sampling effort would be equal during the two sampling periods and

underlying fish density would be constant. However, sampling effort was lower in non-retention fisheries, but extending the time period to include comparable levels of effort data increases the chance that local changes in underlying fish abundance would be unequal between the two time-periods (an untestable assumption with the current dataset). The non-retention sampling period length we selected was thus a compromise between having sufficient data for statistical analysis (which is somewhat subjective) and time-period equalization. The fifth model was run with data collected during June non-retention and July retention fisheries from area FB. Statistical significance was evaluated using an uncorrected p-value of 0.05. Data from OR were not considered because, with the exception of the non-retention September fishery, only the KO had both types of sampling, and data collection in that area was minimal early in the season. Furthermore, previous retention sampling off OR showed a decrease in CPUE from August to September, which suggests that the fall time-period coincides with fish movement out of the area (Bellinger and Banks 2007). Second, we used a Chi-square test to evaluate whether the two fisheries differed in the proportion of "successful" (at least one fish caught) versus "unsuccessful" (zero fish caught) days. This analysis was performed using the same data inputs as described above but, because area differences in CPUE are minimized by grouping days with catch into binary categories, the area data were pooled into a single retention versus non-retention data set.

2.3.7. Comparisons between stock composition and SSCPUE measures

Discord in the relationship between stock composition and corresponding values of SSCPUE is expected to occur because, for a target stock with constant

abundance in a given area, changes in the distribution and abundance of other stocks over time affect stock composition but not SSCPUE results. Here, strengths of associations between SSCPUE and stock composition values were examined by calculating the non-parametric test statistic Kendall's tau rank correlation coefficient (τ), which is used to evaluate the similarity of the orderings of the data ranked by each of the quantities and tests the data set against the null hypothesis of $\tau = 0$ with a two-sided p-value of ≤ 0.05. Correlations between stock composition and SSCPUE values were first assessed using all pairs of non-zero data ("full data set") from retention and non-retention month-area strata. Then, correlations were re-evaluated after considering only data points above a range of threshold percent stock composition values ("threshold data set"), iterating to find the interval at which Kendall's τ correlations were reduced to non-significant levels. This threshold data set was created to account for the presence of rare stocks which inflate τ values because, for those stocks, SSCPUE and stock composition measures will always be ranked low relative to the full range of available values. Moreover, rare stocks exert little influence on catch composition values of other relatively more abundant stocks encountered in mixed stock fishery samples. Data were ranked by stock composition because this is the value most widely reported in the literature. These analyses were also performed on individual stock-fishery bases for the five most frequently encountered stocks across all strata to permit exploration of patterns over a variety of stock richness and abundance conditions. Retention and non-retention fishery data sets were analyzed separately because of potential differences in catch efficiency. Scatterplots with linear trends were created to visually assess the data and aid

interpretation of results. The statistical package Wessa (Wessa 2012) was used for these analyses.

2.4 RESULTS

2.4.1. At-sea data collection, sampling, and mixed stock fishery analysis

Fisheries data were collected from 38 of the 40 pre-defined month-area sampling strata (Tables 2.1, 2.2), with only KO/May and KC-n/May lacking data, but samples and effort were unevenly distributed over space (Figure 2.1) and time. The total number of vessel-days of fishing effort was 2,651. Fisheries sampling effort in CA was 73% non-retention (1,079 of 1,477 total vessel-days), while in OR only 10% (119 of 1,174 total vessel-days). The number of vessel-days of effort expended in each month-area stratum varied widely, ranging from 7 to 205 (mean = 62, median = 55). Over the entire data set, the number of non-retention (1,198, 45%) and retention (1,453, 55%) sampling days was similar. The number of vessels used for sampling in CA and OR was also similar, 88 and 89, respectively.

A total of 9,584 Chinook salmon captures were recorded by participating fishermen (Table 2.2), and biological samples were obtained from 9,554 of these fish. Eight samples were collected by fishermen participating in a similar project; these fish were included in the fish encounter and GSI data sets but excluded from CPUE analysis, because compatible effort data were not available. The number of legal-sized fish encounters per month-area stratum varied widely from 2 to 1,102 (mean = 207, median = 91; includes 28 fish with missing length data estimated as legal-sized). The sampling goal of 200 legal-sized fish per week was rarely achieved due to overall

low catch rates (Table 2.3), non-retention permit constraints, and because sufficient numbers of sampling vessels were not always available. Larger sample sizes were generally obtained in areas (NO, CO, and FB) having sizeable fleets and open to fishing at least part of the season. After removal of sub-legal sized fish, fish that failed genetic amplification, duplicate genotypes, and some sampled fish that were a different species (mostly coho salmon), 8,240 individual assignments for legal-sized fish (n = 3, 866 in OR and n = 4374 in CA) were available for stock composition estimates (supplementary Appendix 2.2) and used for SSCPUE calculations.

Catch and fishing effort for this project provided good coverage relative to the commercial fishery (Table 2.4). At-sea catch locations represented 21.4% of the total commercial harvest and vessel-days effort were 20.6% of total commercial fishing effort conducted May – August, 2010. The project CPUEs calculated with inclusion and exclusion of zero-catch vessel-days effort (6.11 and 7.98 fish / day, respectively) bracketed that of the commercial fishery (7.69 fish / day). Approximately 24.1% of the total commercial fleet that made landings in 2010 participated in project sampling.

Genetic stock assignments were mostly concordant with stock of origin as identified by recovery of CWTs in OR (51 total, supplementary Appendix 2.3), although the small sample size of physically tagged fish limits strength of inference from this data set. Correct assignment to region of origin was made for 35 of the 38 fish (92%) that met the posterior probability criteria of ≥ 90%. Eight reporting regions were represented in the 38 fish sample, with 100% correct allocation to six (Central Valley fall, Lower Columbia fall, Mid Columbia Tule, Rogue, Snake fall, and Upper Columbia summer/fall runs) of those eight regions. For the other two

regions, two Snake River fall hatchery and one N. OR Coast fish mis-assigned. No tagged fish were available for comparison in CA, but the concordance rate between assignments with the SNP baseline and CWT recoveries was 98.95% for a sample of over 1,000 fish landed in CA fisheries in 2010 (Clemento et al 2014).

2.4.2. Stock richness, distribution, and CPUE-based abundance patterns

The SSCPUE for 22 stock groups encountered in the study area indicated consistent patterns in stock richness and diverse patterns in CPUE-based abundance across strata (Figure 2.2). Stock richness was highest in the northernmost two areas, reflecting overlap in distribution of stocks originating from northern (e.g., British Columbia, Puget Sound, Columbia, and Snake) and southern (e.g., Coastal Oregon and California) regions. The southernmost sampling areas, MO-n and MO-s, had lowest stock richness levels, and most of the detected stocks originated from CA Rivers. The GSI sample sizes in the MO were small, which was in part due to low effort, but likely also low underlying abundance. The small sample size reduces the probability of detecting stocks that contribute at low rates to the fishery. In contrast, despite the low level of sampling effort in the KO area during July, stock richness was moderate. The Central Valley fall run stock was widely distributed, peaking in areas FB and KC-n, and was the only stock present across nearly all sampled strata. Throughout nearly the entire study region, this stock's SSCPUE was approximately equal to or greater than that of all other individual stocks. Stocks originating from near the OR-CA border (e.g., Rogue, Klamath and CA Coastal), had highest SSCPUE in areas proximal to their natal river mouths, although individual stock patterns were

variable across months and areas. The Central Valley winter run was encountered only in southern sampling areas, with a peak in SSCPUE during September. In contrast, Columbia River stocks were distributed primarily to the north, and their SSCPUEs decreased towards the end of the sampling season. Greater fishing effort occurred in the north where fisheries were open, and were lower in the south where non-retention sampling predominated. In CA, sampling effort trended higher during the open or partially-open time-area strata (FB and southward/July; FB/August). At the time of manuscript submission, maps of stock-specific distribution and catch per unit effort data are publicly available from https://fp.pacificfishtrax.org/portal/.

Spatial and temporal changes in overall CPUE and individual differences in stock distribution and CPUE-based abundance are easily visualized from log CPUE contour plots (Figure 2.3). Comparisons between the all-stock panel to individual stock panels helps to identify when and where these individual stocks contributed to the total CPUE. For example, the highest observed CPUE was in FB/May, and both Klamath and Central Valley fall run stocks contributed to that peak. In the northernmost areas, the Columbia River stocks contributed to the seasonal change in overall CPUE, along with other stocks not shown in the figure. The Central Valley fall run was widely distributed, and its abundance exceeded that of the other individual stocks. The cool (low CPUE) contours between KO and KC-n /May does not reflect stock distribution, as the contour plot smoothing algorithm implemented in SigmaPlot v11 fills in missing data.

2.4.3. Catch per unit effort and statistical modeling

2.4.3.1. Data description and model selection

The majority of vessel-days effort (n = 2,580 vessel-days, ~97% days fished) met the criteria of sufficient GPS tracks and ≥ 85% time spent in a single sampling area for inclusion in the CPUE modeling data set. Data were overdispersed (likelihood ratio test Chi-square test statistic for overdispersion = 7400.98, p-value = < 2.2 e -16), and the negative binomial model was better supported than the Poisson model (Vuong non-nested hypothesis test statistic -18.49, p-value = 1.2 e - 76). Therefore, log-linear negative binomial models were used to estimate mean CPUE and to assess the effects of exogenous variables on fish catch.

2.4.3.2. Time-area variability in CPUE

The CPUE of legal-sized Chinook salmon was highly variable across time-area strata, indicating complex patterns of fish movement. On a month-area basis, observed mean CPUE ranged widely, from a low of 0.24 to a high of 10.11 (Table 2.3). This variability was consistent with mean CPUE from the month x area negative binomial model (Figure 2.4; ΔAIC -475, residual deviance = 2741 on 2537 df) and ANOVA analysis indicated that time and area terms, and their interaction term, had significant effects on the model (residual deviance decreases: month = 20.3, p-value = 0.0005; area = 279, p-value < 2.2 e -16; interaction term = 333, p-value < 2.2 e -16). The week x area model (ΔAIC = -637, residual deviance = 2719.4 on 2415 df) was a better fit to the data than the month x area model, and all terms again had significant effects (residual deviance decreases: week = 128, area = 302, interaction term = 682; all p-values < 2.2 e -16). In both models, the interaction

between time and area terms explained more variance than either of the explanatory variables individually, and area had a stronger effect than time. The time and area interaction term can be conceptualized as the movement of a body of fish throughout time and space. The month-area CPUE results are convenient for considering fisheries management applications and aid in interpreting complex SSCPUE results across strata, we therefore focus on results at a monthly-time scale. Trends in CPUE gradually swung up or down over time within areas but, in contrast, individual months lacked persistent temporal patterns. In the northernmost two strata, mean CPUEs were higher early in the season and then trended downward (NO) or stabilized (CO), while in the KO and KC-n, mean CPUE started low and then generally increased as the season progressed (Figure 2.4). Catch was unusually high in FB in May, and mean CPUE fluctuated at moderately high levels (relative to other areas) over the rest of the season. The area MO-s had lowest overall CPUE of any strata. Comparisons between SSCPUE results (Figure 2.4) to modeled mean CPUE enables the identification of stocks that contribute to increases or decreases in catch. The shorter week x area model results (Supplementary Figure 2.1) reveal additional insight into mean CPUE data patterns. The week-area CPUE results mostly correspond to monthly patterns within areas but, at a weekly time-scale, moderate fluctuations and occasional abrupt changes in CPUE are more apparent. Although the week x area model was technically a better fit to the data, the broader month time-scale provides larger, more representative sample sizes (both for fisheries sampling and GSI) and balances the effect of outlier weeks on CPUE results.

2.4.4. Effect of fishery sampling method on fish catchability

An overall effect of fisheries sampling methodology on catch rates was not strongly supported by statistical analyses (Table 2.5, Figure 2.5). For four of the five analyses, estimated mean CPUE did not differ statistically between retention and non-retention fisheries. However, in the SF-n area, estimated mean CPUE was statistically higher in the retention fishery. Non-significant trends within areas were variable: retention fishery CPUE was higher in area FB, lower in area MO-n, and similar in SF-s and MO-s. There was not a significant difference in proportions of observed and expected successful (catch > 0) and unsuccessful days (catch = 0) between fisheries (Chi-square test statistic 3.18, df = 1, p-value = 0.0744). However, since this p-value was marginally low, the data were reanalyzed after partitioning into two data sets that minimized temporal mismatch: 1) MO-n, MO-s, and SF-n, SF-s/July, and 2) FB/June, July. For the FB area, a greater proportion than expected of observed days with zero-catch was identified for the non-retention fishery (Chi-square = 8.68, df = 1, p-value = 0.0032). Results for the MO and SF areas were consistent with the previous finding, with no evidence of a difference between fishery sampling methodology on catch success (Chi-square = 0.0098, df = 1, p-value = 0.9211). For the FB area, fishery technique had some effect on catch efficiency, with a greater proportion of unsuccessful days in the non-retention fishery, but there was not a significant difference in mean CPUE. Due to weak evidence for an overall effect of fisheries technique on catch efficiency, it was not included in statistical modeling of CPUE and SSCPUE, or visual representations of CPUE-based abundance.

Individual fisherman CPUE was variable, but this result is confounded by sampling under conditions of non-uniform fish abundance across strata, as inferred by general trends in CPUE. The model with fishermen effect as the only independent variable (ΔAIC = -648, residual deviance = 2741 on 2401 degrees of freedom) was a slightly better fit to the data than the week x area model. An ANOVA of the model indicated that the term fishermen effect had a significant effect ($p < 2.2e-16$). In most cases (n = 131 of 172), estimated CPUE was not significantly different between fishermen, although a small number had higher (n = 19, uncorrected p-values ranging from 0.0004 to 0.0488) or lower catch rates (n = 22, uncorrected p-values 0.0000 to 0.0400), relative to the arbitrarily set reference that was required by the model. Although individual fisherman power may result in higher or lower vessel-day catch on some days, its overall effect on mean CPUE estimates for each strata is unlikely to be large because of the many fishermen that participated in this study.

2.4.5. Correlations between stock composition data and SSCPUE

While stock composition values are likely to be poor representatives of relative stock abundance estimates, the degree of discord between these two measures was not previously described or statistically evaluated using empirical data. The stock composition and SSCPUE measures were significantly correlated for retention and non-retention data sets evaluated over their full range of values, but when data were limited to pairs of values above a range of threshold percent stock contributions, the Kendall's τ values decreased and p-values increased to non-significant levels (Table 2.6). The threshold point at which this occurred was fairly low, with

minimum stock composition values of 13.2% or 22.3% (non-retention and retention fishery, respectively). Concordant with these results, scatterplots show that spread between data points was greater as each measure increased (Figure 2.6a). Results for the five most frequently encountered stocks - CA Central Valley fall run, Rogue, Klamath, CA Coastal, and Northern CA / Southern Oregon Coastal (Figure 2.6b – f) - were consistent with the overall fishery data sets (Table 2.6). Again, scatterplots show data points clustered at low values and scatter was greater as values increased. These multiple stock-fishery threshold points were provided as examples of variability in results found for data collected under heterogeneous stock abundance and richness conditions, and do not imply that a specific stock can be linked to a particular threshold value. A primary driver of scatter in the non-retention data set was a combination of stock richness and abundances across multiple time-area fisheries, resulting in extreme ranges in stock composition values. This was revealed by visually comparing the full non-retention data set (Figure 2.6a) to the Central Valley fall run, non-retention fishery data set (Figure 2.6b), which shows that scatter in the full non-retention data set can mostly be attributed to Central Valley fall run contributions. The proportion of this stock ranged above 90% in some of the southern mixed stock fishery samples (e.g., SF-s/August, September; MO-n/July), and were below 15% in other strata (e.g., MO-s/May and June; KO/July, September; Figure 2.2 and Supplementary Appendix 2.2). Similarly, Central Valley fall run SSCPUE values ranged nearly 10-fold across strata (Figure 2.2, Supplementary Appendix 2.4). Consequently, this wide-ranging, relatively abundant stock was the only one for which Kendall's τ correlation analysis failed to show support for an

association between SSCPUE and stock composition over the entire range of non-retention fishery values. Results for the southern non-retention samples, which happened to be comprised predominantly of Central Valley fall run, drove the slope of the linear trend line in the full data set. For other stocks, slopes of linear trend lines for the two sampling methodologies were nearly equal. Thus, the non-retention fishery scatterplot trend line was lower than the retention trend line due to fishery mixture results for samples collected in areas with wide ranges of stock richness (and variable CPUE), and was not specifically related to retention versus non-retention sampling methodology. Overall, findings support the assertion that stock distribution and abundance are better represented by SSCPUE measures because of independence from the presence of other stocks in mixed stock fishery samples. In contrast, wide fluctuations of dominant stocks (CA Central Valley fall) have overarching influence on stock composition estimates across multiple time-area strata.

2.5 DISCUSSION

We provide the first comprehensive assessment of fine-scale ocean distribution patterns among genetically distinct Chinook salmon stocks through large-scale, simultaneous sampling conducted throughout the southern California Current large marine ecosystem. Such a perspective enables unique insight into dynamic spatial and temporal shifts of relative abundance, as indexed by CPUE, of multiple stocks at a scale that approaches the size of typical migratory patterns for southerly orienting Chinook stocks. This individual-based stock, region and time specific approach makes significantly more dense and focused information available for the

study of Chinook migration behavior than previously possible using physical tags (Weitkamp 2010) or genetic stock composition data alone (Winans et al 2001).

Our results show a clear increase in stock richness from south to north, a trend that held true even with low sample sizes for some time-area strata (e.g., KO/July). For example, the difference between distribution maps for two stocks (Figure 3; Klamath River and Central Valley fall) shows areas of stock separation and overlap. This information suggests how SSCPUE results might be applied to finer-scale strategies for fishery-stock conservation or harvest management. Consistent fisheries, or an understanding of how to compare dissimilar fisheries, would also enhance the usefulness of contour plot data presentations.

Determining that measures of stock composition were not correlated with SSCPUE when moderate to high abundance stocks were considered in a mixed stock fishery sample has important implications for interpreting stock composition data. This discredits the tendency to assume that an increase in stock composition represents an increase in abundance. The greatest discord between stock composition and SSCPUE was observed for stocks in areas with unusually low stock richness relative to other sampled areas, and wide ranges of SSCPUE-based abundance across all sampled strata. Central Valley fall run, for example, had composition values that were near 100% in some southern sampling regions, but stock density estimated from SSCPUE was substantially higher in other regions. When possible, effort data should be incorporated into measures of stock composition to clarify whether a change in the contribution of a stock to a fishery mixture sample is caused by an actual change in stock abundance, is related to changes in the underlying distributions of other stocks,

or is a consequence of both processes. The SSCPUE metric is clearly a more informative measure than stock composition because findings are unbiased by other stocks present in fishery samples and are more easily interpreted across multiple time-area strata.

Although SSCPUE is useful for inferring stock distributions, CPUE-based estimates of fish density carry the assumption that fish contact rates are proportional to abundance. Catch can, however, fluctuate for reasons unrelated to population levels (Maunder et al 2006). Potential influences include changes in the efficiency of the fleet, the environment, and dynamics of the fish population. A concern regarding the use of such fishery-dependent data is that fish are more susceptible to capture when they are actively feeding, and local feeding conditions and temperatures may affect this behavior. Fish may also scatter when fishermen target a feeding aggregation. In spite of these concerns regarding the use of CPUE data, the CPUE method (of calibration) presented here provides new understanding of migration and perspective for improving stock-specific management and conservation of salmon.

2.5.1 The effect of fishery sampling method on fish catchability

Fishery sampling methodology had little effect on catch rates as measured by CPUE and successful days fished, but some indication of lower CPUE in non-retention sampling was present and therefore warrants consideration for future study sampling design. Mean CPUE was not statistically higher in the FB retention fishery, but there was a greater proportion of successful days of fishing than expected. This contrasted with no difference in the proportion of successful days fished in the other

four area data sets. The SF and MO retention fisheries were only open for short durations, punctuated by closures in a manner similar to non-retention sampling periods. The lack of difference in zero-catch days observed for retention and non-retention fisheries in the SF and MO areas may have been due to the similarity of sampling time-windows. In contrast, the FB retention fishery had two short open periods and then was open for nearly two consecutive weeks in July. More zero-catch days in the FB non-retention fishery could be explained by difficulties in locating fish feeding aggregations due to short open periods and fewer fishermen searching for fish, and possibly less cooperation involved in non-retention fishery sampling. Furthermore, financial compensation incentives that usually drive increased fishing success were absent because all non-retention vessels were compensated a set rate each day regardless of the catch and fish could not be retained for sale. Combining data from two sampling methods is not ideal, but doing so provides insights into stock-specific distributions and general patterns of SSCPUE at fine sampling scales. Moreover, non-retention sampling can provide unbiased measures of CPUE and therefore holds potential for test fisheries or other scientific sampling.

2.5.2 Application of genetic data to fisheries management

At-sea collection of geo-referenced catch and effort data coupled with GSI provide higher resolution stock composition (Winther and Beacham 2009) and novel SSCPUE-based abundance estimates for Chinook salmon. However, direct

application of GSI data to fisheries management remains difficult because stock assessment and harvest models are built around data derived from CWTs. These models rely heavily on age-specific cohort reconstruction, but GSI alone does not provide age data. This obstacle could be overcome using scale-based aging or through alternative genetic methodologies including pedigree-based ones (e.g. parentage-based tagging) that yield cohort and stock data similar to those obtained through CWTs (Garza and Anderson 2007; Steele et al 2013). Further hindrances to use of GSI in fisheries management are the uncertainty in the error structure of genetic stock composition estimates (Waples et al 2008) and that some GSI reporting groups do not coincide with management units defined in fishery management plans. Previous evaluation of the microsatellite baseline using 100% mixture simulations and leave-one-out tests of proportional allocation indicate that most regional groups are highly identifiable (Seeb et al 2007), although some Columbia and Snake River stocks could not be distinguished at the 90% correct assignment threshold (Hess et al 2011) commonly considered to be adequate for delineation of baseline populations. The SNP baseline, however, adequately distinguished all major stock groups found in these fisheries. Concordance of GSI and CWT results were consistent with those power analyses, with mis-allocation for Snake River fall fish to other populations using the microsatellite baseline.

Management and conservation applications of geo-referenced, SSCPUE extend beyond direct incorporation into fisheries assessment models. For example, Satterthwaite and colleagues (Satterthwaite et al 2014) used some of the same data to assess the performance of the data-rich Klamath fall run stock as a proxy for the data-

poor California Coastal stock by inferring ocean distribution from spatial variation in CPUE. The overall level of sampling in this project relative to the commercial fishery (e.g., > 20%) and GSI results provide proof in principle for implementation of a coast-wide sampling program for fisheries management purposes. There is also a current effort to implement ecosystem-based fishery management which uses models that consider larger marine and land ecosystems, climate change impacts, and predator/food web dynamics, instead of single-species harvest limits. Ecosystem based fishery management will require improved spatial representation of stocks, better knowledge of climate-ocean interactions, and refined definitions of habitat (Marasco et al 2007). Using GSI or other types of biochemical markers (e.g., isotopes and trace elements) for stock provenance, geo-referenced methods presented herein are widely applicable for modeling populations of any species where pulses of different life history types or strategies are expressed over time.

2.6 CONCLUSIONS

Coordinated, geo-referenced sampling on a large spatial and temporal scale enabled high-resolution assessment of stock-specific abundance and distribution of migrating Chinook salmon in the coastal California Current ecosystem. Discord between stock composition and SSCPUE estimates highlights the importance of collecting effort data in tandem with fisheries sampling when interpreting mixed stock analysis to understand migratory distribution. Improved understanding of stock-specific abundance and migration patterns achieved by continued sampling and ongoing analyses of such data will contribute to a growing body of knowledge about

the ocean behavior of salmon that can be applied to harvest management so as to allow maximum sustainable harvest while achieving conservation objectives. Accumulation of such information over time may permit the elucidation of which biogeographic and oceanographic factors affect migratory decisions and local abundance of salmon in the coastal ocean.

2.7 ACKNOWLEDGEMENTS

We thank the multitude of fisherman and port representatives who participated in sample collection and routing, and A. Longton, OR Fleet Manager, and C. Faulk for CA sample management. We thank V. Apkenas, A. Clemento, C. Columbus, E. Gilbert-Horvath and D. Pearse for assistance with data generation and sample handling in CA and A. Whitcomb for genotyping and sample handling in OR. The OR Department of Fish and Wildlife recorded barcodes of fish that tested positive for CWTs. N. Sard and W. Satterthwaite provided helpful suggestions for statistical analyses. This manuscript was improved by technical review from M. O'Farrell, W. Satterthwaite and L. Weitkamp. This project was guided by members of the West Coast Salmon Genetic Stock Identification Collaboration. For project leadership and management we thank N. Fitzpatrick, OR Salmon Commission, and D. Goldenberg, CA Salmon Council. Oregon acknowledges the Pacific Salmon Commission for supporting development of the standardized GAPS microsatellite baseline and GAPS consortium members for their baseline contributions. The project was supported by fishery disaster funds NA07NMF4540337, Saltonstall-Kennedy Grant

NA08NMF4270421, and Federal Appropriations NA08NMF4720662 and

NA09NMF4720381.

2.8 REFERENCES

Allison PD (2008) Convergence failures in logistic regression. Proceedings of the SAS Global Forum 2008 Conference 360-2008: 1–11.

Anderson EC, Waples RS, Kalinowski ST (2008) An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65: 1475–1486. doi:10.1139/F08-049.

Bellinger MR, Banks MA (2007). Genetics. Project CROOS Collaborative Research on Oregon Ocean Salmon. Final Report to the Oergon Watershed Enhancement Board. Available http://www.pacificfishtrax.org/media/final-rpt-10307_fleet-genetics.pdf. Oregon Salmon Commission. pp. 1–30. Accessed 20 June 2014.

Banks MA, Rashbrook VK, Calavetta MJ, Dean CA, Hedgecock D (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. doi:10.1139/f00-034.

Beacham TD, Lapointe M, Candy JR, Miller KM, Withler RE (2004) DNA in Action: Rapid Application of DNA Variation to Sockeye Salmon Fisheries Management. Conservation Genetics 5: 411–416. doi:10.1023/B:COGE.0000031140.41379.73.

Beacham TD, Winther I, Jonsen KL, Wetklo M, Deng L, et al. (2008) The Application of Rapid Microsatellite-Based Stock Identification to Management of a Chinook Salmon Troll Fishery off the Queen Charlotte Islands, British Columbia. North American Journal of Fisheries Management 28: 849–855. doi:10.1577/M06-167.1.

Burke BJ, Peterson WT, Beckman BR, Morgan C, Daly EA, et al. (2013) Multivariate Models of Adult Pacific Salmon Returns. PLoS ONE 8: e54134. doi:10.1371/journal.pone.0054134.

Clemento AJ, AbadíA-Cardoso A, Starks HA, Garza JC (2011) Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, Oncorhynchus tshawytscha. Molecular Ecology Resources 11: 50–66. doi:10.1111/j.1755-0998.2010.02972.x.

Clemento AJ, Crandall ED, Garza JC, Anderson EC (2014.) Evaluation of a baseline for genetic stock identification of Pacific Chinook salmon (Oncorhynchus tshawytscha) using SNPs. 112:112-130. Fishery Bulletin

Dann TH, Habicht C, Baker TT, Seeb JE (2013) Exploiting genetic diversity to balance conservation and harvest of migratory salmon. Canadian Journal of Fisheries and Aquatic Sciences 70: 785–793.

Garza JC, Anderson EC (2007) Large scale parentage inference as an alternative to coded-wire tags for salmon fishery management. PSC Genetic Stock Identification Workshop (May and September 2007): Logistics Workgroup Final Report and Recommendations. Available from http://www.psc.org/info_genetic_stock_id.htm. Vancouver, British Columbia, Canada: Pacific Salmon Commission. pp. 45–55. Accessed 20 June 2014.

Grant WS, Milner GB, Krasnowski P, Utter FM (1980) Use of Biochemical Genetic Variants for Identification of Sockeye Salmon (Oncorhynchus nerka) Stocks in Cook Inlet, Alaska. Canadian Journal of Fisheries and Aquatic Sciences 37: 1236–1247. doi:10.1139/f80-159.

Heard WR, Shevlyakov E, Zikunova OV, McNicol RE (2007) Chinook Salmon - Trends in Abundance and Biological Characteristics. North Pacific Anadromous Fish Commission Bulletin 4: 77–91.

Hedgecock D, Banks MA, Rashbrook VK, Dean CA, Blankenship SM (2001) Applications of population genetics to conservation of chinook salmon diversity in the Central Valley. Contributions to the Biology of Central Valley Salmonids, California Department of Fish and Game. 179. UC San Diego: Scripps Institution of Oceanography Library: Fish Bulletin. pp. 45–69.

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources 11: 137–149. doi:10.1111/j.1755-0998.2010.02958.x.

Hinke JT, Watters GM, Boehlert GW, Zedonis P (2005) Ocean habitat use in autumn by Chinook salmon in coastal waters of Oregon and California. Marine Ecology Progress Series 285: 181–192. doi:10.3354/meps285181.

Hobday AJ, Boehlert GW (2001) The role of coastal ocean variation in spatial and temporal patterns in survival and size of coho salmon (Oncorhynchus kisutch). Canadian Journal of Fisheries and Aquatic Sciences 58: 2021–2036. doi:10.1139/cjfas-58-10-2021.

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes 6: 998–1002. doi:10.1111/j.1471-8286.2006.01428.x.

Jefferts KB, Bergman PK, Fiscus HF (1963) A Coded Wire Identification System for Macro-Organisms. Nature 198: 460–462. doi:10.1038/198460a0.

Johnson KJ (2004) Regional Overview of Coded Wire Tagging of Anadromous Salmon and Steelhead in northwest America. Regional Mark Processing Center, Pacific States Marine Fisherie s Commission 205 SE Spokane Street, Suite 100, Portland, Oregon 97202-6413, USA.

.

Kalinowsi ST, Manlove KR, Taper ML (2007) ONCOR A computer program for Genetic Stock Identification. Department of Ecology, Montana State University, Bozeman MT 59717 Available for download from http://www.montana.edu/kalinowski.

Kallio-Nyberg I, Peltonen H, Rita H (1999) Effects of stock-specific and environmental factors on the feeding migration of Atlantic salmon (Salmo salar) in the Baltic Sea. Canadian Journal of Fisheries and Aquatic Sciences 56: 853–861. doi:10.1139/cjfas-56-5-853.

Marasco RJ, Goodman D, Grimes CB, Lawson PW, Punt AE, et al. (2007) Ecosystem-based fisheries management: some practical suggestions. Canadian Journal of Fisheries and Aquatic Sciences 64: 928–939. doi:10.1139/f07-062.

Maunder MN, Sibert JR, Fonteneau A, Hampton J, Kleiber P, et al. (2006) Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES Journal of Marine Science 63: 1373–1385. doi:10.1016/j.icesjms.2006.05.008.

McDonald J (1981) The Stock Concept and its Application to British Columbia Salmon Fisheries. Canadian Journal of Fisheries and Aquatic Sciences 38: 1657–1664. doi:10.1139/f81-214.

Milner GB, Teel DJ, Utter FM, Winans GA (1985) A Genetic Method of Stock Identification in Mixed Populations of Pacific Salmon, Oncorhynchus spp. Marine Fisheries Review 47: 1–8.

Moran P, Teel DJ, Banks MA, Beacham TD, Bellinger MR, et al. (2013) Divergent life-history races do not represent Chinook salmon coast-wide: the importance of scale in Quaternary biogeography. Canadian Journal of Fisheries and Aquatic Sciences 70: 415–435. doi:10.1139/cjfas-2012-0135.

Natural Resources Consultants, Inc. (1998) Northwest Emergency Assistance Plan at-sea Research Programs. Final Report. Prepared for Pacific States Marine Fisheries Commission. Natural Resources Consultants, Inc, Seattle, Washington

Pacific Fishery Management Council (2010) Preseason Report III: Analysis of Council Adopted Management Measures for 2010 Ocean Salmon Fisheries. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 101, Portland, Oregon 97220-1384.

Pacific Fishery Management Council A (2007) Fisheries Regulation Assessment Model (FRAM) An Overview for Coho and Chinook. Pacific Fishery Management Council, 7700 NE Abassador Place, Suite 101, Portland, OR.

Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. Doctoral Thesis, University of Dublin, Dublin.

Pearcy WG (1992) Ocean ecology of North Pacific salmonids. University of Washington, Washington Sea Grant, Seattle. 179 p.

Peyronnet A, Friedland KD, Ó Maoileidigh N (2008) Different ocean and climate factors control the marine survival of wild and hatchery Atlantic salmon Salmo salar in the north-east Atlantic Ocean. Journal of Fish Biology 73: 945–962. doi:10.1111/j.1095-8649.2008.01984.x.

Satterthwaite WH, Mohr MS, O'Farrell MR, Anderson EC, Banks MA, et al. (2014) Use of Genetic Stock Identification Data for Comparison of the Ocean Spatial Distribution, Size at Age, and Fishery Exposure of an Untagged Stock and Its Indicator: California Coastal versus Klamath River Chinook Salmon. 143: 117–133.

Satterthwaite WH, Mohr MS, O'Farrell MR, Wells BK (2013) A comparison of temporal patterns in the ocean spatial distribution of California's Central Valley Chinook salmon runs. Canadian Journal of Fisheries and Aquatic Sciences 70: 574–584. doi:10.1139/cjfas-2012-0395.

Seeb LW, Antonovich A, Banks MA, Beacham TD, Bellinger MR, et al. (2007) Development of a Standardized DNA Database for Chinook Salmon. Fisheries 32: 540–552. doi:10.1577/1548-8446(2007)32[540:DOASDD]2.0.CO;2.

Shaklee JB, Beacham TD, Seeb L, White BA (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fisheries Research 43: 45–78. doi:10.1016/S0165-7836(99)00066-1.

Shaul L, Weitkamp L, Simpson K, Sawada J (2007) Trends in abundance and size of coho salmon in the Pacific Rim. North Pacific Anadromous Fish Commission Bulletin 4: 93–107.

Steele CA, Anderson EC, Ackerman MW, Hess MA, Campbell NR, et al. (2013) A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Canadian Journal of Fisheries and Aquatic Sciences 70: 1046–1054. doi:10.1139/cjfas-2012-0451.

Thériault V, Moyer GR, Jackson LS, Blouin MS, Banks MA (2011) Reduced reproductive success of hatchery coho salmon in the wild: insights into most likely mechanisms. Molecular Ecology 20: 1860–1869. doi:10.1111/j.1365-294X.2011.05058.x.

Tufte ER (1986) The Visual Display of Quantitative Information. 7th printing. Cheshire, CT: Graphics Press. 197 p.

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypothesis. Econometrica 57: 307–333.

Waples RS, Teel DJ, Myers JM, Marshall AR (2004) Life-history divergence in Chinook salmon: historic contingency and parallel evolution. Evolution 58: 386–403. doi:10.1111/j.0014-3820.2004.tb01654.x.

Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9: 423–449. doi:10.1111/j.1467-2979.2008.00303.x.

Waples R (2001) Characterizing diversity in salmon from the Pacific Northwest. Journal of Fish Biology 59: 1–41. doi:10.1006/jfbi.2001.1764.

Wells BK, Grimes CB, Sneva JG, McPherson S, Waldvogel JB (2008) Relationships between oceanic conditions and growth of Chinook salmon (Oncorhynchus tshawytscha) from California, Washington, and Alaska, USA. Fisheries Oceanography 17: 101–125. doi:10.1111/j.1365-2419.2008.00467.x.

Wessa P (2012). Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7, URL http://www.wessa.net/.

Weitkamp LA (2010) Marine Distributions of Chinook Salmon from the West Coast of North America Determined by Coded Wire Tag Recoveries. Transactions of the American Fisheries Society 139: 147–170. doi:10.1577/T08-225.1.

Winans GA, Viele D, Grover A, Palmer-Zwahlen M, Teel D, et al. (2001) An Update of Genetic Stock Identification of Chinook Salmon in the Pacific Northwest:

Test Fisheries in California. Reviews in Fisheries Science 9: 213–237. doi:10.1080/20016491101753.

Winther I, Beacham TD (2009) The application of Chinook salmon stock composition data to management of the Queen Charlotte Islands troll fishery. Pacific Salmon: Ecology and Management of Western Alaska's Populations. American Fisheries Society, Symposium 70. American Fisheries Society, Symposium 70. pp. 977–1004.

Table 2.1. Monthly numbers of non-retention and retention vessel-days of salmon troll fishing effort during 2010 in nine spatial strata from Cape Falcon, Oregon (OR) to Santa Barbara, California (CA). Two OR areas, North and South Oregon Coast (NO, CO) were open May though August while all other areas were either closed or periodically open and then closed to commercial fisheries. The month of September was closed for all areas. Regional boundaries for each area and table abbreviations are as follows: Cape Falcon to Florence south jetty (NO, lat 45.767º to 44.015º N); Florence south Jetty to Humbug Mountain (CO, lat 44.015º to 42.667º N); Humbug Mountain to CA/OR border (KO, lat 42.667º to 42.000º N); CA/OR border to Humboldt south jetty (KC-n, lat 42.000º to 40.765º N); Horse Mountain to Point Arena (FB, lat 40.083º to 38.958º); Point Arena to Point Reyes (SF-n, lat 38.958º to 37.996º N); Point Reyes to Pigeon Point (SF-s, lat 37.996º to 37.183º N); Pigeon Point to Point Sur (MO-n, 37.183º to 36.300º N) and Point Sur to Mexican Border (MO-s, lat 36.300º to 32.584º N). The MO area was sampled as a single area but data were partitioned into separate strata for reporting purposes.

	May		June		July		August		September	Totals	
	non-retention	retention	non-retention	retention	non-retention	retention	non-retention	retention	non-retention	non-retention	retention
NO		75.29		176.96		73.86		181.35	33.00	33.00	507.45
CO		108.71		166.08		34.14		204.65	30.07	30.07	513.58
KO		0.00	34.96			7.00		27.00	20.93	55.89	34.00
KC-n	0.00		37.00		55.00		60.00		60.00	212.00	0.00
FB[1]	9.00		47.00			91.57		120.00	70.00	126.00	211.57
SF-n[2]	24.00		59.00		32.00	60.20	60.00		59.73	234.73	60.20
SF-s[2]	52.00		60.00		38.00	48.19	58.00		67.27	275.27	48.19
MO-n[2]	35.69		42.11		22.00	60.04	40.00		28.00	167.80	60.04
MO-s[2]	8.31		14.89		11.00	18.00	17.00		12.00	63.20	18.00
totals	129.00	184.00	294.96	343.04	158.00	393.00	235.00	533.00	381.00	1197.96	1453.04

[1] Open July 1-4, 8-11, 15-29, and all of August

[2] Open July 1-4, 8-11

Table 2.2. Monthly numbers of sub-legal and legal-sized Chinook salmon encounters recorded at-sea in project area and period. Retention sampling was conducted during months and areas open to commercial troll fishing and non-retention sampling was conducted when the fishery was closed. Time/area strata information is provided in Figure 2.1 and Table 2.1 (also see text). Biological samples were obtained from all but 30 fish.

