Identification, Evolution, and Expression of the Trace Amine-Associated Receptor (TAAR) Gene Family in Atlantic Salmon (*Salmo salar*)

by

Jordan Anthony Tessarolo

B.Sc. (Hons.) University of Victoria, 2011

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2014

Approval

Name:	**Jordan Anthony Tessarolo**
Degree:	**Master of Science**
Title:	**Identification, Evolution, and Expression of the Trace Amine-Associated Receptor (TAAR) Gene Family in Atlantic Salmon (*Salmo salar*)**
Examining Committee:	**Chair:** Dr. Sharon Gorski, Associate Professor

Dr. William S. Davidson
Senior Supervisor
Professor, Department of MBB _____

Dr. Erika Plettner
Supervisor
Professor, Department of Chemistry _____

Dr. Mark Paetzel
Supervisor
Professor, Department of MBB _____

Dr. Jack Chen
Internal Examiner
Professor, Department of MBB _____

Date Defended/Approved: September 12, 2014

Partial Copyright Licence

Abstract

It is widely hypothesized that Atlantic salmon are imprinted at a young age with olfactory cues, which they use as a guide in order to return to their natal streams to spawn. However, the molecular mechanism(s) behind this biological phenomenon remain unknown. Therefore, in order to better understand imprinting and homing in Atlantic salmon, it is important to characterize the repertoire of olfactory receptors in this species. A search of the first assembly of the Atlantic salmon genome revealed 27 putatively functional trace amine-associated receptor (TAAR) genes and 25 putative TAAR pseudo-genes. Genetic mapping, phylogenetic analysis, binding-site prediction, and quantitative PCR were performed using the Atlantic salmon TAAR genes. The identification of this gene family in Atlantic salmon will facilitate additional studies involving olfaction and homing such as determining the range of allelic variation in olfactory receptors genes of different salmon populations.

Keywords: Atlantic salmon; salmonid fishes; homing; olfaction; olfactory receptors; trace amine

Acknowledgements

First of all I would like to thank my senior supervisor, Dr. Willie Davidson, for your patience and guidance throughout my project. I also really appreciate your generosity in allowing me to miss a number of days to participate in some of my extra-curricular activities. To the rest of the members of the Davidson lab, it has been a pleasure getting to know you all and I feel very fortunate to have been a part of such a knowledgeable, passionate, and kind-hearted group of people. Thank you to my parents and the rest of my family for their unconditional love and support. Lastly, and most importantly, thank you to my beautiful fiancée Kimberley. I am so grateful that you have been beside me during these past two years. You have really kept me together and I could not have done this without you. I can't wait to start the next chapter of our lives together!

Table of Contents

Approval ii

Partial Copyright Licence iii

Abstract iv

Acknowledgements v

Table of Contents vi

List of Tables viii

List of Figures ix

List of Acronyms xii

Glossary xiii

Chapter 1. Introduction 1

1.1. Atlantic Salmon life cycle 3

1.2. Homing, Imprinting, and Straying 9

1.3. Olfactory System in Fish 12

1.4. Odorant Receptors 16

1.5. Trace amines and their receptors 20

1.6. Gustatory System in Fish 21

1.7. Signal Transduction 22

1.8. Olfactory and Gustatory Coding 26

1.9. Whole Genome Duplication and the Atlantic Salmon Genome 28

Chapter 2. Materials and Methods 31

2.1. Identification and mapping of the Atlantic salmon TAAR genes 31

2.2. Identification of TAARs in other vertebrates 35

2.3. Alignment and Phylogenetic Analysis TAARs 35

2.4. Secondary Structure and Ligand-Binding Prediction 36

2.5. RNA Isolation and cDNA synthesis 36

2.6. Quantitative Real-Time PCR (qPCR) of TAAR genes in wild Atlantic salmon at different life stages from three distinct populations 37

2.7. Quantitative Real-Time PCR (qPCR) Data Collection 41

Chapter 3. Results 42

3.1. Introduction 42

3.2. The Atlantic Salmon TAAR Repertoire 42

3.3. Genomic Organization of Atlantic Salmon TAARs 43

3.4. Evolution of Vertebrate TAARs 44

3.5. Phylogenetic Analysis of Atlantic Salmon TAARs 46

3.6. Predicted Structure and Binding Sites of Atlantic Salmon TAARs 48

3.7. Expression of TAAR genes at different life stages and in different populations of wild Atlantic salmon 54

Chapter 4. Discussion 56

4.1. The Atlantic Salmon TAAR Repertoire 56

4.2. Genomic organization of Atlantic Salmon TAARs 58

4.3. Evolution of Vertebrate TAARs 58

4.4. Phylogenetic Analysis of Atlantic salmon TAARs 61

4.5. Predicted Structure and Binding Sites of Atlantic Salmon TAARs 61

4.6. Expression of TAAR genes at different life stages and in different populations of wild Atlantic salmon 62

Chapter 5. Conclusions and Future Directions 65

References 67

Supplementary Material 80

Appendix A 80

List of Tables

Table 1: Microsatellite markers used to map the Atlantic salmon contigs containing TAAR genes in the Br5 and Br6 mapping families. An M13 tag has been added to the forward primers. 34

Table 2: Summary of samples used for qPCR (Johnstone *et al*., 2011). 38

Table 3: Primer sequences for genes used in qPCR analysis 40

Table 4: Number of putatively functional TAAR genes included in this report by species compared to the previously report number of putatively functional TAAR receptors 57