	May		June		July		Aug		Sept		Totals		
	sub-legal	legal	sub-legal	legal	sub-legal	legal	sub-legal	legal	sub-legal	legal	sub-legal	legal	Grand total
NO	-	404	-	1102	-	403	-	[1]532	7	23	7	2464	2471
CO	-	453	-	616	-	75	-	601	31	64	31	1809	1840
KO	-	0	-	44	-	10	-	69	49	86	49	209	258
KC-n	0	0	6	64	7	127	88	382	121	247	222	820	1042
FB[2]	6	91	10	159	2	483	8	533	49	441	75	1707	1782
SF-n[3]	10	37	23	87	9	395	22	138	13	37	77	694	771
SF-s[3]	27	86	78	198	17	99	5	114	11	70	138	567	705
MO-n[3]	6	11	3	17	16	377	34	114	42	50	101	569	670
MO-s[3]	0	2	0	5	0	13	0	9	0	16	0	45	45
Totals	49	1084	120	2292	51	1982	157	2492	323	1034	700	8884	9584

[1] Eight fish encounters were excluded from CPUE calculations (see text for details)
[2] Open July 1-4, 8-11, 15-29, and all of August
[3] Open July 1-4, 8-11

Table 2.3. Observed mean catch per unit effort, measured as legal-sized fish encounters per vessel-day of effort, calculated for nine strata from Cape Falcon, Oregon (OR), to Santa Barbara, California (CA) sampling during 2010. Retention fishery sampling is indicated by bold, mixed retention/non-retention fisheries sampling by italic, and non-retention fisheries by regular text. No data were collected during May in the Klamath management zones (OR & CA). Time/area strata information is provided in Figure 2.1 and Table 2.1 (also see text).

	May	June	July	August	September
North Oregon Coast	**5.37**	**6.23**	**5.46**	[1]**2.89**	0.70
Central Oregon Coast	**4.17**	**3.71**	**2.20**	**2.94**	2.13
Klamath Zone-OR		1.26	**1.43**	**2.56**	4.11
Klamath Zone-CA, north		1.68	2.31	6.37	4.12
Fort Bragg	10.11	3.38	**5.27**	**4.44**	6.30
San Francisco – north	1.54	1.47	***4.28***	2.30	0.62
San Francisco – south	1.65	3.30	***1.15***	1.97	1.04
Monterey – north	0.31	0.40	***4.60***	2.85	1.79
Monterey – south	0.24	0.34	***0.45***	0.53	1.33

[1] CPUE calculation exclude 8 samples (see text for details)

Table 2.4. Legal-sized catch, vessel-days effort, vessel participation, and catch per vessel day effort (CPUE) for the 2010 coast-wide sampling program relative to the Oregon (OR) and California (CA) commercial fishery conducted south of Cape Falcon, OR (excludes OR fall area fisheries data) from May – August 2010.

	At-sea study	Commercial fishery	% At-sea study / commercial fishery
Landed fish, OR	4,482	26,454	16.9%
Landed fish, CA	4,402	15,088	29.2%
Landed fish, total	8,884	41,542	21.4%
Vessel effort, OR	1055	3428	30.8%
Vessel effort, CA	398	1,975	20.2%
Vessel effort, total	1,453	5,403	26.9%
Vessel-days effort excluding days with zero-catch	1,113	as above	20.6%
N participating vessels, OR[1]	78	370	21.1%
N participating vessels, CA	63	215	29.3%
N participating vessels (retention only)	141	585	24.1%
CPUE (legal-sized fish/vessel day effort)	6.11	7.69	n/a
CPUE (excluding zero-catch days)	7.98	n/a	n/a

[1] Includes Astoria

Table 2.5. Estimated mean daily catch per unit effort and 95% confidence intervals (CI) generated with log-linear negative binomial models. The null hypothesis of no difference between estimated daily catch for retention (Fishery R) and non-retention fisheries is rejected at a probability of $z > 0.05$ (shown in bold). In San Francisco (SF-n, -s) and Monterey (MO-n, -s) areas retention sampling was conducted July 1 – 4 and 8 – 11, and non-retention during weeks in July. The Fort Bragg (FB) area retention fishery (July 1 – 4, 8 – 11, and 15 – 29) was compared to the June non-retention fishery.

	Estimated Model Coefficients						Negative Binomial Model Results		
Area	Fishery	N days	Mean	Lower CI	Upper CI		Std. Error	z value	Pr (> | z |)
FB	Non-retention	47	3.38	2.24	5.09	(Intercept)	0.208	5.847	5.00 e - 09
	Retention	83	5.14	3.81	6.94	Fishery R	0.259	1.622	0.105
SF-n	Non-retention	32	1.44	0.83	2.48	(Intercept)	0.278	1.307	0.191
	Retention	57	6.04	4.21	8.66	Fishery R	0.333	4.306	**1.66 e – 05**
SF-s	Non-retention	38	1.24	0.77	1.99	(Intercept)	0.247	0.862	0.389
	Retention	45	1.15	0.74	1.8	Fishery R	0.338	-0.192	0.848
MO-n	Non-retention	22	6.27	4.09	9.62	(Intercept)	0.218	8.419	< 2 e - 16
	Retention	59	4.05	3.08	5.32	Fishery R	0.258	-1.692	0.091
MO-s	Non-retention	10	0.40	0.088	1.82	(Intercept)	0.775	-1.182	0.237
	Retention	18	0.39	0.12	1.21	Fishery R	0.969	-0.029	0.977

Table 2.6. Strengths of associations, evaluated using Kendall's τ correlation coefficients, between pairs of estimated genetic stock composition values and stock-specific catch per unit effort (SSCPUE), an effort-calibrated measure of stock abundance. Data were comprised of mixed stock fishery samples collected in nine strata sampled from May – September, 2010, using either retention or non-retention fishing protocols. The full data set includes all non-zero pairs of data, in contrast to the threshold data set which includes only pairs of data above a minimum stock composition value (min %) to determine the threshold point above which stock composition and SSCPUE values failed to correlate as measured by Kendall's τ with p-values < 0.05. Analyses were performed using data from all stocks and then on an individual-stock basis for each of the five stocks that were encountered most frequently and with greatest abundance across all month-area-fishery strata. CA = California, OR = Oregon.

	Full data set					Threshold data set				
	Retention		Non-retention			Retention			Non-retention	
	τ	p-value	τ	p-value	min %	τ	p-value	min %	τ	p-value
Retention – all stocks	0.85	0.000	----	----	22.3	0.33	0.079	----	----	----
Non-retention – all stocks	----	----	0.67	0.000	----	----	----	13.9	0.16	0.111
CA Central Valley	0.58	0.011	0.22	0.113	13.2	0.42	0.107	0.0	0.22	0.108
Rogue	0.69	0.001	0.74	0.000	3.3	0.44	0.119	14.3	0.39	0.178
Klamath	0.82	0.000	0.74	0.000	5.16	0.73	0.060	10.9	0.50	0.107
CA Coastal	0.73	0.004	0.54	.0020	1.7	1.00	0.089	3.5	0.42	0.064
Northern CA/Southern OR Coast	0.82	<0.001	0.65	0.001	2.8	0.52	0.133	8.7	0.43	0.174

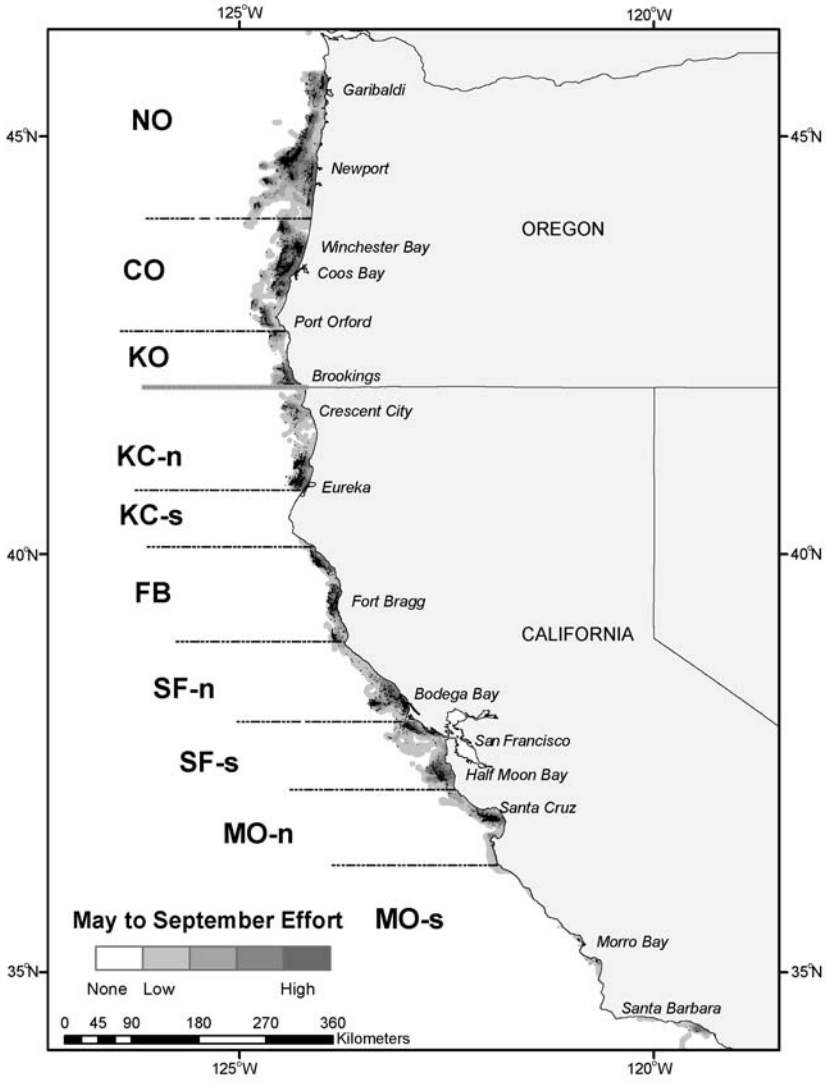


Figure 2.1. Troll fishing effort and Chinook salmon catch locations. Catch locations (n = 9,854) are shown as black dots while effort (n = 2651 vessel days effort) is conveyed using shaded contour of fishing vessels locations logged by GPS units in five-minute intervals. Area codes and latitudes are as follows: North Oregon Coast (NO, lat 45.767º to 44.015º N), Central OR Coast (CO, lat 44.015º to 42.667º N), Klamath Zone OR (KO, lat 42.667 º to 42.000º N), Klamath Zone California (KC-n, lat 42.000º to 40.765º N, with no sampling permitted in the KC-s between 40.765º N to 40.083º N), Fort Bragg (FB, lat 40.083º to 38.958º N), San Francisco north (SF-n, lat 38.958º to 37.996º N), San Francisco south (SF-s, lat 37.996º to 37.183º N), Monterey north (MO-n, lat 37.183º N to lat 36.300º), and Monterey south (MO-s, lat 36.300º to 32.584º N).

Figure 2.2. Log stock-specific catch per vessel day of fishing effort. Twenty-two stock groups were sampled within nine area strata May – September, 2010. Vertical green (retention) and magenta (non-retention) bars on left axis shows effort in total days fished. Stocks are ordered north to south. Area abbreviations are: North Oregon Coast, NO; South Oregon Coast, CO; Oregon Klamath Zone, KO; California Klamath Zone north, KC-n; Fort Bragg, FB; San Francisco north, SF-n; San Francisco south, SF-s; Monterey Bay north MO-n; and Monterey Bay south, MO-s (see Figure 1 for spatial details). Sampling was not conducted during the month of May in KO and KC-n.

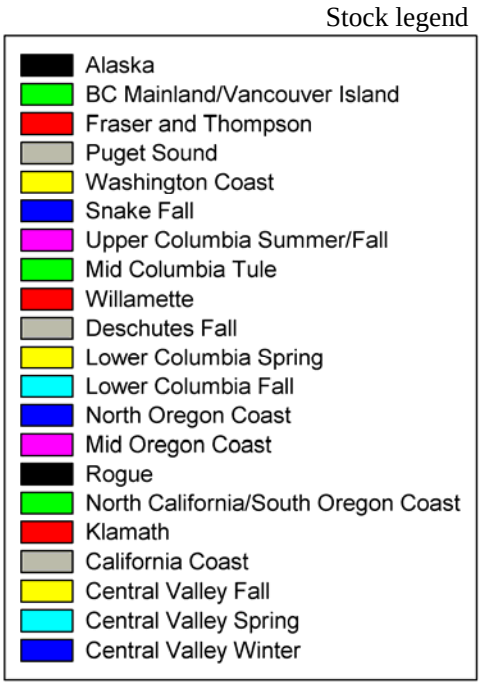


Figure 2.2. Continued

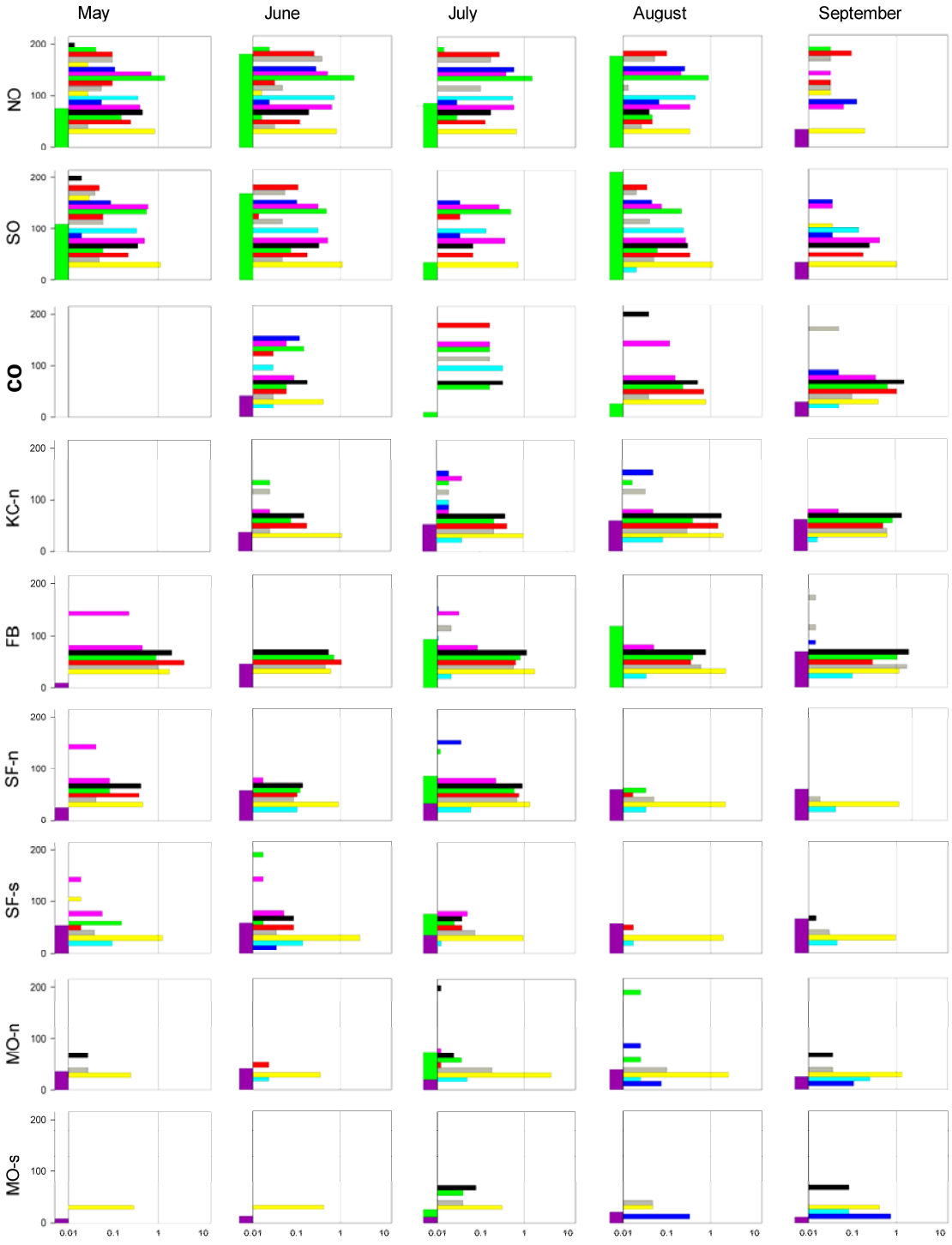


Figure 2, continued

Log catch per unit effort

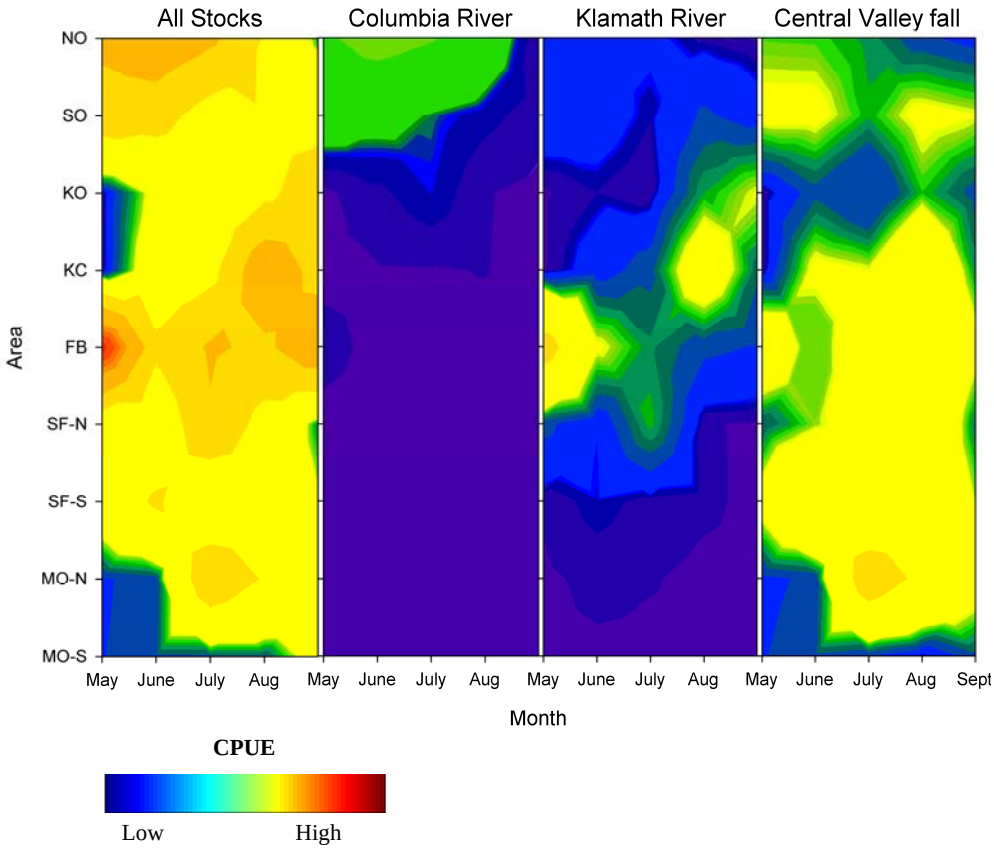


Figure 2.3. Catch per unit effort (CPUE) contour plots of legal-sized Chinook salmon across month (x-axis) and area (y-axis) strata. The CPUE results are shown for all Chinook salmon stocks, and for stock groupings Columbia River, Klamath, and California Central Valley fall run. Areas range from Cape Falcon, Oregon to Morro Bay, California (see text for area abbreviations and details). No sampling was conducted in KO/May and KC-n/May and KC-s, all season.

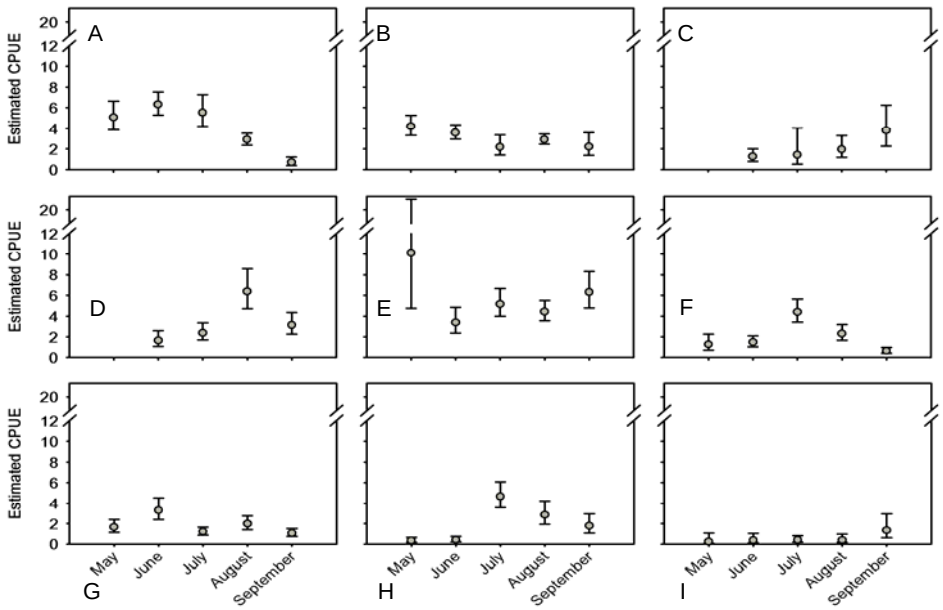


Figure 2.4. Estimated mean catch per unit effort (CPUE, with vessel-day effort) and 95% confidence intervals. The CPUE was calculated for nine month-area strata using a log-linear negative binomial model that included terms sampling month, area, and their interaction term. Area abbreviations are: A) North Oregon Coast, B) South Oregon Coast, C) Oregon Klamath Zone, D) California Klamath Zone-north (KC-n), E) Fort Bragg, F) San Francisco north and G) south, H) Monterey north and I) south. See Chapter 2 text and Figure 2.2 for details on sample sizes and data collection.

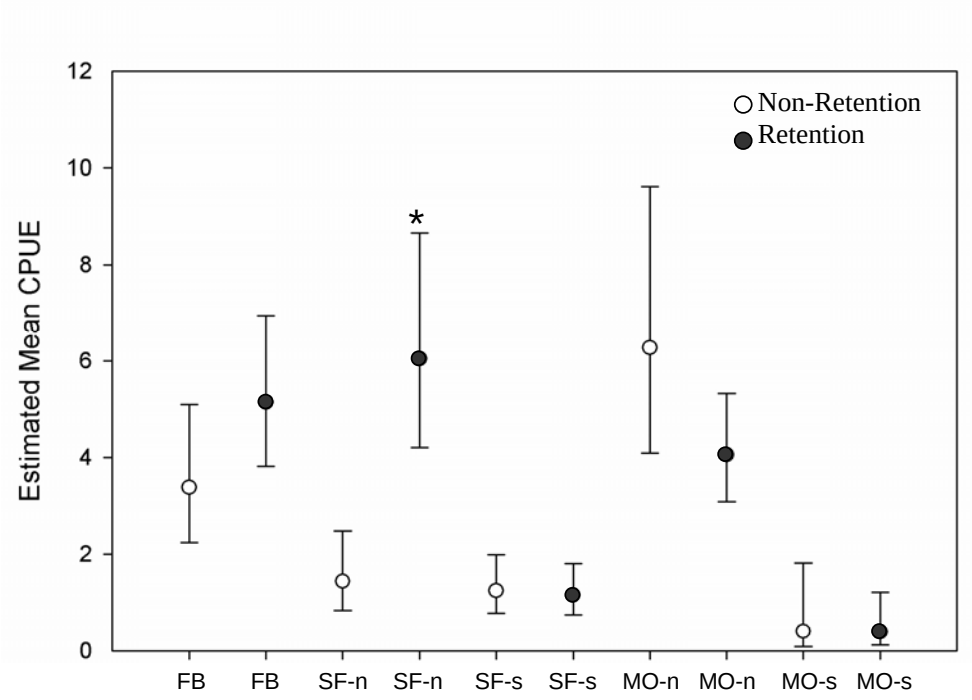


Figure 2.5. Estimated mean catch per vessel-day effort (CPUE) for non-retention and retention fisheries sampling. A log-linear negative binomial model was used to estimate CPUE and 95% confidence intervals for retention and non-retention sampling conducted over five areas: Fort Bragg (FB), San Francisco north (SF-n) and south (SF-S), Monterey north (MO-n) and south (MO-s). Non-retention and retention sampling in FB was performed during June and July, 2010. In the other areas, retention sampling was conducted July 1 – 4 and 8 – 11, and non-retention sampling performed for the remainder of the month. The null hypothesis of no difference between CPUE for retention fishery is rejected at a probability of $z > 0.5$, denoted by *.

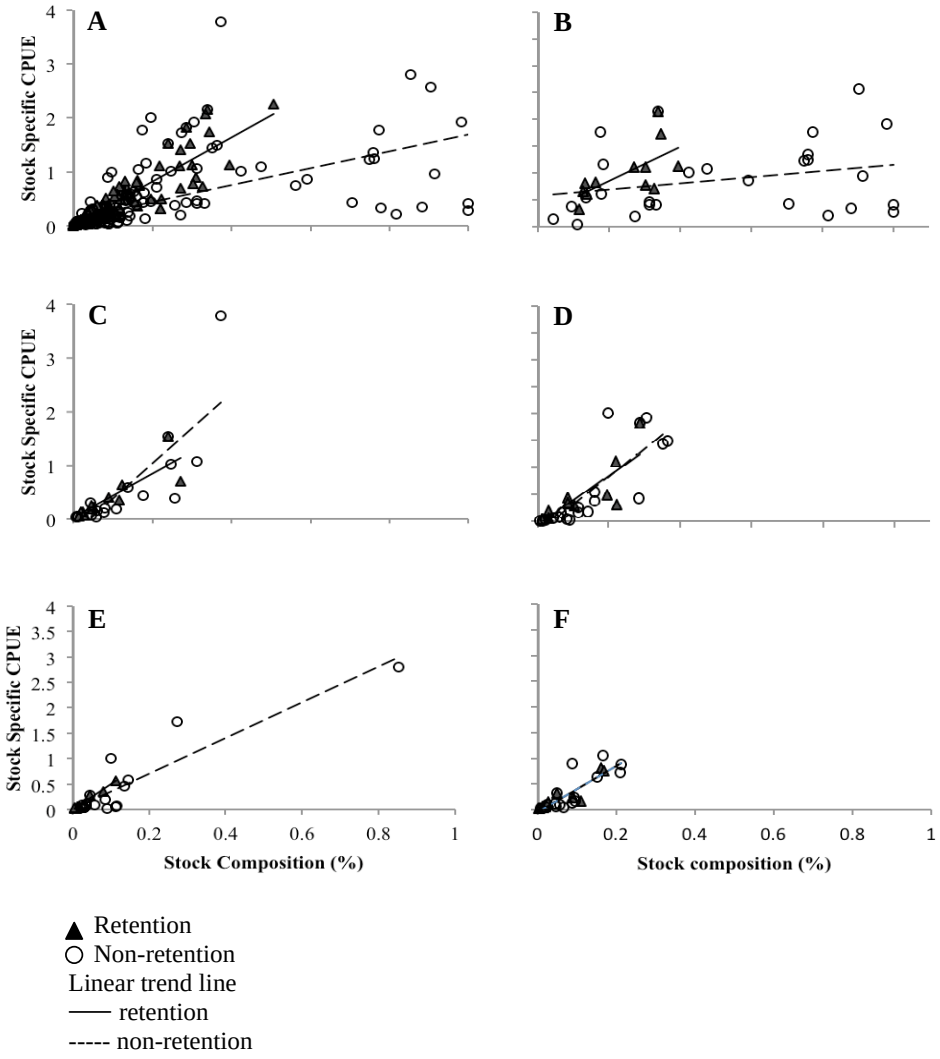
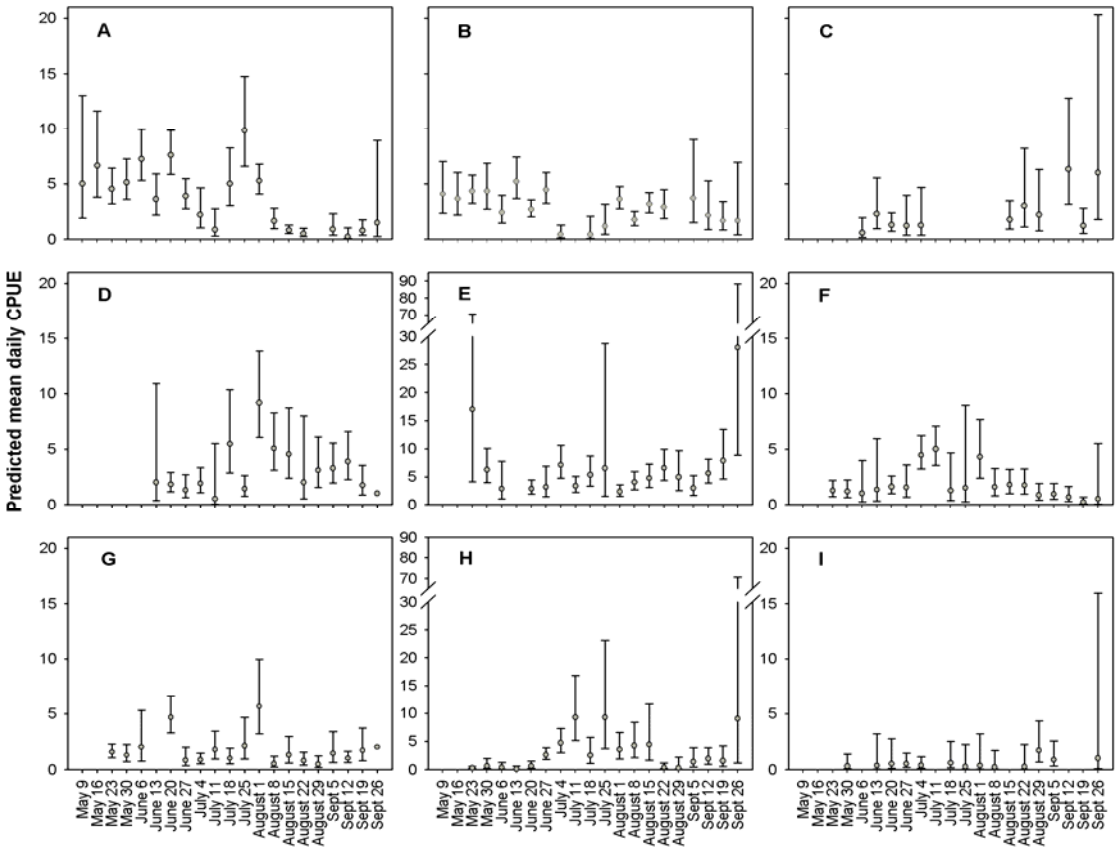


Figure 2.6. Scatterplots of stock-specific catch per unit effort (SSCPUE) and stock composition values. Paired values indicate stock composition (x-axis) and SSCPUE (y-axis) results for each stock encountered across eight spatial strata. Areas were sampled on a monthly from May through September 2010; retention and non-retention fisheries were treated separately: (a) all data, (b) California Central Valley Fall, (c) Klamath, (d) Rogue, (e) California Coastal, and (f) Northern California/Southern Oregon Coastal stocks.



Supplemental Figure 2.1. Predicted mean daily catch per unit effort (CPUE) and 95% confidence intervals on a weekly basis for 2010. The CPUE was modeled using a generalized linear model with a log-link negative binomial distribution. Model terms included exogenous variables week and area, and a term for their interaction. A total of 2,580 days of vessel effort and catch data were included in analyses. Areas are: North Oregon Coast (A), South Oregon Coast (B), Oregon Klamath Zone (C), California Klamath Zone-north (D), Fort Bragg (E), San Francisco area north of Point Reyes (F), San Francisco area south of Point Reyes (G), Monterey, Monterey Bay (H), and Monterey, Santa Barbara and Morro Bay (I).

3- Next generation sequencing reveals genomic signatures of candidate magnetoreceptor cells of salmonids

3.1 ABSTRACT

The ability of animals to detect the Earth's magnetic field is scientifically well established, but exactly which proposed receptor(s) provide magnetic sense has been "hotly contested" and no one knows with certainty how any animal perceives magnetic information (Johnsen and Lohmann 2008). A leading hypothesis is that magnetic sense is mediated by crystals of the highly magnetic iron-mineral magnetite connected to ion channels in receptor cells where geomagnetic information is transduced into neural signals. Magnetite crystals found in salmonid tissues are superficial to those found in bacteria, for which genetic underpinnings of crystal formation are well understood, but lacking is knowledge of any eukaryotic genes involved in magnetite production. Here we identified a list of candidate magnetite genes by RNA-seq transcriptome profiling of magnetic and non-magnetic olfactory epithelial cells of salmonids. Complementary muscle, blood, and additional olfactory transcriptome data sets yield further insights into candidate gene characteristics. This study provides the first ever list of candidate magnetoreceptor genes in vertebrates, and is foundational to formulating methods for affirming the "magnetite hypothesis". The impact of identifying the genetic basis of crystal formation in eukaryotes has significance beyond resolving bionavigation with potential to fuel invention of materials with lasting societal impact in applications such as neuroscience, medicine, nano-electronics, and navigation.

3.2 INTRODUCTION

Animals demonstrate behavioral sensitivity to magnetic fields, yet the biophysical transduction mechanism(s) that underpin magnetosensation have been difficult to elucidate. One plausible hypothesis is that crystals of magnetite are contained inside receptors cells where they interact with Earth-strength magnetic fields and transduce geomagnetic information into neural signals (Gould, Kirschvink, and Deffeyes 1978; Kirschvink and Gould 1981; Walker et al. 1997b; Kirschvink, Winklhofer, and Walker 2010). This mechanism is proposed to be highly sensitive to magnetic intensity (Kirschvink and Gould 1981; Walker et al. 1997b), in contrast to the hypothesized quantum-level magnetic compass (Ritz, Adem, and Schulten 2000; Wiltschko and Wiltschko 2007). Fish, turtles, and birds are capable of orientation in complete darkness, absent of visual cues (Lohmann and Lohmann 1993; Schlegel 2007; Stapput et al. 2008; Putman et al. 2014). The presence of a "universal" magnetite-based magnetoreceptor is further supported by behavioral responses to pulse magnetization and/or impairment of the trigeminal nerve or nasal tissue in birds (Beason and Semm 1987; Mora et al. 2004; Heyers et al. 2010; Kishkinev et al. 2013), sea turtles (Irwin and Lohmann 2005), and bats (Holland et al. 2008).

The biogenic production of magnetite was first documented in chiton (Mollusca, Polyplacophora, Lowenstam 1962), followed by the discovery that some bacteria produce a unique intracellular organelle, termed magenetosome, comprised of a nanometer-sized crystal of permanently magnetized magnetite (Fe_3O_4) or greigite

(Fe_3S_4) surrounded by a lipid membrane (Frankel, Blakemore, and Wolfe 1979; Balkwill, Maratea, and Blakemore 1980; Bazylinski and Frankel 2004). Magnetosomes are arranged in well-organized intracellular chains that generate a magnetic force sufficiently strong to physically align the organism to the geomagnetic field, aiding bacteria to maintain position in anoxic layers of sediment (Blakemore and Frankel 1981). The genetic basis of crystal and membrane formation in bacteria is well understood (Richter et al. 2007; Schüler 2008), with magnetosome assembly occurring as a biologically controlled, step-wise process (Murat et al. 2010). Genes are co-localized in operon-like clusters, including a well conserved "magnetosome gene island" (Schübbe et al. 2003a) and additional non-conserved genes involved in crystal formation (Murat et al. 2010). The most recent estimate of proteins known to be involved in magnetosome formation in prokaryotes is ~28 (Nudelman and Zarivach 2014) , although more are likely to play important roles. Many of these magnetosome genes share homologous domains with eukaryotes, but none of those genes have been associated with magnetite crystal formation.

Intensive searches in vertebrate animals for magnetite have revealed its presence in diverse taxa, e.g., bees, birds, and fish, (Gould, Kirschvink, and Deffeyes 1978; Walcott, Gould, and Kirschvink 1979; Fleissner et al. 2003; Mann et al. 1988), however, crystals have not yet been directly associated with a sensory transduction mechanism. Moreover, in most cases these deposits are unlikely to represent the receptor site of magnetic perception (Mann et al. 1988; Treiber et al. 2012), with the important exception that magnetite-containing cells have consistenly been observed in olfactory tissue of salmonid fish (Diebel et al. 2000; Eder et al. 2012) near a nerve

that responds to magnetic intensity treatments (Walker et al. 1997b). Confirming magnetite's involvement in signal transduction is currently hindered by cell scarcity (~1 to 10,000 cells, Eder et al. 2012), lack of obvious physical characteristics visible through light microscopy, and no methodology yet available to locate physiologically responsive cells for assessment of magnetic sensitivity.

Behavioral studies of Pacific salmon and trout (genus *Oncorhynchus*) have established that they possess magnetic sense (T. P. Quinn 1980; Walker et al. 1997b), and magnetic perception is hypothesized to be a component of an innate guidance mechanism utilized for long-distance migration and homing to natal river mouths (K. J. Lohmann, Putman, and Lohmann 2008; Putman et al. 2013). Experimental treatment of juvenile salmon using simulated magnetic displacements revealed that magnetic intensity and inclination geomagnetic field cues are used in tandem during long-distant migration (Putman et al. 2014).

The magnetite crystals extracted from Sockeye salmon dermethmoid tissue (Mann et al. 1988) and olfactory rosettes (Wei 2009) are within theoretical size limits of single domain magnetite (Butler and Banerjee 1975), consistent with the hypothesized size of energetically favorable magnetite-based functional receptors (Kirschvink and Gould 1981). The shape, size, and purity of these crystals indicates strict genetic control of crystal formation, possibly one that shares common origins with bacteria consequent to an ancient lateral gene transfer event (Mann et al. 1988; Kirschvink and Hagadorn 2000). Understanding exactly how salmon and other organisms receive magnetic field cues and use them for migration would possibly enhance conservation efforts, including animal rearing practices that aim to

supplement populations for conservation or harvest, and provide insights into magnetoreceptor mechanisms used by other vertebrate taxa. Here we identify candidate components of magnetite formation in eukaryotes using high-throughput Illumina RNA-seq sequencing and transcriptome profiling of magnetic and non-magnetic cells of Chinook salmon olfactory epithelium, blood, and muscle tissues. Through comparison to prokaryotes, we provide a novel, plausible genetic model of the genomic basis of magnetic perception in vertebrates.

3.3 METHODS

3.3.1 Biological samples

Biological samples were obtained from South Santiam hatchery Chinook salmon stock, a long-distant migrant, reared at the Fish Research Laboratory in Corvallis, Oregon. Fish were housed indoors in a single well-water tank (date in tank 6/26/09). Pairs of olfactory capsules ("rosettes") were rapidly dissected from fish following humane termination by decapitation and pithing (Oregon State University ACUP 4421; samples obtained March and May 2012). Rosettes were quickly rinsed in nanopure water, briefly cut to smaller pieces, and placed in 350 uL of buffer solution (200 mM sucrose, 20 mM Hepes, 4 mM EDTA, pH to 7.4 (unpublished magnetite buffer recipe courtesy H. Cadiou)) containing five uL 2-mercaptoethanol to aid RNA preservation. Enzymatic dissociation of tissues into free cells was achieved by adding 100 uL of 2.5 mg/mL papain (Sigma Aldrich P4762) activated by five mM L-cysteine and 20-40 uL of Trypsin-EDTA solution 0.25% (Sigma). Trials 1 and 3 incorporated 150 ul of antibiotic/antimycotic buffer (Invitrogen) diluted to 1x in

Hanks Balanced Salt Solution pH 7. Enzymatic dissociation proceeded for up to two hours at 12 to 15˚ C, followed by trituration and transfer of the cell slurry into a 50 mL beaker. Cells were spun for ~30 minutes in the presence of 1% triton-X at 4˚ to decrease solution viscosity. Approximately 3 mL of the final cell slurry was transferred to a 4 mL glass vial having a strong neodymium magnet with a pointed tip positioned at mid-point on the vial's upper side. Overnight, magnetic cells collected at the tip of the magnet while opposing gravitational forces pulled non-magnetic cells to the vial's bottom. Plastic wrap secured over the top of the vial was used to prevent dust from contacting the solution. The magnetic pellet, only visible under a dissecting microscope, was aspirated and placed in an RNAse free vial, followed by transfer of non-magnetic cells to a separate vial. Finally, total RNA was extracted using a Qiagen glass-silica based column kit following manufacturer's protocols, and submitted to Oregon State University's Center for Genome Research and Biocomputing (CGRB) for sample processing and sequencing.