List of Figures

Figure 1: Atlantic salmon life cycle © Atlantic Salmon Federation and photo credit to the artist J. O. Pennanen. 5

Figure 2: Schematic of the North Atlantic subpolar gyre (→) with average velocities (km/day) of the major surface currents. BIC, Baffin Island Current; EGC, East Greenland Current; FC, Faroe Current; IC, Irminger Current; LC, Labrador Current; NAC, North Atlantic Current; NCC, Norwegian Coastal Current; NwAT, Norwegian Atlantic Current; SC, Slope Current; WGC, West Greenland Current. Figure adapted from Dadswell *et al.* (2010). 7

Figure 3: Schematic of the fish olfactory system. Crypt cells are in red, microvillous sensory neurons in green, and ciliated sensory neurons in blue. LOT, Lateral Olfactory Tract; l MOT, lateral Medial Olfactory Tract; m MOT, medial Medial Olfactory Tract. Diagram retrieved from Hamandi and Døving (2007). 15

Figure 4: Depicts the number of ORs genes (Intact, Truncated, or Pseudogenes) identified in whole genome sequences of the various species examined. The percentage of pseudogenes (P%) is calculated from the number of pseudogenes divided by the total number of ORs. Some truncated genes may in fact be pseudogenes, and the P% may be inaccurate for low-coverage genomes. Figure adapted from Niimura (2012). 19

Figure 5: Schematic of the cAMP (A) and IP3 (B) secondary messenger signal transduction pathways utilized in the primary sensory neurons of the vertebrate olfactory system. R, olfactory receptor; G, G protein; cAMP, cyclic AMP; AC, adenylyl cyclase; CNG, cyclic nucleotide gated channel; GRK, G-protein-coupled receptor kinase; PKA, protein kinase A; PDE, phosphodiesterase; CAM, camodulin; PI3K, phosphoinositide 3-OH kinase; PLC, phospholipase-C; DAG, diacylglycerol; IP3, inositol 1,4,5-trisphosphate; PIP3, phosphatidylinositol 3,4,5- trisphosphate; PIP2, phosphatidylinositol 4,5-bisphosphate; IP3R, plasma membrane IP3 25

Figure 6: Schematic representation of the labelled-line (A) and across-fiber pattern (B) hypotheses in the coding of olfactory and gustatory information. In the labelled-line hypothesis, the receptors respond to a very narrow range of stimuli. Therefore, odour A and odour B activate different receptors with no overlap. This allows for more detailed information regarding a limited number of stimuli to be detected. In the across-fiber pattern hypothesis, the receptors are said to respond to a wide range of stimuli. Odour A and odour B both activate some of the same receptors, but to a different degree. Therefore, the combinatorial effects allow a much larger repertoire of stimuli that can be detected, but at the cost of sensitivity. Figure adapted from Sandoz (2007). 27

Figure 7: Bayesian phylogenetic analysis of 246 full-length putatively functional TAAR amino acid sequences from fifteen vertebrate species including lamprey (*Petromyzon marinus*), elephant shark (*Callorhinchus milii*), coelacanth (*Latimeria chalumnae*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), zebrafish (*Danio rerio*), frog (*Xenopus tropicalis*), human (*Homo sapiens*), mouse (*Mus musculus*), lizard (*Anolis carolinensis*), alligator (*Alligator mississippiensis*), chicken (*Gallus gallus*) and Atlantic salmon (*Salmo salar*) (see supporting Table A2 for summary). Rhodopsin from human was used as an out-group. The clades containing Atlantic salmon TAARs are indicated by an asterisk..... 45

Figure 8: Bayesian phylogenetic analysis of 27 full-length putatively functional Atlantic salmon (*Salmo salar*) TAAR amino acid sequences using the histamine receptor H2 from *Danio rerio* as an out-group. Information regarding physical and genetic location has been added beside the contig name when available. The clades correspond to those in Figure 1. Alignment provided in File S3. 47

Figure 9: Amino acid alignment of putatively functional Atlantic salmon TAARs. Secondary structures were predicted using TMHMM2.0. The predicted extracellular segments have been highlighted blue, transmembrane segments red, and intracellular segments green. The TAAR fingerprint motif has been coloured yellow about the alignment. Predicted ligand-binding site residues from RaptorX-*binding* software have been highlighted brown below the alignment. TM, Transmembrane segment; IL, Intracellular Loop; EL, Extracellular Loop. 49

Figure 10: Schematic 2D structure of an Atlantic salmon TAAR based on the consensus sequence from the multiple alignment of putatively functional TAARs shown in Figure 9. The secondary structure was predicted using TMHMM2.0 and ligand-binding site residues predicted by RaptorX-*binding* (also shown in figure 9). The predicted binding sites have been coloured green. The TAAR fingerprint motif has been coloured blue. 53

Figure 11: Relative Quantification (RQ) of two TAAR genes in three populations of wild Atlantic salmon at different life stages. 55