The mRNA was isolated from total RNA using poly-A labeled magnetic beads, individual sample were indexed (labeled), and mRNA were deep sequenced on an Illumina HiSeq2000 using 101 cycles and paired-end protocols. Three biological replicate trials were run (n = 6 samples, 3 magnetic and 3 non-magnetic) to permit statistical modeling of gene count data. Each trial was sequenced in an individual Illumina lane, with the exception that the Trial 3 lane included indexed mRNA obtained from a pair of whole olfactory rosettes rapidly dissected, snap-frozen in a dry-ice ethanol slurry, and stored at -80 until RNA extraction.

Contamination by RNAses and non-target magnetic particles were minimized through vigorous cleaning of supplies and tools. Work was performed inside a HEPA filtered hood as protocols permitted. Using an overnight soak in 6N HCl followed by thorough rinse in nanopure water, potential iron-dust contaminants were stripped from ceramic, glass, and plastic tools used for terminating fish, dissecting, and processing fish tissues. Non-magnetic titanium forceps for tissue handling were subjected to 30 minute sonication in nanopure water followed by a brief soak in 6N HCl and final nanopure water rinse. All reagents were purchased as molecular biology grade solutions or filtered through 0.22 micron membrane filter (papain and L-Cysteine) using a HCL-cleaned syringe. The HCL treatment was not applied to filter tips used for RNA liquid handling.

Gene expression profiles of non-olfactory tissues can provide context for candidate gene specificity and possible functions. The mRNA from blood, muscle, and additional olfactory rosettes of three adult Chinook salmon (collection date June 2013) from the same brood and reared in the same tank as fish used for the magnetic/non-magnetic sequencing experiment were isolated and sequenced using RNA-seq. Biological samples were placed in magnetite buffer and immediately flash-frozen by immersion in an ethanol and dry ice slurry, followed by RNA isolation and transfer to the CGRB for final sample processing. All samples were indexed to match individual tissue with fish and then sequenced in a single Illumina HiSeq 2000 lane using a single-end 50-cycle protocol. This shallow-sequencing yielded 10 to 15 million mRNA transcripts per fish-tissue sample type (total of 9 indexed samples, Table 3 1). Whole blood gene expression profiles were intended to

identify and exclude genes associated with macrophages, a type of white blood cell, that can be iron-rich and potentially magnetic (Treiber et al. 2012). Magnetoreceptors are presumed to absent from muscle tissue, however data are lacking to support this assertion.

3.3.2. RNA-Seq quality processing and alignment to reference data sets

The RNA-seq sequences ("reads") were quality processed using FASTX Toolkit programs fastq_quality_trimmer, fastx_clipper, and fastq_quality_filter (0.0.13, Blankenberg et al. 2010; analysis pipeline available from Github X). After low-quality nucleotides (phred score < 20) were trimmed from the ends of sequences and Illumina adapters removed, reads of minimum length 25 nucleotides were retained only if at least 90% of all cycles had a minimum base pair quality of phred score 20. Paired-end reads were re-matched using the script PE_Combiner (De Wit et al. 2012) with singleton reads retained as "orphans".

Gene expression analysis requires matching reads to a reference data set, ideally one that is well annotated and suited for revealing information about gene function and pathways. Owing to the size and complexity of salmonid genomes, and lack of a sequenced reference genome from a closely related fish (N. L. Quinn et al. 2008; Davidson et al. 2010), development of salmonid "–omic" reference data sets have lagged behind those of model organisms. Thus given the emergent state of non-model fish transcriptomics (Qian et al. 2014), in this study the gene expression patterns of candidate magnetoreceptor cells were evaluated using references compiled from diverse methods: SalmonDB, a well-annotated "unigene" fish dataset developed

from *Oncorhynchus* and *Salmo* Sanger-sequenced expressed sequence tags (using only contigs >= 736 nucleotides, Di Genova et al. 2011), a reference guided *O. mykiss* transcriptome assembled from Illumina short reads and expressed sequence tags (Fox et al. 2014), an *O. mykiss* transcriptome derived by annotation from the first ever published salmonid genome (Berthelot et al. 2014), two types of de novo references, and a novel genome-guided olfactory transcriptome (this study). The de novo assemblies, assembled with Trinity (Grabherr et al. 2011), were constructed using two data sets: (~248 million reads, Trials 1 and 3 data; computational constraints limited the use of all available data), and only reads from Trials 1-3 that failed to align to the SalmonDB data set (~342 million reads). The genome-guided assembly was constructed by alignment of all olfactory epithelial reads (~568 million reads Table 3. 1) to the *O. mykiss* genome (Berthelot et al. 2014) using Tuxedo Tools (Trapnell et al. 2012) (n = 406,355,615 reads aligned, 72% alignment rate; similar to alignment of Atlantic salmon RNA-seq data to draft genome (Wang et al. 2014)). Attributes for -omic references are provided in Table 3. 2.

3.3.3. Differential gene expression analyses

Quality processed RNA-Seq reads were aligned to reference data sets using Bowtie2 v2.1.0 (Trapnell et al. 2012) with the pre-set parameters --very-sensitive. For analyses of differential gene expression, the RNA-seq reads that matched each reference, except the genome-guided transcriptome, were tallied using a custom "count" script (available from Github XX) and organized into a numerical (p x n) matrix (M) with p representing the number of reads that aligned to a contig in the

transcriptome and *n* the number of samples (total n = 6, 3 magnetic and 3 non-magnetic samples; 6 M matrixes, references listed in Table 3. 2) for statistical analyses. Data were initially explored with scatterplot matrixes to assess correlation among treatment groups and biological samples. The correlation plots demonstrated that magnetic and non-magnetic samples from Trial 2, which excluded antibiotic buffer, were more correlated with each other than samples of the same treatment. Based on this evidence, data from Trial 2 were excluded from differential gene expression analyses.

3.3.4 Statistical analysis

Tests for differential gene expression between magnetic and non-magnetic cell types were performed separately for each numeric matrix M using the statistical package NPBSeq (Di et al. 2011). This program incorporates a negative binomial distribution to model gene counts (Y_{ijk}) from RNA sequencing as independent NBP negative binomial random variables $Y_{ijk} \sim$ NBP ($\pi_{ik} m_{jk}$, Φ, α) (Di et al. 2011). The dispersion parameters (Φ, α) were estimated for each M by maximizing the conditional log-likelihood over all of the data and assumed to be the same for all genes *i = 1, …. N* within each M. A statistical test of the null hypothesis H_0: $\pi_{i1} = \pi_{i2}$ for each gene *i* was performed to identify differentially expressing genes based on the evidence provided by sequence read counts. Two sided p-values were computed by Robinson and Smyth's exact test adapted to the NBP distribution (Robinson and Smyth 2007, 2008; Di et al. 2011) using a nominal false discovery rate (*q*-value) of 0.05 as a cutoff for statistical significance. Genes with less than one mean counts or variance were

excluded from DEG consideration. To accommodate the combination of paired-end reads and "orphans", only forward reads that aligned to the reference, or the reverse if the forward failed to align, were counted as a gene match. The genome-guided olfactory transcriptome data were analyzed with Cuffdiff (Trapnell et al. 2012) using settings --compatible-hits-norm, --frag-bias correct, and --multi-read-correct. Blood and muscle tissue-specific expression profiles were modeling, using the three individual fish-tissue samples, relative to the magnetic cell-type using same methods as for the magnetic and non-magnetic cell type.

3.3.5. Candidate gene annotations

Candidate genes identified as differentially expressed in the magnetic cell type (cDEGs) were annotated by extracting from each reference data set the full nucleotide sequence and matching those to National Center for Biotechnology Information (NCBI) non-redundant protein (download date June 15, 2014), UniProt:Swiss-Prot (download September 23, 2014), and European Bioinformatics Institute *Danio rerio* protein (downloaded October 14, 2014) databases using the Basic Local Alignment Search Tool (BLAST, S. F. Altschul et al. 1990). All query matches were considered "significant" if they met the criteria of minimum 65 percent identity, E-value threshold <1 x e -05, and a high-scoring segment pairs cut-off length of 50.

3.3.6 Candidate gene functional classification

The cDEGs were classified into functional categories through import of BLAST annotations into the Protein Analysis Through Evolutionary Relationships

(PANTHER) classification system (Thomas et al. 2003; Mi et al. 2004), version 9.0 which contains 7180 protein families divided into 52,768 functionally distinct protein sub-families. A statistical overrepresentation test was implemented to compare the number of observed genes expected for gene ontology functional classifications based on their prevalence in a model organism's genome. For this analysis, the whole zebrafish genome was the most suitable reference. Used as input data were the cDEGs identified by RNA-seq alignment to the Rainbow trout transcriptome (Berthelot et al. 2014) and also having homology to the Zebrafish EMBL genome. The ratio of observed to expected genes for each PANTHER category was calculated to guide the interpretation of statistical results. The PANTHER statistical enrichment test was performed with a sequential Bonferroni correction and a p-value < 0.05 for statistical significance of enrichment. Grouping of putative ortholog sequences for all cDEGs, using amino acid sequences from NCBI non-redundant protein database matches, was performed using OrthoMCL (Li, Stoeckert, and Roos 2003) version 2.0.

3.3.7 Eukaryotic model for the genetic basis of magnetoreception

A possible model for the genomic basis of magnetite formation in eukaryotes was guided by comparison of cDEGs to genes of prokaryotes that are known or possibly involved in magnetosome formation. The cDEGs were assigned to potential magnetotactic bacteria homologues on the basis of domain commonalities and/or functional similarity (e.g., magnetosome arrangement into chain structure; iron transport). Unique to a vertebrate receptor system, sensory transduction was placed

in a separate category. Some cDEGs were not assignable to categories. Category apportions were premised by: (1) match between non-redundant protein BLASTx domain annotations (all cDEGs) to those of known or predicted magnetite protein homologues (2) PANTHER GO Slim classification categories of cDEGs with BLASTx match to Uniprot:Swiss-Prot or Danio rerio genome (required for ontology classification), and (3) PANTHER statistical enrichment test results.

3.3.8 Are cDEG results consistent with the lateral gene transfer hypothesis?

Support for the hypothesis that magnetite crystal formation in eukaryotes results from an ancient lateral gene transfer event was evaluated through multiple lines of reasoning. The magnetosome gene island (Schübbe et al. 2003a) is a cluster of known and putative operons believed to encode a number of magnetosome-associated proteins required for membrane formation and magnetite biomineralization. Genomic islands are expected to contain genes linked to a metabolic activity or function and include flanking regions with signatures of mobile genetic elements, e.g., direct repeats, integrases, and transposases, that aid in lateral gene transfer (Bazylinski and Frankel 2004). Here, cDEGs were mapped to chromosomes using the Bertholet et al. genome to evaluate physical grouping of cDEGs, recognizing a priori that characteristics of magnetosome islands would by nature promote frequent rearrangement (Ullrich et al. 2005). Special focus was applied to identify possible functional analogues of mamAB operon magnetosome genes, the smallest genetic unit identified as functionally sufficient for magenetite biomineralization (Lohße et al. 2011). To evaluate if any of the individual cDEGs

might have evolved from a common ancestor to magnetotactic bacteria, protein sequences of cDEGs (obtained from match to NCBI non-redundant database) and magnetosome protein sequences (n = 357) extracted from UnipProtKB (search date 8/31/2014) were subjected to clustering analyses on the basis of shared, conserved domains using OrthoMCL (Li, Stoeckert, and Roos 2003). For a sub-set of genes, matches between conserved magnetosome gene motifs and DEGs were searched by translation of cDEG nucleotide sequences into all possible six reading frames as implemented in tBLASTn (S. Altschul et al. 1997), retaining matches with a percent identity threshold of 50%. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 (Tamura et al. 2013) or through built-in functions available through the Uniprot server. The expression of candidate DEGs compared to non-magnetic cell, blood and muscle transcriptomes provides complementary support for candidate receptor-specific expression.

3.4 RESULTS

3.4.1 Candidate gene annotations

Candidate magnetoreceptor cell transcriptome profiles resulted in a total of 1,006 cDEGs, n = 1674 isotigs (Table 3.3) identified by alignment of RNA-seq reads to the four reference data sets. Some identical cDEGs were in multiple data sets (data not shown). The numbers of cDEGs that were assigned gene annotations that matched PANTHER gene ontology classification was limited (Table 3.4). The gene enrichment analyses was most useful for identifying general gene function patterns,

which were consistent with expectations for genomic constituents of an iron-mineral crystal formation system.

3.4.2. Gene enrichment analysis

The PANTHER gene ontology functional enrichment for cDEGs identified by alignment to the trout transcriptome (Bertholet et al. 2014) and tested against the *Danio rerio* genome revealed expression of genes with diverse functions: assembly of multiprotein complexes, protein sorting, lipid and protein transport, anion transport, redox, and as expected for sensory cells, signatures of synaptic transmission and response to stimulus (Figure 3.1, Supplementary Appendix 3.1 lists gene ontology results for a sub-set of broad categories; Supplementary Appendixes 3.2, 3.3, and 3.4 list all gene enrichment test results for Biological Process, Cellular Function, and Molecular Process). Strongly enriched were genes in GO Slim categories RNA/DNA processing. Notably lacking were expression of genes attributed to sensory perception of chemical stimulus, as would be expected if functional olfactory receptor cells were a component of the sample. Also lacking were genes associated with immune defense, including apoptosis or pathways relating to MHC Class II, which supports the premise that the magnetic cell pellet was not (at least primarily) comprised of macrophages. PANTHER functional classifications for all cDEGs (data not shown) were largely consistent with the enrichment analysis.

3.4.3 Eukaryotic model for the genetic basis of magnetoreception

Contained in cDEGs were domains and motifs characteristic of homologues previously identified in magnetosome proteins: tetratricopeptide repeats (TPR), PDZ proteins, cation diffusion facilitator (CDF) transporters, and proteases (Nudelman and Zarivach 2014). Through domain annotation, motif analysis, and homology search we were able to determine which of the candidate genes are most similar to those known for involvement in prokaryotic iron crystal formation.

3.4.3.1 Trout genes with TPR motifs as possible homologues to MamA

Common to many proteins and characteristic of magnetosome protein MamA (= Mam22), tetra-trico-peptide (TPR) motifs are present in cell division-cycle gene products and in proteins involved in the regulation of RNA synthesis (Okuda, Denda, and Fukumori 1996). They are common in the salmon genome, with n = 351 TPR motifs annotated in the Bertholet et al. 2004 Rainbow trout genome. As one of the most abundant magnetosome-associated proteins (Grünberg et al. 2001), MamA contains five sequential TPR motifs that are believed to act as a multi-protein assembly site on the magnetosome (Zeytuni et al. 2011). It localizes to the magnetosome chain (Zeytuni et al. 2011) and possibly contributes to magnetosome chain structure stabilization (Yamamoto et al. 2010; Nudelman and Zarivach 2014). Repetitive TPR domains or signatures of motifs were indicated for 11 cDEGs (Table 3.5), with motifs identified as single-occurrences in an additional 62 cDEGs. The gene GSONMT00076779001, identified by alignment of the protein sequence with highest BLASTx match to cDEG salmon_1629, bore reasonable resemblance to MamA (Figure 3.2). Sequence homology was most conserved in alpha-helix regions

6 – 8, with an insertion present in the hinge between the NTD and CTD. Attempts to align the remaining genes with TPR repeats resulted in poor quality alignments. At least five of the 62 cDEGs having matches to the mamA N-terminal amino acid motif MSSKPSN-MLDEV (Grunberg et al. 2004) were annotated as (1) non-long terminal repeat retrotransposon and non-LTR retrovirus reverse transcriptase and/or (2) reverse transcriptase, both of which are attributable to mobile elements. For perspective, a BLASTP search for the MSSKPSN-MLDEV motif in the entire Trout genome returned 68 hits, while the number of motifs in cDEGs identified by alignment to that reference was 28. This means that of all available motif-containing genes in the mRNA reference, ~41.2% of those were detected as differentially expressed in the 215 cDEG sample (14% of the cDEGs). A connection between this mamA sequence motif and retrotransposons to the cDEGs could somehow in part explain the enriched categories RNA/DNA processing relative to the *Danio rerio* genome. The relationship between the mamA motif identified by Grunberg and colleagues to cDEGs with mobile elements warrants further exploration to because of the possibility of functional exaptation (Feschotte 2008) (Santangelo et al. 2007).

3.4.3.2 Trout genes with PDZ domains as possible homologues to MamE

The magnetosome protein MamE, predicted to fold as a putative serine protease based on modeled structure and containing two PDZ domains, is important for protein localization to the magnetosome membrane (Nudelman and Zarivach 2014). Two cDEGs with PDZ domain annotations were identified among those genes in the trout genome also containing PDZ domains (n = 346), however neither of these

contained also the serine proteases sequence (genome n = 7). While cDEG PDZ domain containing proteins and those of MamE magnetosomes yielded poor results, alignment the Trout protein sequences with PDZ domain and Trypsin-like serine protease yielded highly conserved segments (Figure 3.4).

In general, the magnetosome protein alignments indicated high variability within this gene class. The double cytochrome c signature motif CXXCH, which suggests the existence of an electron transport chain and possibly contributes to the process of biomineralization in magnetotactic bacteria, is present in mamE (plus mamP, mamT; Siponen et al. 2012). This double-motif was present in cDEG locus tag GSONMT00041848001 (CDQ83993), annotated as containing five zinc-finger domains, however PDZ domains were lacking. The genome-guided cDEG isotig XP_003558178 was annotated as a cytochrome C peroxidase with heme binding site.

3.4.3.3. Trout genes possibly homologues to MamY

In magnetotactic bacteria, a membrane deformation protein, MamY, is believed to bend and constrict the magnetosome membrane during magnetosome vesicle formation (Tanaka, Arakaki, and Matsunaga 2010). In eukaryotes, the bin/amphiphysin/Rvs (BAR) family members bend the steady state membrane architecture of organelles in intact cells (Suarez et al. 2014). The *O. mykiss* transcriptome (Berthelot et al. 2014) contains a total of 102 annotated BAR proteins (total n genes 46,271), of which two were identified by differential gene expression analysis to that reference. A third BAR protein (the same one) was detected as differentially expressed in both the SalmonDB (Di Genova et al. 2011) and whole

ground fish transcriptome (Fox et al. 2014) data sets. Based on the frequency BAR proteins in the Trout mRNA reference, 215 cDEGs would be expected to yield ~0.44 BAR genes.

3.4.3.4. Trout genes possibly homologues to MamU

The novel genome-guided reference gene XLOC_171228 contains a conserved diacylglycerol kinases catalytic domain (presumed) and accessory domain, which are members of the diacylglycerol kinase family of enzymes to which MamU is homologous. The DGK phosphorylates the second-messenger diacylglycerol (DAG). An unnamed protein product from the trout mRNA transcriptome DEG dataset contains four potentially interesting regions (of five total): a DAG region; EF-hand calcium binding motif, which is within a superfamily of calcium sensors and calcium signal modulators); RasGEF of which the superfamily function as molecular switches in signal transduction, cytoskeleton dynamics, and intracellular trafficking (from annotation); and a Protein kinase C region, which is a cysteine-rich zinc binding domain.

3.4.4 Are cDEG results consistent with the lateral gene transfer hypothesis?

3.4.4.1. Signatures of mobile genetic elements

One signature characteristic of a lateral gene transfer event is presence of mobile elements. As previously described, numerous cDEGs had matches to the mamA N-terminal amino acid motif MSSKPSN-MLDEV (Grunberg et al. 2004) that can be attributed to mobile elements retrotranspons. The bacterial MamR amino acid

protein, which is important for crystal number and size control, contains a predicted DNA-binding domain similar to the helix-turn-helix (HTH) 17 superfamily along with an exicisionase domain (Nudelman and Zarivach 2014). Of 22 cDEG sequences with HTH domains, six (27%) also contained homeobox domains. Examination of the self-blast database (top 5 hits, including de novo assembled genes) revealed that 8.7% (n = 88 of 1,006 cDEGs; 8.5%) of all cDEGs contained high-scoring segment pairs for HTH domains. For context, these HTH domains are present in only 1.07% of all gene sequences annotations (476 occurrences) in the Trout genome. The cDEG annotations revealed ten transposase protein products, of which nine contained HTH domains. Transposase genes account for 20% of the coding region in the MAI of Magnetospirillum gryphiswaldens (Schübbe et al. 2003b). Alignment of HTH containing cDEGs to MamR amino acid sequences revealed a fairly short conserved stretch (data not shown).

3.4.4.2 Do cDEGs cluster into operons, indicating lateral gene transfer?

Results for assessment of co-localization of cDEGs in operon-like gene clusters on chromosomes by chromosomal mapping are intriguing. The organization of the Rainbow trout genome includes 38 pairs of large duplicated regions distributed over 30 chromosomes (Berthelot et al. 2014). A total of 30 chromosomes are characterized as “known” and 30 more as “unknown” chromosomes. One “general unknown” sequence appears to have been designated as a catchall for un-assignable segments. Eight-seven of 215 cDEGs (Table 3. 3) identified by alignment to the Trout transcriptome could be assigned to chromosome segments. Under a random model,

one would expect these to allocate at a frequency of 1.52 genes per chromosome, however, chromosome unknown 12 and 2 each have 6 hits, Unk. 1, 11, 5 and 8 have 5 hits which is not consistent with the random model. It may be worthwhile to assess the BLAST homology of cDEGs to magentosome genes (in progress).

3.4.4.3. Orthologous gene clustering results

One differentially expressed gene (SalmonDB 25581 = GSONMT00018588001) was identified as a putative ortholog (category OG5_126876, Laminin) to a magnetosome associated protein, "mad21" (Swissprot id U5IGN4_9DELT, described by Lefèvre et al. 2013). This cDEG was annotated as containing a Laminin G domain. The mad21 magnetosome gene also has a Laminin domain, but the function of this gene is unknown. Laminin is a common component of extracellular and receptor proteins (Timpl et al. 2000; Nishimune, Sanes, and Carlson 2004), and in olfactory epithelium this gene is known to stimulate and guide neuronal cell migration (Calof and Lander 1991). An amino acid alignment between these two Laminan proteins (Figure 3.4) indicated fairly low percent identity, 21.43, as expected for distantly related eukaryotes and prokaryotes. No other trout genes clustered with magnetosome proteins.

The number of OrthoMCL groups that were identified from Chinook salmon cDEG annotations was 286. The magnetosome proteins clustered in 26 categories; along with two unassigned proteins, this is a close match to current estimates of magnetosome protein families, approximately 28 (Nudelman and Zarivach 2014).

Although de novo assembled genes were excluded from clustering analyses, an interesting pattern was observed whereby de novo assembled cDEGs (Trial 1, 3 gene 276_c2, with 12 isotigs; redundant to denovo assembly 2 gene 971) matched with surprisingly high sequence percent identity (61%) to a gene annotated as a hypothetical protein sequenced from *Magnetospirillum gryphiswaldense* (WP_024080208). Whether this result is meaningful needs further evaluation.

3.4.4.4. Summary of results for the genetic basis of magnetoreception in Eukaryotes

The genetic annotations and clustering results are consistent with having identified genes involved in magnetite formation in vertebrates. Gene-based, fluorescent RNA probes were developed and tested for hybridization to candidate magnetoreceptor cells visualized using the laser scanning mode of confocal reflectance (work performed in collaboration with Herve Cadiou). Unfortunately, background fluorescence precluded the ability to attribute hybridization signal to the candidate magnetoreceptor cell (data not shown). Additional bioinformatics analysis is needed for thorough assessment of cDEG results.

3.4.5 Notes on alignment of RNA-seq data to reference data sets

References selected for alignment of RNA-seq data varied by the number of assembled bases, numbers of contigs, length of N50, and assembly method (Table 3.2). The most complete reference was the *O. mykiss* genome, with 2.13 billion nucleotides, while and the Fox et al. O. mykiss transcriptome and Bertholet et al. (2014) were similar in numbers of assembled nucleotides. However the N50 of the

Bertholet et al. (2014) transcriptome was twice the length of the Fox et al. (2014) transcriptome. That RNA-seq reads aligned to both of these transcriptomes at simlar rates was surprising.

RNA-seq transcript match alignment rates were greatest for the genomic reference dataset (Table 3.2) and relatively low (~30%) for alignment to O. mykiss transcriptome and de novo references. Interestingly, difference in percent alignment rates between magnetic and non-magnetic samples was lowest for the partial steelhead genome reference, probably because genomes, unlike de novo constructed transcriptomes, are unbiased by transcript expression levels. Merits of the SALMON:db "unigene" data set are high quality annotation, long reads, and tie-in to a long history of gene expression research through the cGRASP consortium. The references assembly constructed from RNA isolated from whole-ground fish was prone to missing low-expressed transcripts because of low sequencing depth, yet it still provided a valuable reference. The percent read match to expression levels inherent to RNA-based data sets, however, the only available full-genome reference for Oncorhynchus is based on the closely related *O. mykiss*.

3.4.6. Data archive

Transcriptome profiles (magnetic cell, non-magnetic cell, whole olfactory rosettes, muscle and blood) were made publicly available by submitting them to the National Center for Biotechnology Information Genbank data repository short read archive (placeholder).

3.4 DISCUSSION

The novel list of candidate vertebrate magnetoreceptor genes provided by this study, the first data set of its kind, heralds a new era for understanding magnetic sensory transduction mechanisms. While further research is required to confirm how specific genes are involved in crystal formation, this rich data set provisions exploration of numerous future research avenues to assess magnetite crystal involvement in sensory transduction: in situ hybridization, single-cell qPCR, genetic engineering of green fluorescent probe labeled cells, and utilizing behavioral experiments of vertebrates with gene knock-outs to assess loss of magnetoreception ability. Our findings have direct implications for revealing how animals are able to sense magnetic fields and use geomagnetic information to orient themselves and navigate.

Although impossible to tell for sure, the magnetic cell samples sequenced in this study was possibly comprised of heterogeneous cell types, candidate magnetoreceptor cells and macrophages. Comparisons between candidate magnetoreceptor cell and blood transcriptomes could be applied as an exclusion criteria for macrophages, but one must consider potential for homology between these owing to possibility for conserved genetic pathways common to iron-transporting red blood cells or macrophages.

Obtaining direct evidence for magnetic signal transduction by magnetite-cells is hindered by cell scarcity (Walker et al. 1997a; Eder et al. 2012) and lack of characteristics visible through light microscopy. At present cell-identification methods are limited to reflectance mode of confocal microscopy (Green, Holloway,

and Walker 2001) or cell dissociation and a "magnetoscope" that spins the cells in the presence of a rotating magnetic field (Eder et al. 2012). Trangenic approaches could obviate these difficulties. One could genetically engineer green fluorescent proteins into a magnetosome vesicle to rapidly and efficiently identify physiologically responsive cells and test them for response to magnetic treatment. Alternatively, tests for lack of response to magnetic field treatment consequent to gene knock-outs could be conducted to assess a particular gene's involvement in magnetic reception, as was performed for the fly and cryptochrome gene putatively utilized as a component of a light-dependent receptor system (Gegear et al. 2010).

In salmon, many proteins are likely to be contribute to the formation of a magnetosome vesicle, some of which are likely to be magnetosome-specific as is the case for magnetotactic bacteria (Okamura, Takeyama, and Matsunaga 2001). The remaining challenge is to sift through the available list of candidate genes and identify those that are unique to vesicle formation. For this task, in situ hybridization or single-cell qPCR tests of single cells could be used to narrow the list. Moreover, qPCR could be used as a tool to interrogate tissues of other species to identify the general location of their magnetite-based magnetoreceptor.

Magnetite formation in fish is not limited to olfactory tissues. Crystals of magnetite in chain formation, similar to bacteria, have been isolated from connective tissue from the ethmoid region of the skull (Mann et al. 1988; Walker et al. 1988) and associated with lateral lines of Atlantic salmon and lateral line mandibular canal of eels. The crystals in the ethmoid region lacked association with sensory nerves which suggests they are not a component of a receptor system. Magnetization

measurements of the lateral line in Atlantic salmon (Moore, Freake, and Thomas 1990) and mandibular canals of the lateral line systems of eels (Moore and Riley 2009) indicates presence of magnetic material, but single domain crystals have never been isolated from specimens. This work will enable novel single cell genetic and fluorescence tools, and improved microscopy methods for ultra-structure analysis, to identify key components of the genetic, physiological, and structural basis of magnetite-containing cells of salmon and zebrafish. Regardless of whether the olfactory cells underpin magnetic sensory perception, these results will contribute to the understanding of magnetite biomineralization in fish and eukaryotes, providing a foundation for nanoparticle research.

3.6 ACKNOWLEDGEMENTS:

I thank the following people for access to space to raise fish or for providing grown hatchery specimens for experiments: Staff from the Oregon Department of Fish and Wildlife Nestucca Hatchery, Mary Arkoosh, Stacy Strickland, Rob Chitwood, David Noakes, Joseph O'Neil. I am indebted to Herve Cadiou who provided unpublished protocols, stimulating conversations, hosted my visit at University of Strasbourg, and taught me how to produce thin slices of fish tissue from ceramic vegetable peelers. Stephen Eder and Michael Winklehofer graciously hosted a visit during which we attempted to measure magnetic properites of cells. Teresa Sawyer and Michael Nesson provided invaluable instruction for Transmission Electron Scanning Microscopy. Mike served as informal adviser, guiding research

ideas and presenting solutions for the many challenges. Nick Sard and Dalton Hance provided valuable feedback on initial statistical analysis.

3.7 REFERENCES

Allendorf, Fred W., and Gary H. Thorgaard. 1984. "Tetraploidy and the Evolution of Salmonid Fishes." In *Evolutionary Genetics of Fishes*, edited by Bruce J. Turner, 1–53. Boston, MA: Springer US. http://link.springer.com/10.1007/978-1-4684-4652-4_1.

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 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Altschul, S., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25 (17): 3389–3402. doi:10.1093/nar/25.17.3389.

Balkwill, D L, D Maratea, and R P Blakemore. 1980. "Ultrastructure of a Magnetotactic Spirillum." *Journal of Bacteriology* 141 (3): 1399–1408.

Bazylinski, Dennis A., and Richard B. Frankel. 2004. "Magnetosome Formation in Prokaryotes." *Nature Reviews Microbiology* 2 (3): 217–30. doi:10.1038/nrmicro842.

Beason, R C, and P Semm. 1987. "Magnetic Responses of the Trigeminal Nerve System of the Bobolink (Dolichonyx Oryzivorus)." *Neuroscience Letters* 80 (2): 229–34.

Berthelot, Camille, Frédéric Brunet, Domitille Chalopin, Amélie Juanchich, Maria Bernard, Benjamin Noël, Pascal Bento, et al. 2014. "The Rainbow Trout Genome Provides Novel Insights into Evolution after Whole-Genome Duplication in Vertebrates." *Nature Communications* 5 (April). doi:10.1038/ncomms4657.

Blakemore, Richard P., and Richard B. Frankel. 1981. "Magnetic Navigation in Bacteria." *Scientific American* 245 (6): 58–65. doi:10.1038/scientificamerican1281-58.

Blankenberg, Daniel, Assaf Gordon, Gregory Von Kuster, Nathan Coraor, Galaxy Team, James Taylor, and Anton Nekrutenko. 2010. "Manipulation of FASTQ Data with Galaxy." *Bioinformatics (Oxford, England)* 26 (14): 1783–85. doi:10.1093/bioinformatics/btq281.

Butler, Robert F., and Subir K. Banerjee. 1975. "Theoretical Single-Domain Grain Size Range in Magnetite and Titanomagnetite." *Journal of Geophysical Research* 80 (29): 4049–58. doi:10.1029/JB080i029p04049.

Calof, A. L., and A. D. Lander. 1991. "Relationship between Neuronal Migration and Cell-Substratum Adhesion: Laminin and Merosin Promote Olfactory

Neuronal Migration but Are Anti-Adhesive." *The Journal of Cell Biology* 115 (3): 779–94.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14 (6): 1188–90. doi:10.1101/gr.849004.

Davidson, William S., Ben F. Koop, Steven J. M. Jones, Patricia Iturra, Rodrigo Vidal, Alejandro Maass, Inge Jonassen, Sigbjorn Lien, and Stig W. Omholt. 2010. "Sequencing the Genome of the Atlantic Salmon (Salmo Salar)." *Genome Biology* 11 (9): 403. doi:10.1186/gb-2010-11-9-403.

De Wit, Pierre, Melissa H. Pespeni, Jason T. Ladner, Daniel J. Barshis, François Seneca, Hannah Jaris, Nina Overgaard Therkildsen, Megan Morikawa, and Stephen R. Palumbi. 2012. "The Simple Fool's Guide to Population Genomics via RNA-Seq: An Introduction to High-Throughput Sequencing Data Analysis." *Molecular Ecology Resources* 12 (6): 1058–67. doi:10.1111/1755-0998.12003.

Diebel, C E, R Proksch, C R Green, P Neilson, and M M Walker. 2000. "Magnetite Defines a Vertebrate Magnetoreceptor." *Nature* 406 (6793): 299–302. doi:10.1038/35018561.

Di Genova, A., A. Aravena, L. Zapata, M. Gonzalez, A. Maass, and P. Iturra. 2011. "SalmonDB: A Bioinformatics Resource for Salmo Salar and Oncorhynchus Mykiss." *Database* 2011 (0): bar050–bar050. doi:10.1093/database/bar050.

Di, Yanming, Daniel W Schafer, Jason S Cumbie, and Jeff H Chang. 2011. "The NBP Negative Binomial Model for Assessing Differential Gene Expression from RNA-Seq." *Statistical Applications in Genetics and Molecular Biology* 10 (1): 1–28. doi:10.2202/1544-6115.1637.

Eder, S. H. K., H. Cadiou, A. Muhamad, P. A. McNaughton, J. L. Kirschvink, and M. Winklhofer. 2012. "Magnetic Characterization of Isolated Candidate Vertebrate Magnetoreceptor Cells." *Proceedings of the National Academy of Sciences* 109 (30): 12022–27. doi:10.1073/pnas.1205653109.

Feschotte, Cédric. 2008. "Transposable Elements and the Evolution of Regulatory Networks." *Nature Reviews. Genetics* 9 (5): 397–405. doi:10.1038/nrg2337.

Fleissner, Gerta, Elke Holtkamp-Rötzler, Marianne Hanzlik, Michael Winklhofer, Günther Fleissner, Nikolai Petersen, and Wolfgang Wiltschko. 2003. "Ultrastructural Analysis of a Putative Magnetoreceptor in the Beak of Homing Pigeons." *The Journal of Comparative Neurology* 458 (4): 350–60. doi:10.1002/cne.10579.

Fox, Samuel E., Mark R. Christie, Melanie Marine, Henry D. Priest, Todd C. Mockler, and Michael S. Blouin. 2014. "Sequencing and Characterization of the Anadromous Steelhead (Oncorhynchus Mykiss) Transcriptome." *Marine Genomics* 15 (June): 13–15. doi:10.1016/j.margen.2013.12.001.

Frankel, R. B., R. P. Blakemore, and R. S. Wolfe. 1979. "Magnetite in Freshwater Magnetotactic Bacteria." *Science* 203 (4387): 1355–56. doi:10.1126/science.203.4387.1355.

Gegear, Robert J., Lauren E. Foley, Amy Casselman, and Steven M. Reppert. 2010. "Animal Cryptochromes Mediate Magnetoreception by an Unconventional

Photochemical Mechanism." *Nature* 463 (7282): 804–7. doi:10.1038/nature08719.

Gould, J. L., J. L. Kirschvink, and K. S. Deffeyes. 1978. "Bees Have Magnetic Remanence." *Science* 201 (4360): 1026–28. doi:10.1126/science.201.4360.1026.

Grabherr, Manfred G, Brian J Haas, Moran Yassour, Joshua Z Levin, Dawn A Thompson, Ido Amit, Xian Adiconis, et al. 2011. "Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome." *Nature Biotechnology* 29 (7): 644–52. doi:10.1038/nbt.1883.

Green, C. R., H. Holloway, and M. M. Walker. 2001. "Detection of Submicroscopic Magnetite Particles Using Reflectance Mode Confocal Laser Scanning Microscopy." *Cell Biology International* 25 (10): 985–90. doi:10.1006/cbir.2001.0773.

Grunberg, K., E.-C. Muller, A. Otto, R. Reszka, D. Linder, M. Kube, R. Reinhardt, and D. Schuler. 2004. "Biochemical and Proteomic Analysis of the Magnetosome Membrane in Magnetospirillum Gryphiswaldense." *Applied and Environmental Microbiology* 70 (2): 1040–50. doi:10.1128/AEM.70.2.1040-1050.2004.

Grünberg, K., C. Wawer, B. M. Tebo, and D. Schüler. 2001. "A Large Gene Cluster Encoding Several Magnetosome Proteins Is Conserved in Different Species of Magnetotactic Bacteria." *Applied and Environmental Microbiology* 67 (10): 4573–82.

Heyers, D., M. Zapka, M. Hoffmeister, J. M. Wild, and H. Mouritsen. 2010. "Magnetic Field Changes Activate the Trigeminal Brainstem Complex in a Migratory Bird." *Proceedings of the National Academy of Sciences* 107 (20): 9394–99. doi:10.1073/pnas.0907068107.

Holland, Richard A., Joseph L. Kirschvink, Thomas G. Doak, and Martin Wikelski. 2008. "Bats Use Magnetite to Detect the Earth's Magnetic Field." Edited by Sarah Frances Brosnan. *PLoS ONE* 3 (2): e1676. doi:10.1371/journal.pone.0001676.

Irwin, William P, and Kenneth J Lohmann. 2005. "Disruption of Magnetic Orientation in Hatchling Loggerhead Sea Turtles by Pulsed Magnetic Fields." *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology* 191 (5): 475–80. doi:10.1007/s00359-005-0609-9.

Johnsen, Sönke, and Kenneth J. Lohmann. 2008. "Magnetoreception in Animals Feature Article." *Physics Today* 61 (3): 29–35. doi:10.1063/1.2897947.

Kirschvink, J. L., and J. L. Gould. 1981. "Biogenic Magnetite as a Basis for Magnetic Field Detection in Animals." *BioSystems* 13: 181–201.

Kirschvink, J L, and J. W. Hagadorn. 2000. "A Grand Unified Theory of Biomineralization." In *The Biomineralization of Nano- and Micro-Structures*, Bäuerlein E., 139–50. Weinheim, Germany: Wiley-VCH.

Kirschvink, J. L., M. Winklhofer, and M. M. Walker. 2010. "Biophysics of Magnetic Orientation: Strengthening the Interface between Theory and Experimental Design." *Journal of The Royal Society Interface* 7 (Suppl_2): S179–91. doi:10.1098/rsif.2009.0491.focus.

Kishkinev, Dmitry, Nikita Chernetsov, Dominik Heyers, and Henrik Mouritsen. 2013. “Migratory Reed Warblers Need Intact Trigeminal Nerves to Correct for a 1,000 Km Eastward Displacement.” Edited by M. Fabiana Kubke. *PLoS ONE* 8 (6): e65847. doi:10.1371/journal.pone.0065847.

Leask, M. J. M. 1977. “A Physicochemical Mechanism for Magnetic Field Detection by Migratory Birds and Homing Pigeons.” *Nature* 267 (5607): 144–45. doi:10.1038/267144a0.

Lefèvre, Christopher T., Denis Trubitsyn, Fernanda Abreu, Sebastian Kolinko, Christian Jogler, Luiz Gonzaga Paula de Almeida, Ana Tereza R. de Vasconcelos, et al. 2013. “Comparative Genomic Analysis of Magnetotactic Bacteria from the *Deltaproteobacteria* Provides New Insights into Magnetite and Greigite Magnetosome Genes Required for Magnetotaxis: New Insights in Genes Required for Magnetotaxis.” *Environmental Microbiology*, May, n/a – n/a. doi:10.1111/1462-2920.12128.