List of Acronyms

BAC	Bacterial Artificial Chromosome
CAM	calmodulin
cAMP	cyclic 3',5-adenosine monophosphate
CNG	Cyclic-nucleotide gated ion channel
CNS	Central nervous system
DAG	diacylglycerol
DDC	duplication-degeneration-complementation
GPCR	G-protein coupled receptor
GRK	G-protein-coupled receptor kinase
IP3	inositol 1,45-triphosphate
LOT	Lateral olfactory tract
MHC	Major Histocompatibility Complex
MOR	Main olfactory receptor
MOT	Medial olfactory tract
MYA	Million years ago
OR	Odorant Receptor
OSN	Olfactory sensory neuron
PDE	Phosphodiesterase
PI3K	phosphoinositide 3-OH kinase
PIP2	phosphatidylinositol 4,5-bisphosphate
PIP3	phosphatidylinositol 3,4,5- trisphosphate
PLC	phospholipase C
PST	Parr-smolt transformation
qPCR	quantitative polymerase chain reaction
RQ	Relative Quantification
TAAR	Trace amine-associated receptor
TM	Transmembrane
V1R	Vomeronasal type I receptor
V2R	Vomeronasal type II receptor
WGD	Whole Genome Duplication

Glossary

Alevin	Salmon life stage from hatching to end of dependence on the yolk sac
Anadromous	Fish that migrate from the marine environments to freshwater systems to spawn
BAC	Bacterial artificial chromosome; a large fragment of a genome has been cloned into a bacterial plasmid
Binding Site	The portion of a protein responsible for interacting with a particular ligand.
Contig	A DNA sequence based on the overlapping of many reads
EST	Expressed sequence tag; a DNA sequence known to be within the coding region of a gene
Fry	Salmon life stage from independence of yolk sac and dispersal from the redd
Gustation	Sense of taste
Grilse	Salmon which spawn after only one winter in the ocean
gyre	rotating ocean current
Homing	Refers to salmon returning to their natal streams to spawn
Homeologous Chromosomes	Two chromosomes where one is the result of a whole genome duplication event.
Imprinting	A phase-sensitive learning process
Linkage Map	A map which depicts the relative position of genes or genetic markers in terms of recombination frequency
Microsatellite	A series of di- and tri- nucleotides repeats in DNA
Multi-sea-winter salmon	Salmon which spend more than one winter in the ocean prior to spawning
Neuron	A cell of the nervous system that processes and transmits information via electrical and chemical signals
Odorant	A chemical than can be sensed through an organisms olfactory system
Olfaction	Sense of smell
Parr	Salmon life stage from dispersal from the redd to parr-smolt transformation (PST)
Phylogeny	The evolutionary development and history of a species or higher taxonomic grouping of organisms.

Pseudo-gene	A segment of DNA that has lost its ability to code for a functional protein. Often through the accumulation of many mutations.
PST	Parr-smolt transformation; A period of time when salmon undergo morphological and physiological stages in preparation for life in a marine environment
Redd	Location where female salmon deposit their eggs
Signal Transduction	Conversion of an extracellular stimulus to a intracellular signal that can be responded to
Smolt	A fully silvered juvenile salmon migrating to the ocean
Straying	Refers to salmon which do not return to their natal stream to spawn.
Vertebrate	a species belonging to the phylum chordata, which possess a true backbone.

Chapter 1.

Introduction

Atlantic salmon (*Salmo salar*) is one of the most extensively studied species within the Salmonidae family due to its significant importance both commercially and recreationally. Atlantic salmon dominate the world salmonid aquaculture industry with leading producers in Norway, Chile, the United Kingdom, and Canada, respectively. In British Columbia, Canada, farmed salmon is the largest agricultural export.

Salmon also have unique life histories that make them models for studying behaviour, ecology, and evolution. However, wild Atlantic salmon populations have been in decline since the early 19th century (NOAA Fisheries, Office of Protected Resources). A 2010 assessment by The Committee On the Status of Endangered Wildlife in Canada (COSEWIC, www.cosewic.gc.ca) reported five populations of Atlantic salmon as endangered, one as threatened, four as of special concern, one as extinct, four as not at risk and one as data deficient. Despite restoration initiatives and extensive research into the biology of this species, there has been limited success to conserve wild Atlantic salmon populations.

Part of this can be attributed to the lack of understanding regarding the molecular mechanisms behind salmon olfaction, imprinting and homing. There is significant evidence indicating salmon rely on olfactory cues learned as juveniles to guide their homeward migration for spawning as adults. Therefore, fully understanding genetic,

physiological, and environmental components behind these processes would be critical to the success of conservation efforts.

The first step in understanding the olfactory system in Atlantic salmon is characterizing the repertoire of odorant receptors (ORs) this species. The Davidson lab has previously shown that Atlantic salmon have seven functional and one pseudo- Ora genes (Johnstone *et al*., 2008), which correspond to the vomeronasal type I genes in mammals (Saraiva and Korsching, 2007), and 29 functional and 26 pseudo- OlfC genes (Johnstone *et al*., 2009), which correspond to mammalian vomeronasal type II genes (Alioto and Ngai, 2006). To date, 24 putatively functional and 24 pseudo- main olfactory receptor genes have been identified (Johnstone *et al*., 2012). However, no attempt has been made to identify and characterize Atlantic salmon genes encoding trace amine-associated receptors (TAARs), the fourth class of olfactory receptors, which appears to have emerged in the common ancestor of vertebrates (Grus and Zhang, 2009; Libants *et al*., 2009). The goals of this study were: (1) to identify genes encoding TAARs by screening the first assembly of the Atlantic salmon genome (Davidson *et al*., 2010); (2) to determine their genomic organization; (3) to predict key ligand binding site residues; and (4) to examine the evolution of the TAAR gene family by phylogenetic analysis with their counterparts in other teleosts and tetrapods. The molecular genetic and bioinformatic research methodologies used to carry out this research are described in Chapter 2. The results are presented in Chapter 3 and their implications are discussed in Chapter 4. Lastly, Chapter 5 provides a brief summary and suggests possible future directions.

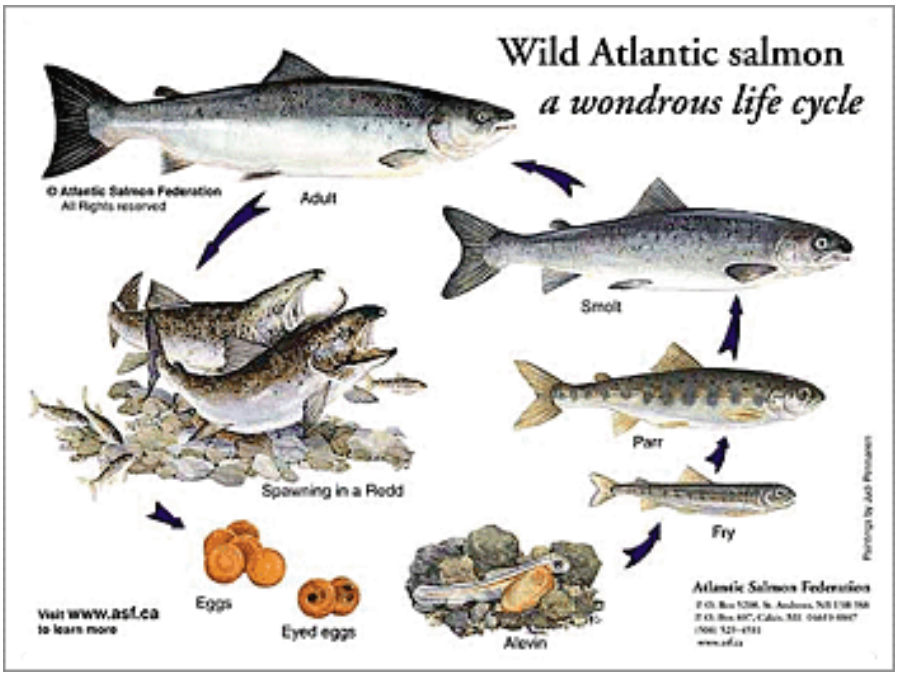
1.1. Atlantic Salmon life cycle

Atlantic salmon display two main life history strategies: anadromous and non-anadromous. Anadromous Atlantic salmon refers to populations that migrate from the ocean to freshwater for spawning whereas non-anadromous Atlantic salmon (a.k.a landlocked or ouananiche) remain in freshwater systems for their entire life.

Anadromous Atlantic salmon hatch in freshwater where they spend a minimum of two years (reviewed in Klemetsen *et al*., 2003). As they grow from alevin to fry to parr, salmon venture further from their redd, which is located in gravel beds with moderate current velocity and depth, to faster flowing riffles (reviewed in Fleming, 1996). After anywhere between 2-8 years, depending on the population, salmon undergo parr-smolt transformation (Figure 1). This involves drastic morphological, physiological, and behavioural changes required for adapting to life in a marine environment. These morphological changes include the loss of parr marks and the development of a silver/reflective appearance resulting from the deposition of guanine and hypoxanthine in the skin and scales. There is also a darkening of fin margins and an increase in their length to weight ratio. The physiological changes prepare their bodies for the change from freshwater to seawater. Salinity tolerance is achieved by increased Na^+,K^+-ATPase activity, number and size of gill chloride cells, and intestinal water permeability. However, there are also metabolic changes required for increased growth, a shift from porphyropsin to rhodopsin for vision, the use of adult isoforms of haemoglobin, and increased buoyancy. Behavioural changes such as increased negative rheotaxis and