Li, L., C. J. Jr. Stoeckert, and D. S. Roos. 2003. “OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes.” *Genome Research* 13 (9): 2178–89. doi:10.1101/gr.1224503.

Lohmann, Kenneth J., and Catherine M. Fittinghoff Lohmann. 1993. “A Light-Independent Magnetic Compass in the Leatherback Sea Turtle.” *Biological Bulletin* 185 (1): 149–51. doi:10.2307/1542138.

Lohmann, K. J., N. F. Putman, and C. M. F. Lohmann. 2008. “Geomagnetic Imprinting: A Unifying Hypothesis of Long-Distance Natal Homing in Salmon and Sea Turtles.” *Proceedings of the National Academy of Sciences* 105 (49): 19096–101. doi:10.1073/pnas.0801859105.

Lohße, Anna, Susanne Ullrich, Emanuel Katzmann, Sarah Borg, Gerd Wanner, Michael Richter, Birgit Voigt, Thomas Schweder, and Dirk Schüler. 2011. “Functional Analysis of the Magnetosome Island in Magnetospirillum Gryphiswaldense: The mamAB Operon Is Sufficient for Magnetite Biomineralization.” *PLoS ONE* 6 (10): e25561. doi:10.1371/journal.pone.0025561.

Lowenstam, Heinz A. 1962. “Magnetite in Denticle Capping in Recent Chitons (Polyplacophora).” *Geological Society of America Bulletin* 73 (4): 435–38. doi:10.1130/0016-7606(1962)73[435:MIDCIR]2.0.CO;2.

Mann, S, N H Sparks, M M Walker, and J L Kirschvink. 1988. “Ultrastructure, Morphology and Organization of Biogenic Magnetite from Sockeye Salmon, Oncorhynchus Nerka: Implications for Magnetoreception.” *The Journal of Experimental Biology* 140 (November): 35–49.

Meyer, A., and M. Schartl. 1999. “Gene and Genome Duplications in Vertebrates: The One-to-Four (-to-Eight in Fish) Rule and the Evolution of Novel Gene Functions.” *Current Opinion in Cell Biology* 11 (6): 699–704.

Mi, H., B. Lazareva-Ulitsky, R. Loo, A. Kejariwal, J. Vandergriff, S. Rabkin, N. Guo, et al. 2004. “The PANTHER Database of Protein Families, Subfamilies, Functions and Pathways.” *Nucleic Acids Research* 33 (Database issue): D284–88. doi:10.1093/nar/gki078.

Moore, A., S. M. Freake, and I. M. Thomas. 1990. “Magnetic Particles in the Lateral Line of the Atlantic Salmon (Salmo Salar L.).” *Philosophical Transactions of*

the Royal Society B: Biological Sciences 329 (1252): 11–15. doi:10.1098/rstb.1990.0145.

Moore, A., and W. D. Riley. 2009. "Magnetic Particles Associated with the Lateral Line of the European Eel *Anguilla Anguilla*." *Journal of Fish Biology* 74 (7): 1629–34. doi:10.1111/j.1095-8649.2009.02197.x.

Mora, Cordula V., Michael Davison, J. Martin Wild, and Michael M. Walker. 2004. "Magnetoreception and Its Trigeminal Mediation in the Homing Pigeon." *Nature* 432 (7016): 508–11. doi:10.1038/nature03077.

Murat, D., A. Quinlan, H. Vali, and A. Komeili. 2010. "Comprehensive Genetic Dissection of the Magnetosome Gene Island Reveals the Step-Wise Assembly of a Prokaryotic Organelle." *Proceedings of the National Academy of Sciences* 107 (12): 5593–98. doi:10.1073/pnas.0914439107.

Nishimune, Hiroshi, Joshua R. Sanes, and Steven S. Carlson. 2004. "A Synaptic Laminin-Calcium Channel Interaction Organizes Active Zones in Motor Nerve Terminals." *Nature* 432 (7017): 580–87. doi:10.1038/nature03112.

Nudelman, Hila, and Raz Zarivach. 2014. "Structure Prediction of Magnetosome-Associated Proteins." *Frontiers in Microbiology* 5. doi:10.3389/fmicb.2014.00009.

Okamura, Y., H. Takeyama, and T. Matsunaga. 2001. "A Magnetosome-Specific GTPase from the Magnetic Bacterium Magnetospirillum Magneticum AMB-1." *The Journal of Biological Chemistry* 276 (51): 48183–88. doi:10.1074/jbc.M106408200.

Okuda, Y., K. Denda, and Y. Fukumori. 1996. "Cloning and Sequencing of a Gene Encoding a New Member of the Tetratricopeptide Protein Family from Magnetosomes of Magnetospirillum Magnetotacticum." *Gene* 171 (1): 99–102.

Putman, Nathan F., Kenneth J. Lohmann, Emily M. Putman, Thomas P. Quinn, A. Peter Klimley, and David L.G. Noakes. 2013. "Evidence for Geomagnetic Imprinting as a Homing Mechanism in Pacific Salmon." *Current Biology* 23 (4): 312–16. doi:10.1016/j.cub.2012.12.041.

Putman, Nathan F., Michelle M. Scanlan, Eric J. Billman, Joseph P. O'Neil, Ryan B. Couture, Thomas P. Quinn, Kenneth J. Lohmann, and David L.G. Noakes. 2014. "An Inherited Magnetic Map Guides Ocean Navigation in Juvenile Pacific Salmon." *Current Biology* 24 (4): 446–50. doi:10.1016/j.cub.2014.01.017.

Qian, Xi, Yi Ba, Qianfeng Zhuang, and Guofang Zhong. 2014. "RNA-Seq Technology and Its Application in Fish Transcriptomics." *OMICS: A Journal of Integrative Biology* 18 (2): 98–110. doi:10.1089/omi.2013.0110.

Quinn, Nicole L., Natasha Levenkova, William Chow, Pascal Bouffard, Keith A. Boroevich, James R. Knight, Thomas P. Jarvie, et al. 2008. "Assessing the Feasibility of GS FLX Pyrosequencing for Sequencing the Atlantic Salmon Genome." *BMC Genomics* 9: 404. doi:10.1186/1471-2164-9-404.

Quinn, Thomas P. 1980. "Evidence for Celestial and Magnetic Compass Orientation in Lake Migrating Sockeye Salmon Fry." *Journal of Comparative Physiology□? A* 137 (3): 243–48. doi:10.1007/BF00657119.

Richter, Michael, Michael Kube, Dennis A Bazylinski, Thierry Lombardot, Frank Oliver Glöckner, Richard Reinhardt, and Dirk Schüler. 2007. "Comparative Genome Analysis of Four Magnetotactic Bacteria Reveals a Complex Set of Group-Specific Genes Implicated in Magnetosome Biomineralization and Function." *Journal of Bacteriology* 189 (13): 4899–4910. doi:10.1128/JB.00119-07.

Ritz, T., S. Adem, and K. Schulten. 2000. "A Model for Photoreceptor-Based Magnetoreception in Birds." *Biophysical Journal* 78 (2): 707–18. doi:10.1016/S0006-3495(00)76629-X.

Santangelo, Andrea M., Flávio S. J. de Souza, Lucía F. Franchini, Viviana F. Bumaschny, Malcolm J. Low, and Marcelo Rubinstein. 2007. "Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene." *PLoS Genetics* 3 (10): e166. doi:10.1371/journal.pgen.0030166.

Schlegel, Peter A. 2007. "Spontaneous Preferences for Magnetic Compass Direction in the American Red-Spotted Newt, Notophthalmus Viridescens (Salamandridae, Urodela)." *Journal of Ethology* 25 (2): 177–84. doi:10.1007/s10164-006-0016-x.

Schneider, T. D., and R. M. Stephens. 1990. "Sequence Logos: A New Way to Display Consensus Sequences." *Nucleic Acids Research* 18 (20): 6097–6100.

Schübbe, Sabrina, Michael Kube, André Scheffel, Cathrin Wawer, Udo Heyen, Anke Meyerdierks, Mohamed H. Madkour, Frank Mayer, Richard Reinhardt, and Dirk Schüler. 2003a. "Characterization of a Spontaneous Nonmagnetic Mutant of Magnetospirillum Gryphiswaldense Reveals a Large Deletion Comprising a Putative Magnetosome Island." *Journal of Bacteriology* 185 (19): 5779–90.

———. 2003b. "Characterization of a Spontaneous Nonmagnetic Mutant of Magnetospirillum Gryphiswaldense Reveals a Large Deletion Comprising a Putative Magnetosome Island." *Journal of Bacteriology* 185 (19): 5779–90.

Schüler, D. 2008. "Genetics and Cell Biology of Magnetosome Formation in Magnetotactic Bacteria." *FEMS Microbiology Reviews* 32 (4): 654–72. doi:10.1111/j.1574-6976.2008.00116.x.

Siponen, Marina I., Géraldine Adryanczyk, Nicolas Ginet, Pascal Arnoux, and David Pignol. 2012. "Magnetochrome: A c-Type Cytochrome Domain Specific to Magnetotatic Bacteria." *Biochemical Society Transactions* 40 (6): 1319–23. doi:10.1042/BST20120104.

Stapput, Katrin, Peter Thalau, Roswitha Wiltschko, and Wolfgang Wiltschko. 2008. "Orientation of Birds in Total Darkness." *Current Biology* 18 (8): 602–6. doi:10.1016/j.cub.2008.03.046.

Suarez, Allison, Tasuku Ueno, Robert Huebner, J. Michael McCaffery, and Takanari Inoue. 2014. "Bin/Amphiphysin/Rvs (BAR) Family Members Bend Membranes in Cells." *Scientific Reports* 4 (May). doi:10.1038/srep04693.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." *Molecular Biology and Evolution* 30 (12): 2725–29. doi:10.1093/molbev/mst197.

Tanaka, Masayoshi, Atsushi Arakaki, and Tadashi Matsunaga. 2010. “Identification and Functional Characterization of Liposome Tubulation Protein from Magnetotactic Bacteria.” *Molecular Microbiology* 76 (2): 480–88. doi:10.1111/j.1365-2958.2010.07117.x.

Thomas, P. D., M. J. Campbell, A. Kejariwal, Huaiyu Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, and A. Narechania. 2003. “PANTHER: A Library of Protein Families and Subfamilies Indexed by Function.” *Genome Research* 13 (9): 2129–41. doi:10.1101/gr.772403.

Timpl, R., D. Tisi, J. F. Talts, Z. Andac, T. Sasaki, and E. Hohenester. 2000. “Structure and Function of Laminin LG Modules.” *Matrix Biology: Journal of the International Society for Matrix Biology* 19 (4): 309–17.

Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. 2012. “Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks.” *Nature Protocols* 7 (3): 562–78. doi:10.1038/nprot.2012.016.

Treiber, Christoph Daniel, Marion Claudia Salzer, Johannes Riegler, Nathaniel Edelman, Cristina Sugar, Martin Breuss, Paul Pichler, et al. 2012. “Clusters of Iron-Rich Cells in the Upper Beak of Pigeons Are Macrophages Not Magnetosensitive Neurons.” *Nature*, April. doi:10.1038/nature11046.

Ullrich, Susanne, Michael Kube, Sabrina Schübbe, Richard Reinhardt, and Dirk Schüler. 2005. “A Hypervariable 130-Kilobase Genomic Region of Magnetospirillum Gryphiswaldense Comprises a Magnetosome Island Which Undergoes Frequent Rearrangements during Stationary Growth.” *Journal of Bacteriology* 187 (21): 7176–84. doi:10.1128/JB.187.21.7176-7184.2005.

Walcott, C, J. Gould, and J. Kirschvink. 1979. “Pigeons Have Magnets.” *Science* 205 (4410): 1027–29. doi:10.1126/science.472725.

Walker, M M, C E Diebel, C V Haugh, P M Pankhurst, J C Montgomery, and C R Green. 1997a. “Structure and Function of the Vertebrate Magnetic Sense.” *Nature*, no. 390: 371–76.

———. 1997b. “Structure and Function of the Vertebrate Magnetic Sense.” *Nature* 390 (6658): 371–76.

Walker, M. M., T. P. Quinn, J. L. Kirschvink, and C. Groot. 1988. “Production of Single-Domain Magnetite throughout Life by Sockeye Salmon, Oncorhynchus Nerka.” *The Journal of Experimental Biology* 140 (November): 51–63.

Wang, Shou, Tomasz Furmanek, Harald Kryvi, Christel Krossøy, Geir K Totland, Sindre Grotmol, and Anna Wargelius. 2014. “Transcriptome Sequencing of Atlantic Salmon (Salmo Salar L.) Notochord prior to Development of the Vertebrae Provides Clues to Regulation of Positional Fate, Chordoblast Lineage and Mineralisation.” *BMC Genomics* 15 (1): 141. doi:10.1186/1471-2164-15-141.

Wei, Jiandong. 2009. “Nanoscale Analysis of Natural and Artificial Magnetic Objects: Particles, Thin Films and Recording Heads.” Saarbrücken, Germany: Universität des Saarlandes. http://d-nb.info/1001100956/34.

Wiltschko, Wolfgang, and Roswitha Wiltschko. 2007. “Magnetoreception in Birds: Two Receptors for Two Different Tasks.” *Journal of Ornithology* 148 (S1): 61–76. doi:10.1007/s10336-007-0233-2.

Yamamoto, D., A. Taoka, T. Uchihashi, H. Sasaki, H. Watanabe, T. Ando, and Y. Fukumori. 2010. “Visualization and Structural Analysis of the Bacterial Magnetic Organelle Magnetosome Using Atomic Force Microscopy.” *Proceedings of the National Academy of Sciences* 107 (20): 9382–87. doi:10.1073/pnas.1001870107.

Zeytuni, N., E. Ozyamak, K. Ben-Harush, G. Davidov, M. Levin, Y. Gat, T. Moyal, A. Brik, A. Komeili, and R. Zarivach. 2011. “Self-Recognition Mechanism of MamA, a Magnetosome-Associated TPR-Containing Protein, Promotes Complex Assembly.” *Proceedings of the National Academy of Sciences* 108 (33): E480–87. doi:10.1073/pnas.1103367108.

Table 3.1. Attributes of "-omic" reference data sets applied to study transcriptome profiles of Chinook salmon olfactory epithelium candidate magnetoreceptor cells, non-magnetic olfactory epithelium cells, blood, and muscle tissues.

Reference	Assembly method	Citation	n contigs (isotigs)	n assembled bases, using longest isoform per contig	N50	Min contig length	Max contig length	n contigs with hits
Oncorynchus mykiss Transcriptome, whole ground fishes	Short reads/Reference guided	Fox et al. (2014)	86,402	77,750,399	1043	80	7,928	79,059
O. mykiss transcriptome	454, short reads, BAC/ Reference guided	Bertholet et al (2014)	46,585	76,202,098	2156	60	64,013	46,271
O. mykiss genome	454, Illumina, BAC/ Reference guided	Bertholet et al (2014)	58	2,134,686,837	1110957068	134600	41,402,622	n/a
O. tshawytcha reference-guided transcriptome (assembled with n 568,176,843 reads)	Short reads/reference guided *to O. mykiss* genome	This study	186,899 (186,732)	52,294,560	390	23	94,150	110,643
Trinity de novo transcriptome (1059, 1079i5i6, 247mi reads)	Short reads/Trinity	This study	46,909 (54,472)	13,835,619 (16,743,860)	296 (538)	201	5,177	79,059
SalmonidDB:Unigenes with min contig length 736	Sanger sequencing	___	42,604	43,352,072	988	736	7,055	40,556
Trinity de novo w/ reads that failed to align to OM2Umin737	Short reads/Trinity	This study	273,972 (277,747)	76,168,415 (77,463,655)	267 (268)	201	4,174	243,654

Table 3.2. Numbers of Illumina RNA-seq reads and alignment rates of blood, muscle, olfactory rosette (OR), candidate magnetoreceptors (mag) and non-magnetic cells (non-mag) of Chinook salmon (Oncorhynchus tshawytscha) aligned to two *O. mykiss* reference data sets (from Bertholet et al. 2014). Each experiment was performed in triplicate

Sample Type	Total reads	Single-end or orphan reads (separated from mate)	n paired-end reads (total n is x 2)	N reads aligned to transcriptome	% reads aligned to O. mykiss transcriptome	N reads aligned to O. mykiss genome	% reads aligned to O. mykiss genome	Difference between Genome - Transcriptome (%)
Bood 1	15,572,090	15,572,090		11,336,384	72.80%	12776446	82.05%	9.25%
Blood 2	16,151,385	16,151,385		11,093,075	68.68%	13199197	81.72%	13.04%
Blood 4	16,038,258	16,038,258		11,642,839	72.59%	13215825	82.40%	9.81%
Muscle 1	14,899,160	14,899,160		12,146,583	81.53%	12077427	81.06%	-0.46%
Muscle 2	16,081,043	16,081,043		12,792,810	79.55%	13084691	81.37%	1.82%
Muscle 4	14,871,474	14,871,474		11,908,749	80.08%	11987305	80.61%	0.53%
OR 1	15,171,725	15,171,725		8,409,887	55.43%	12102035	79.77%	24.34%
OR 2	16,284,939	16,284,939		8,716,113	53.52%	13025640	79.99%	26.46%
OR 4	15,289,567	15,289,567		7,240,020	47.35%	12209806	79.86%	32.50%
OR, deep sequence	87,691,817	21,515,109	33,088,354	47,921,496	54.65%	66483573	75.82%	21.17%
Trial 1 non-mag	61,984,437	33,114,699	14,434,869	24,000,507	38.72%	41927292	67.64%	28.92%
Trial 1 mag	31,280,838	17,193,252	7,043,793	4,160,004	13.30%	16459012	52.62%	39.32%
Trial 2 non-mag	124,431,944	36,543,416	43,944,264	25,740,231	20.69%	90379246	72.63%	51.95%
Trial 2 mag	61,640,941	20,384,293	20,628,324	11,583,048	18.79%	41062982	66.62%	47.83%
Trial 3 non-mag	72,304,591	20,369,593	25,967,499	16,816,600	23.26%	54861597	75.88%	52.62%
Trial 3 mag	82,096,044	23,104,774	29,495,635	15,137,653	18.44%	57844432	70.46%	52.02%
Sum	661,790,253			240,645,999				

Table 3.3. Salmonid reference data sets (also see Table 3.1 and text) were used for alignment of RNA-seq reads from magnetic and non-magnetic transcriptomes of Chinook salmon. The number of differentially expressed genes is shown for magnetic genes (cDEGs) and isotigs. In the next column, the number of cDEGs that could be annotated by BLAST homology search against National Center for Biotechnology Information non-redundant protein (NR), Uniprot:Swiss-Prot (Swissprot), and European Bioinformatics Institute *Danio rerio* (Danio) databases are listed. The number of differentially expressed genes for the non-magnetic cell type is presented in the last column.

Reference for Gene Expression Evaluation	Reference Citation	N cDEGs (isotigs)	N NR (isotigs)	N Swissprot (isotigs)	N Danio	N non-magnetic DEGs
Trout transcriptome, reference guided	Fox et al. (2014)	342	152	35	36	173
Trout transcriptome, based on genome	Bertholet et al. (2014)	215	210	80	95	558
O. tshawytscha genome-guided (to Bertholet et al. (2014)	This study	52	3	0	0	125
SalmonDB	Di Genova et al. (2011)	243	161	39	55	111
O. tshawytscha, de novo, Trials 1 and 3	This study	84 (747)	54 (414)	37 (94)	7 (9)	65 (85)
O. tshawytscha, de novo, no match	This study	70 (75)	41 (42)	30 (30)	8 (8)	92 (106)
Totals		1006 (1674)	621 (982)	221 (278)	201 (203)	1124 (1158)

* some redundant matches collapse into categories; includes isotigs

Table 3.4. Numbers of differentially expressed Chinook salmon (*O. tshawytscha*) candidate magnetoreceptor genes (cDEGs) with BLAST identifiers that could be imported into the Protein Analysis Through Evolutionary Relationships (PANTHER) database (Thomas et al. 2003, Mi et al. NAR 2005).

Reference Description	Reference Citation	Magnetic DEG (isotigs)	NCBI NR protein (gene + isotig)	Uniprot:Swiss-Prot match (gene + isotig)	EMBL Danio rerio matches (gene + isotig)	n Non-magnetic DEG (isotigs)
Trout transcriptome, reference guided	Fox et al. (2014)	342	27	26	28	173
Trout transcriptome, based on genome	Bertholet et al. (2014)	215	54	70	67	558
O. tshawytscha genome-guided (to Bertholet et al. (2014)	This study	52	16	0	0	125
SalmonDB	Di Genova et al. (2011)	243	32	34	39	111
O. tshawytscha, de novo, Trials 1 and 3	This study	84 (747)	46	28	65	65 (85)
O. tshawytscha, de novo, no match	This study	70 (75)	19	10	92	92 (106)
Totals		1006 (1674)	98	134	126	1124 (1158)

Table 3.5. Genes with tetratricopeptide repeat (TPR) and helix-turn-helix (HTH) domains identified through mRNA gene annotations and motif search using BLASTx against the NCBI Ref-Seq non-redundant gene archive (percent identity threshold 50%).

General domain or specific region	Identified by BLASTx or motif search	Reference	cDEG	Locus	N TPR repeats
TPR	BLASTx	BLAST-NR	Salmon_68079 (=GSONMT00076779001)	CDQ57999	7
TPR	BLASTx	BLAST-NR	GSONMT00050724001	CDQ83139	6
TPR	msskpsnXmldev	Grunberg 2004	salmon_1629		2
TPR	xskpsdildevtlyahygl	Okuda et al. 1996	GSONMT00040386001		3
TPR	xskpsdildevtlyahygl	Okuda et al. 1996	salmon_18517		2
helix 1	evtlyahyglsvak	Zeytuni et al. 2011	GSONMT00001009001		2
C-terminal binding site H11	elalvpr	Zeytuni et al. 2011	GSONMT00026599001		2
helix 2	gmnmvdafraafsv	Zeytuni et al. 2011	OM2U39250	NP_001118010.1	2
C-terminal binding site H11	elalvpr	Zeytuni et al. 2011	salmon_22961		2
TPR	xskpsdildevtlyahygl	Okuda et al. 1996	OM2U043487		1
helix 1	evtlyahyglsvak	Zeytuni et al. 2011	"		1
TPR	msskpsnXmldev	Grunberg 2004	GSONMT00067511001		1
helix 1	evtlyahyglsvak	Zeytuni et al. 2011	"		1

* 62 genes with single TPR domains were identified by motif search

Figure 3.1. Statistical enrichment analysis of Chinook salmon candidate magnetoreceptor genes (cDEGs) implemented in PANTHER (Thomas et al. 2003). The number of gene ontology terms in the cDEG sample were compared to the number of occurrences in the Zebrafish genome. Enriched genes were grouped by general categories: (a) localization and transport; (b) protein biogenesis, organization, folding, and localization; (c) sensory perception; (d) miscellaneous functions, including adhesion or macrophage signatures, and (e) RNA/DNA processing.

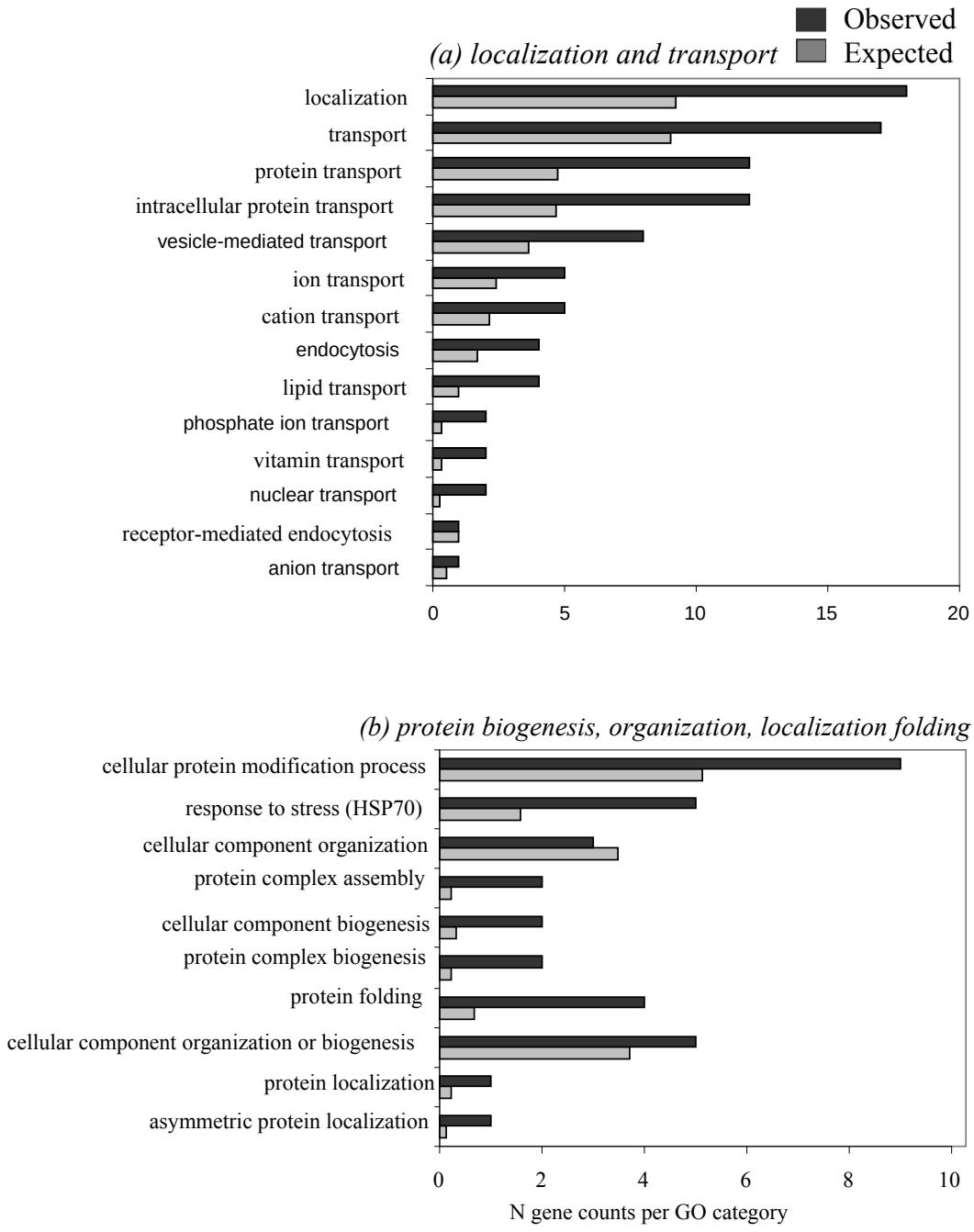


Figure 3.1, continued

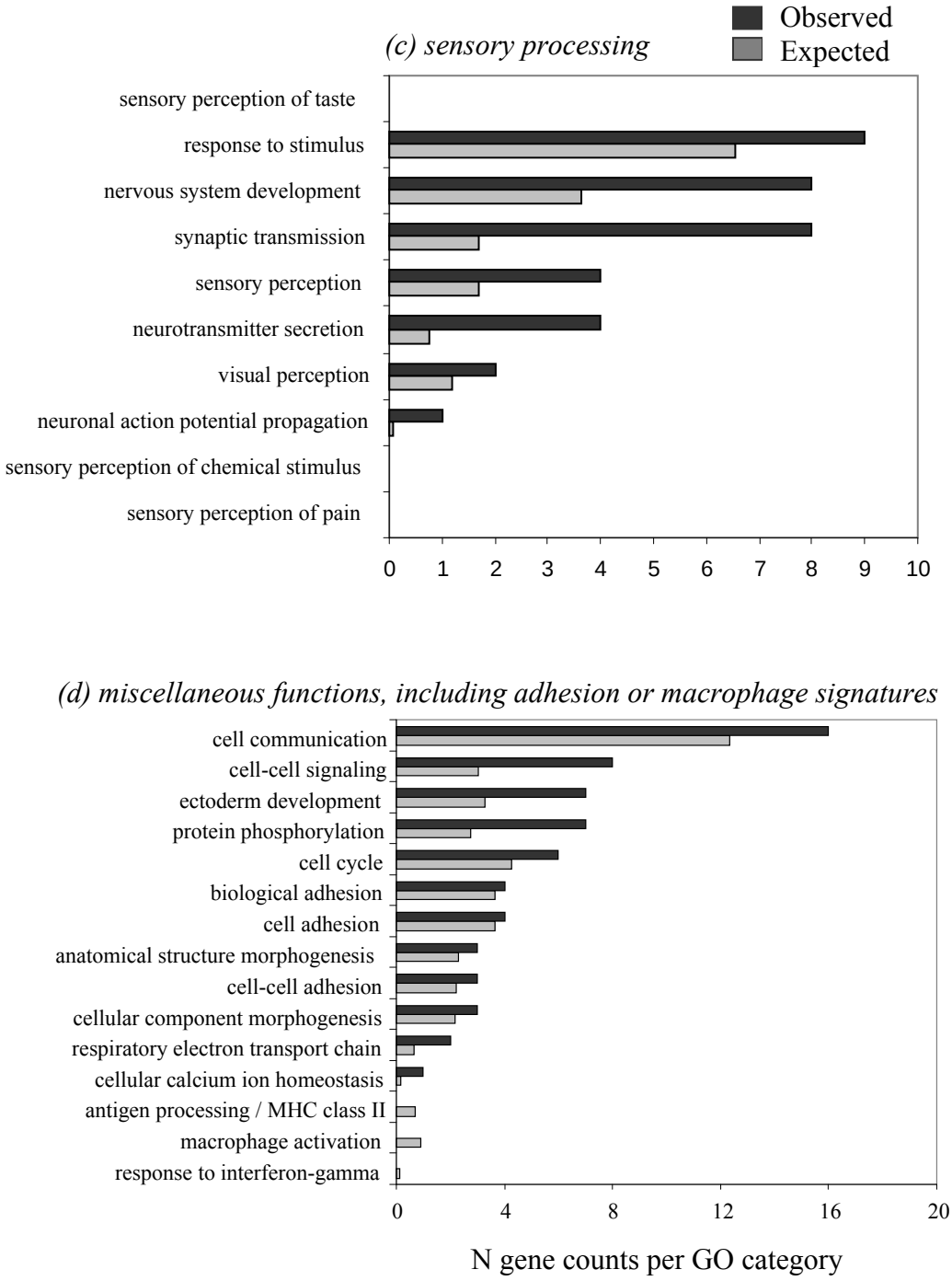


Figure 3.1, continued

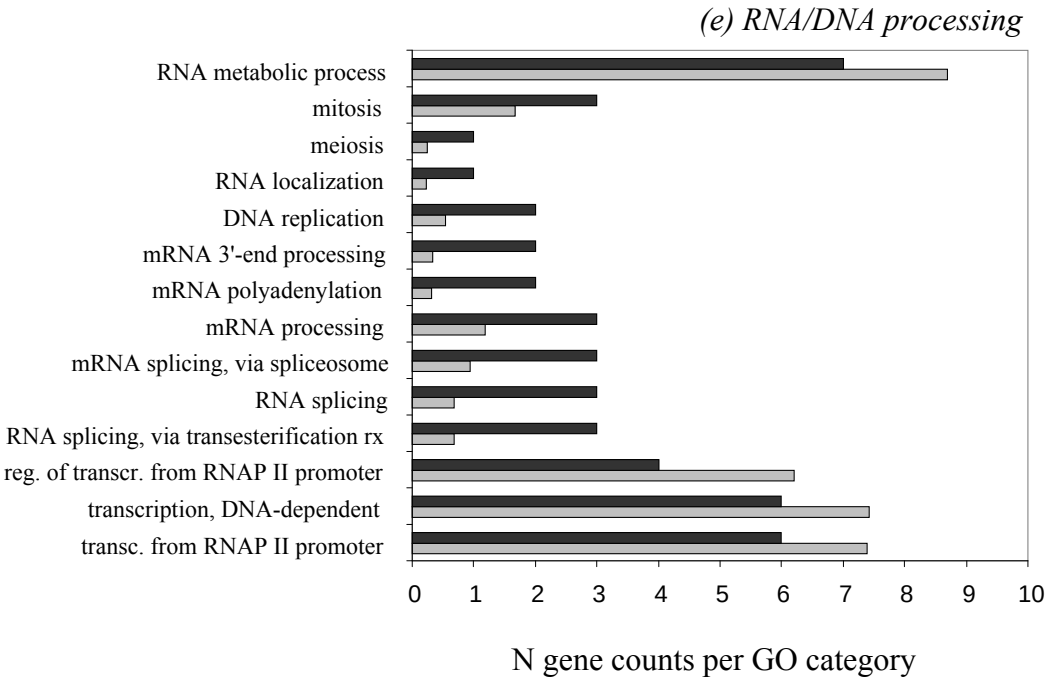


Figure 3.1, continued

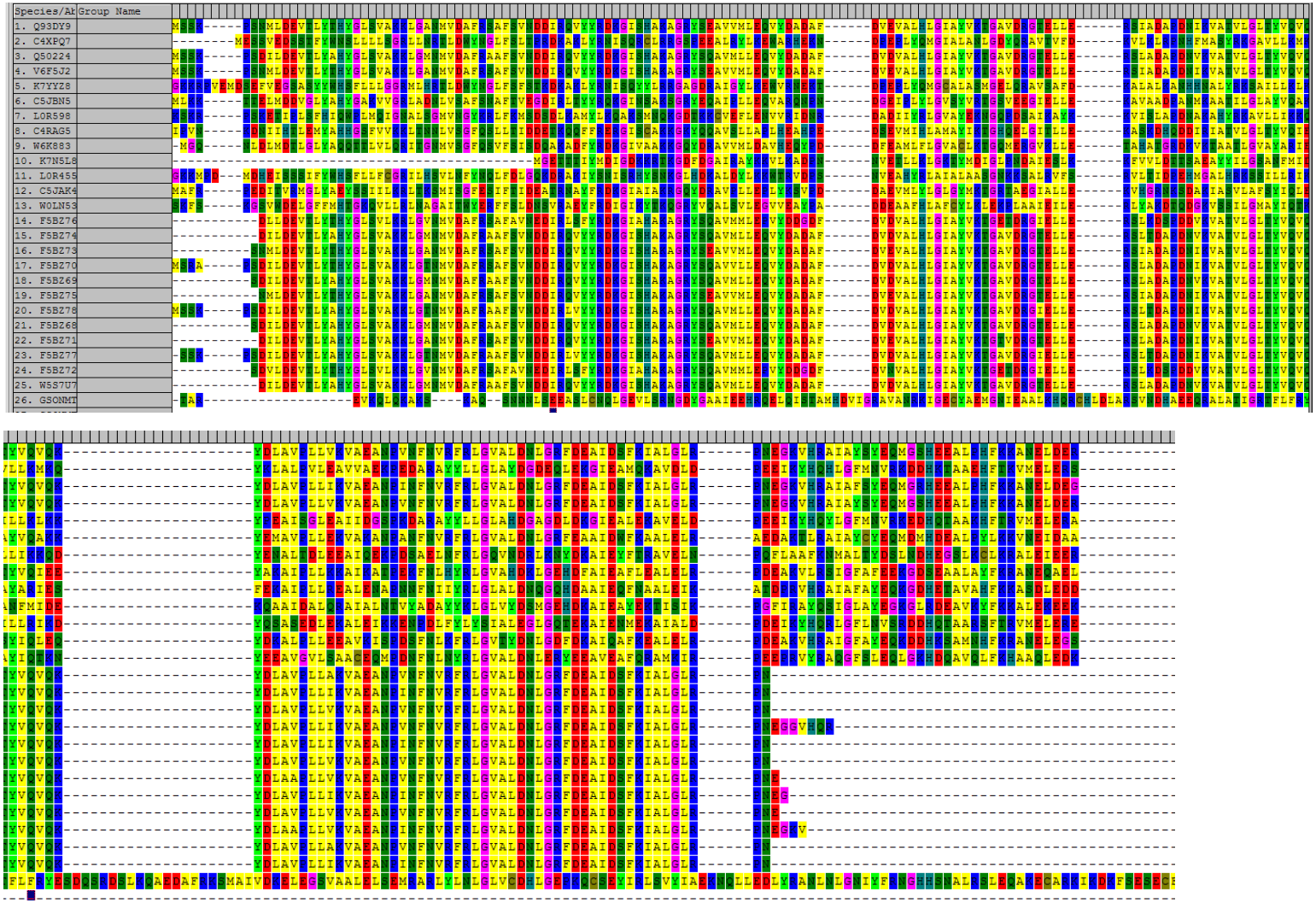


Figure 3.2. Magnetotactic bacteria MamA proteins aligned to a putative Trout protein homologue identified by TPR domain annotations and motif search. The Trout protein homologue is the bottommost sequence. [Mock-up figure; I just found a program JarView, that can achieve what I need].

Figure 3.3. Trout protein sequences with PDZ-domains and serine proteases aligned against the magnetotactic bacteria protein MamE.

```
tr|U5IGN4|U5IGN4_9DELT          MHGMTHEDLLRDPDVAALLERREELDCTHRELVRSIEELRSEHVALSERLRLLNREQAVA
tr|A0A060YJ67|A0A060YJ67_ONCMY  ------------------------------------------------------------

tr|U5IGN4|U5IGN4_9DELT          RHDRNGLRNETASLKTRLQTLTEQMRAGGPEYMARKETLRKLTEKRRAMRERVDELDQEC
tr|A0A060YJ67|A0A060YJ67_ONCMY  ------------------------------------------------------------

tr|U5IGN4|U5IGN4_9DELT          VAIGSRVTALEMDLIQAQEQLGSLEEQKQAIVDEISGRLSNTSLDRDEVEKHLNEVALEF
tr|A0A060YJ67|A0A060YJ67_ONCMY  ----------------------------------------------------MTVAXXXLEF
                                                                                    :    ***

tr|U5IGN4|U5IGN4_9DELT          RDAVTRRDEVLEEHGQVSEAMAAVGAEVKDY---------AERIDVLERMRALDQELRKL
tr|A0A060YJ67|A0A060YJ67_ONCMY  RSTAP--DGVLL--GVSSAKVDAIGLELASGQAVFHVNNGAGRVSATSRLGR--------
                                *.:.   * **   *  *  : *:* *: .          * *:.. .*:

tr|U5IGN4|U5IGN4_9DELT          TKRAEGERRETGQAEQNL-SEARAALALSQEMLAERVEKTEA-LERRSTEIQTLVGSYQA
tr|A0A060YJ67|A0A060YJ67_ONCMY  -------WLCDGRWHTLLAKKTKNALSLSVDGVTVLTDNPHPQSTSAETKDPIYVGGFPV
                                      *: .  * .::: **:** : ::  .:: .      .*:    **.: .

tr|U5IGN4|U5IGN4_9DELT          AKSAMEAKEESLRLARESLAGFEKELRE----------ILVVLPLVGSDVRALRQGLERV
tr|A0A060YJ67|A0A060YJ67_ONCMY  GV-----KQNC-LTSRSPFRGCMRNIRLIKGHVIDILDFSTAFTLQGVSPHSCPAT----
                                .      *::.   :*. : *  :::*           : ..: * * . ::

tr|U5IGN4|U5IGN4_9DELT          ARGG        343
tr|A0A060YJ67|A0A060YJ67_ONCMY  ----        161
```

Figure 3.4. Alignment of Rainbow trout Laminin-G gene (SalmonDB 25581 = GSONMT00018588001) to its putative magnetosome associated ortholog ("mad21" U5IGN4_9DELT Lefèvre et al. 2013) identified through a clustering analysis of differentially expressed genes in candidate magnetoreceptor cells. Clustering analysis was implemented in OrthoMCL (Li, Stoeckert, and Roos 2003).