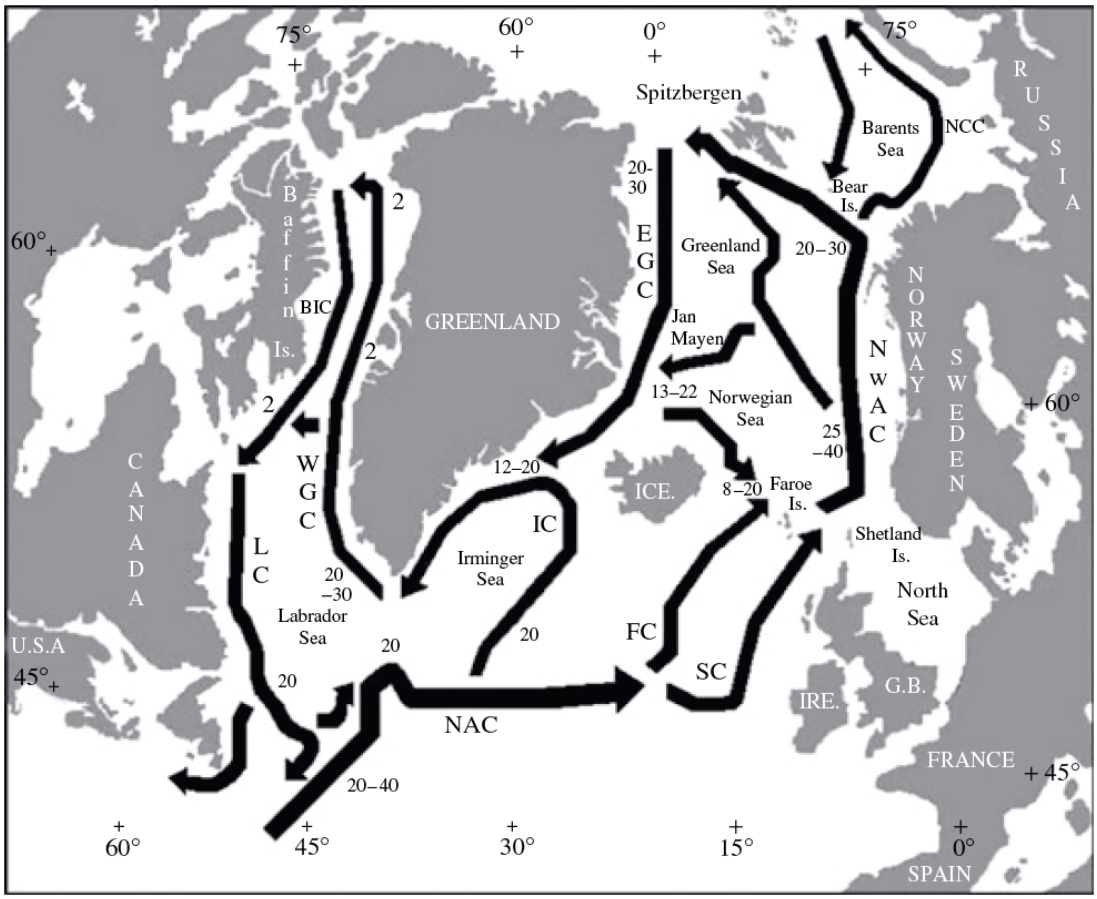
schooling, decreased agonistic and territorial behaviour, and preference for increased salinity have also been observed (McCormick *et al*., 1998).

Figure 1: Atlantic salmon life cycle © Atlantic Salmon Federation and photo credit to the artist J. O. Pennanen.



Compared to the freshwater phase, relatively little is known about the marine phase of the Atlantic salmon life cycle. Once they reach the ocean, Atlantic salmon migrate to feeding grounds where they spend the majority of their time near the surface and feed on anything from small invertebrates, such as insects, to fish and squid (reviewed in Hansen and Quinn, 1998). These feeding grounds can be several thousand kilometers away. One hypothesis states the fish migrate in relatively straight lines from their natal rivers to feeding grounds whereas another, known as the 'Merry-Go-Round Hypothesis,' suggests salmon migrate counter clockwise as they follow surface currents of the North Atlantic Subpolar gyre (NASpG) (Figure 2) (Dadswell *et al*., 2010). There are a wide range of marine survival rates reported (5 – 40%) depending on the population (reviewed in Hansen and Quinn, 1998). Survival rates are influenced by a variety of factors including disease, predation, food resources, and other environmental factors. During the ocean phase Atlantic salmon experience rapid growth and become sexually mature at which point they migrate to their natal streams to spawn in a process called homing. The initiation of homing is highly variable. For example, Atlantic salmon known as grilse only spend one winter at sea whereas others, multi-sea-winter salmon, can spend two to four years. Little is known regarding the biological mechanisms responsible for initiating the homing process; however the circannual rhythms of reproductive hormones synchronized by photoperiod are likely involved (reviewed in Hansen and Quinn, 1998).

Figure 2: Schematic of the North Atlantic subpolar gyre (→) with average velocities (km/day) of the major surface currents. BIC, Baffin Island Current; EGC, East Greenland Current; FC, Faroe Current; IC, Irminger Current; LC, Labrador Current; NAC, North Atlantic Current; NCC, Norwegian Coastal Current; NwAT, Norwegian Atlantic Current; SC, Slope Current; WGC, West Greenland Current. Figure adapted from Dadswell *et al.* (2010).



Not all populations of Atlantic salmon are migratory. There are also non-anadromous salmon which spend their entire lives in freshwater systems. It is thought that these populations, also known as ouananiche salmon, were originally formed due to physical barriers, either natural or man-made, which resulted in geographic isolation and thus prevented these salmon from ever reaching the ocean (Power, 1958; Berg, 1958). However, there are a number non-anadromous salmon populations that do have access to the ocean. In this situation it is possible to have both anadromous and non-anadromous populations existing sympatrically, as is the case in Big Triangle Pond in Newfoundland. Although these populations are in close contact, they remain genetically and behaviourally distinct suggesting that inter-mixing between the two populations does not occur despite being capable. This is also supported by the fact their spawning grounds are physically separated (Couturier *et al*., 1986). There are also physiological and morphological differences between these populations with these two life history strategies. Non-anadromous salmon do not undergo the full parr-smolt transformation as they are not exposed to a marine environment. As a result, there is not the increase in Na^+-K^+ ATPase activity or in the number of gill chloride cells. They also do not experience the same degree of body silvering and as a result their parr marks remain visible. Compared with anadromous salmon, non-anadromous salmon are much smaller. At maturity they are often smaller than 15 cm and hence they have earned the nickname “dwarf” salmon. Like anadromous salmon, juveniles spend two to three years in slow moving streams. Seasonal changes and other environmental and physiological factors initiate their migration to ponds or lakes where they mature sexually and spawn.

1.2. Homing, Imprinting, and Straying

Adult salmon are marvelled for their ability to accurately return to their natal streams to spawn in a process known as homing. Homing can be thought of as a two stage process: the first phase being an ocean migration where salmon locate the freshwater river system and the second stage where salmon decipher between various tributaries and identify their natal stream.

Putman *et al*. (2013) have recently shown that Pacific salmon utilize a geomagnetic imprinting mechanism as a means of navigating through the open ocean during the first stage of their migration, but there is strong evidence indicating that olfaction plays a key role in salmon homing during the freshwater phase (Putman *et al.,* 2013). Two main hypotheses exist around olfaction and homing in salmonids. The first, from studies involving Arctic char (*Salvelinus alpinus*) and Atlantic salmon (*Salmo* salar), suggests that the release of a population-specific pheromone by juveniles guides returning adults (Nordeng, 1971, 1977). However, this cannot explain homing in chum (*Oncorhynchus keta*) and pink (*Oncorhynchus gorbuscha*) salmon. These species mature relatively quickly and as a result there are no juveniles present in the natal streams when the adults return. The second, and more widely accepted hypothesis, was introduce by Hasler and Wisby in 1951 (Hasler and Wisby, 1951). Their hypothesis was based on three main tenets: (1) odorant cues are derived from unique water chemistries that are influenced by geographic differences in soil and vegetation; (2) juvenile salmon are imprinted with this odour prior to their migration to the sea; (3) adult salmon are able to respond to this odorant cue during their upstream migration (Hasler and Wisby, 1951).

The sensitivity of the fish olfactory system was first observed in some fundamental experiments (Hasler and Wisby, 1951). Using classical conditioning, the authors trained bluntnose minnows (*Pimephales notatus*) and coho salmon (*Oncorhynchus kisutch*) to distinguish between two water sources (Hasler and Wisby, 1951). The fishes' ability to decipher between the two water sources was disrupted when either the nares were cauterized or the organic fraction of the water sample was removed. As a control, the authors found blinded fish with an intact olfactory organ could still discriminate between the two water sources confirming these fish were relying on olfaction rather than visual cues. Although bluntnose minnows are non-anadromous, these experiments demonstrate fishes' ability to discriminate between different water sources using olfactory cues. These findings led to numerous studies regarding the potential link between homing and olfaction, which became better known as the odour hypothesis.

One such field study supporting the odour hypothesis involved the capture of migrating adult coho salmon after they had navigated a fork in Issaquah Creek near Seattle, WA (Wisby and Hasler, 1954). The authors found that if they released the fish 1.6 km downstream of the fork and allowed them to repeat their upstream migration, 65 of the 73 control fish returned to the branch where they were originally captured. However, if prior to release the salmon had their nares plugged with Vaseline coated cotton, only 42 of the 70 salmon chose the same path. These sensory-impairment studies have been highly criticized. Opponents argue the disrupted homing ability can explained as a side effect of the traumatic events such as cauterization or plugging of nares the animals went through in these experiments.

Further evidence implicating the role of olfaction in homing comes from electrophysiological studies (Hara *et al*., 1965; Ueda *et al*., 1967). When the nares of salmon are flushed with water from their home stream, there is a characteristically high intensity electroencephalographic (EEG) reading compared to when other water sources are used. When fish from other populations are tested, the EEG reading is the strongest when water from their respective natal streams is used.

In another critical field study that supported the odour hypothesis, the authors exposed juvenile coho salmon to either morpholine or phenethyl alcohol over a six week period (Scholz *et al., 1*976). The salmon were then released into Lake Michigan. Eighteen months later, the artificial odorants were released into two rivers that feed into Lake Michigan. Overall, 92% of salmon returned to the river with the scent they were exposed to as juveniles. For example, fish that had been previously exposed to morpholine were found to return to the morpholine-scented river whereas phenethyl alcohol exposed fish returned to the phenethyl alcohol scented river.

More recent two-tank tests by some Japanese research groups have indicated that Pacific salmon can be imprinted with a single odorant or a mixture of odorants and that parr-smolt transformation is a particularly sensitive period for imprinting. By reconstituting components of natal stream water, these researchers have determined that the odorant cues utilized by returning adult salmon are at least in part derived from the amino acid composition of stream water (Shoji *et al*., 2000, 2003; Yamamoto and Ueda, 2009; Yamamoto *et al*., 2010, 2013).

Salmon are well known for their precise homing. This behaviour allows salmon to return to appropriate spawning grounds and helps to sustain the population. Since salmon return to their natal streams with such high fidelity, homing results in the formation of genetically distinct populations. The alternative to homing, where salmon do no return to their natal stream and spawn elsewhere, is known as straying (Bams, 1976). Straying has two main advantages. It allows for the colonization of new habitats, which may be important when a particular water system is unstable (ie. drought). Secondly, straying allows for gene flow through interbreeding between populations. However, excessive straying can have negative impacts by disrupting the gene pool in locally adapted populations and homogenizing genetic diversity between populations. These are major concerns for the introduction of hatchery fish into river systems as well as escapees from sea based aquaculture facilities (Vasemägi *et al.*, 2005). It is estimated that the straying rates in salmonids of hatchery origin are greater than with their wild counterparts, but the biological basis and environmental aspects impacting straying have not been studied extensively (Grant, 1997). Factors such as population size, stream size, stream stability, and smolt age are likely players.