4 - Evolution of olfactory subgenome repertoires of the jawless fish, sea lamprey (*Petromyzon marinus*) and jawed fish, Rainbow trout (*Oncorhynchus mykiss*), with comparison to infraclass Teleostei

4.1. ABSTRACT

The olfactory receptor subgenome enables the detection of scents vitally important for navigation, identification of food, mates, and danger avoidance. Salmonids use olfactory cues for natal homing, but few studies have investigated whole olfactory sub-genome with complementary transcriptome profiling. In this study, evolution of the salmonid olfactory genomic repertoire was assessed by phylogenetic analysis of olfactory receptor genes identified from the recently published genomes of Rainbow Trout, sea lamprey, and 15 other Teleost genomes and through transcriptome profiling by RNA-seq. Transcriptomes of olfactory tissue, whole blood, and muscle revealed a greater number of genes are uniquely expressed in the olfactory transcriptome data set. A simultaneous examination of all classes of olfactory gene super-families indicates (1) vomeronasal 1 like genes are the least diversified gene family, yet salmon have retained four gene copies; (2) for vomeronasal 2 like genes, more duplicate copies are retained in salmon relative to other species; (3) for trace amine-associated receptors, salmon have retained gene diversity but in a matter similar to other teleosts; and (4) main olfactory receptors are highly diverse among all teleosts, and zebrafish have higher rates of duplication than salmon. In salmon, the vomeronasal 2 like genes show the greatest difference relative to the other fifteen teleost species examined.

4.2. INTRODUCTION

The olfactory receptor subgenome imparts the ability to discriminate scents vitally important for navigation, identification of food, mates, and danger avoidance. In vertebrates, odor detection operates by interaction between a ligand and a G protein-coupled receptor (GPCR) with seven transmembrane-spanning regions (7tm). The ligand-GPCR interaction induces a conformational change which initiates a signaling pathway that ultimately results in chemosensory signal transduction (Mombaerts 1999). Olfactory stimulants can be divided into two major classes: general odorants, small molecules derived from food or the environment, and pheromones, molecules released from an organism to convey sexual or social cues (Touhara and Vosshall 2009). Four evolutionarily unrelated, large superfamilies of GPCRs function as odorant and pheromone chemoreceptors (Mombaerts 2004). These receptors are generally classified as: vomeronasal type-1 (V1R; also called Ora, VNR), vomeronasal type-2 (V2R) or V2-like (sometimes termed Olfactory C family, OlfC, in fish), trace amine-associated (TAAR), and main olfactory receptors (MOR). The vomeronasal receptors have been proposed to recognize pheromonal cues (Dulac and Axel 1995), but the main olfactory receptor system can also perceive those types of signals (Baum 2012). Moreover, at least in fish, the V2R like genes recognize amino acids not classified as pheromones (Speca et al. 1999). Categorization of olfactory gene superfamilies by function thus presents a somewhat confounded picture despite proposed independent evolutionary histories of these gene families.

Olfactory gene diversity is thought to reflect the ability of olfaction of the species and is tied to their environment (Niimura 2009). Since Buck and Axel (1991) provisioned a foundation for a molecular understanding of odor recognition in vertebrates in 1991, the genomics era has enabled the deduction of complete olfactory gene repertoires for taxonomically divergent species (Niimura and Nei 2005; Nguyen et al. 2012; Glusman et al. 2000; Robertson, Warr, and Carlson 2003). Extensive expansions have occurred within some olfactory gene families, as best studied in the MOR superfamily. Drawing vertebrate data from publically searchable genomic databases, Glusman and colleagues (Glusman et al. 2000) determined that all MOR-like sequences cluster into two basic groups: Class I ("fish-like") and Class II (tetrapodan), consistent with previous findings by Freitag and colleagues (Freitag et al. 1995). A phylogenetic analysis of this superfamily in zebrafish, pufferfish, frogs, chickens, humans, and mice indicated the most recent common ancestor between fishes and tetrapods carried at least nine ancestral MOR genes (Niimura and Nei 2005). Fish retained eight of these nine groups, while mammalian genomes contain signatures of only two. One of those two groups, group γ genes (in Class II), expanded in land-dwelling organisms but is nearly absent in fish. That diversification pattern is consistent with group γ genes having a functional role in airborne odorant detection. For scale, the number of predicted MORs varies enormously by taxa, from ~44 in pufferfish and ~150 in zebrafish, to ~400 in humans and ~1,200 in rats (Niimura 2009; Niimura and Nei 2005). In contrast to the observed MOR pattern, recent work in fish has shown that TAAR and V2R gene families have expanded and diversified (Hashiguchi and Nishida 2007; Hashiguchi and Nishida 2006; Tessarolo et

al. 2014), a pattern that suggests these genes have important roles for olfaction in aquatic environments. Findings across olfactory gene studies are inadequate to quantify the relative diversity of olfactory gene families within single species because each study tends to focus on a single olfactory family across single or multiple taxa and rarely takes into account the full genomic olfactory repertoire.

The olfactory repertoire of salmonids represents an interesting case for study because these fish possess remarkable ability to learn odors associated with their natal site as juveniles and later use these odors to guide their homing migration. Their long-distant migratory journey actually involves a myriad of navigation cues that are for the most part poorly understood, but the instrumental role of olfaction in homing once fish have reached fresh water is well established (Wisby and Hasler 1954; Dittman and Quinn 1996). The physiological mechanisms of olfactory imprinting have yet to be elucidated (Ueda 2012). River-specific chemosensory cues might involve odor substances dissolved as free amino acids (Sato 2001; Shoji et al. 2003) or population-specific pheromones released by juveniles (Nordeng 1971). To provide background bioinformatics data to help inform this quest, here we use a functional genomics approach to identify the complete expressed repertoire of this gene family. Olfactory transcriptome profiles generated by RNA-seq were compared to transcriptomes profiles of non-olfactory tissues to identify differentially expressed genes. The complete olfactory subgenome of salmonids was identified through homology search of novel Chinook salmon transcriptome (this study) and Rainbow Trout genome (Berthelot et al. 2014) transcripts, against GPCR and Teleostei olfactory gene databases. To provide context for findings, phylogenetic assessment

was performed for all Teleostei olfactory genes and those identified from the Rainbow Trout genome (Berthelot et al. 2014) and, as an outgroup to the jawless vertebrate, sea lamprey *Petromyzon marinus* genome (Smith et al. 2013).

4.3. METHODS

4.3.1. Tissue sampling for RNA-seq

Biological samples for reference assembly and transcriptome profiles were obtained from three adult Chinook salmon, South Santiam hatchery spring run stock, raised at the Corvallis Fish Research Laboratory, Corvallis, Oregon. Fish were reared in an indoor well-water tank (date in tank 6/26/09, sampled June 2013). Following humane termination by decapitation and pithing (Oregon State University ACUP 3595, 4421), three types of tissues were sampled from each fish: pairs of olfactory rosettes (the olfactory epithelium), ~1 cm x 1 cm muscle dissected from the anterior of the fish mid-way between the lateral line and dorsal fin, and approximately 200 uL of whole blood. Upon dissection, each sample was placed in buffer and immediately flash-frozen in an ethanol-dry ice slurry and then stored in a -80° C freezer until RNA isolation. The total RNA was isolated using glass-silica based columns (RNAeasy Minikit, Qiagen), following manufacturer protocols, and submitted to Oregon State University's Center for Genome Research and Biocomputing (CGRB) for mRNA isolation using poly-A labeled magnetic beads. At the CGRB, the mRNA samples were indexed for multiplexing and sequenced in a single lane of an Illumina HiSeq2000 using 50 cycles and single-end protocols. To augment the relatively shallow (~15 million reads per sample) RNA-seq data set, an additional pair of

olfactory rosettes (sampled from the same brood stock and tank, sample date May 2012) was deep sequenced, yielding ~87 millions gene transcripts, using a 101 cycle, paired-end protocol. In total, 228,051,458 RNA-seq reads from these 10 samples passed quality filters (Table 4.1) and were available for reference assembly and transcriptome profiling.

4.3.2. RNA-seq data processing

Raw gene transcripts were processed using the FASTX Toolkit (0.0.13, (Blankenberg et al. 2010). After trimming low-quality nucleotides (phred score < 20) from the ends of sequences and removing Illumina adapters, fastq_quality_filter was applied to remove gene transcripts that failed to meet at least 90% of all cycles of the read having a minimum base pair quality of phred score 20 and those shorter than 25 nucleotides. Paired-end reads were matched to their mate using the script PE_Combiner (De Wit et al. 2012), retaining single reads as "orphans".

Transcriptome diversity is expected to be higher for olfactory tissue because of the large olfactory receptor gene family. The average numbers of reads per tissue type that aligned to the Rainbow trout genome were assigned to chromosomal locations. Transcripts counts were not adjusted for library size differences. The length of the chromosomes was taken into account by divided by the number of reads that aligned to a chromosome, dividing by nucleotide length, and multiplied that result by 1000. This standardized raw counts to the average number of reads per 1000 nucleotides of chromosomal sequence.

To estimate tissue-specific gene expression patterns using RNA-seq data requires the alignment of expressed sequence reads to a reference gene sequence data set. Here we developed, for Chinook salmon, a novel genome-guided transcriptome reference by aligning the ~228 million RNA-seq reads to the genome of a closely related species Rainbow Trout/Steelhead Salmon (*O. mykiss*) (Berthelot et al. 2014). Following alignments performed with Bowtie2 (Langmead and Salzberg 2012) using the setting '--very-sensitive, reference transcripts were assembled and extracted using Tuxedo Tools (Trapnell et al. 2012). Identification of differentially expressed genes and estimates of gene expression levels for the olfactory transcriptome relative to those of blood and muscle transcriptomes were performed using program CuffDiff. For comparative purposes, the RNA-seq reads were also aligned to predicted gene transcripts deduced from Trout genome (Oncorhynchus_mykiss_mRNA.fa and Oncorhynchus_mykiss_pep.fa) (Berthelot et al. 2014).

4.3.3. Olfactory gene identification and annotation

Olfactory genes were identified from the Chinook salmon genome-guided transcriptome (this study), mRNA and peptide sequences predicted for the Rainbow Trout genome (Berthelot et al. 2014), and for predicted peptide sequences from the sea lamprey genome (Smith et al. 2013) (P_marinus 7.0, available from http://uswest.ensembl.org/Petromyzon_marinus/Info/Index), a species with a most recent common ancestor that pre-dates the split with jawed vertebrates, genome. Olfactory genes were identified by Basic Local Alignment Tool (BLAST) (Altschul et al. 1990) homology search of gene transcripts against two databases: "Uniprot

GPCR" and "RefSeq Teleostei olfactory gene". The Uniprot GPCR database was created by download of 7-transmembrane G-linked receptor protein sequences listed in the 7tmrlist.txt file maintained by UniProt - Swiss-Prot Protein Knowledgebase (download date October 27, 2014, http://www.uniprot.org/docs/7tmrlist) and having protein sequences available through ENTREZ Batch by record match (n = 3082, Table 4.2). In the 7tmrlist.txt file, these olfactory receptor like genes are listed under groupings "Odorant/olfactory and gustatory receptors", "Pheromone" and "Trace amine". This GPCR gene list is heavily weighted towards vertebrate mammalian species (minimum 73% of all genes). The RefSeq Teleostei olfactory gene database was developed through search for olfactory receptor like gene sequences contained in the National Center for Biotechnology Information (NCBI) Reference Sequence gene (RefSeq, vs. 67) and having sequences available from the linked NCBI protein database (download date October 23, 2014; n = 2,722 sequences after removal of 16 duplicates). Search terms included "olfactory", "odorant", "vomeronasal", and "trace amine-associated", and results were filtered for "Teleostei". This genetic reference contains olfactory sequences from fourteen fish species (Table 4.3), but lacks representation from salmonids. Noteworthy here is that teleosts comprise constitute >99% of all living species of fish (Venkatesh 2003), thus the Teleost reference database is broadly representative of olfactory variation present in fish. For both databases, sequences were sub-divided into the olfactory receptor superfamily categories vomeronasal type-1 (V1), vomeronasal type-2 (V2), pheromone (likely similar to V1 or V2), trace amine-associated receptor (TAAR), and olfactory or odorant receptor (Tables 4.1, 4.3). Non-olfactory receptor genes (e.g., extracellular

calcium sensing receptor, marker proteins, guanylyl cyclase) were placed into separate "other" category. Predicted gene transcripts were classified as olfactory receptors if they met BLASTx or BLASTp criteria of e-value <= 1 e -05 and minimum alignment length of amino acid matches >= 250 to olfactory gene categories. This length stringency was decreased to a minimum of 50 for the Chinook salmon genome-guided transcriptome assembly.

4.3.4. Evolution of salmonid olfactory repertoire

To understand evolution of the salmonid olfactory genomic repertoire, phylogenetic analysis was performed for all olfactory receptor genes identified from the Rainbow Trout (Berthelot et al. 2014) and sea lamprey genomes (Smith et al. 2013), along with all olfactory receptors identified for Teleostei. Following methods of Niimura and Nei (Niimura and Nei 2005), outgroup non-OR GPCR genes included bovine adenosine A1 receptor and rat α2B-adrenergic receptor (GenBank accession no. X63592, AF366899, respectively). Attempts (see results) were made to include olfactory genes identified from the genome-guided Chinook salmon transcriptome and translated through TransDecoder (Haas et al. 2013). Multiple alignments of olfactory gene sequences were obtained using ClustalW implemented on a 40 Gb processor with 24 cores. Together, these analyses provide multi-scaled context for the olfactory repertoire of salmonids, the sea lamprey, and teleosts.

4.4. RESULTS

BLAST homology searches to identify the full Rainbow Trout olfactory repertoire revealed numerous trace amine-associated receptors (n = 227), moderate numbers of V2R like (n = 41) and MOR (n = 61) genes, and a limited number of V1R-like genes (n = 4) (non-redundant numbers of genes listed in Tables 4.1 and 4.3). The majority of these olfactory genes were identified through BLAST homology search against the RefSeq Teleostei olfactory gene database compared to against the Uniprot GPCR database. The sea lamprey exhibited the same basic patterns, but returned fewer BLASTx hits per olfactory gene superfamily.

The number of olfactory genes identified from the Chinook salmon genome-guided assembly was surprisingly less than those predicted from the Rainbow Trout transcriptome (Tables 4.1, 4.3). Relaxing the minimum alignment length criteria to >= 50 for the Chinook transcriptome assembly had little affect on the inclusion of more sequences. Of 188 olfactory genes identified from the genome-guided assembly, only 35 contained viable coding regions. The olfactory transcriptome as a whole was much more diverse than that of blood and muscle tissue, with six times more unique genes than the blood and muscle transcriptome data set. When the deep-sequenced olfactory data set was considered another 552 genes were identified as unique in the olfactory data set (Table 4.4).

The number of MORs was greater than TAARs (Table 4.2), in contrast to the ratio of genes predicted from the Rainbow Trout transcriptome. That the reference guided assembly did not yield nucleotide sequences with high translation rates into amino acids is puzzling because the number of reads used for the transcriptome

assembly, ~228 million (Table 4.1), should have been sufficient for transcriptome assembly. The number of RNA-seq reads for muscle and blood tissues that aligned against the transcriptome and genome was similar (average 6%, 0 – 13%), but for the olfactory data set more than 21 to 33% reads (average 26%) aligned against the genome relative to the transcriptome. But, because results for the translated, genome-guided assembly were inconsistent with those predicted from the Rainbow Trout transcriptome, we instead focus on predicted Rainbow Trout olfactory genes.

Phylogenetic reconstruction of the olfactory subgenome of teleosts and the complete subgenomes of the Rainbow trout and sea lamprey indicates strikingly different evolutionary pressures for olfactory receptor superfamily genes (Tables 4.1, 4.3, Figure 1). For both jawed and jawless fish, the V1R superfamily contained the fewest number of genes, and the TAAR superfamily contained the greatest number of genes. The "mirrored" branching pattern of V1R genes is consistent with two sequential gene (or genome) duplication events in Rainbow trout, while the sea lamprey contains only two V1R genes that are likely closely related genes. For the V2R like gene family, only two single-copy genes were predicted from the sea lamprey genome. In contrast, Rainbow trout genes are distributed throughout most groups contained in the star-like, loosely clustered phylogeny. While the numbers of V1R and V2R genes are modest, the TAAR superfamily demonstrates a signature of substantial expansion. Especially notable within this gene superfamily is a single group that contains all sea lamprey and the majority of Rainbow trout genes. A second, smaller expansion within Rainbow trout is indicated by a distinct cluster for the Rainbow trout. The fourth gene family, MORs, contains the greatest olfactory

gene representation for teleosts in the NCBI database yet does not represent the greatest diversity amongst the various olfactory families. Nearly all groups of MORs contain one or more Rainbow trout genes, but in contrast genes from sea lampreys were restricted to one group with the exception of a single gene in a different cluster. In sum, the diversity of genes within olfactory superfamilies indicates substantially different evolutionary pressures for each group. For TAAR superfamily, evolutionary pressures appear to be similar for Rainbow trout and the sea lamprey.

4.5. DISCUSSION

Standardized comparisons among olfactory gene superfamilies identified from jawed and jawless fish genomes, and teleost fish obtained from genetic repositories, provide striking contrasts for the evolutionary dynamics of olfactory gene families. Among all, the TAAR and MOR superfamilies show the greatest duplication and expansion. This TAAR expansion, and to a lesser degree the V2Rs and MORs expansion, stands in sharp contrast to the limited V1R gene repertoire. The retention and expansion of the TAAR superfamily in genomes of jawed and jawless fish suggests these genes serve important functions, but at present these functions are poorly understood. The TAARs were initially considered neurotransmitter receptors, but were recently reclassified as a class of vertebrate chemosensory receptors by Hashiguchi and Nishida (Hashiguchi and Nishida 2007). In rodents, the ligands for olfactory TAARs include volatile and highly aversive amines and have therefore been proposed to bind specific odors that activate hardwired neural circuits and ultimately elicit stereotyped, innate behavior ((Johnson et al. 2012) and references therein).

Both salmon and sea lamprey have anadromous life histories, but their migratory behaviors are vastly different: salmonids are famous for their precise homing to natal grounds for spawning (T. P. Quinn and Dittman 1990), in contrast to sea lampreys that appear to simply find a "suitable river" (Waldman, Grunwald, and Wirgin 2008). Tessarolo and colleagues (Tessarolo et al. 2014) recently performed a phylogenetic assessment of the TAAR repertoire of Atlantic salmon (*Salmo salar*), sea lamprey, and numerous other vertebrate species, and concluded that the lamprey TAARs were so distinct from Atlantic salmon and the other species that perhaps they were perhaps actually some other type of aminergic receptor. Their approach of identifying TAARs was described as BLASTx of contig sequences using *Danio rerio* TAAR1a as the query. That is a substantially different approach from BLASTx search of predicted transcriptomes using as references the Uniprot GPCR 7tm list and all annotated teleost TAAR genes from the NCBI RefSeq database (described as "comprehensive, non-redundant, and well-annotated). In both Tessarolo and this study the TAAR genes were highly diverse for salmon and we reached the same conclusion: the large expansion of this family is consistent with these genes having an important role in the biology of this species.

Whole genome duplication events provide gene diversity and enables functional specialization that may be important for key evolutionary processes such as adaptation and speciation. The salmonid genome underwent multiple whole genome duplication events (WGD) (Allendorf and Thorgaard 1984), the most recent of which took place approximately 25 to 100 million years ago (Meyer and Schartl 1999; Allendorf and Thorgaard 1984). Berthelot and colleagues (Berthelot et al.

2014), who performed the impressive task of sequencing the entire Rainbow trout genome, concluded that salmonids retained collinear ancestral sub-genomes as evidenced by retention of gene copies across diverse gene families. Since olfactory genes are known for duplications and expansions, and salmonids are known for multiple WGD events, finding that excessive gene copies were not more prevalent in their olfactory gene repertoire was surprising. Given the retention of single gene copies across broad MOR and TAAR groupings, those olfactory genes may function as "generalists" with rapid degradation or neofunctionalization of the alternate gene copies.

Why did the genome-guided assembly return so few transcripts that could be translated into amino acids? One possible explanation is related to the nature of gene expression patterns of olfactory neurons. First, olfactory receptor superfamilies are known for variation in their amino acid sequences which enables ligand binding of 7Tm proteins to odorant stimulant having vastly different structures. Second, high sequence divergence would result in reduced sequencing coverage for any given gene. Last, olfactory sensory neurons in fish are expressed in dispersed and punctuate patterns, thus the expression of any single receptor is likely to be low relative to the total olfactory (or any) transcriptome (see Mombaerts review (Mombaerts 1999). The sequencing coverage in this experimental set-up may not have yielded sufficient depth for transcript assembly. That the genomes of salmonids are highly complex and intractable to short read assembly (N. L. Quinn et al. 2008; Davidson et al. 2010) also is a potential complicating factor for assembly.

In conclusion, a simultaneous examination of all olfactory genes super-families reveals that salmon do not show exceptional gene duplication or retention in any one group, except for perhaps the V2R like genes. Across all teleosts, the TAAR and MORs genes are highly diversified, with patterns of retention and duplication mostly similar among salmon and other teleosts. One exception is that within the MOR super-family, zebrafish display an exceptional number of gene duplication events. A few salmon TAAR genes are highly divergent, but within groupings duplicate copies have been lost. The diversity of V2R like genes is low relative to MORs and TAARs, but for the genes that are present salmon appear to have retained multiple duplicate copies. Of all olfactory super-families the V1Rs are the least diversified, yet four copies are present in the salmon genome. The V2Rs appear to be most different for salmon relative to other teleost fish.

4.6. REFERENCES

Allendorf, Fred W., and Gary H. Thorgaard. 1984. "Tetraploidy and the Evolution of Salmonid Fishes." In *Evolutionary Genetics of Fishes*, edited by Bruce J. Turner, 1–53. Boston, MA: Springer US. http://link.springer.com/10.1007/978-1-4684-4652-4_1.

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 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Baum, Michael J. 2012. "Contribution of Pheromones Processed by the Main Olfactory System to Mate Recognition in Female Mammals." *Frontiers in Neuroanatomy* 6: 20. doi:10.3389/fnana.2012.00020.

Berthelot, Camille, Frédéric Brunet, Domitille Chalopin, Amélie Juanchich, Maria Bernard, Benjamin Noël, Pascal Bento, et al. 2014. "The Rainbow Trout Genome Provides Novel Insights into Evolution after Whole-Genome Duplication in Vertebrates." *Nature Communications* 5 (April). doi:10.1038/ncomms4657.

Blankenberg, Daniel, Assaf Gordon, Gregory Von Kuster, Nathan Coraor, Galaxy Team, James Taylor, and Anton Nekrutenko. 2010. "Manipulation of FASTQ Data with Galaxy." *Bioinformatics (Oxford, England)* 26 (14): 1783–85. doi:10.1093/bioinformatics/btq281.

Buck, L., and R. Axel. 1991. "A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition." *Cell* 65 (1): 175–87.

Davidson, William S., Ben F. Koop, Steven J. M. Jones, Patricia Iturra, Rodrigo Vidal, Alejandro Maass, Inge Jonassen, Sigbjorn Lien, and Stig W. Omholt. 2010. "Sequencing the Genome of the Atlantic Salmon (Salmo Salar)." *Genome Biology* 11 (9): 403. doi:10.1186/gb-2010-11-9-403.

De Wit, Pierre, Melissa H. Pespeni, Jason T. Ladner, Daniel J. Barshis, François Seneca, Hannah Jaris, Nina Overgaard Therkildsen, Megan Morikawa, and Stephen R. Palumbi. 2012. "The Simple Fool's Guide to Population Genomics via RNA-Seq: An Introduction to High-Throughput Sequencing Data Analysis." *Molecular Ecology Resources* 12 (6): 1058–67. doi:10.1111/1755-0998.12003.

Dittman, Andrew H., and Thomas P. Quinn. 1996. "Homing in Pacific Salmon: Mechanisms and Ecological Basis." *Journal of Experimental Biology* 199 (1): 83–91.

Dulac, C., and R. Axel. 1995. "A Novel Family of Genes Encoding Putative Pheromone Receptors in Mammals." *Cell* 83 (2): 195–206.

Freitag, J., J. Krieger, J. Strotmann, and H. Breer. 1995. "Two Classes of Olfactory Receptors in Xenopus Laevis." *Neuron* 15 (6): 1383–92.

Glusman, G., A. Bahar, D. Sharon, Y. Pilpel, J. White, and D. Lancet. 2000. "The Olfactory Receptor Gene Superfamily: Data Mining, Classification, and Nomenclature." *Mammalian Genome: Official Journal of the International Mammalian Genome Society* 11 (11): 1016–23.

Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. "De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis." *Nature Protocols* 8 (8): 1494–1512. doi:10.1038/nprot.2013.084.

Hashiguchi, Yasuyuki, and Mutsumi Nishida. 2006. "Evolution and Origin of Vomeronasal-Type Odorant Receptor Gene Repertoire in Fishes." *BMC Evolutionary Biology* 6: 76. doi:10.1186/1471-2148-6-76.

———. 2007. "Evolution of Trace Amine Associated Receptor (TAAR) Gene Family in Vertebrates: Lineage-Specific Expansions and Degradations of a Second Class of Vertebrate Chemosensory Receptors Expressed in the Olfactory Epithelium." *Molecular Biology and Evolution* 24 (9): 2099–2107. doi:10.1093/molbev/msm140.

Johnson, Mark A., Lulu Tsai, Dheeraj S. Roy, David H. Valenzuela, Colleen Mosley, Angeliki Magklara, Stavros Lomvardas, Stephen D. Liberles, and Gilad Barnea. 2012. "Neurons Expressing Trace Amine-Associated Receptors Project to Discrete Glomeruli and Constitute an Olfactory Subsystem." *Proceedings of the National Academy of Sciences of the United States of America* 109 (33): 13410–15. doi:10.1073/pnas.1206724109.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4): 357–59. doi:10.1038/nmeth.1923.

Meyer, A., and M. Schartl. 1999. "Gene and Genome Duplications in Vertebrates: The One-to-Four (-to-Eight in Fish) Rule and the Evolution of Novel Gene Functions." *Current Opinion in Cell Biology* 11 (6): 699–704.

Mombaerts, . 1999. "Molecular Biology of Odorant Receptors in Vertebrates." *Annual Review of Neuroscience* 22: 487–509. doi:10.1146/annurev.neuro.22.1.487.

Mombaerts, Peter. 2004. "Genes and Ligands for Odorant, Vomeronasal and Taste Receptors." *Nature Reviews. Neuroscience* 5 (4): 263–78. doi:10.1038/nrn1365.

Nguyen, Dinh Truong, Kyooyeol Lee, Hojun Choi, Min-kyeung Choi, Minh Thong Le, Ning Song, Jin-Hoi Kim, et al. 2012. "The Complete Swine Olfactory Subgenome: Expansion of the Olfactory Gene Repertoire in the Pig Genome." *BMC Genomics* 13: 584. doi:10.1186/1471-2164-13-584.

Niimura, Yoshihito. 2009. "Evolutionary Dynamics of Olfactory Receptor Genes in Chordates: Interaction between Environments and Genomic Contents." *Human Genomics* 4 (2): 107–18.

Niimura, Yoshihito, and Masatoshi Nei. 2005. "Evolutionary Dynamics of Olfactory Receptor Genes in Fishes and Tetrapods." *Proceedings of the National Academy of Sciences of the United States of America* 102 (17): 6039–44. doi:10.1073/pnas.0501922102.

Nordeng, H. 1971. "Is the Local Orientation of Anadromous Fishes Determined by Pheromones□?" *Nature* 233 (5319): 411–13. doi:10.1038/233411a0.

Quinn, Nicole L., Natasha Levenkova, William Chow, Pascal Bouffard, Keith A. Boroevich, James R. Knight, Thomas P. Jarvie, et al. 2008. "Assessing the

Feasibility of GS FLX Pyrosequencing for Sequencing the Atlantic Salmon Genome." *BMC Genomics* 9: 404. doi:10.1186/1471-2164-9-404.

Quinn, Thomas P., and Andrew H. Dittman. 1990. "Pacific Salmon Migrations and Homing: Mechanisms and Adaptive Significance." *Trends in Ecology & Evolution* 5 (6): 174–77. doi:10.1016/0169-5347(90)90205-R.

Robertson, Hugh M., Coral G. Warr, and John R. Carlson. 2003. "Molecular Evolution of the Insect Chemoreceptor Gene Superfamily in Drosophila Melanogaster." *Proceedings of the National Academy of Sciences of the United States of America* 100 Suppl 2 (November): 14537–42. doi:10.1073/pnas.2335847100.

Sato, K. 2001. "Whole-Cell Response Characteristics of Ciliated and Microvillous Olfactory Receptor Neurons to Amino Acids, Pheromone Candidates and Urine in Rainbow Trout." *Chemical Senses* 26 (9): 1145–56. doi:10.1093/chemse/26.9.1145.

Shoji, T., Y. Yamamoto, D. Nishikawa, K. Kurihara, and H. Ueda. 2003. "Amino Acids in Stream Water Are Essential for Salmon Homing Migration." *Fish Physiology and Biochemistry* 28 (1-4): 249–51. doi:10.1023/B:FISH.0000030544.64774.f6.

Smith, Jeramiah J, Shigehiro Kuraku, Carson Holt, Tatjana Sauka-Spengler, Ning Jiang, Michael S Campbell, Mark D Yandell, et al. 2013. "Sequencing of the Sea Lamprey (Petromyzon Marinus) Genome Provides Insights into Vertebrate Evolution." *Nature Genetics* 45 (4): 415–21. doi:10.1038/ng.2568.

Speca, D. J., D. M. Lin, P. W. Sorensen, E. Y. Isacoff, J. Ngai, and A. H. Dittman. 1999. "Functional Identification of a Goldfish Odorant Receptor." *Neuron* 23 (3): 487–98.

Tessarolo, J. A., M. J. Tabesh, M. Nesbitt, and W. S. Davidson. 2014. "Genomic Organization and Evolution of the Trace Amine-Associated Receptor (TAAR) Repertoire in Atlantic Salmon (Salmo Salar)." *G3: Genes|Genomes|Genetics*, April. doi:10.1534/g3.114.010660.

Touhara, Kazushige, and Leslie B. Vosshall. 2009. "Sensing Odorants and Pheromones with Chemosensory Receptors." *Annual Review of Physiology* 71 (1): 307–32. doi:10.1146/annurev.physiol.010908.163209.

Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. 2012. "Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks." *Nature Protocols* 7 (3): 562–78. doi:10.1038/nprot.2012.016.

Ueda, H. 2012. "Physiological Mechanisms of Imprinting and Homing Migration in Pacific Salmon Oncorhynchus Spp." *Journal of Fish Biology* 81 (2): 543–58. doi:10.1111/j.1095-8649.2012.03354.x.

Venkatesh, Byrappa. 2003. "Evolution and Diversity of Fish Genomes." *Current Opinion in Genetics & Development* 13 (6): 588–92.

Waldman, J., C. Grunwald, and I. Wirgin. 2008. "Sea Lamprey Petromyzon Marinus: An Exception to the Rule of Homing in Anadromous Fishes." *Biology Letters* 4 (6): 659–62. doi:10.1098/rsbl.2008.0341.

Wisby, Warren J., and Arthur D. Hasler. 1954. “Effect of Olfactory Occlusion on Migrating Silver Salmon (*O . kisutch*).” *Journal of the Fisheries Research Board of Canada* 11 (4): 472–78. doi:10.1139/f54-031.

Table 4.1. Numbers and percentages of Chinook salmon Illumina RNA-seq reads aligned against the recently published Rainbow Trout genome and predicted transcriptome (Berthelot et al. 2014). The mRNA samples were isolated from whole blood, muscle, and olfactory epithelium.

Sample Type	Total reads	N reads aligned to genome	N reads aligned to transcriptome	% reads aligned to genome	% reads aligned to transcriptome	% difference in alignment rate - genome - transcriptome (%)
Blood 1	15,572,090	12,776,446	11,336,384	0.82	0.73	0.09
Blood 2	16,151,385	13,199,197	11,093,075	0.82	0.69	0.13
Blood 4	16,038,258	13,215,825	11,642,839	0.82	0.73	0.10
Muscle 1	14,899,160	12,077,427	12,146,583	0.81	0.82	0.00
Muscle 2	16,081,043	13,084,691	12,792,810	0.81	0.80	0.02
Muscle 4	14,871,474	11,987,305	11,908,749	0.81	0.80	0.01
Olfactory 1	15,171,725	12,102,035	8,409,887	0.80	0.55	0.24
Olfactory 2	16,284,939	13,025,640	8,716,113	0.80	0.54	0.26
Olfactory 4	15,289,567	12,209,806	7,240,020	0.80	0.47	0.33
Olfactory, deep sequence	87,691,817	66,483,573	47,921,496	0.76	0.55	0.21
Totals	228,051,458	180,161,945	143,207,956			

Table 4.2. Counts of olfactory receptor like receptor genes identified by BLASTx homology to the "GPCR + RefSeq Teleostei database". The GPCR genes are seven-transmembrane G-linked receptors protein sequences (n = 3088) identified from the 7tmrlist.txt record file maintained by UniProt - Swiss-Prot Protein Knowledgebase (download date October 27, 2014). The Teleostei olfactory genes were obtained by search of records containing keywords "olfactory", "trace amine" or "vomeronasal" from the National Center for Biotechnology Information Reference Sequence ("RefSeq") gene database and having available protein sequences. Results are presented for olfactory genes identified from the Lamprey genome (version 77, downloaded from ENSEMBL), Rainbow trout (*O. mykiss*) transcriptome as predicted from the genome (both peptide and mRNA sequences) (Berthelot et al. 2014), and the Chinook salmon genome-guided transcriptome (this study). Genes were classified as homologous if they met criteria minimum e-value 1e-05 and alignment length 250 or 50 (Chinook salmon genome-guided transcriptome only).

Olfactory gene category	7tm GPCR database	Teleostei olfactory receptor reference database	Lamprey (v 77, pep, BLASTp)	Trout transcriptome (pep or mRNA)	Trout transcriptome predicted translated pep (BLASTp)	Trout transcriptome mRNA (BLASTx)	Genome-guided Chinook reference (BLASTx)
				>=250			>= 50
Vomeronasal type-1, *ora*	40	41	4	4	4	3	0
Vomeronasal type-2, *OlfC*	2	56	1	15	11	13	18
Trace amine-associated receptor	43	477	18	29	28	29	62
Main olfactory receptor	554	1692	24	54	52	51	93
Calcium-sensing	5	40	2	21	18	18	4
GPCR, other category	2443	415	542	1985	1928	1828	642
Total	3082	2721	591	2108	2041	1942	819

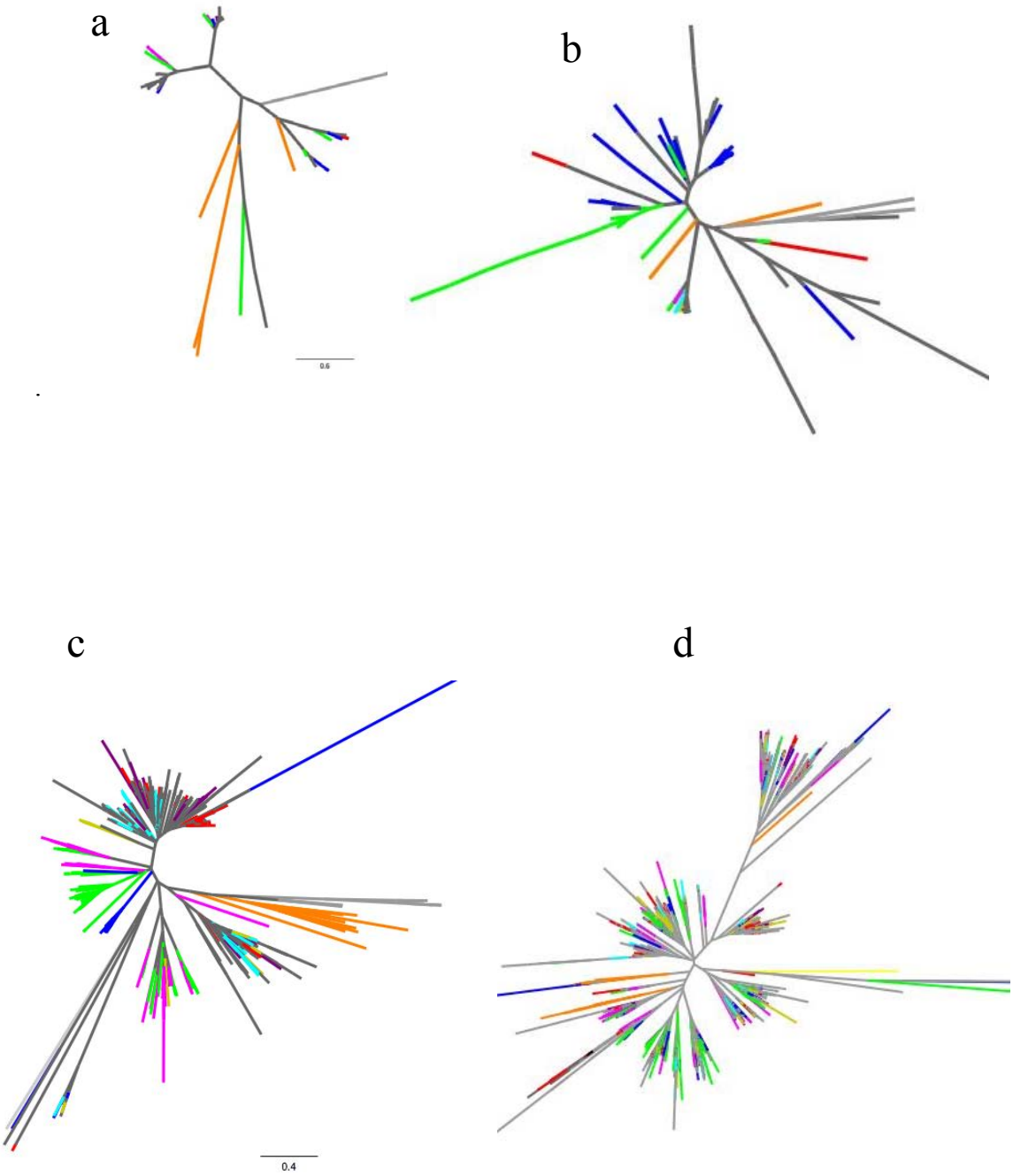
Table 4.3. List of records by species included in the Teleostei fish olfactory database obtained by search for olfactory receptor genes in the National Center for Biotechnology Information Reference Sequence ("RefSeq") gene database. Records identified through keywords "olfactory", "trace amine" nd "vomeronasal" and with protein sequences were compiled into the reference database.

Species	Common name	Main olfactory receptors	Trace amine-associated	V1R-like	V2R-like
Astyanax mexicanus	Mexican tetra, blind cave fish	108	40	2	3
Cynoglossus semilaevis	Half-smooth tongue sole	92	22	2	1
Danio rerio	Zebrafish	427	122	8	18
Haplochromis burtoni	Burton's mouthbrooder (cichlid)	98	20	3	1
Maylandia zebra	Zebra Mbuna (cichlid)	100	25	2	3
Neolamprologus brichardi	Cichlid sp.	76	14	2	2
Oreochromis niloticus	Nile tilapia	183	43	6	3
Oryzias latipes	Japanese rice fish	76	31	3	1
Poecilia fo6tazrmosa	Amazon molly	107	35	3	6
Poecilia reticulate	Guppy	98	30	4	9
Pundamilia nyererei	Cichlid sp.	95	18	2	2
Stegastes partitus	Bicolor damselfish	67	25	2	1
Takifugu rubripes	Japanese puffer	78	16	0	4
Xiphophorus maculatus	Platyfish	87	37	2	2
Totals		1692	478	41	56

Table 4.4. Numbers of unique genes (raw counts) of genes identified by alignment of RNA-seq data to a predicted Rainbow Trout transcriptome (Berthelot et al. 2014). Shallow sequencing yielded ~11 to 15 million reads per transcriptome same, while deep sequencing yielded 87 million reads.

Tissue	Shallow sequencing, all samples	Shallow sequencing blood and muscle; deep sequencing OR
Blood	507	302
Muscle	432	211
OR	3112	3664
Grand Total	4051	4177

Figure 4.1. Predicted olfactory gene repertoire for olfactory gene superfamilies (a) vomeronasal type-1 like, (b) vomeronasal type-2 like, (c) trace amine-associated receptors, and (d) main olfactory receptors. Data are shown for fourteen species of teleost, Rainbow trout, and sea lamprey.


a
b
c
d
0.6
0.4
Chinook
sea lamprey
zebrafish
mexican tetra
tilapia
puffer

5 – CONCLUSION

In this dissertation I applied diverse approaches to investigate animal migration. The results presented in Chapter two, fine-scale, stock-specific distribution patterns calibrated using catch-per unit effort abundance, are part of a broader collaborative effort by fishermen and scientists to collect information applicable for increasing the precision of local stock abundance estimates. These combined genetic and fisheries effort data provide a richer and denser in-season representation of fish distribution and abundance than what is possible to obtain from CWT mark-recapture information. Moreover, the use of genetic tags to assess stock distributions overcomes limitations inherent to fish abundance models that are unable to account for wild, unmarked fish. Findings from this study stimulated my interest in developing a deeper understand of the sensory systems utilized by fish to identify their geographic position and home to spawn in their natal river. In general, cue use by fish for navigation and orientation during long-distant migration is not well understood. As discussed in Chapter four, experimental evidence indicates that fish possess an innate “geographic positioning system” derived from geomagnetic field information. Understanding whether or not magnetic field information is used for navigation has important conservation implications because of plans to install electromagnetic-emitting wave energy buoys on the coast margins of the Eastern Pacific Ocean. At present we do not know if the magnetic anomalies caused by these devices might somehow interfere with fish location determination. The work described in chapter four paves the way for developing tools for in situ tests of sensory cells for response to magnetic treatment. Cells could be probed to understanding electromagnetic sensitivity thresholds and the physical effect of magnetic anomalies on sensory cells.

During the final stage of salmonid migration at some point fish switch to olfaction as a primary navigation cue, but when, and which exact receptors and their odorant cues are involved, is unknown. The comparative genomics approach used in Chapter three indicates substantial diversification in two classes of olfactory receptors, trace amine-associated receptors and V2R like receptors. The main olfactory receptors are also highly diversified, but the patterns for that gene class are similar to those of non-migratory teleost fish. The V2R like family is more diversified in salmonids relative to other classes of fish, suggesting this gene plays vital chemosensory role. In conclusion, this study contributes to understanding salmon migration at multiple hierarchical levels, from population-level migratory distributions to the genetic basis of olfactory and putative magnetic sensory systems that organisms use for orientation and navigation during migration.

BIBLIOGRAPHY

Alerstam T, Hedenstrom A, Akesson S (2003) Long-distance migration: evolution and determinants. Oikos 103:247–260. doi: 10.1034/j.1600-0706.2003.12559.x

Allendorf FW, Thorgaard GH (1984) Tetraploidy and the Evolution of Salmonid Fishes. In: Turner BJ (ed) Evolutionary Genetics of Fishes. Springer US, Boston, MA, pp 1–53

Allison PD (2008) Convergence failures in logistic regression. Proceedings of the SAS Global Forum 2008 Conference 360-2008:1–11.

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi: 10.1016/S0022-2836(05)80360-2

Altschul S, Madden TL, Schäffer AA, et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389–3402. doi: 10.1093/nar/25.17.3389

Alves JA, Gunnarsson TG, Hayhow DB, et al (2012) Costs, benefits and fitness consequences of different migratory strategies. Ecology 94:11–17. doi: 10.1890/12-0737.1

Anderson EC, Waples RS, Kalinowski ST (2008) An improved method for predicting the accuracy of genetic stock identification. Canadian Journal of Fisheries and Aquatic Sciences 65:1475–1486. doi: 10.1139/F08-049

Baker CS, Herman LM, Perry A, et al (1986) Migratory movement and population structure of humpback whales (Megaptera novaeangliae) in the central and eastern North Pacific. Marine Ecology Progress Series 31:105–119.

Baker CS, Palumbi SR, Lambertsen RH, et al (1990) Influence of seasonal migration on geographic distribution of mitochondrial DNA haplotypes in humpback whales. Nature 344:238–240. doi: 10.1038/344238a0

Balkwill DL, Maratea D, Blakemore RP (1980) Ultrastructure of a magnetotactic spirillum. J Bacteriol 141:1399–1408.

Banks MA, Rashbrook VK, Calavetta MJ, et al (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.

Baum MJ (2012) Contribution of pheromones processed by the main olfactory system to mate recognition in female mammals. Front Neuroanat 6:20. doi: 10.3389/fnana.2012.00020

Bazylinski DA, Frankel RB (2004) Magnetosome formation in prokaryotes. Nature Reviews Microbiology 2:217–230. doi: 10.1038/nrmicro842

Beacham TD, Lapointe M, Candy JR, et al (2004) DNA in Action: Rapid Application of DNA Variation to Sockeye Salmon Fisheries Management. Conservation Genetics 5:411–416. doi: 10.1023/B:COGE.0000031140.41379.73

Beacham TD, Winther I, Jonsen KL, et al (2008) The Application of Rapid Microsatellite-Based Stock Identification to Management of a Chinook Salmon Troll Fishery off the Queen Charlotte Islands, British Columbia. North American Journal of Fisheries Management 28:849–855. doi: 10.1577/M06-167.1

Beason RC, Semm P (1987) Magnetic responses of the trigeminal nerve system of the bobolink (Dolichonyx oryzivorus). Neurosci Lett 80:229–234.

Bellinger MR, Banks MA (2007) Genetics. Oregon Salmon Commission, pp 1–30

Berthelot C, Brunet F, Chalopin D, et al (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature Communications. doi: 10.1038/ncomms4657

Blakemore RP, Frankel RB (1981) Magnetic Navigation in Bacteria. Scientific American 245:58–65. doi: 10.1038/scientificamerican1281-58

Blankenberg D, Gordon A, Von Kuster G, et al (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics 26:1783–1785. doi: 10.1093/bioinformatics/btq281

Both C, Visser ME (2001) Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. Nature 411:296–298. doi: 10.1038/35077063

Braemer W (1960) A Critical Review of the Sun-Azimuth Hypothesis. Cold Spring Harbor Symposia on Quantitative Biology 25:413–427. doi: 10.1101/SQB.1960.025.01.044

Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell 65:175–187.

Burke BJ, Peterson WT, Beckman BR, et al (2013) Multivariate Models of Adult Pacific Salmon Returns. PLoS ONE 8:e54134. doi: 10.1371/journal.pone.0054134

Butler RF, Banerjee SK (1975) Theoretical single-domain grain size range in magnetite and titanomagnetite. Journal of Geophysical Research 80:4049–4058. doi: 10.1029/JB080i029p04049

Calof AL, Lander AD (1991) Relationship between neuronal migration and cell-substratum adhesion: laminin and merosin promote olfactory neuronal migration but are anti-adhesive. J Cell Biol 115:779–794.

Clemento AJ, AbadíA-Cardoso A, Starks HA, Garza JC (2011) Discovery and characterization of single nucleotide polymorphisms in Chinook salmon, <i>Oncorhynchus tshawytscha<i>. Molecular Ecology Resources 11:50–66. doi: 10.1111/j.1755-0998.2010.02972.x

Clemento AJ, Crandall ED, Garza JC, Anderson EC (2014) Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Pacific Chinook salmon (Oncorhynchus tshawytscha) in the California Current large marine ecosystem. Fishery Bulletin 112–130.

Dann TH, Habicht C, Baker TT, Seeb JE (2013) Exploiting genetic diversity to balance conservation and harvest of migratory salmon. Canadian Journal of Fisheries and Aquatic Sciences 70:785–793.

Davidson WS, Koop BF, Jones SJM, et al (2010) Sequencing the genome of the Atlantic salmon (Salmo salar). Genome Biol 11:403. doi: 10.1186/gb-2010-11-9-403

Delmore KE, Fox JW, Irwin DE (2012) Dramatic intraspecific differences in migratory routes, stopover sites and wintering areas, revealed using light-level geolocators. Proceedings of the Royal Society B: Biological Sciences 279:4582–4589. doi: 10.1098/rspb.2012.1229

De Wit P, Pespeni MH, Ladner JT, et al (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Molecular Ecology Resources 12:1058–1067. doi: 10.1111/1755-0998.12003

Diebel CE, Proksch R, Green CR, et al (2000) Magnetite defines a vertebrate magnetoreceptor. Nature 406:299–302. doi: 10.1038/35018561

Di Genova A, Aravena A, Zapata L, et al (2011) SalmonDB: a bioinformatics resource for Salmo salar and Oncorhynchus mykiss. Database 2011:bar050–bar050. doi: 10.1093/database/bar050

Dittman AH, Quinn TP (1996) Homing in Pacific salmon: mechanisms and ecological basis. Journal of Experimental Biology 199:83–91.

Di Y, Schafer DW, Cumbie JS, Chang JH (2011) The NBP Negative Binomial Model for Assessing Differential Gene Expression from RNA-Seq. Statistical Applications in Genetics and Molecular Biology 10:1–28. doi: 10.2202/1544-6115.1637

Dulac C, Axel R (1995) A novel family of genes encoding putative pheromone receptors in mammals. Cell 83:195–206.

Eder SHK, Cadiou H, Muhamad A, et al (2012) Magnetic characterization of isolated candidate vertebrate magnetoreceptor cells. Proceedings of the National Academy of Sciences 109:12022–12027. doi: 10.1073/pnas.1205653109

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9:397–405. doi: 10.1038/nrg2337

Fleissner G, Holtkamp-Rötzler E, Hanzlik M, et al (2003) Ultrastructural analysis of a putative magnetoreceptor in the beak of homing pigeons. J Comp Neurol 458:350–360. doi: 10.1002/cne.10579

Fox SE, Christie MR, Marine M, et al (2014) Sequencing and characterization of the anadromous steelhead (Oncorhynchus mykiss) transcriptome. Marine Genomics 15:13–15. doi: 10.1016/j.margen.2013.12.001

Frankel RB, Blakemore RP, Wolfe RS (1979) Magnetite in Freshwater Magnetotactic Bacteria. Science 203:1355–1356. doi: 10.1126/science.203.4387.1355

Freake MJ, Muheim R, Phillips JB (2006) Magnetic maps in animals: a theory comes of age? The Quarterly Review of Biology 81:327–347.

Freitag J, Krieger J, Strotmann J, Breer H (1995) Two classes of olfactory receptors in Xenopus laevis. Neuron 15:1383–1392.

Garza JC, Anderson EC (2007) Large scale parentage inference as an alternative to coded-wire tags for salmon fishery management. PSC Genetic Stock Identification Workshop (May and September 2007): Logistics Workgroup Final Report and Recommendations. Available from http://www.psc.org/info_genetic_stock_id.htm. Pacific Salmon Commission, Vancouver, British Columbia, Canada, pp 45–55

Gegear RJ, Foley LE, Casselman A, Reppert SM (2010) Animal cryptochromes mediate magnetoreception by an unconventional photochemical mechanism. Nature 463:804–807. doi: 10.1038/nature08719

Glusman G, Bahar A, Sharon D, et al (2000) The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mamm Genome 11:1016–1023.

Gould JL, Kirschvink JL, Deffeyes KS (1978) Bees Have Magnetic Remanence. Science 201:1026–1028. doi: 10.1126/science.201.4360.1026

Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644–652. doi: 10.1038/nbt.1883

Grant WS, Milner GB, Krasnowski P, Utter FM (1980) Use of Biochemical Genetic Variants for Identification of Sockeye Salmon (<i>Oncorhynchus nerka<i>) Stocks in Cook Inlet, Alaska. Canadian Journal of Fisheries and Aquatic Sciences 37:1236–1247. doi: 10.1139/f80-159

Green CR, Holloway H, Walker MM (2001) Detection of submicroscopic magnetite particles using reflectance mode confocal laser scanning microscopy. Cell Biol Int 25:985–990. doi: 10.1006/cbir.2001.0773

Grunberg K, Muller E-C, Otto A, et al (2004) Biochemical and Proteomic Analysis of the Magnetosome Membrane in Magnetospirillum gryphiswaldense. Applied and Environmental Microbiology 70:1040–1050. doi: 10.1128/AEM.70.2.1040-1050.2004

Grünberg K, Wawer C, Tebo BM, Schüler D (2001) A large gene cluster encoding several magnetosome proteins is conserved in different species of magnetotactic bacteria. Appl Environ Microbiol 67:4573–4582.

Haas BJ, Papanicolaou A, Yassour M, et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. doi: 10.1038/nprot.2013.084

Hashiguchi Y, Nishida M (2006) Evolution and origin of vomeronasal-type odorant receptor gene repertoire in fishes. BMC Evol Biol 6:76. doi: 10.1186/1471-2148-6-76

Hashiguchi Y, Nishida M (2007) Evolution of trace amine associated receptor (TAAR) gene family in vertebrates: lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium. Mol Biol Evol 24:2099–2107. doi: 10.1093/molbev/msm140

Hasler AD, Schwassman HO (1960) Sun orientation of fish at different latitudes. Cold Spring Harb Symp Quant Biol 25:429–441.

Hasler AD, Wisby WJ (1951) Discrimination of Stream Odors by Fishes and Its Relation to Parent Stream Behavior. The American Naturalist 85:223–238. doi: 10.1086/281672

Healey MC (1983) Coastwide distribution and ocean migration patterns of stream- and ocean-type Chinook salmon, Oncorhynchus tshawytscha. Canadian Field Naturalist 97:427–433.

Heard WR, Shevlyakov E, Zikunova OV, McNicol RE (2007) Chinook Salmon - Trends in Abundance and Biological Characteristics. North Pacific Anadromous Fish Commission Bulletin 4:77–91.

Hedgecock D, Banks MA, Rashbrook VK, et al (2001) Applications of population genetics to conservation of chinook salmon diversity in the Central Valley. Contributions to the Biology of Central Valley Salmonids, California Department of Fish and Game, Brown, R. L. Fish Bulletin, UC San Diego: Scripps Institution of Oceanography Library, pp 45–69

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Molecular Ecology Resources 11:137–149. doi: 10.1111/j.1755-0998.2010.02958.x

Heyers D, Zapka M, Hoffmeister M, et al (2010) Magnetic field changes activate the trigeminal brainstem complex in a migratory bird. Proceedings of the National Academy of Sciences 107:9394–9399. doi: 10.1073/pnas.0907068107

Hinke JT, Watters GM, Boehlert GW, Zedonis P (2005) Ocean habitat use in autumn by Chinook salmon in coastal waters of Oregon and California. Marine Ecology Progress Series 285:181–192. doi: 10.3354/meps285181

Hobday AJ, Boehlert GW (2001) The role of coastal ocean variation in spatial and temporal patterns in survival and size of coho salmon (<i>Oncorhynchus kisutch<i>). Canadian Journal of Fisheries and Aquatic Sciences 58:2021–2036. doi: 10.1139/cjfas-58-10-2021

Holland RA, Kirschvink JL, Doak TG, Wikelski M (2008) Bats Use Magnetite to Detect the Earth's Magnetic Field. PLoS ONE 3:e1676. doi: 10.1371/journal.pone.0001676

Irwin WP, Lohmann KJ (2005) Disruption of magnetic orientation in hatchling loggerhead sea turtles by pulsed magnetic fields. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 191:475–480. doi: 10.1007/s00359-005-0609-9

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Molecular Ecology Notes 6:998–1002. doi: 10.1111/j.1471-8286.2006.01428.x

Jefferts KB, Bergman PK, Fiscus HF (1963) A Coded Wire Identification System for Macro-Organisms. Nature 198:460–462. doi: 10.1038/198460a0

Johnsen S, Lohmann KJ (2008) Magnetoreception in animals feature article. Physics Today 61:29–35. doi: 10.1063/1.2897947

Johnson KJ (2004) Regional Overview of Coded Wire Tagging of Anadromous Salmon and Steelhead in northwest America. Regional Mark Processing Center, Pacific States Marine Fisherie s Commission 205 SE Spokane Street, Suite 100, Portland, Oregon 97202-6413, USA

Johnson MA, Tsai L, Roy DS, et al (2012) Neurons expressing trace amine-associated receptors project to discrete glomeruli and constitute an olfactory subsystem. Proc Natl Acad Sci USA 109:13410–13415. doi: 10.1073/pnas.1206724109

Kalinowsi ST, Manlove KR, Taper ML (2007) ONCOR A computer program for Genetic Stock Identification. Department of Ecology, Montana State University, Bozeman MT 59717. Available for download from http://www.montana.edu/kalinowski

Kallio-Nyberg I, Koljonen M-L, Saloniemi I (2000) Effect of maternal and paternal line on spatial and temporal marine distribution in Atlantic salmon. Anim Behav 60:377–384. doi: 10.1006/anbe.2000.1465

Kallio-Nyberg I, Peltonen H, Rita H (1999) Effects of stock-specific and environmental factors on the feeding migration of Atlantic salmon (*Salmo salar*) in the Baltic Sea. Canadian Journal of Fisheries and Aquatic Sciences 56:853–861. doi: 10.1139/cjfas-56-5-853

Kirschvink JL, Gould JL (1981) Biogenic magnetite as a basis for magnetic field detection in animals. BioSystems 13:181–201.

Kirschvink JL, Hagadorn JW (2000) A Grand Unified Theory of Biomineralization. The Biomineralization of Nano- and Micro-structures, Bäuerlein E. Wiley-VCH, Weinheim, Germany, pp 139–150

Kirschvink JL, Winklhofer M, Walker MM (2010) Biophysics of magnetic orientation: strengthening the interface between theory and experimental design. Journal of The Royal Society Interface 7:S179–S191. doi: 10.1098/rsif.2009.0491.focus

Kishkinev D, Chernetsov N, Heyers D, Mouritsen H (2013) Migratory Reed Warblers Need Intact Trigeminal Nerves to Correct for a 1,000 km Eastward Displacement. PLoS ONE 8:e65847. doi: 10.1371/journal.pone.0065847

Lagerquist BA, Mate BR, Ortega-Ortiz JG, et al (2008) Migratory movements and surfacing rates of humpback whales (Megaptera novaeangliae) satellite tagged at Socorro Island, Mexico. Marine Mammal Science 815–830. doi: 10.1111/j.1748-7692.2008.00217.x

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359. doi: 10.1038/nmeth.1923

Leask MJM (1977) A physicochemical mechanism for magnetic field detection by migratory birds and homing pigeons. Nature 267:144–145. doi: 10.1038/267144a0

Lefèvre CT, Trubitsyn D, Abreu F, et al (2013) Comparative genomic analysis of magnetotactic bacteria from the *Deltaproteobacteria* provides new insights into magnetite and greigite magnetosome genes required for magnetotaxis: New insights in genes required for magnetotaxis. Environmental Microbiology n/a–n/a. doi: 10.1111/1462-2920.12128

Li L, Stoeckert CJJ, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Research 13:2178–2189. doi: 10.1101/gr.1224503

Lohmann KJ, Lohmann CMF (1993) A Light-Independent Magnetic Compass in the Leatherback Sea Turtle. Biological Bulletin 185:149–151. doi: 10.2307/1542138

Lohmann KJ, Putman NF, Lohmann CMF (2008) Geomagnetic imprinting: A unifying hypothesis of long-distance natal homing in salmon and sea turtles. Proceedings of the National Academy of Sciences 105:19096–19101. doi: 10.1073/pnas.0801859105

Lohße A, Ullrich S, Katzmann E, et al (2011) Functional Analysis of the Magnetosome Island in Magnetospirillum gryphiswaldense: The mamAB Operon Is Sufficient for Magnetite Biomineralization. PLoS ONE 6:e25561. doi: 10.1371/journal.pone.0025561

Lopes RJ, Marques JC, Wennerberg L (2006) Migratory connectivity and temporal segregation of dunlin (Calidris alpina) in Portugal: evidence from morphology, ringing recoveries and mtDNA. Journal of Ornithology 147:385–394. doi: 10.1007/s10336-005-0048-y

Lowenstam HA (1962) Magnetite in Denticle Capping in Recent Chitons (Polyplacophora). Geological Society of America Bulletin 73:435–438. doi: 10.1130/0016-7606(1962)73[435:MIDCIR]2.0.CO;2

Mann S, Sparks NH, Walker MM, Kirschvink JL (1988) Ultrastructure, morphology and organization of biogenic magnetite from sockeye salmon, Oncorhynchus nerka: implications for magnetoreception. J Exp Biol 140:35–49.

Marasco RJ, Goodman D, Grimes CB, et al (2007) Ecosystem-based fisheries management: some practical suggestions. Canadian Journal of Fisheries and Aquatic Sciences 64:928–939. doi: 10.1139/f07-062

Maunder MN, Sibert JR, Fonteneau A, et al (2006) Interpreting catch per unit effort data to assess the status of individual stocks and communities. ICES Journal of Marine Science 63:1373–1385. doi: 10.1016/j.icesjms.2006.05.008

McDonald J (1981) The Stock Concept and its Application to British Columbia Salmon Fisheries. Canadian Journal of Fisheries and Aquatic Sciences 38:1657–1664. doi: 10.1139/f81-214

Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol 11:699–704.

Mi H, Lazareva-Ulitsky B, Loo R, et al (2004) The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Research 33:D284–D288. doi: 10.1093/nar/gki078

Milner GB, Teel DJ, Utter FM, Winans GA (1985) A Genetic Method of Stock Identification in Mixed Populations of Pacific Salmon, <i>Oncorhynchus<i> spp. Marine Fisheries Review 47:1–8.

Mombaerts . (1999) Molecular biology of odorant receptors in vertebrates. Annu Rev Neurosci 22:487–509. doi: 10.1146/annurev.neuro.22.1.487

Mombaerts P (2004) Genes and ligands for odorant, vomeronasal and taste receptors. Nat Rev Neurosci 5:263–278. doi: 10.1038/nrn1365

Moore A, Freake SM, Thomas IM (1990) Magnetic Particles in the Lateral Line of the Atlantic Salmon (Salmo salar L.). Philosophical Transactions of the Royal Society B: Biological Sciences 329:11–15. doi: 10.1098/rstb.1990.0145

Moore A, Riley WD (2009) Magnetic particles associated with the lateral line of the European eel *Anguilla anguilla*. Journal of Fish Biology 74:1629–1634. doi: 10.1111/j.1095-8649.2009.02197.x

Mora CV, Davison M, Martin Wild J, Walker MM (2004) Magnetoreception and its trigeminal mediation in the homing pigeon. Nature 432:508–511. doi: 10.1038/nature03077

Moran P, Teel DJ, Banks MA, et al (2013) Divergent life-history races do not represent Chinook salmon coast-wide: the importance of scale in Quaternary biogeography. Canadian Journal of Fisheries and Aquatic Sciences 70:415–435. doi: 10.1139/cjfas-2012-0135

Murat D, Quinlan A, Vali H, Komeili A (2010) Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. Proceedings of the National Academy of Sciences 107:5593–5598. doi: 10.1073/pnas.0914439107

Natural Resources Consultants, Inc. (1998) Northwest Emergency Assistance Plan at-sea Research Programs. Final Report. Prepared for Pacific States Marine Fisheries Commission. Natural Resources Consultants, Inc, Seattle, Washington.

Neave F (1964) Ocean migration of Pacific salmon. Journal of Fisheries Research Board Canada 21:1227–1244.

Nguyen DT, Lee K, Choi H, et al (2012) The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome. BMC Genomics 13:584. doi: 10.1186/1471-2164-13-584

Nicholas JW, Hankin DG (1988) Chinook salmon populations in Oregon coastal river basins: description of life histories and assessment of recent trends in run strengths. Oregon Department of Fish and Wildlife, Fish Division Information Reports 88-1, Corvallis.

Niimura Y (2009) Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. Hum Genomics 4:107–118.

Niimura Y, Nei M (2005) Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. Proc Natl Acad Sci USA 102:6039–6044. doi: 10.1073/pnas.0501922102

Nishimune H, Sanes JR, Carlson SS (2004) A synaptic laminin-calcium channel interaction organizes active zones in motor nerve terminals. Nature 432:580–587. doi: 10.1038/nature03112

Nordeng H (1971) Is the local orientation of anadromous fishes determined by pheromones□? Nature 233:411–413. doi: 10.1038/233411a0

Nudelman H, Zarivach R (2014) Structure prediction of magnetosome-associated proteins. Frontiers in Microbiology. doi: 10.3389/fmicb.2014.00009

Okamura Y, Takeyama H, Matsunaga T (2001) A magnetosome-specific GTPase from the magnetic bacterium Magnetospirillum magneticum AMB-1. J Biol Chem 276:48183–48188. doi: 10.1074/jbc.M106408200

Okuda Y, Denda K, Fukumori Y (1996) Cloning and sequencing of a gene encoding a new member of the tetratricopeptide protein family from magnetosomes of Magnetospirillum magnetotacticum. Gene 171:99–102.

Pacific Fishery Management Council (2010) Preseason Report III: Analysis of Council Adopted Management Measures for 2010 Ocean Salmon Fisheries. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 101, Portland, Oregon 97220-1384.

Pacific Fishery Management Council A (2007) Fisheries Regulation Assessment Model (FRAM) An Overview for Coho and Chinook. Pacific Fishery Management Council, 7700 NE Abassador Place, Suite 101, Portland, OR

Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. Doctoral Thesis, University of Dublin, Dublin.

Pearcy WG (1992) Ocean ecology of North Pacific salmonids. University of Washington, Washington Sea Grant, Seattle.

Peyronnet A, Friedland KD, Ó Maoileidigh N (2008) Different ocean and climate factors control the marine survival of wild and hatchery Atlantic salmon <i>Salmo salar<i> in the north-east Atlantic Ocean. Journal of Fish Biology 73:945–962. doi: 10.1111/j.1095-8649.2008.01984.x

Putman NF, Lohmann KJ, Putman EM, et al (2013) Evidence for Geomagnetic Imprinting as a Homing Mechanism in Pacific Salmon. Current Biology 23:312–316. doi: 10.1016/j.cub.2012.12.041

Putman NF, Scanlan MM, Billman EJ, et al (2014a) An Inherited Magnetic Map Guides Ocean Navigation in Juvenile Pacific Salmon. Current Biology 24:446–450. doi: 10.1016/j.cub.2014.01.017

Putman NF, Scanlan MM, Billman EJ, et al (2014b) An inherited magnetic map guides ocean navigation in juvenile Pacific salmon. Current Biology 24:446–450. doi: 10.1016/j.cub.2014.01.017

Qian X, Ba Y, Zhuang Q, Zhong G (2014) RNA-Seq Technology and Its Application in Fish Transcriptomics. OMICS: A Journal of Integrative Biology 18:98–110. doi: 10.1089/omi.2013.0110

Quinn NL, Levenkova N, Chow W, et al (2008) Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. BMC Genomics 9:404. doi: 10.1186/1471-2164-9-404

Quinn TP (1980) Evidence for celestial and magnetic compass orientation in lake migrating sockeye salmon fry. Journal of Comparative Physiology□? A 137:243–248. doi: 10.1007/BF00657119

Quinn TP, Brannon EL (1982) The use of celestial and magnetic cues by orienting sockeye salmon smolts. Journal of Comparative Physiology A 147:547–552. doi: 10.1007/BF00612020

Quinn TP, Chamberlin J, Brannon EL (2011) Experimental evidence of population-specific marine spatial distributions of Chinook salmon, Oncorhynchus tshawytscha. Environmental Biology of Fishes 92:313–322. doi: 10.1007/s10641-011-9841-z

Quinn TP, Dittman AH (1990) Pacific salmon migrations and homing: mechanisms and adaptive significance. Trends in Ecology & Evolution 5:174–177. doi: 10.1016/0169-5347(90)90205-R

Richter M, Kube M, Bazylinski DA, et al (2007) Comparative genome analysis of four magnetotactic bacteria reveals a complex set of group-specific genes

implicated in magnetosome biomineralization and function. J Bacteriol 189:4899–4910. doi: 10.1128/JB.00119-07

Ritz T, Adem S, Schulten K (2000) A model for photoreceptor-based magnetoreception in birds. Biophys J 78:707–718. doi: 10.1016/S0006-3495(00)76629-X

Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proc Natl Acad Sci USA 100 Suppl 2:14537–14542. doi: 10.1073/pnas.2335847100

Santangelo AM, de Souza FSJ, Franchini LF, et al (2007) Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene. PLoS Genetics 3:e166. doi: 10.1371/journal.pgen.0030166

Sato K (2001) Whole-cell Response Characteristics of Ciliated and Microvillous Olfactory Receptor Neurons to Amino Acids, Pheromone Candidates and Urine in Rainbow Trout. Chemical Senses 26:1145–1156. doi: 10.1093/chemse/26.9.1145

Satterthwaite WH, Mohr MS, O'Farrell MR, et al (2014) Use of Genetic Stock Identification Data for Comparison of the Ocean Spatial Distribution, Size at Age, and Fishery Exposure of an Untagged Stock and Its Indicator: California Coastal versus Klamath River Chinook Salmon. 143:117–133.

Satterthwaite WH, Mohr MS, O'Farrell MR, Wells BK (2013) A comparison of temporal patterns in the ocean spatial distribution of California's Central Valley Chinook salmon runs. Canadian Journal of Fisheries and Aquatic Sciences 70:574–584. doi: 10.1139/cjfas-2012-0395

Schlegel PA (2007) Spontaneous preferences for magnetic compass direction in the American red-spotted newt, Notophthalmus viridescens (Salamandridae, Urodela). Journal of Ethology 25:177–184. doi: 10.1007/s10164-006-0016-x

Schübbe S, Kube M, Scheffel A, et al (2003a) Characterization of a spontaneous nonmagnetic mutant of Magnetospirillum gryphiswaldense reveals a large deletion comprising a putative magnetosome island. J Bacteriol 185:5779–5790.

Schübbe S, Kube M, Scheffel A, et al (2003b) Characterization of a spontaneous nonmagnetic mutant of Magnetospirillum gryphiswaldense reveals a large deletion comprising a putative magnetosome island. J Bacteriol 185:5779–5790.

Schüler D (2008) Genetics and cell biology of magnetosome formation in magnetotactic bacteria. FEMS Microbiology Reviews 32:654–672. doi: 10.1111/j.1574-6976.2008.00116.x

Schweiger O, Settele J, Kudrna O, et al (2008) Climate change can cause spatial mismatch of trophically interacting species. Ecology 89:3472–3479. doi: 10.1890/07-1748.1

Seeb LW, Antonovich A, Banks MA, et al (2007) Development of a Standardized DNA Database for Chinook Salmon. Fisheries 32:540–552. doi: 10.1577/1548-8446(2007)32[540:DOASDD]2.0.CO;2

Shaklee JB, Beacham TD, Seeb L, White BA (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fisheries Research 43:45–78. doi: 10.1016/S0165-7836(99)00066-1

Shaul L, Weitkamp L, Simpson K, Sawada J (2007) Trends in abundance and size of coho salmon in the Pacific Rim. North Pacific Anadromous Fish Commission Bulletin 4:93–107.

Shoji T, Yamamoto Y, Nishikawa D, et al (2003) Amino acids in stream water are essential for salmon homing migration. Fish Physiology and Biochemistry 28:249–251. doi: 10.1023/B:FISH.0000030544.64774.f6

Siponen MI, Adryanczyk G, Ginet N, et al (2012) Magnetochrome: a c-type cytochrome domain specific to magnetotatic bacteria. Biochem Soc Trans 40:1319–1323. doi: 10.1042/BST20120104

Smith JJ, Kuraku S, Holt C, et al (2013) Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nature Genetics 45:415–421. doi: 10.1038/ng.2568

Speca DJ, Lin DM, Sorensen PW, et al (1999) Functional identification of a goldfish odorant receptor. Neuron 23:487–498.

Stapput K, Thalau P, Wiltschko R, Wiltschko W (2008) Orientation of Birds in Total Darkness. Current Biology 18:602–606. doi: 10.1016/j.cub.2008.03.046

Steele CA, Anderson EC, Ackerman MW, et al (2013) A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. Canadian Journal of Fisheries and Aquatic Sciences 70:1046–1054. doi: 10.1139/cjfas-2012-0451

Suarez A, Ueno T, Huebner R, et al (2014) Bin/Amphiphysin/Rvs (BAR) family members bend membranes in cells. Scientific Reports. doi: 10.1038/srep04693

Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30:2725–2729. doi: 10.1093/molbev/mst197

Tanaka M, Arakaki A, Matsunaga T (2010) Identification and functional characterization of liposome tubulation protein from magnetotactic bacteria. Molecular Microbiology 76:480–488. doi: 10.1111/j.1365-2958.2010.07117.x

Tessarolo JA, Tabesh MJ, Nesbitt M, Davidson WS (2014) Genomic Organization and Evolution of the Trace Amine-Associated Receptor (TAAR) Repertoire in Atlantic Salmon (Salmo salar). G3: Genes|Genomes|Genetics. doi: 10.1534/g3.114.010660

Thériault V, Moyer GR, Jackson LS, et al (2011) Reduced reproductive success of hatchery coho salmon in the wild: insights into most likely mechanisms. Molecular Ecology 20:1860–1869. doi: 10.1111/j.1365-294X.2011.05058.x

Thomas PD, Campbell MJ, Kejariwal A, et al (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Research 13:2129–2141. doi: 10.1101/gr.772403

Timpl R, Tisi D, Talts JF, et al (2000) Structure and function of laminin LG modules. Matrix Biol 19:309–317.

Touhara K, Vosshall LB (2009) Sensing Odorants and Pheromones with Chemosensory Receptors. Annual Review of Physiology 71:307–332. doi: 10.1146/annurev.physiol.010908.163209

Trapnell C, Roberts A, Goff L, et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7:562–578. doi: 10.1038/nprot.2012.016

Treiber CD, Salzer MC, Riegler J, et al (2012) Clusters of iron-rich cells in the upper beak of pigeons are macrophages not magnetosensitive neurons. Nature. doi: 10.1038/nature11046

Tufte ER (1986) The Visual Display of Quantitative Information., 7th printing. Graphics Press, Cheshire, CT

Ueda H (2012) Physiological mechanisms of imprinting and homing migration in Pacific salmon Oncorhynchus spp. Journal of Fish Biology 81:543–558. doi: 10.1111/j.1095-8649.2012.03354.x

Ullrich S, Kube M, Schübbe S, et al (2005) A hypervariable 130-kilobase genomic region of Magnetospirillum gryphiswaldense comprises a magnetosome island which undergoes frequent rearrangements during stationary growth. J Bacteriol 187:7176–7184. doi: 10.1128/JB.187.21.7176-7184.2005

Venkatesh B (2003) Evolution and diversity of fish genomes. Curr Opin Genet Dev 13:588–592.

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypothesis. Econometrica 57:307–333.

Walcott C, Gould J, Kirschvink J (1979) Pigeons have magnets. Science 205:1027–1029. doi: 10.1126/science.472725

Waldman J, Grunwald C, Wirgin I (2008) Sea lamprey Petromyzon marinus: an exception to the rule of homing in anadromous fishes. Biology Letters 4:659–662. doi: 10.1098/rsbl.2008.0341

Walker MM, Diebel CE, Haugh CV, et al (1997a) Structure and function of the vertebrate magnetic sense. Nature 390:371–376.

Walker MM, Diebel CE, Haugh CV, et al (1997b) Structure and function of the vertebrate magnetic sense. Nature 371–376.

Walker MM, Quinn TP, Kirschvink JL, Groot C (1988) Production of single-domain magnetite throughout life by sockeye salmon, Oncorhynchus nerka. J Exp Biol 140:51–63.

Wang S, Furmanek T, Kryvi H, et al (2014) Transcriptome sequencing of Atlantic salmon (Salmo salar L.) notochord prior to development of the vertebrae provides clues to regulation of positional fate, chordoblast lineage and mineralisation. BMC Genomics 15:141. doi: 10.1186/1471-2164-15-141

Waples R (2001) Characterizing diversity in salmon from the Pacific Northwest. Journal of Fish Biology 59:1–41. doi: 10.1006/jfbi.2001.1764

Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9:423–449. doi: 10.1111/j.1467-2979.2008.00303.x

Wei J (2009) Nanoscale Analysis of Natural and Artificial Magnetic Objects: Particles, Thin Films and Recording Heads. Universität des Saarlandes

Weitkamp LA (2010) Marine Distributions of Chinook Salmon from the West Coast of North America Determined by Coded Wire Tag Recoveries. Transactions of the American Fisheries Society 139:147–170. doi: 10.1577/T08-225.1

Weitkamp L, Neely K (2002) Coho salmon (Oncorhynchus kisutch) ocean migration patterns: insight from marine coded-wire tag recoveries. Canadian Journal of Fisheries and Aquatic Sciences 59:1100–1115. doi: 10.1139/f02-075

Wells BK, Grimes CB, Sneva JG, et al (2008) Relationships between oceanic conditions and growth of Chinook salmon (<i>Oncorhynchus tshawytscha<i>) from California, Washington, and Alaska, USA. Fisheries Oceanography 17:101–125. doi: 10.1111/j.1365-2419.2008.00467.x

Wessa P (2012)Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7, URL http://www.wessa.net/

Wiltschko W, Wiltschko R (2007) Magnetoreception in birds: two receptors for two different tasks. Journal of Ornithology 148:61–76. doi: 10.1007/s10336-007-0233-2

Winans GA, Viele D, Grover A, et al (2001) An Update of Genetic Stock Identification of Chinook Salmon in the Pacific Northwest: Test Fisheries in California. Reviews in Fisheries Science 9:213–237. doi: 10.1080/20016491101753

Winther I, Beacham TD (2009) The application of Chinook salmon stock composition data to management of the Queen Charlotte Islands troll fishery. Pacific Salmon: Ecology and Management of Western Alaska's Populations, In C. C. Krueger, and C. E. Zimmerman, editors. American Fisheries Society, Symposium 70, pp 977–1004

Wisby WJ, Hasler AD (1954) Effect of Olfactory Occlusion on Migrating Silver Salmon (*O . kisutch*). Journal of the Fisheries Research Board of Canada 11:472–478. doi: 10.1139/f54-031

Yamamoto D, Taoka A, Uchihashi T, et al (2010) Visualization and structural analysis of the bacterial magnetic organelle magnetosome using atomic force microscopy. Proceedings of the National Academy of Sciences 107:9382–9387. doi: 10.1073/pnas.1001870107

Zeytuni N, Ozyamak E, Ben-Harush K, et al (2011) Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. Proceedings of the National Academy of Sciences 108:E480–E487. doi: 10.1073/pnas.1103367108

Appendix 2.1. List of regions and populations in GAPS (Genetic Analysis of Pacific Salmonids) baseline v 3. Run time, H. (H) or wild (W) origin, life stage, collection data, and analysis laboratory are given (from Seeb et al. 2007 and unpublished data).