1.3. Olfactory System in Fish

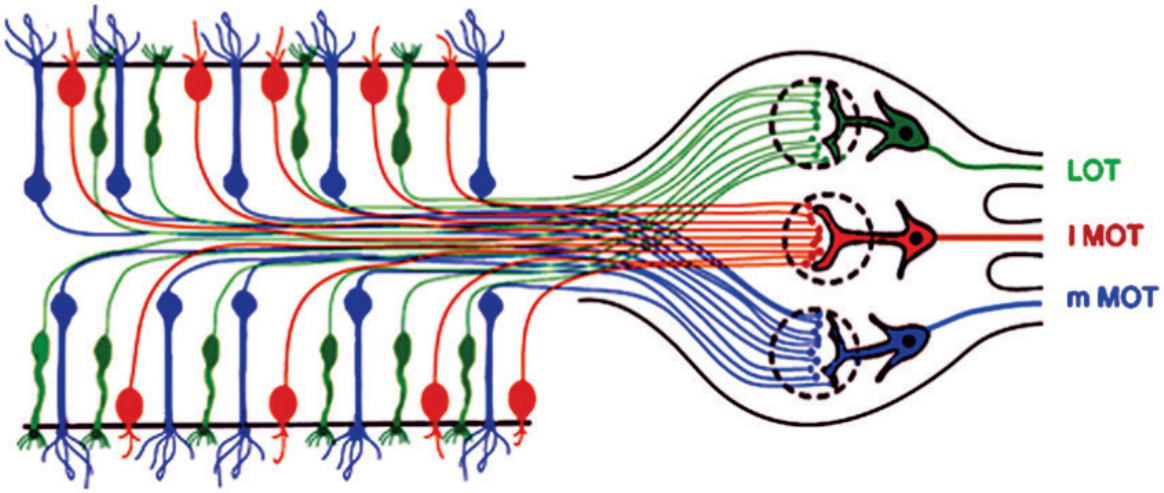
The role of olfaction in homing has already been introduced, but it is also involved in other fundamental processes in the life of a fish such as feeding, predation avoidance, and reproduction. Fish have a single olfactory organ called the olfactory rosette. The highly folded structure serves to maximize surface area of the sensory epithelium, where the olfactory sensory neurons (OSNs) are embedded. These OSNs can be grouped into

three main types based on their cellular morphology: ciliated sensory neurons, microvillous sensory neurons, and crypt cells. Ciliated sensory neurons have few cilia and long dendrites. Microvillous sensory neurons have shorter dendrites and microvillae visible on their surface. Crypt cells are close to the epithelial surface and have both cilia and microvillae. The staggered lengths of dendrites create a pseudo-stratified structure in the sensory epithelium (reviewed in Hamdani and Døving, 2007). OSNs are randomly distributed throughout the olfactory epithelium and each OSN expresses a single OR.

The thin axons of these primary sensory neurons, along with other OSN expressing the same receptor, terminate in stereotypically organized structures called glomeruli within the olfactory bulb (Ressler *et al.*, 1994). At this point, the electrical impulses are processed and sent to the appropriate higher brain centers via olfactory tracts (Figure 3). These olfactory tracts are named based on their anatomical positions and each is associated with distinct behaviours. The medial portion of the medial olfactory tract (m MOT) serves as an alarm reaction, whereas the lateral part of the medial olfactory tract (l MOT) is implicated in reproduction. The lateral olfactory tract (LOT) is involved in feeding (reviewed in Hamdani and Døving, 2007). Interestingly, it has been shown that ciliated sensory neurons terminate on secondary neurons that are a part of the m MOT (Hamdani and Døving, 2002). Crypt cells on the other hand terminate on secondary neurons of the lMOT (Hamdani and Døving, 2006). Lastly, the microvillus sensory neurons terminate on secondary neurons of the LOT (Hamdani *et al*., 2001). Therefore, it is not difficult to see how different odours are able to cue different behavioural responses depending on which receptor(s) are activated, and which olfactory tract these signals are relayed on.

Similar to how the visual and touch senses create topographic maps, the precise and highly conserved organization of glomeruli within the olfactory bulb allows odorants to be spatially organized in the brain based on the pattern of activated glomeruli. This is supported by evidence that odorants with similar chemical properties activate glomeruli that are near each other whereas different odorants activate glomeruli in distinct locations (reviewed in Johnson and Leon, 2007).

Figure 3: Schematic of the fish olfactory system. Crypt cells are in red, microvillous sensory neurons in green, and ciliated sensory neurons in blue. LOT, Lateral Olfactory Tract; l MOT, lateral Medial Olfactory Tract; m MOT, medial Medial Olfactory Tract. Diagram retrieved from Hamandi and Døving (2007).



1.4. Odorant Receptors

Odorant receptors (ORs) are one of the largest known multi-gene families. They are G-protein coupled receptors (GPCRs) and initiate signal transduction by associating with G-proteins upon ligand binding (Buck and Axel, 1991). GPCRs are hallmarked by their seven alpha-helical transmembrane (TM) segments, but can be subdivided into six classes based on sequence similarity. ORs belong to the largest GPCR superfamily: the Class A (rhodopsin-like) gene family, which includes the opsin genes and neurotransmitter receptors (Niimura, 2012). ORs have many conserved motifs. One such motif shared between all rhodopsin-like GPCRs is the 'DRY' motif found between TM3 and IL2 that is thought to be important in G-protein coupling (Rovati *et al*., 2007). Other motifs include an N-linked glycosylation site, conserved cysteine residues and variable residues within TM3 to TM7 that may be important in ligand binding (Man *et al*., 2004).