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
1	Central Valley fall	Battle Creek	Fa	W	Adult	2002, 2003	SWFSC
		Feather H. fall	Fa	H	Adult	2003	SWFSC
		Stanislaus River	Fa	W	Adult	2002	SWFSC
		Tuolumne River	Fa	W	Adult	2002	SWFSC
2	Central Valley spring	Butte Creek	Sp	W	Adult	2002, 2003	SWFSC
		Deer Creek spring	Sp	W	Adult	2002	SWFSC
		Feather H. spring	Sp	H	Adult	2003	SWFSC
		Mill Creek spring	Sp	W	Adult	2002, 2003	SWFSC
3	Central Valley winter	Sacramento River winter	Wi	W/H	Adult	1992 - 1995, 1997, 1998, 2001, 2003, 2004	SWFSC
4	California Coast	Eel River	Fa	W	Adult	2000, 2001	SWFSC
		Russian River	Fa	W	Juvenile	2001	SWFSC
5	Klamath River	Klamath River fall	Fa	W	Adult	2004	SWFSC

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Trinity H. fall	Fa	H	Adult	1992	SWFSC
		Trinity H. spring	Sp	H	Adult	1992	SWFSC
6	N California/S Oregon Coast	Chetco	Fa	W	Adult	2004	OSU
7	Rogue River	Applegate	Fa	W	Adult	2004	OSU
		Cole Rivers H.	Sp	H	Adult	2004	OSU
8	Mid Oregon Coast	Coquille	Fa	W	Adult	2000	OSU
		Siuslaw	Fa	W	Adult	2001	OSU
		North Umpqua	Sp	W	Adult	2004	OSU
		Coos[3]	Fa	H/W	Adult	2000, 2005	OSU
		Millicoma[3]	Fa	H/W	Adult	2000, 2005	OSU
		Sixes[3]	Fa	W	Adult	2005	OSU
		Elk[3]	Fa	H	Adult	2004	OSU
		South Umpqua[3]	Fa	H/W	Adult	2002	OSU
9	North Oregon Coast	Alsea	Fa	W	Adult	2004	OSU

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Nehalem	Fa	W	Adult	2000, 2002	OSU
		Siletz	Fa	W	Adult	2000	OSU
		Salmon[3]	Fa	W	Adult	2003	OSU
		Yaquina[3]	Fa	W	Adult	2005	OSU
		Necanicum[3]	Fa	W	Adult	2005	OSU
		Trask[3]	Fa	W	Adult	2005	OSU
		Wilson[3]	Fa	W	Adult	2005	OSU
		Kilchis[3]	Fa	W	Adult	2005	OSU
10	Lower Columbia R. spring	Cowlitz H. spring	Sp	H		2004	CRITFC
		Kalama H. spring	Sp	H		2004	CRITFC
		Lewis H. spring	Sp	H		2004	CRITFC
11	Lower Columbia R. fall	Cowlitz H. fall	Fa	H		2004	CRITFC
		Elochoman River	Fa	W	Adult	1995	WDFW
		Green River	Fa	W	Adult	2000	WDFW
		Lewis fall	Fa	W	Adult	2003	WDFW

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Lewis North Fork Su[3]	Fa	W	Adult	2004	WDFW
		Sandy	Fa	W	Adult	2002, 2004	OSU
		Washougal River	Fa	W	Adult	2005	WDFW
12	Willamette River	McKenzie	Sp	H	Adult	2002, 2004	OSU
		North Santiam	Sp	H	Adult	2002, 2004	OSU
13	Mid Columbia R. tule fall	Spring Creek	Fa	H		2001, 2002	CRITFC
14	Mid and Upper Columbia R. spring	American River	Sp	W	Adult	2003	WDFW
		Carson H.	Sp	H		2001, 2004	CRITFC
		Carson NFH	Sp/Su	H	Juvenile	2006	CRITFC
		Entiat NFH	Sp	H	Juvenile	2002	CRITFC
		Granite Creek	Sp	W	Adult	2005	CRITFC

Region #	Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
	Granite Creek	Sp	W	Adult	2006	NWFSC
	John Day	Sp	W	Juvenile	2000	OSU
				Adult	2004	OSU
	Klickitat H.	Sp	H	Adult	2002, 2006	CRITFC
	Klickitat River	Sp	W	Adult	2005	CRITFC
	Little Naches	Sp	W	Adult	2004	WDFW
	Little White Salmon	Sp/Su	H	Juvenile	2005	CRITFC
	Methow River	Sp	H	Juvenile	1998, 2000	CRITFC
	Middle Fork John Day	Sp	W	Adult	2004	OSU
		Sp	W	Adult	2005	CRITFC
		Sp	W	Adult	2006	NWFSC
	Naches River	Sp	W	Adult	1989, 1993	WDFW
	North Fork John Day	Sp	W	Adult	2004	OSU
		Sp	W	Adult	2005	CRITFC
		Sp	W	Adult	2006	NWFSC

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Shitike Creek	Sp	H	Juvenile	2003, 2004	CRITFC
		Twisp River	Sp	W	Adult	2001, 2005	WDFW
		Upper John Day	Sp	W	Adult	2004	OSU
			Sp	W	Adult	2005	CRITFC
			Sp	W	Adult	2006	NWFSC
		Upper Yakima	Sp	H	Adult, Mixed	1992, 1997 1998	WDFW
		Warm Springs H.	Sp	H		2002, 2003	CRITFC
		Wenatchee spring	Sp	W	Adult	1993, 1998, 2000	WDFW
15	Deschutes River fall	Lower Deschutes R.	Fa	W		1999, 2001, 2002	CRITFC
		Upper Deschutes R.[3]	Su/Fa	W	Juvenile		
16	Upper Columbia R. summer/fall	Hanford Reach CR	Su/Fa	W		1999 - 2001	CRITFC
		Klickitat River	Su	W	Adult	1994	WDFW
		Klickitat River	Su/Fa	W	Adult	2005	CRITFC

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Little White Salmon NFH	Fa	H	Juvenile	2006	CRITFC
		Lower Yakima River	Su/Fa	W	Adult	1990, 1993, 1998	WDFW
		Marion Drain	U/Fa	W	Adult	1998, 1992	WDFW
		Methow R. summer	Su	W		1992 - 1994	CRITFC
		Priest Rapids H.	Su/Fa	H	Adult	1998	WDFW
		Priest Rapids H.	Fa	H	Juvenile	1998 - 2001	CRITFC
		Umatilla H.	Su/Fa	H	Adult	2003	WDFW
			Fa	H	Adult	2006	CRITFC
		Wells Dam	Su/Fa	H		1993	CRITFC
		Wenatchee[3]	Su	W	Adult	1993	WDFW
17	Snake River fall	Clearwater River	Fa	W	Adult	2000 - 2002	CRITFC
		Lyons Ferry	Fa	W	Adult	2002, 2003	WDFW
		Nez Perce Tribal H.	Fa	H	Adult	2003, 2004	CRITFC
18	Snake River spring/summer	Bear Valley	Sp	W	Juvenile	2006	IDFG
		Big Creek	Sp/Su	W	Adult	2001 - 2003	CRITFC

Region #	Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
	Camas Creek	Sp	W	Juvenile	2006	IDFG
	Capehorn Creek	Sp/Su	W	Juvenile	2006	CRITFC
	Catherine Creek	Sp/Su	W	Adult	2002, 2003	CRITFC
	Chamberlin Creek	Sp	W	Juvenile	2006	IDFG
	Crooked Fork Creek	Sp/Su	W	Juvenile	2005, 2006	CRITFC
	Dworshak H.	Sp/Su	H	Adult	2005	CRITFC
	EF Salmon River		W	Adult	2004, 2005	IDFG
	Imnaha R.	Sp/Su	W		1998, 2002, 2003	CRITFC
	Johnson Creek	Sp/Su	W	Adult	2002, 2003	CRITFC
		Sp/Su	H	Juvenile	2002 - 2004	CRITFC
	Lochsa River (Powell Satellite)	Sp/Su	H	Adult	2005	CRITFC
	Lolo Creek	Sp/Su	W	Adult	2001, 2002	CRITFC
		Sp/Su	W	Juvenile	2001	CRITFC
	Looking Glass H.	Sp/Su	H	Juvenile	1994, 1995, 1998	CRITFC

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Pahsimeroi River	Sp/Su	W	Adult	2002	CRITFC
		Minam R.	Sp/Su	W		1994, 2002, 2003	CRITFC
		Rapid River H.	Sp	H		1997, 1999, 2002	CRITFC
		Red River	Sp/Su	H	Adult	2005	CRITFC
		Sawtooth Hatchery	Sp/Su	H	Adult	2002, 2003	CRITFC
		Sesech R.	Sp/Su	W		2001, 2002, 2003	CRITFC
		Newsome Creek	Sp/Su	W	Adult	2001, 2002	CRITFC
		Tucannon	Sp/Su	H/W	Adult	2003	WDFW
		Wenaha Creek	Sp	W	Juvenile	2002	IDFG
		WF Yankee Fork[3]		W		2005	IDFG
19	Washington Coast	Chehalis River	Fa	W	Adult	1999	WDFW
		Forks Creek H.	Fa	H	Adult	2005	WDFW
		Hoh River	Fa	W	Adult	2004, 2005	WDFW
			Sp	W	Adult	1995 - 1998, 2005, 2006	WDFW

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Hoko H.	Fa	H/W	Adult	2004	WDFW
			Fa	W	Adult	2006	WDFW
		Humtulips H.	Fa	H	Adult	1990	WDFW
		Makah NFH	Fa	H	Adult	2001, 2003	WDFW
		Queets	Fa	W	Adult	1996, 1997	WDFW
		Quillayute/ Bogachiel	Fa	W	Adult	1995, 1996	WDFW
		Quinalt River	Fa	W	Adult	1995, 1997, 1998	WDFW
		Quinalt NFH	Fa	H	Adult	2001	WDFW
		Sol Duc	Sp	H	Adult	2003	WDFW
20	South Puget Sound	Bear Creek	Su/Fa	W	Adult	1998, 1999, 2003, 2004	WDFW
		Cedar river	Su/Fa	W	Adult	1994, 2003, 2004	WDFW
		Clear Creek (Nisqually)[3]	Fa	H	Adult	2005	WDFW
		Grovers Creek H.	Su/Fa	H	Adult	2004	WDFW
		Hupp Springs H.[3]	Sp	H	Adult	2002	WDFW

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Issaquah Creek	Su/Fa	W	Adult	1999	WDFW
			Su/Fa	H	Adult	2004	WDFW
		Nisqually R	Su/Fa	W	Adult	1998, 1999, 2000, 2006	WDFW
		Soos Creek	Fa	H	Adult	1998, 2004	WDFW
		South Prairie Creek[3]	Fa	W	Adult	1998, 1999, 2002	WDFW
		University of Washington H.	Su/Fa	H	Adult	2004	WDFW
		Voights Creek[3]	Fa	H	Adult	1998	WDFW
		White River	Sp	H	Adult	1998, 2002	WDFW
21	North Puget Sound	Lower Sauk River	Su	W		1998	NWFSC
		Lower Skagit	Fa	W	Adult	1998, 2006	WDFW
		Marblemount[3]	Sp	H		1997	NWFSC
			Sp	H		2006	WDFW
			Su	H		1997	NWFSC

Region #	Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
	NF Nooksack	Sp	H	Juvenile	1998	WDFW
			H/W	Adult	1999	
	NF Stilliguamish	Su	H/W	Adult	1996, 2001	WDFW
	Samish	Fa	H	Adult	1998	NWFSC
	Skagit	Su	W	Adult	1994, 1995	WDFW
	Skykomish River	Su	W		2004, 2005	NWFSC
		Su	W	Adult	1996, 2000	WDFW
	Snoqualmie		W		2005	NWFSC
	Stilliguamish	Su	H	Adult	2004	NWFSC
	Suiattle (Skagit)	Sp	W	Adult	1989, 1998, 1999	WDFW
	Suiattle River	Sp	W	Adult	1998	NWFSC
	Upper Cascade River	Sp	W		1998	NWFSC
		Sp	W	Adult	1998, 1999	WDFW
	Upper Sauk River	Sp	W		1998	NWFSC

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Upper Sauk River	Sp/Su	W	Adult	1994, 1998, 1999, 2006	WDFW
		Upper Skagit[3]	Su	W		1998	NWFSC
			Su	H	Adult	1998	WDFW
		Wallace	Su	H		2004, 2005	NWFSC
			Su	H	Adult	1996	WDFW
22	Lower Fraser River	Birkenhead River	Sp	H	Adult	1996, 1997, 1999, 2001 - 2003	SWFSC
		W Chilliwack	Fa	H	Adult	1998, 1999	DFO
		Maria Slough	Su	W	Adult	1999 - 2001	DFO
23	Lower Thompson River	Nicola	Sp	H		1998, 1999	OSU
		Spius River	Sp	H	Adult	1996 - 1998	SWFSC
24	South Thompson River	Lower Adams	Fa	H	Adult	1996	DFO
		Lower Thompson	Fa	W	Adult	2001	DFO
		Middle Shuswap	Fa	H	Adult	1997	DFO

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
25	North Thompson River	Clearwater	Fa	W	Adult	1997	DFO
		Deadman[3]	Sp	H	Adult	1996 - 1999	DFO
		Louis River	Fa	W	Adult	2001	DFO
		Raft[3]	Su	W	Adult	2001, 2002	DFO
26	Mid Fraser River	Chilko	Fa	W	Adult	1995, 1996, 1999, 2002	DFO
		Nechako	Fa	W	Adult	1996	DFO
		Quesnel	Fa	W	Adult	1996	DFO
		Stuart	Fa	W	Adult	1996	DFO
		Upper Chilcotin	Fa	W	Adult	2001	DFO
27	Upper Fraser River	Morkill River	Fa	W	Adult	2001	DFO
		Salmon River (Fraser)	Sp	W	Adult	1997	SWFSC
		Swift	Fa	W	Adult	1996	DFO
		Torpy River	Fa	W	Adult	2001	DFO
28	East Vancouver Island	Big Qualicum	Fa	H	Adult	1996	DFO

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Quinsam	Fa	H	Adult	1996, 1998	DFO
		Cowichan	Fa	H	Adult	1999, 2000	DFO
		Nanaimo	Fa	H	Adult	1998, 2002	DFO
		Puntledge	Fa	H	Adult	2000, 2001	DFO
29	West Vancouver Island	Conuma	Fa	H	Adult	1997	DFO
		Marble at NVI	Fa	H	Adult	1996, 1999, 2000	DFO
		Nitinat	Fa	H	Adult	1996	DFO
		Robertson	Fa	H	Adult	1996, 2003	DFO
		Sarita	Fa	H	Adult	1997, 2001	DFO
		Tahsis	Fa	W	Adult	1996, 2002, 2003	DFO
		Tranquil	Fa	W	Adult	1996, 1999	DFO
30	S BC Mainland	Klinaklini	Fa	W	Adult	1997	DFO
		Porteau Cove	Fa	H	Adult	2003	DFO
31	Central BC Coast	Atnarko	Fa	H	Adult	1996	DFO

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Kitimat	Fa	H	Adult	1997	DFO
		Wannock	Fa	H	Adult	1996	DFO
32	Lower Skeena River	Ecstall	Fa	W	Adult	2000- 2002	DFO
		Lower Kalum	Fa	W	Adult	2001	DFO
33	Upper Skeena River	Babine	Fa	H	Adult	1996	DFO
		Bulkley	Fa	W	Adult	1999	DFO
		Sustut	Fa	W	Adult	2001	DFO
34	Nass River	Damdochax	Fa	W	Adult	1996	DFO
		Kincolith	Fa	W	Adult	1996	DFO
		Kwinageese	Fa	W	Adult	1996	DFO
		Owegee	Fa	W	Adult	1996	DFO
35	Upper Stikine River	Little Tahltan River	Sp	W	Adult	1989, 1990	OSU
36	Taku River	Kowatua Creek		W	Adult	1989, 1990	ADFG
		Nakina River		W	Adult	1989, 1990	ADFG
		Tatsatua Creek			Adult	1989, 1990	ADFG

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
		Upper Nahlin River		W	Adult	1989, 1990, 2004	ADFG
37	Southern Southeast Alaska	Chikamin River (West Behm Canal)		W	Adult	1990, 1993	ADFG
		Chikamin River Whitman Lake H.		H		2005	ADFG
		Clear Creek		W	Adult	1989, 2003, 2004	ADFG
		Cripple Creek		W	Adult	1988, 2003	ADFG
		Keta River		W	Adult	1989, 2003	ADFG
		King Creek		W	Adult	2003	ADFG
38	Southeast Alaska Stikine R.	Andrew Creek, Crystal Lake H.		H		2005	ADFG
		Andrew Creek, Macaulay H.		H		2005	ADFG
		Andrew Creek, Medvejie H.		H		2005	ADFG
		Andrews Creek		W	Adult	1989, 2004	ADFG

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
39	N. Southeast Alaska King Salmon River	King Salmon River		W	Adult	1989, 1990, 1993	ADFG
40	N Southeast Alaska Chilkat River	Big Boulder Creek		W	Adult	1992, 1995, 2004	ADFG
		Tahini River		W	Adult	1992, 2004	ADFG
		Tahini River, Macaulay H.		H		2005	ADFG
41	Alsek River	Klukshu River		W	Adult	1989, 1990	ADFG
42	Situk River	Situk River		W	Adult	1988, 1990, 1991, 1992	ADFG
43	Hood Canal[3]	George Adams H.	Fa	H	Adult	2005	WDFW
		Hamma Hamma River	Fa	W	Adult	1999 - 2001	WDFW
		NF Skokomish River	Fa	W	Adult	1998 - 2000, 2004 - 2006	WDFW
		SF Skokomish River	Su/Fa	H/W	Adult	2005	WDFW

Region #		Population	Run time[1]	Origin	Life Stage	Collection Date	Analysis Laboratory[2]
44	Juan de Fuca[3]	Dungeness River		W	Adult	2004	WDFW
		Elwha H.	Fa		Adult	1996	
				H/W	Mixed	2004, 2005	NWFSC

[1] Run time abbreviations: spring (Sp), summer (Su), fall (Fa), unknown (U) and winter (Wi)
[2] Laboratory abbreviations: OSU, Oregon State University; SWFSC, Southwest Fisheries Science Center – National Marine Fisheries Service; DFO, Department of Fisheries and Oceans Canada; CRITFC, Columbia River Inter-Tribal Fish Commission; ADFG, Alaska Department of Fish & Game; WDFW, Washington Department of Fish & Wildlife.
[3] Reporting regions additional to those published in Seeb et al. 2007.
H = Hatchery, NFH = National fish Hatchery

Appendix 2.2. Genetic stock identification results for n = 3,866 and 4,374 legal-sized Chinook salmon that were genotyped according to a standardized set of microsatellite or single-nucleotide markers, respectively, and compared to corresponding genetic baselines (see Chapter 2 text for details and Table 2.2 for approximate sample sizes).

Stock composition results, Oregon fishery and microsatellite baseline

	Northern Oregon					Central Oregon					Klamath, Oregon				
	May	June	July	Aug	Sept	May	June	July	Aug	Sept	May	June	July	Aug	Sept
Alaska	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00		.00	.00	.02	.00
BC Mainland and Vancouver Island	.01	.00	.00	.00	.05	.00	.00	.00	.00	.00		.00	.00	.00	.00
Fraser and Thompson Rivers	.02	.04	.05	.03	.14	.01	.03	.00	.01	.00		.00	.11	.00	.00
Puget Sound	.02	.06	.03	.02	.05	.01	.01	.00	.01	.00		.00	.00	.00	.01
Juan de Fuca	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00		.00	.00	.00	.00
Washington Coast	.01	.00	.00	.00	.00	.01	.00	.00	.00	.00		.00	.00	.00	.00
Snake R fall	.02	.04	.11	.09	.00	.02	.03	.01	.02	.02		.10	.00	.00	.00
Mid and Upper Columbia R spring	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00		.00	.00	.00	.00
U Columbia R summer/fall	.14	.08	.07	.07	.05	.14	.08	.12	.03	.02		.05	.11	.05	.00
Mid Columbia R tule	.27	.33	.30	.31	.00	.13	.13	.22	.08	.00		.12	.11	.00	.00
Willamette R	.02	.01	.00	.00	.05	.01	.00	.01	.00	.00		.02	.00	.00	.00
Deschutes R fall	.01	.01	.02	.00	.05	.01	.01	.00	.01	.00		.00	.11	.00	.00
L Columbia R spring	.01	.00	.00	.00	.05	.00	.00	.00	.00	.02		.00	.00	.00	.00
L Columbia R fall	.07	.12	.11	.16	.00	.08	.08	.06	.08	.07		.02	.22	.00	.00
N Oregon Coast	.01	.00	.01	.02	.18	.00	.00	.01	.00	.02		.00	.00	.00	.01
Mid Oregon Coast	.08	.10	.11	.12	.09	.12	.14	.16	.09	.20		.07	.00	.06	.08
Rogue R	.09	.03	.03	.01	.00	.08	.09	.03	.11	.11		.14	.22	.20	.36
N California S Oregon Coast	.03	.00	.01	.02	.00	.01	.02	.00	.02	.00		.05	.11	.09	.15
Klamath R	.05	.02	.02	.02	.00	.05	.05	.03	.12	.08		.05	.00	.27	.25
California Coast	.01	.01	.00	.01	.00	.01	.01	.00	.02	.00		.02	.00	.02	.02
Central Valley fall	.16	.13	.13	.12	.27	.27	.30	.33	.40	.48		.33	.00	.30	.09
Central Valley spring	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00		.02	.00	.00	.01
Central Valley winter	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00		.00	.00	.00	.00

Stock compositions, CA fishery and SNP baseline, part I

	Klamath, north, California					Fort Bragg					San Francisco, north				
	June	July	Aug	Sept	May	June	July	Aug	Sept	May	June	July	Aug	Sept
Alaska	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
BC Mainland and Vancouver Island	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Fraser and Thompson Rivers	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Puget Sound	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Juan de Fuca	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Washington Coast	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Snake R fa	.00	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00
Mid and Upper Columbia R sp	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
U Columbia R sufa	.00	.02	.00	.00	.02	.00	.01	.00	.00	.03	.00	.00	.00	.00
Mid Columbia R tule	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Willamette R	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Deschutes R fa	.02	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
L Columbia R sp	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
L Columbia R fa	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
N Oregon Coast	.00	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Mid Oregon Coast	.02	.01	.01	.01	.04	.00	.02	.01	.00	.06	.01	.05	.00	.00
Rogue R	.09	.16	.29	.34	.20	.16	.22	.17	.31	.28	.09	.20	.00	.00
N California S Oregon Coast	.05	.09	.06	.21	.09	.21	.16	.09	.17	.06	.08	.13	.01	.00
Klamath R	.11	.17	.24	.13	.37	.31	.12	.08	.05	.25	.07	.16	.01	.00
California Coast	.02	.09	.05	.15	.10	.13	.11	.14	.28	.03	.06	.15	.02	.03
Central Valley fa	.69	.42	.32	.16	.18	.18	.35	.50	.18	.31	.62	.29	.94	.92
Central Valley sp	.00	.02	.01	.00	.00	.00	.00	.01	.02	.00	.07	.01	.01	.05
Central Valley wi	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Stock compositions, CA fishery and SNP baseline, part II

	San Francisco, south					Monterey Bay, north					Monterey Bay, south				
	May	June	July	Aug	Sept	May	June	July	Aug	Sept	May	June	July	Aug	Sept
Alaska	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
BC Mainland and Vancouver Island	.00	.01	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00
Fraser and Thompson Rivers	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Puget Sound	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Juan de Fuca	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Washington Coast	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Snake R fa	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Mid and Upper Columbia R sp	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
U Columbia R su/fa	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Mid Columbia R tule	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Willamette R	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Deschutes R fa	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
L Columbia R sp	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
L Columbia R fa	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
N Oregon Coast	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00
Mid Oregon Coast	.03	.02	.04	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Rogue R	.00	.03	.03	.00	.01	.09	.00	.01	.00	.02	.00	.00	.17	.00	.06
N California S Oregon Coast	.09	.01	.02	.00	.00	.00	.00	.01	.01	.00	.00	.00	.08	.00	.00
Klamath R	.01	.03	.03	.01	.00	.00	.06	.00	.00	.00	.00	.00	.00	.00	.00
California Coast	.02	.01	.06	.00	.03	.09	.00	.04	.04	.02	.00	.00	.08	.11	.00
Central Valley fa	.76	.85	.80	.98	.91	.82	.88	.93	.90	.76	1.00	1.00	.67	.11	.31
Central Valley sp	.06	.04	.01	.01	.04	.00	.06	.01	.01	.14	.00	.00	.00	.00	.06
Central Valley wi	.00	.01	.00	.00	.00	.00	.00	.00	.03	.06	.00	.00	.00	.78	.56

Abbreviations: L = Lower, N = North, R = River, sp = spring, su = summer, wi = winter

Appendix 2.3. Data for Chinook salmon marked with a coded wire tag (CWT) as juveniles and recovered in Oregon's 2010 commercial salmon troll fishery. Chinook salmon were genotyped using a standardized set of microsatellite markers and matched to a genetic baseline to obtain posterior probabilities for most likely stock origin.

Project Barcode	CWT ID	Life history type	CWT Release Date	Juvenile fish release site (hatchery code)	Hatchery	Stock	Agency	Date of CWT Recovery	**GSI Posterior Probabilty**	GSI Stock assignment1	Correct?
35413	10j6147	Fall	09-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	AMERICAN RIVER	CDFG	03-Aug-10	**1.00**	Central Valley fall	yes
29012	10J7709	Fall	09-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	AMERICAN RIVER	CDFG	26-Jul-10	**0.99**	Central Valley fall	yes
18398	10J7754	Fall	09-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	AMERICAN RIVER	CDFG	02-Aug-10	**1.00**	Central Valley fall	yes
35940	10J6160	Fall	01-Sep-08	TIBURON NET PENS	MOKELUMNE R FISH INS	AMERICAN RIVER	CDFG	09-Aug-10	**1.00**	Central Valley fall	yes
30090	10J7760	Fall	01-Sep-08	TIBURON NET PENS	MOKELUMNE R FISH INS	AMERICAN RIVER	CDFG	12-Aug-10	**1.00**	Central Valley fall	yes
35881	10J5546	Fall	21-May-08	WEST SACRAMENTO	FEATHER R H	FEATHER RIVER	FWS	18-Aug-10	**1.00**	Central Valley fall	yes
35921	10J6159	Spring	07-Apr-08	SAN PABLO BAY NET PEN	FEATHER R H	FEATHER RIVER	CDWR	09-Aug-10	**1.00**	Central Valley fall	yes
35393	10J6143	Fall	12-Jun-07	WICKLAND OIL TERMINAL	FEATHER R H	FEATHER RIVER	CDFG	30-Jun-10	**1.00**	Central Valley fall	yes
33223	10J3000	Fall	25-Apr-08	MARE ISLAND NET PEN	FEATHER R H	FEATHER RIVER	CDFG	30-Jun-10	**1.00**	Central Valley fall	yes
35870	10J5547	Fall	25-Apr-08	MARE ISLAND NET PEN	FEATHER R H	FEATHER RIVER	CDFG	18-Aug-10	**1.00**	Central Valley fall	yes
33124	10J3204	Fall	22-May-08	MARE ISLAND NET PEN	FEATHER R H	FEATHER RIVER	CDFG	06-Jun-10	**1.00**	Central Valley fall	yes
18478	10J3193	Fall	08-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	MOKELUMNE RIVER	CDFG	07-Jun-10	**0.99**	Central Valley fall	yes
33185	10J3202	Fall	08-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	MOKELUMNE RIVER	CDFG	06-Jun-10	0.99	Central Valley fall	yes
18395	10J3229	Fall	08-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	MOKELUMNE RIVER	CDFG	20-Jun-10	1.00	Central Valley fall	yes
35883	10J5545	Fall	08-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	MOKELUMNE RIVER	CDFG	18-Aug-10	1.00	Central Valley fall	yes
18784	10J6152	Fall	08-May-08	SAN PABLO BAY NET PEN	MOKELUMNE R FISH INS	MOKELUMNE RIVER	CDFG	06-Aug-10	0.95	Central Valley fall	yes
35426	10j6148	Fall	19-Jun-06	COWLITZ R 26.0002	COWLITZ SALMON HATCH	COWLITZ R 26.0002	WDFW	03-Aug-10	1.00	L Columbia fall	yes
33437	10J2939	Fall	20-May-08	R-CHILLIWACK	H-CHILLIWACK	S-	CDFO	23-Jun-10	1.00	L Fraser	yes

				R	R	CHILLIWACK R					
18492	10J6151	Fall	20-May-08	R-CHILLIWACK R	H-CHILLIWACK R	S-CHILLIWACK R	CDFO	05-Aug-10	1.00	L Fraser	yes
32892	10J2940	URB L-Fall	14-May-08	BIG CR (LWR COL R)	BIG CR H	BIG CR H	ODFW	23-Jun-10	1.00	Mid Columbia tule	yes
18329	10J3195	Fall	02-May-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	06-Jun-10	1.00	Mid Columbia tule	yes
18327	10J3196	Fall	02-May-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	06-Jun-10	1.00	Mid Columbia tule	yes
33520	10J7755	Fall	10-Apr-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	01-Aug-10	0.98	Mid Columbia tule	yes
5121	10J7712	Fall	10-Apr-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	26-Jul-10	1.00	Mid Columbia tule	yes
5415	10J2998	URB L-Fall	26-May-08	BIG CANYON ACCL POND	LYONS FERRY H	LYONS FERRY H	NEZP	01-Jul-10	0.99	Snake fall	yes
33258	10J3316	URB L-Fall	28-May-08	CAPTAIN JOHNS PD	LYONS FERRY H	LYONS FERRY H	NEZP	19-Jul-10	0.95	Snake fall	yes
32324	10J7756	Fall	06-May-08	SNAKE@ HLLS CNYON DM	OXBOW H	SNAKE R	IDFG	03-Aug-10	0.99	Snake fall	yes
33463	10J3315	Fall	07-Apr-08	SNAKE R-LOWR 33.0002	LYONS FERRY H	SNAKE R-LOWR 33.0002	WDFW	18-Jul-10	0.97	Snake fall	yes
29016	10J7711	Fall	07-Apr-08	SNAKE R-LOWR 33.0002	LYONS FERRY H	SNAKE R-LOWR 33.0002	WDFW	27-Jul-10	0.91	Snake fall	yes
32199	10J2857	Summer	16-Apr-08	SIMILKAMEEN R 490325		METHOW & OKANOGAN	WDFW	06-Jun-10	0.99	U Columbia su fall	yes
30546	10J3320	Fall	05-Mar-08	UMATILLA R	UMATILLA H	KALAMA R 27.0002	ODFW	27-Jul-10	0.95	U Columbia su fall	yes
35442	10J6115	Summer	12-May-06	COLUMBIA R – GENERAL		WELLS H	WDFW	27-May-10	1.00	U Columbia su fall	yes
18325	10J3197	Summer	12-May-07	CHELAN R + COLUMBIA		WELLS H	WDFW	06-Jun-10	1.00	U Columbia su fall	yes
18472	10J3194	Summer	06-Apr-08	COLUMBIA R – GENERAL		WELLS H	WDFW	08-Jun-10	0.99	U Columbia su fall	yes

35347	10J5538	Spring	16-Sep-08	ROGUE R 4	COLE RIVERS H	COLE RIVERS H	ODFW	26-Jun-10	0.98	Rogue	yes
33376	10J3234	Spring	18-Jul-07	TRASK R	TRASK R H	TRASK R (TRASK HT)	ODFW	20-Jun-10	0.99	Mid Oregon Coast	no
6143	10J0001	URB L-Fall	26-May-08	BIG CANYON ACCL POND	LYONS FERRY H	LYONS FERRY H	NEZP	08-Jun-10	0.91	N Puget Sound	no
30539	10J3322	Fall	07-Apr-08	SNAKE R-LOWR 33.0002	LYONS FERRY H	SNAKE R-LOWR 33.0002	WDFW	27-Jul-10	0.99	Deschutes fall	no
35323	10J2901	URB L-Fall	01-Jul-08	YOUNGS R & BAY	CEDC YOUNGS BAY NET	COLE RIVERS H	ODFW	12-Jun-10	0.44	L Columbia fall	no, out of basin transfer
18389	10J3228	Late Fall	02-Jan-08	COLEMAN NFH	COLEMAN NFH	COLEMAN NFH	FWS	19-Jun-10	0.65	U Columbia su fall	n (n/a)
35322	10J2902	Fall	07-May-08	SAN PABLO BAY NET PE	FEATHER R H	FEATHER RIVER	CDFG	12-Jun-10	0.81	Klamath	n (n/a)
33351	10J2999	Spring	01-Mar-07	GOBAR CR 27.0073	GOBAR POND (27)	KALAMA R 27.0002	WDFW	29-Jun-10	0.70	L Columbia fall	n (n/a)
29109	10J3350	URB L-Fall	14-Apr-08	CAPTAIN JOHNS PD	LYONS FERRY H	LYONS FERRY H	NEZP	04-Aug-10	0.79	Mid Oregon Coast	n (n/a)
7682	10J7713	URB L-Fall	10-Jun-08	NPT H	NPT H	LYONS FERRY H	NEZP	26-Jul-10	0.87	Mid Oregon Coast	n (n/a)
18178	10J7723	URB L-Fall	14-Apr-08	CAPTAIN JOHNS PD	LYONS FERRY H	LYONS FERRY H	NEZP	25-Jul-10	0.86	Snake fall	y (n/a)
32276	10J3321	Fall	07-Apr-08	SNAKE R-LOWR 33.0002	LYONS FERRY H	SNAKE R-LOWR 33.0002	WDFW	26-Jul-10	0.73	Central Valley fall	n (n/a)
32011	10J2938	Fall	07-Apr-08	SNAKE R-LOWR 33.0002	LYONS FERRY H	SNAKE R-LOWR 33.0002	WDFW	23-Jun-10	0.79	Snake fall	y (n/a)
33377	10J3235	Fall	02-Jun-08	SNAKE R-LOWR 33.0002	LYONS FERRY H	SNAKE R-LOWR 33.0002	WDFW	20-Jun-10	0.80	U Columbia su fall	n (n/a)
7687	10J7710	Fall	10-Apr-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	26-Jul-10	0.85	L Columbia fall	n (n/a)
18387	10J3227	Fall	10-Apr-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	19-Jun-10	0.58	Mid Columbia tule	y (n/a)
35249	10J6136	Fall	10-Apr-08	SPRING CR 29.0159	SPRING CR NFH	SPRING CR 29.0159	FWS	27-Jun-10	0.70	Mid Columbia tule	y (n/a)

35247	10J6135	Summer	21-Apr-08	WENATCHEE R 45.0030	WENATCHE E R 45.0030	WDFW	27-Jun-10	0.73	Mid Oregon Coast	n (n/a)

Abbreviations: CDFG = California Department of Fish and Game, CDFO = Canadian Department of Fisheries and Oceans, CR = Creek, FWS = Fisheries and Wildlife Service, h = hatchery, L = lower, n = no, NEZP = Nez Perce, NFH = National Fish Hatchery, ODFW = Oregon Department of Fish and Wildlife, R = River, U = Upriver Bright, WDFW = Washington Department of Fish and Wildlife, y = yes.

Appendix 2.4 Stock-specific catch per unit effort results for Chinook salmon stocks harvested during the 2010 non-retention and retention fishery conducted off the coasts of Oregon and California.

Stock	North Oregon Coast					Central Oregon Coast					Klamath Zone - OR				
	May	June	July	Aug	Sept	May	June	July	Aug	Sept	May	June	July	Aug	Sept
Alaska	0.01	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00
BC Mainland and Vancouver Island	0.04	0.02	0.01	0.01	0.03	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Fraser and Thompson Rivers	0.10	0.25	0.28	0.10	0.10	0.05	0.11	0.00	0.04	0.00	0.00	0.00	0.16	0.00	0.00
Puget Sound	0.10	0.38	0.18	0.05	0.03	0.04	0.05	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.05
Juan de Fuca	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Washington Coast	0.03	0.01	0.00	0.00	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Snake R fall	0.11	0.28	0.61	0.26	0.00	0.09	0.10	0.03	0.05	0.03	0.00	0.12	0.00	0.00	0.00
Snake R spring summer	0.00	0.01	0.00	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mid and Upper Columbia R spring	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U Columbia R summer/fall	0.73	0.51	0.40	0.21	0.03	0.60	0.31	0.26	0.08	0.03	0.00	0.06	0.16	0.12	0.00
Mid Columbia R tule	1.46	2.08	1.62	0.90	0.00	0.55	0.48	0.49	0.22	0.00	0.00	0.15	0.16	0.00	0.00
Willamette R	0.10	0.03	0.00	0.01	0.03	0.06	0.01	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Deschutes R fall	0.06	0.05	0.10	0.01	0.03	0.06	0.05	0.00	0.04	0.00	0.00	0.00	0.16	0.00	0.00
L Columbia R spring	0.03	0.02	0.00	0.00	0.03	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
L Columbia R fall	0.37	0.73	0.58	0.45	0.00	0.33	0.31	0.13	0.25	0.14	0.00	0.03	0.32	0.00	0.00
N Oregon Coast	0.06	0.02	0.03	0.07	0.13	0.02	0.01	0.03	0.01	0.03	0.00	0.00	0.00	0.00	0.05
Mid Oregon Coast	0.41	0.64	0.61	0.34	0.06	0.50	0.52	0.36	0.28	0.42	0.00	0.09	0.00	0.15	0.34
Rogue R	0.46	0.19	0.18	0.04	0.00	0.35	0.32	0.07	0.31	0.24	0.00	0.18	0.32	0.50	1.50
N California S Oregon Coast	0.15	0.02	0.03	0.05	0.00	0.06	0.07	0.00	0.06	0.00	0.00	0.06	0.16	0.23	0.63
Klamath R	0.25	0.12	0.13	0.05	0.00	0.22	0.18	0.07	0.35	0.17	0.00	0.06	0.00	0.70	1.02
California Coast	0.03	0.03	0.00	0.03	0.00	0.05	0.05	0.00	0.05	0.00	0.00	0.03	0.00	0.04	0.10
Central Valley fall	0.87	0.82	0.70	0.34	0.19	1.12	1.12	0.72	1.16	1.01	0.00	0.42	0.00	0.77	0.39
Central Valley spring	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.00	0.00	0.03	0.00	0.00	0.05
Central Valley winter	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
All stock CPUE	5.37	6.23	5.46	2.89	0.70	4.17	3.71	2.20	2.94	2.13	0.00	1.26	1.43	2.56	4.11

Appendix 2.4, continued

Stock	SF-n					MO-n (Santa Cruz)					MO-s (Santa Barbara)				
	May	June	July	Aug	Sept	May	June	July	Aug	Sept	May	June	July	Aug	Sept
Alaska	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BC Mainland and Vancouver Island	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Fraser and Thompson Rivers	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Puget Sound	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Juan de Fuca	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Washington Coast	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Snake R fa	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Snake R spring summer	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mid and Upper Columbia R sp	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U Columbia R sufa	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mid Columbia R tule	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Willamette R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Deschutes R fa	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L Columbia R sp	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L Columbia R fa	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N Oregon Coast	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
Mid Oregon Coast	0.06	0.05	0.05	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rogue R	0.00	0.09	0.04	0.00	0.01	0.03	0.00	0.02	0.00	0.04	0.00	0.00	0.07	0.00	0.08
N California S Oregon Coast	0.15	0.02	0.02	0.00	0.00	0.00	0.00	0.04	0.03	0.00	0.00	0.00	0.04	0.00	0.00
Klamath R	0.02	0.09	0.04	0.02	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
California Coast	0.04	0.03	0.07	0.00	0.03	0.03	0.00	0.18	0.10	0.04	0.00	0.00	0.04	0.06	0.00
Central Valley fa	1.25	2.82	0.92	1.93	0.95	0.25	0.36	4.26	2.58	1.36	0.24	0.34	0.30	0.06	0.42
Central Valley sp	0.10	0.14	0.01	0.02	0.04	0.00	0.02	0.05	0.03	0.25	0.00	0.00	0.00	0.00	0.08
Central Valley wi	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.11	0.00	0.00	0.00	0.41	0.75
Al l stock CPUE	1.65	3.30	1.15	1.97	1.04	0.31	0.40	4.60	2.85	1.79	0.24	0.34	0.45	0.53	1.33

Appendix 3.1. For gene ontology classifications, the number of genes differentially expressed in candidate magnetoreceptor cells of Chinook salmon compared to numbers expected based on presence in the Zebrafish genome. The statistical overrepresentation test was implemented in PANTHER. Results are organized by broad category and are not a full list of possible categories.

Basic category (biological processes)	Biological Process	Danio rerio - REFLIST (25708)	Expected matches to PANTHER category	Observed matches to PANTHER category	Over or under-expressed	P-value	Ratio observed / expected
sensory	neuronal action potential propagation	24	0.06	1	+	1.00E+00	16.67
	visual perception	469	1.2	2	+	1.00E+00	1.67
	neurotransmitter secretion	301	0.77	4	+	1.00E+00	5.19
	sensory perception	657	1.69	4	+	1.00E+00	2.37
	synaptic transmission	660	1.69	8	+	5.03E-02	4.73
	nervous system development	1418	3.64	8	+	1.00E+00	2.20
	response to stimulus	2545	6.53	9	+	1.00E+00	1.38
	sensory perception of pain	14	0.04	0	-	1.00E+00	
	sensory perception of chemical stimulus	11	0.03	0	-	1.00E+00	
	sensory perception of taste	8	0.02	0	-	1.00E+00	
Protein, protein localization	asymmetric protein localization	51	0.13	1	+	1.00E+00	7.69
	protein localization	87	0.22	1	+	1.00E+00	4.55
	cellular component organization or biogenesis	1444	3.71	5	+	1.00E+00	1.35
	protein folding	266	0.68	4	+	8.68E-01	5.88
	protein complex biogenesis	86	0.22	2	+	1.00E+00	9.09
	cellular component biogenesis	130	0.33	2	+	1.00E+00	6.06
	protein complex assembly	86	0.22	2	+	1.00E+00	9.09

	cellular component organization	1359	3	3.49	-	1.00E+00	0.86
	response to stress (HSP70)	618	5	1.59	+	1.00E+00	3.14
	cellular protein modification process	2003	9	5.14	+	1.00E+00	1.75
localization/transport	anion transport	192	1	0.49	+	1.00E+00	2.04
	receptor-mediated endocytosis	379	1	0.97	+	1.00E+00	1.03
	nuclear transport	99	2	0.25	+	1.00E+00	8.00
	vitamin transport	116	2	0.3	+	1.00E+00	6.67
	phosphate ion transport	137	2	0.35	+	1.00E+00	5.71
	lipid transport	383	4	0.98	+	1.00E+00	4.08
	endocytosis	648	4	1.66	+	1.00E+00	2.41
	cation transport	827	5	2.12	+	1.00E+00	2.36
	ion transport	937	5	2.41	+	1.00E+00	2.07
	vesicle-mediated transport	1408	8	3.61	+	1.00E+00	2.22
	intracellular protein transport	1819	12	4.67	+	3.71E-01	2.57
	protein transport	1842	12	4.73	+	4.12E-01	2.54
	transport	3508	17	9.01	+	1.00E+00	1.89
	localization	3602	18	9.25	+	6.12E-01	1.95
DNA/RNA processing	transc. from RNAP II promoter	2874	6	7.38	-	1.00E+00	0.81
	transcription, DNA-dependent	2893	6	7.43	-	1.00E+00	0.81
	reg. of transcr. from RNAP II promoter	2414	4	6.2	-	1.00E+00	0.65
	RNA splicing, via transesterification rx	265	3	0.68	+	1.00E+00	4.41
	RNA splicing	265	3	0.68	+	1.00E+00	4.41
	mRNA splicing, via spliceosome	368	3	0.94	+	1.00E+00	3.19
	mRNA processing	464	3	1.19	+	1.00E+00	2.52
	mRNA polyadenylation	126	2	0.32	+	1.00E+00	6.25

	mRNA 3'-end processing	129	2	0.33	+	1.00E+00	6.06
	DNA replication	212	2	0.54	+	1.00E+00	3.70
	RNA localization	91	1	0.23	+	1.00E+00	4.35
	meiosis	99	1	0.25	+	1.00E+00	4.00
	mitosis	651	3	1.67	+	1.00E+00	1.80
	RNA metabolic process	3386	7	8.69	-	1.00E+00	0.81
process	metabolic process	11677	43	29.98	+	1.71E-01	1.43
	primary metabolic process	9902	35	25.42	+	1.00E+00	1.38
	cellular process	8246	23	21.17	+	1.00E+00	1.09
	protein metabolic process	3936	21	10.1	+	1.05E-01	2.08
	developmental process	4166	13	10.7	+	1.00E+00	1.21
	single-multicellular organism process	2661	10	6.83	+	1.00E+00	1.46
	multicellular organismal process	2661	10	6.83	+	1.00E+00	1.46
	neurological system process	1828	9	4.69	+	1.00E+00	1.92
	system process	2226	9	5.71	+	1.00E+00	1.58
	nucleobase-containing compound metabolic process	4861	9	12.48	-	1.00E+00	0.72
	regulation of biological process	3210	7	8.24	-	1.00E+00	0.85
	lipid metabolic process	1109	6	2.85	+	1.00E+00	2.11
	proteolysis	1155	6	2.97	+	1.00E+00	2.02
	phosphate-containing compound metabolic process	696	4	1.79	+	1.00E+00	2.23
	immune system process	2678	4	6.88	-	1.00E+00	0.58
	carbohydrate metabolic process	817	3	2.1	+	1.00E+00	1.43
	homeostatic process	185	2	0.47	+	1.00E+00	4.26
	steroid metabolic process	258	2	0.66	+	1.00E+00	3.03
regulation	biological regulation	4516	13	11.59	+	1.00E+00	1.12
	regulation of catalytic activity	1348	5	3.46	+	1.00E+00	1.45

	regulation of molecular function	1367	5	3.51	+	1.00E+00	1.42
	regulation of nucleobase-containing compound metabolic process	2557	4	6.56	-	1.00E+00	0.61
immune	response to interferon-gamma	43	0	0.11	-	1.00E+00	
	macrophage activation	359	0	0.92	-	1.00E+00	
	antigen processing / MHC class II	271	0	0.7	-	1.00E+00	
miscellaneous	cellular calcium ion homeostasis	58	1	0.15	+	1.00E+00	6.67
	respiratory electron transport chain	257	2	0.66	+	1.00E+00	3.03
	cellular component morphogenesis	842	3	2.16	+	1.00E+00	1.39
	cell-cell adhesion	861	3	2.21	+	1.00E+00	1.36
	anatomical structure morphogenesis	898	3	2.31	+	1.00E+00	1.30
	cell adhesion	1426	4	3.66	+	1.00E+00	1.09
	biological adhesion	1426	4	3.66	+	1.00E+00	1.09
	cell cycle	1659	6	4.26	+	1.00E+00	1.41
	protein phosphorylation	1063	7	2.73	+	1.00E+00	2.56
	ectoderm development	1276	7	3.28	+	1.00E+00	2.13
	cell-cell signaling	1189	8	3.05	+	1.00E+00	2.62
	cell communication	4805	16	12.34	+	1.00E+00	1.30
	blood coagulation	287	2	0.74	+	1.00E+00	2.70
	angiogenesis	310	2	0.8	+	1.00E+00	2.50

Appendix 3.2. For gene ontology classifications “Biological Process”, the number of genes differentially expressed in candidate magnetoreceptor cells of Chinook salmon are compared to numbers present in the Zebrafish genome. The statistical overrepresentation test was implemented in PANTHER. Results are shown for all ontology terms; genes not identified as expressed in the candidate magnetoreceptor cell are listed last.

Biological Process	Danio rerio - REFLIST (25708)	Observed matches to PANTHER category	Expected matches to PANTHER category	Over or under-expressed	P-value	Ratio observed / expected
neuronal action potential propagation	24	1	0.06	+	1.00E+00	16.67
fatty acid beta-oxidation	33	1	0.08	+	1.00E+00	12.50
vitamin biosynthetic process	37	1	0.09	+	1.00E+00	11.11
protein complex biogenesis	86	2	0.22	+	1.00E+00	9.09
protein complex assembly	86	2	0.22	+	1.00E+00	9.09
nuclear transport	99	2	0.25	+	1.00E+00	8.00
vitamin metabolic process	52	1	0.13	+	1.00E+00	7.69
asymmetric protein localization	51	1	0.13	+	1.00E+00	7.69
cellular glucose homeostasis	54	1	0.14	+	1.00E+00	7.14
vitamin transport	116	2	0.3	+	1.00E+00	6.67
cellular calcium ion homeostasis	58	1	0.15	+	1.00E+00	6.67
mRNA polyadenylation	126	2	0.32	+	1.00E+00	6.25
mRNA 3'-end processing	129	2	0.33	+	1.00E+00	6.06
cellular component biogenesis	130	2	0.33	+	1.00E+00	6.06
protein folding	266	4	0.68	+	8.68E-01	5.88
phosphate ion transport	137	2	0.35	+	1.00E+00	5.71
neurotransmitter secretion	301	4	0.77	+	1.00E+00	5.19
regulation of vasoconstriction	81	1	0.21	+	1.00E+00	4.76
synaptic transmission	660	8	1.69	+	5.03E-02	4.73
protein localization	87	1	0.22	+	1.00E+00	4.55
RNA splicing, via transesterification reactions	265	3	0.68	+	1.00E+00	4.41
RNA splicing	265	3	0.68	+	1.00E+00	4.41
RNA localization	91	1	0.23	+	1.00E+00	4.35
homeostatic process	185	2	0.47	+	1.00E+00	4.26
lipid transport	383	4	0.98	+	1.00E+00	4.08
meiosis	99	1	0.25	+	1.00E+00	4.00
DNA replication	212	2	0.54	+	1.00E+00	3.70
cell differentiation	115	1	0.3	+	1.00E+00	3.33

cholesterol metabolic process	122	1	0.31	+	1.00E+00	3.23
mRNA splicing, via spliceosome	368	3	0.94	+	1.00E+00	3.19
response to stress	618	5	1.59	+	1.00E+00	3.14
steroid metabolic process	258	2	0.66	+	1.00E+00	3.03
respiratory electron transport chain	257	2	0.66	+	1.00E+00	3.03
nucleobase-containing compound transport	127	1	0.33	+	1.00E+00	3.03
carbohydrate transport	127	1	0.33	+	1.00E+00	3.03
blood coagulation	287	2	0.74	+	1.00E+00	2.70
cell-cell signaling	1189	8	3.05	+	1.00E+00	2.62
intracellular protein transport	1819	12	4.67	+	3.71E-01	2.57
protein phosphorylation	1063	7	2.73	+	1.00E+00	2.56
cellular amino acid biosynthetic process	150	1	0.39	+	1.00E+00	2.56
protein transport	1842	12	4.73	+	4.12E-01	2.54
mRNA processing	464	3	1.19	+	1.00E+00	2.52
angiogenesis	310	2	0.8	+	1.00E+00	2.50
generation of precursor metabolites and energy	317	2	0.81	+	1.00E+00	2.47
endocytosis	648	4	1.66	+	1.00E+00	2.41
sensory perception	657	4	1.69	+	1.00E+00	2.37
cation transport	827	5	2.12	+	1.00E+00	2.36
phosphate-containing compound metabolic process	696	4	1.79	+	1.00E+00	2.23
vesicle-mediated transport	1408	8	3.61	+	1.00E+00	2.22
nervous system development	1418	8	3.64	+	1.00E+00	2.20
extracellular transport	178	1	0.46	+	1.00E+00	2.17
heart development	363	2	0.93	+	1.00E+00	2.15
ectoderm development	1276	7	3.28	+	1.00E+00	2.13
monosaccharide metabolic process	182	1	0.47	+	1.00E+00	2.13
lipid metabolic process	1109	6	2.85	+	1.00E+00	2.11
protein metabolic process	3936	21	10.1	+	1.05E-01	2.08
ion transport	937	5	2.41	+	1.00E+00	2.07
anion transport	192	1	0.49	+	1.00E+00	2.04
response to external stimulus	386	2	0.99	+	1.00E+00	2.02
proteolysis	1155	6	2.97	+	1.00E+00	2.02
fatty acid metabolic process	199	1	0.51	+	1.00E+00	1.96
localization	3602	18	9.25	+	6.12E-01	1.95
neurological system process	1828	9	4.69	+	1.00E+00	1.92
transport	3508	17	9.01	+	1.00E+00	1.89

skeletal system development	412	2	1.06	+	1.00E+00	1.89
chromosome segregation	207	1	0.53	+	1.00E+00	1.89
blood circulation	212	1	0.54	+	1.00E+00	1.85
mitosis	651	3	1.67	+	1.00E+00	1.80
muscle contraction	435	2	1.12	+	1.00E+00	1.79
cellular protein modification process	2003	9	5.14	+	1.00E+00	1.75
visual perception	469	2	1.2	+	1.00E+00	1.67
phospholipid metabolic process	239	1	0.61	+	1.00E+00	1.64
DNA metabolic process	485	2	1.25	+	1.00E+00	1.60
system process	2226	9	5.71	+	1.00E+00	1.58
cytokinesis	253	1	0.65	+	1.00E+00	1.54
single-multicellular organism process	2661	10	6.83	+	1.00E+00	1.46
multicellular organismal process	2661	10	6.83	+	1.00E+00	1.46
regulation of catalytic activity	1348	5	3.46	+	1.00E+00	1.45
metabolic process	11677	43	29.98	+	1.71E-01	1.43
carbohydrate metabolic process	817	3	2.1	+	1.00E+00	1.43
regulation of molecular function	1367	5	3.51	+	1.00E+00	1.42
cell cycle	1659	6	4.26	+	1.00E+00	1.41
cellular component morphogenesis	842	3	2.16	+	1.00E+00	1.39
response to stimulus	2545	9	6.53	+	1.00E+00	1.38
primary metabolic process	9902	35	25.42	+	1.00E+00	1.38
cell-cell adhesion	861	3	2.21	+	1.00E+00	1.36
cellular component organization or biogenesis	1444	5	3.71	+	1.00E+00	1.35
spermatogenesis	301	1	0.77	+	1.00E+00	1.30
anatomical structure morphogenesis	898	3	2.31	+	1.00E+00	1.30
cell communication	4805	16	12.34	+	1.00E+00	1.30
system development	2412	8	6.19	+	1.00E+00	1.29
developmental process	4166	13	10.7	+	1.00E+00	1.21
biological regulation	4516	13	11.59	+	1.00E+00	1.12
cell adhesion	1426	4	3.66	+	1.00E+00	1.09
biological adhesion	1426	4	3.66	+	1.00E+00	1.09
cellular process	8246	23	21.17	+	1.00E+00	1.09
exocytosis	373	1	0.96	+	1.00E+00	1.04
receptor-mediated endocytosis	379	1	0.97	+	1.00E+00	1.03
cellular component organization	1359	3	3.49	-	1.00E+00	0.86
regulation of biological process	3210	7	8.24	-	1.00E+00	0.85

transcription from RNA polymerase II promoter	2874	6	7.38	-	1.00E+00	0.81
cellular amino acid metabolic process	479	1	1.23	-	1.00E+00	0.81
transcription, DNA-dependent	2893	6	7.43	-	1.00E+00	0.81
RNA metabolic process	3386	7	8.69	-	1.00E+00	0.81
Unclassified	8708	17	22.36	-	1.00E+00	0.76
nucleobase-containing compound metabolic process	4861	9	12.48	-	1.00E+00	0.72
regulation of transcription from RNA polymerase II promoter	2414	4	6.2	-	1.00E+00	0.65
gamete generation	607	1	1.56	-	1.00E+00	0.64
regulation of nucleobase-containing compound metabolic process	2557	4	6.56	-	1.00E+00	0.61
cellular component movement	638	1	1.64	-	1.00E+00	0.61
immune system process	2678	4	6.88	-	1.00E+00	0.58
reproduction	705	1	1.81	-	1.00E+00	0.55
mesoderm development	1418	2	3.64	-	1.00E+00	0.55
immune response	765	1	1.96	-	1.00E+00	0.51
death	1112	0	2.85	-	1.00E+00	
cell death	1109	0	2.85	-	1.00E+00	
apoptotic process	1109	0	2.85	-	1.00E+00	
cellular defense response	838	0	2.15	-	1.00E+00	
induction of apoptosis	534	0	1.37	-	1.00E+00	
muscle organ development	525	0	1.35	-	1.00E+00	
translation	464	0	1.19	-	1.00E+00	
organelle organization	361	0	0.93	-	1.00E+00	
macrophage activation	359	0	0.92	-	1.00E+00	
nitrogen compound metabolic process	358	0	0.92	-	1.00E+00	
pattern specification process	319	0	0.82	-	1.00E+00	
antigen processing and presentation	301	0	0.77	-	1.00E+00	
B cell mediated immunity	297	0	0.76	-	1.00E+00	
polysaccharide metabolic process	284	0	0.73	-	1.00E+00	
chromatin organization	280	0	0.72	-	1.00E+00	
antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	271	0	0.7	-	1.00E+00	
embryo development	258	0	0.66	-	1.00E+00	
protein glycosylation	232	0	0.6	-	1.00E+00	
hemopoiesis	222	0	0.57	-	1.00E+00	
negative regulation of apoptotic process	221	0	0.57	-	1.00E+00	

DNA repair	211	0	0.54	-	1.00E+00
segment specification	205	0	0.53	-	1.00E+00
catabolic process	201	0	0.52	-	1.00E+00
female gamete generation	184	0	0.47	-	1.00E+00
synaptic vesicle exocytosis	177	0	0.45	-	1.00E+00
cell-matrix adhesion	164	0	0.42	-	1.00E+00
regulation of translation	160	0	0.41	-	1.00E+00
biosynthetic process	157	0	0.4	-	1.00E+00
protein targeting	156	0	0.4	-	1.00E+00
neuron-neuron synaptic transmission	149	0	0.38	-	1.00E+00
rRNA metabolic process	141	0	0.36	-	1.00E+00
cyclic nucleotide metabolic process	138	0	0.35	-	1.00E+00
purine nucleobase metabolic process	133	0	0.34	-	1.00E+00
sensory perception of sound	125	0	0.32	-	1.00E+00
amino acid transport	125	0	0.32	-	1.00E+00
complement activation	121	0	0.31	-	1.00E+00
glycogen metabolic process	109	0	0.28	-	1.00E+00
phagocytosis	108	0	0.28	-	1.00E+00
DNA recombination	106	0	0.27	-	1.00E+00
sulfur compound metabolic process	100	0	0.26	-	1.00E+00
regulation of phosphate metabolic process	100	0	0.26	-	1.00E+00
coenzyme metabolic process	98	0	0.25	-	1.00E+00
fertilization	90	0	0.23	-	1.00E+00
cellular amino acid catabolic process	84	0	0.22	-	1.00E+00
response to toxic substance	71	0	0.18	-	1.00E+00
tRNA metabolic process	69	0	0.18	-	1.00E+00
cell proliferation	67	0	0.17	-	1.00E+00
pyrimidine nucleobase metabolic process	66	0	0.17	-	1.00E+00
digestive tract mesoderm development	65	0	0.17	-	1.00E+00
regulation of liquid surface tension	64	0	0.16	-	1.00E+00
oxidative phosphorylation	64	0	0.16	-	1.00E+00
response to pheromone	63	0	0.16	-	1.00E+00
dorsal/ventral axis specification	62	0	0.16	-	1.00E+00
anterior/posterior axis specification	61	0	0.16	-	1.00E+00
RNA catabolic process	56	0	0.14	-	1.00E+00
lysosomal transport	51	0	0.13	-	1.00E+00

protein acetylation	46	0	0.12	-	1.00E+00
response to interferon-gamma	43	0	0.11	-	1.00E+00
gluconeogenesis	42	0	0.11	-	1.00E+00
natural killer cell activation	41	0	0.11	-	1.00E+00
porphyrin-containing compound metabolic process	38	0	0.1	-	1.00E+00
peroxisomal transport	38	0	0.1	-	1.00E+00
fatty acid biosynthetic process	36	0	0.09	-	1.00E+00
glycolysis	35	0	0.09	-	1.00E+00
response to abiotic stimulus	31	0	0.08	-	1.00E+00
protein lipidation	29	0	0.07	-	1.00E+00
acyl-CoA metabolic process	28	0	0.07	-	1.00E+00
endoderm development	27	0	0.07	-	1.00E+00
cytoskeleton organization	27	0	0.07	-	1.00E+00
regulation of carbohydrate metabolic process	26	0	0.07	-	1.00E+00
mRNA transcription	26	0	0.07	-	1.00E+00
mitochondrion organization	26	0	0.07	-	1.00E+00
mitochondrial transport	26	0	0.07	-	1.00E+00
tricarboxylic acid cycle	25	0	0.06	-	1.00E+00
neuromuscular synaptic transmission	23	0	0.06	-	1.00E+00
locomotion	22	0	0.06	-	1.00E+00
response to endogenous stimulus	20	0	0.05	-	1.00E+00
ferredoxin metabolic process	16	0	0.04	-	1.00E+00
growth	15	0	0.04	-	1.00E+00
sensory perception of pain	14	0	0.04	-	1.00E+00
protein methylation	14	0	0.04	-	1.00E+00
regulation of cell cycle	13	0	0.03	-	1.00E+00
nitric oxide biosynthetic process	13	0	0.03	-	1.00E+00
regulation of cellular amino acid metabolic process	12	0	0.03	-	1.00E+00
sensory perception of chemical stimulus	11	0	0.03	-	1.00E+00
DNA catabolic process	11	0	0.03	-	1.00E+00
sex determination	10	0	0.03	-	1.00E+00
defense response to bacterium	10	0	0.03	-	1.00E+00
pinocytosis	9	0	0.02	-	1.00E+00
cell growth	9	0	0.02	-	1.00E+00
unsaturated fatty acid biosynthetic process	8	0	0.02	-	1.00E+00
sensory perception of taste	8	0	0.02	-	1.00E+00

pentose-phosphate shunt	8	0	0.02	-	1.00E+00
protein ADP-ribosylation	7	0	0.02	-	1.00E+00
polyphosphate catabolic process	6	0	0.02	-	1.00E+00
disaccharide metabolic process	4	0	0.01	-	1.00E+00
vitamin catabolic process	3	0	0.01	-	1.00E+00
regulation of sequence-specific DNA binding transcription factor activity	3	0	0.01	-	1.00E+00
cytokine production	3	0	0.01	-	1.00E+00
chromatin remodeling	3	0	0.01	-	1.00E+00
response to biotic stimulus	2	0	0.01	-	1.00E+00
cell recognition	2	0	0.01	-	1.00E+00
7-methylguanosine mRNA capping	2	0	0.01	-	1.00E+00
nitrogen utilization	1	0	0	-	1.00E+00
mammary gland development	1	0	0	-	1.00E+00
bile acid metabolic process	1	0	0	-	1.00E+00

Appendix 3.3. For gene ontology classifications “Cellular Function”, the number of genes differentially expressed in candidate magnetoreceptor cells of Chinook salmon compared to numbers expected based on presence in the Zebrafish genome. The statistical overrepresentation test was implemented in PANTHER. Results are shown for all Biological process ontology terms organized by broad category. Genes not identified as expressed in the candidate magnetoreceptor cell but that are present in the Zebrafish genome are presented last.

Cellular function	Danio rerio - REFLIST (25708)	Observed matches to PANTHER category	Expected matches to PANTHER category	Over or under-expressed	P-value	Ratio observed / expected
ribonucleoprotein complex	144	2	0.37	+	1.00E+00	5.41
mitochondrial inner membrane	74	1	0.19	+	1.00E+00	5.26
mitochondrion	89	1	0.23	+	1.00E+00	4.35
microtubule	275	3	0.71	+	1.00E+00	4.23
cell junction	113	1	0.29	+	1.00E+00	3.45
cytoplasm	419	3	1.08	+	1.00E+00	2.78
macromolecular complex	776	5	1.99	+	1.00E+00	2.51
protein complex	632	3	1.62	+	1.00E+00	1.85
extracellular region	723	3	1.86	+	1.00E+00	1.61
cytoskeleton	1005	4	2.58	+	1.00E+00	1.55
extracellular matrix	506	2	1.3	+	1.00E+00	1.54
organelle	1278	5	3.28	+	1.00E+00	1.52
intracellular	1796	7	4.61	+	1.00E+00	1.52
cell part	2077	8	5.33	+	1.00E+00	1.50
membrane	613	2	1.57	+	1.00E+00	1.27
plasma membrane	424	1	1.09	-	1.00E+00	0.92
Unclassified	22610	53	58.05	-	1.00E+00	0.91
actin cytoskeleton	550	1	1.41	-	1.00E+00	0.71
nucleus	78	0	0.2	-	1.00E+00	
intermediate filament cytoskeleton	61	0	0.16	-	1.00E+00	
vesicle coat	50	0	0.13	-	1.00E+00	
integral to membrane	50	0	0.13	-	1.00E+00	
MHC protein complex	46	0	0.12	-	1.00E+00	
immunoglobulin complex	41	0	0.11	-	1.00E+00	
heterotrimeric G-protein complex	30	0	0.08	-	1.00E+00	
cytosol	29	0	0.07	-	1.00E+00	
SNARE complex	26	0	0.07	-	1.00E+00	

cell projection	25	0	0.06	-	1.00E+00
tubulin complex	25	0	0.06	-	1.00E+00
proton-transporting ATP synthase complex	21	0	0.05	-	1.00E+00
chromosome	20	0	0.05	-	1.00E+00
nuclear chromosome	15	0	0.04	-	1.00E+00
neuron projection	15	0	0.04	-	1.00E+00
ribosome	10	0	0.03	-	1.00E+00
synapse	7	0	0.02	-	1.00E+00
cytoplasmic membrane-bounded vesicle	5	0	0.01	-	1.00E+00
vacuole	5	0	0.01	-	1.00E+00
protein-DNA complex	5	0	0.01	-	1.00E+00
tight junction	4	0	0.01	-	1.00E+00
peroxisome	3	0	0.01	-	1.00E+00
dendrite	3	0	0.01	-	1.00E+00
cilium	3	0	0.01	-	1.00E+00
apical part of cell	2	0	0.01	-	1.00E+00
endoplasmic reticulum	2	0	0.01	-	1.00E+00
lysosome	2	0	0.01	-	1.00E+00
microvillus	1	0	0	-	1.00E+00

Appendix 3.4. For gene ontology classifications “Molecular Function”, the number of genes differentially expressed in candidate magnetoreceptor cells of Chinook salmon are compared to numbers present in the Zebrafish genome. The statistical overrepresentation test was implemented in Panther. Results are shown for all ontology terms; gene ontology categories with no gene expression but that are present in the Zebrafish genome are listed last.

Molecular Function	Danio rerio - REFLIST (25708)	Observed matches to PANTHER category	Expected matches to PANTHER category	Over or under-expressed	P-value	Ratio observed / expected
DNA replication origin binding	51	2	0.13	+	1.00E+00	15.38
deacetylase activity	32	1	0.08	+	1.00E+00	12.50
microtubule binding	86	2	0.22	+	1.00E+00	9.09
single-stranded DNA binding	88	2	0.23	+	1.00E+00	8.70
voltage-gated sodium channel activity	45	1	0.12	+	1.00E+00	8.33
poly(A) RNA binding	97	2	0.25	+	1.00E+00	8.00
voltage-gated calcium channel activity	51	1	0.13	+	1.00E+00	7.69
metallopeptidase activity	251	4	0.64	+	6.40E-01	6.25
lipid transporter activity	123	2	0.32	+	1.00E+00	6.25
carbohydrate transmembrane transporter activity	70	1	0.18	+	1.00E+00	5.56
microtubule motor activity	78	1	0.2	+	1.00E+00	5.00
nucleotide kinase activity	83	1	0.21	+	1.00E+00	4.76
mRNA binding	313	3	0.8	+	1.00E+00	3.75
oxidoreductase activity	766	7	1.97	+	5.51E-01	3.55
non-membrane spanning protein tyrosine kinase activity	115	1	0.3	+	1.00E+00	3.33
cation transmembrane transporter activity	508	4	1.3	+	1.00E+00	3.08
transmembrane transporter activity	1483	10	3.81	+	6.97E-01	2.62
methyltransferase activity	156	1	0.4	+	1.00E+00	2.50
amino acid transmembrane transporter activity	157	1	0.4	+	1.00E+00	2.50
transporter activity	1573	10	4.04	+	1.00E+00	2.48
motor activity	171	1	0.44	+	1.00E+00	2.27
protein kinase activity	869	5	2.23	+	1.00E+00	2.24
GTPase activity	354	2	0.91	+	1.00E+00	2.20
cysteine-type peptidase activity	178	1	0.46	+	1.00E+00	2.17
peptidase activity	914	5	2.35	+	1.00E+00	2.13

voltage-gated ion channel activity	192	1	0.49	+	1.00E+00	2.04
cytoskeletal protein binding	394	2	1.01	+	1.00E+00	1.98
structural constituent of ribosome	200	1	0.51	+	1.00E+00	1.96
cation channel activity	212	1	0.54	+	1.00E+00	1.85
serine-type peptidase activity	427	2	1.1	+	1.00E+00	1.82
transferase activity, transferring acyl groups	219	1	0.56	+	1.00E+00	1.79
chromatin binding	218	1	0.56	+	1.00E+00	1.79
kinase activity	1346	6	3.46	+	1.00E+00	1.73
RNA binding	682	3	1.75	+	1.00E+00	1.71
guanyl-nucleotide exchange factor activity	232	1	0.6	+	1.00E+00	1.67
hydrolase activity	2898	12	7.44	+	1.00E+00	1.61
catalytic activity	7245	28	18.6	+	1.00E+00	1.51
enzyme regulator activity	1302	5	3.34	+	1.00E+00	1.50
small GTPase regulator activity	537	2	1.38	+	1.00E+00	1.45
structural constituent of cytoskeleton	1088	4	2.79	+	1.00E+00	1.43
ion channel activity	563	2	1.45	+	1.00E+00	1.38
protein binding	3719	13	9.55	+	1.00E+00	1.36
kinase regulator activity	293	1	0.75	+	1.00E+00	1.33
transferase activity	2343	8	6.02	+	1.00E+00	1.33
phosphoprotein phosphatase activity	297	1	0.76	+	1.00E+00	1.32
calcium ion binding	604	2	1.55	+	1.00E+00	1.29
receptor binding	1212	4	3.11	+	1.00E+00	1.29
structural molecule activity	1527	5	3.92	+	1.00E+00	1.28
calmodulin binding	312	1	0.8	+	1.00E+00	1.25
pyrophosphatase activity	346	1	0.89	+	1.00E+00	1.12
ubiquitin-protein ligase activity	408	1	1.05	-	1.00E+00	0.95
binding	7978	19	20.48	-	1.00E+00	0.93
phosphatase activity	424	1	1.09	-	1.00E+00	0.92
Unclassified	10308	24	26.46	-	1.00E+00	0.91
hydrolase activity, acting on ester bonds	875	2	2.25	-	1.00E+00	0.89
transcription factor binding transcription factor activity	449	1	1.15	-	1.00E+00	0.87
protein binding transcription factor activity	449	1	1.15	-	1.00E+00	0.87
transcription cofactor activity	449	1	1.15	-	1.00E+00	0.87
sequence-specific DNA binding transcription factor activity	2886	6	7.41	-	1.00E+00	0.81

nucleic acid binding transcription factor activity	2896	6	7.43	-	1.00E+00	0.81
G-protein coupled receptor activity	509	1	1.31	-	1.00E+00	0.76
DNA binding	3221	6	8.27	-	1.00E+00	0.73
ligase activity	646	1	1.66	-	1.00E+00	0.60
nucleic acid binding	4653	7	11.95	-	1.00E+00	0.59
receptor activity	2789	4	7.16	-	1.00E+00	0.56
transmembrane receptor protein kinase activity	365	0	0.94	-	1.00E+00	0.00
transmembrane receptor protein serine/threonine kinase activity	337	0	0.87	-	1.00E+00	0.00
enzyme inhibitor activity	335	0	0.86	-	1.00E+00	0.00
transferase activity, transferring glycosyl groups	300	0	0.77	-	1.00E+00	0.00
actin binding	284	0	0.73	-	1.00E+00	0.00
lyase activity	254	0	0.65	-	1.00E+00	0.00
nuclease activity	233	0	0.6	-	1.00E+00	0.00
ligand-gated ion channel activity	212	0	0.54	-	1.00E+00	0.00
peptidase inhibitor activity	198	0	0.51	-	1.00E+00	0.00
growth factor activity	179	0	0.46	-	1.00E+00	0.00
cytokine activity	177	0	0.45	-	1.00E+00	0.00
enzyme activator activity	172	0	0.44	-	1.00E+00	0.00
isomerase activity	171	0	0.44	-	1.00E+00	0.00
translation factor activity, nucleic acid binding	164	0	0.42	-	1.00E+00	0.00
translation regulator activity	157	0	0.4	-	1.00E+00	0.00
helicase activity	156	0	0.4	-	1.00E+00	0.00
calcium-dependent phospholipid binding	152	0	0.39	-	1.00E+00	0.00
hormone activity	134	0	0.34	-	1.00E+00	0.00
acetyltransferase activity	126	0	0.32	-	1.00E+00	0.00
translation initiation factor activity	123	0	0.32	-	1.00E+00	0.00
extracellular matrix structural constituent	119	0	0.31	-	1.00E+00	0.00
kinase activator activity	109	0	0.28	-	1.00E+00	0.00
lipase activity	108	0	0.28	-	1.00E+00	0.00
serine-type endopeptidase inhibitor activity	96	0	0.25	-	1.00E+00	0.00
glutamate receptor activity	95	0	0.24	-	1.00E+00	0.00

kinase inhibitor activity	93	0	0.24	-	1.00E+00	0.00
sequence-specific DNA binding RNA polymerase II transcription factor activity	92	0	0.24	-	1.00E+00	0.00
phospholipase activity	91	0	0.23	-	1.00E+00	0.00
adenylate cyclase activity	89	0	0.23	-	1.00E+00	0.00
RNA helicase activity	87	0	0.22	-	1.00E+00	0.00
DNA helicase activity	86	0	0.22	-	1.00E+00	0.00
voltage-gated potassium channel activity	81	0	0.21	-	1.00E+00	0.00
nucleotidyltransferase activity	81	0	0.21	-	1.00E+00	0.00
cytokine receptor activity	78	0	0.2	-	1.00E+00	0.00
ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity	75	0	0.19	-	1.00E+00	0.00
acetylcholine receptor activity	73	0	0.19	-	1.00E+00	0.00
GABA receptor activity	73	0	0.19	-	1.00E+00	0.00
hydro-lyase activity	65	0	0.17	-	1.00E+00	0.00
racemase and epimerase activity	64	0	0.16	-	1.00E+00	0.00
phosphoric diester hydrolase activity	61	0	0.16	-	1.00E+00	0.00
phosphatase regulator activity	61	0	0.16	-	1.00E+00	0.00
translation elongation factor activity	54	0	0.14	-	1.00E+00	0.00
structural constituent of myelin sheath	53	0	0.14	-	1.00E+00	0.00
nucleotide binding	51	0	0.13	-	1.00E+00	0.00
hydrogen ion transmembrane transporter activity	50	0	0.13	-	1.00E+00	0.00
deaminase activity	49	0	0.13	-	1.00E+00	0.00
hydrolase activity, hydrolyzing N-glycosyl compounds	48	0	0.12	-	1.00E+00	0.00
DNA-directed RNA polymerase activity	48	0	0.12	-	1.00E+00	0.00
neuropeptide hormone activity	48	0	0.12	-	1.00E+00	0.00
nucleotide phosphatase activity	46	0	0.12	-	1.00E+00	0.00
ATPase activity, coupled to transmembrane movement of substances	43	0	0.11	-	1.00E+00	0.00
endoribonuclease activity	42	0	0.11	-	1.00E+00	0.00
antigen binding	41	0	0.11	-	1.00E+00	0.00
anion channel activity	41	0	0.11	-	1.00E+00	0.00
guanylate cyclase activity	40	0	0.1	-	1.00E+00	0.00

gap junction channel activity	40	0	0.1	-	1.00E+00	0.00
chemokine activity	39	0	0.1	-	1.00E+00	0.00
lipid binding	38	0	0.1	-	1.00E+00	0.00
aminoacyl-tRNA ligase activity	38	0	0.1	-	1.00E+00	0.00
cytokine receptor binding	37	0	0.09	-	1.00E+00	0.00
carbohydrate kinase activity	37	0	0.09	-	1.00E+00	0.00
hydrolase activity, hydrolyzing O-glycosyl compounds	36	0	0.09	-	1.00E+00	0.00
transaminase activity	34	0	0.09	-	1.00E+00	0.00
double-stranded DNA binding	34	0	0.09	-	1.00E+00	0.00
tumor necrosis factor-activated receptor activity	33	0	0.08	-	1.00E+00	0.00
exoribonuclease activity	32	0	0.08	-	1.00E+00	0.00
DNA-directed DNA polymerase activity	31	0	0.08	-	1.00E+00	0.00
antioxidant activity	31	0	0.08	-	1.00E+00	0.00
SNAP receptor activity	29	0	0.07	-	1.00E+00	0.00
carboxy-lyase activity	28	0	0.07	-	1.00E+00	0.00
endodeoxyribonuclease activity	27	0	0.07	-	1.00E+00	0.00
peroxidase activity	26	0	0.07	-	1.00E+00	0.00
phosphatase inhibitor activity	23	0	0.06	-	1.00E+00	0.00
transmembrane receptor protein tyrosine kinase activity	22	0	0.06	-	1.00E+00	0.00
proton-transporting ATP synthase activity, rotational mechanism	21	0	0.05	-	1.00E+00	0.00
RNA methyltransferase activity	21	0	0.05	-	1.00E+00	0.00
transforming growth factor beta-activated receptor activity	20	0	0.05	-	1.00E+00	0.00
damaged DNA binding	20	0	0.05	-	1.00E+00	0.00
DNA-methyltransferase activity	19	0	0.05	-	1.00E+00	0.00
glucosidase activity	18	0	0.05	-	1.00E+00	0.00
cysteine-type endopeptidase inhibitor activity	18	0	0.05	-	1.00E+00	0.00
intermediate filament binding	18	0	0.05	-	1.00E+00	0.00
carbohydrate phosphatase activity	18	0	0.05	-	1.00E+00	0.00
protein disulfide isomerase activity	17	0	0.04	-	1.00E+00	0.00
phosphorylase activity	16	0	0.04	-	1.00E+00	0.00

aspartic-type endopeptidase activity	14	0	0.04	-	1.00E+00	0.00
tumor necrosis factor receptor binding	11	0	0.03	-	1.00E+00	0.00
centromeric DNA binding	11	0	0.03	-	1.00E+00	0.00
intramolecular transferase activity	10	0	0.03	-	1.00E+00	0.00
amylase activity	10	0	0.03	-	1.00E+00	0.00
cyclic nucleotide-gated ion channel activity	10	0	0.03	-	1.00E+00	0.00
neurotrophin receptor binding	9	0	0.02	-	1.00E+00	0.00
RNA-directed DNA polymerase activity	9	0	0.02	-	1.00E+00	0.00
DNA photolyase activity	9	0	0.02	-	1.00E+00	0.00
galactosidase activity	8	0	0.02	-	1.00E+00	0.00
amino acid kinase activity	8	0	0.02	-	1.00E+00	0.00
DNA topoisomerase activity	8	0	0.02	-	1.00E+00	0.00
translation release factor activity	7	0	0.02	-	1.00E+00	0.00
exodeoxyribonuclease activity	5	0	0.01	-	1.00E+00	0.00
transketolase activity	5	0	0.01	-	1.00E+00	0.00
metalloendopeptidase inhibitor activity	3	0	0.01	-	1.00E+00	0.00
DNA ligase activity	2	0	0.01	-	1.00E+00	0.00
DNA polymerase processivity factor activity	1	0	0	-	1.00E+00	
phosphatase activator activity	1	0	0	-	1.00E+00	
DNA primase activity	1	0	0	-	1.00E+00	