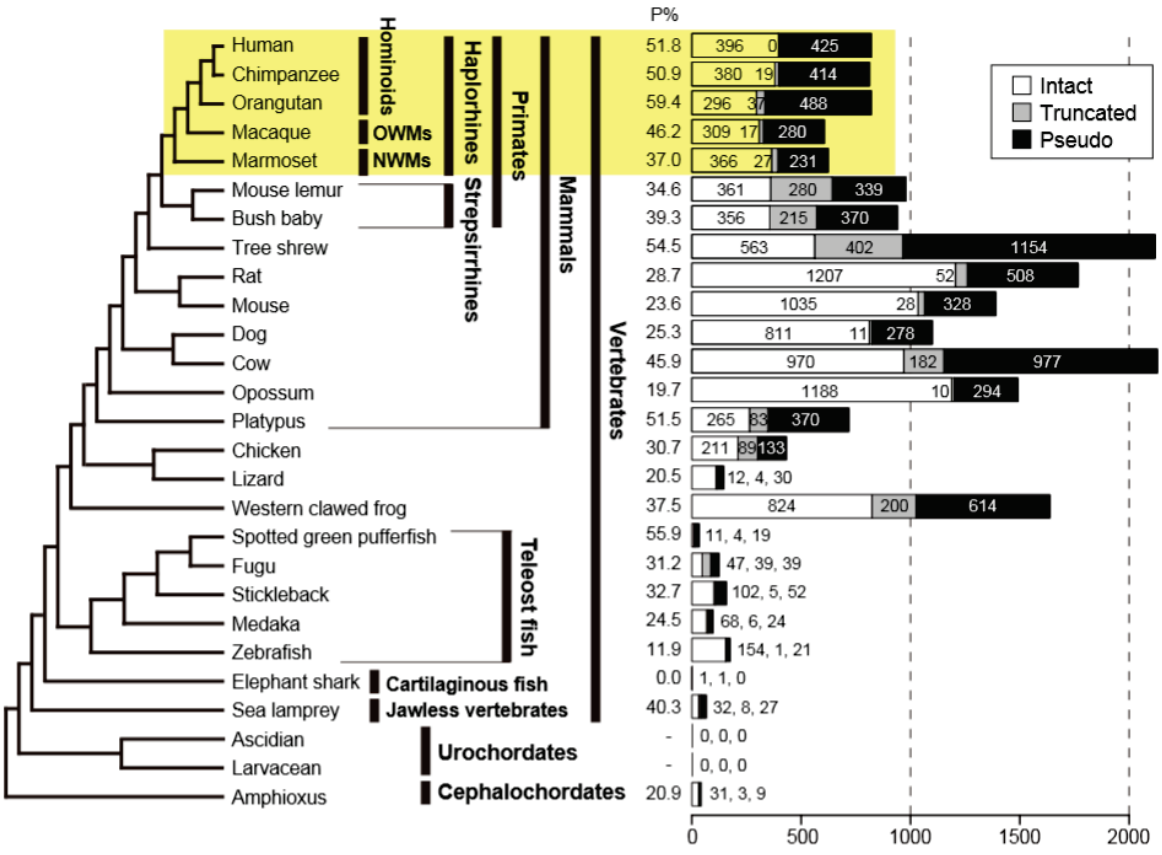
Fish possess four types of ORs. There include the main olfactory receptors (MORs), homologs of the mammalian vomeronasal type I receptors (V1Rs) referred to as Ora receptors, homologs of the mammalian vomeronasal type II odorant receptors (V2Rs) known as OlfC receptors, and the trace amine-associated receptors (TAARs) (Alioto and Ngai, 2006; Saraiva and Korsching, 2007; Johnstone *et al*., 2012). MORs are expressed in ciliated sensory neurons. Ora and OlfC receptors are expressed in crypt cells and microvillious sensory neurons, respectively (reviewed in Hamdani and Døving, 2007). The cells which express the trace amine-associated receptors have not been identified.

The number of functional and pseudogenized ORs in vertebrates varies and is thought to be a reflection of the functional requirements of each species (Niimura and Nei, 2007). For example, land and aquatic vertebrates have drastic differences in their repertoire of ORs since only airborne or dissolved odorants need to be detected. Frogs on the other hand, which live in both terrestrial and aquatic environments, have an OR repertoire that resembles both land and aquatic vertebrates. Another key example is the difference in the OR repertoire between human and the mouse genome. Humans have about a third of the number of ORs and no V2R genes. The vision priority hypothesis suggests this loss is due the heavy reliance on trichromatic vision in humans (Gilad *et al*., 2004). However, there are no significant differences in the number of functional OR genes between New World monkeys and those of Old World monkeys and hominoids indicating that the degeneration of OR genes in primates cannot be solely explained by the acquisition of well-developed trichromatic vision (Matsui *et al*., 2010). Another hypothesis, the brain-function hypothesis, postulates that the higher functioning brain in humans allows the discrimination of more odorants with the same repertoire of ORs (Nei *et al*., 2008).

A single type of OR is expressed on the surface of an OSN: a concept known as the 'one neuron-one receptor rule' (Chess *et al*., 1994; Ishii *et al*., 2001). This notion was first discovered using DNA and RNA fluorescent *in situ* hybridization (FISH). It was found that using DNA probes one could detect the paternal and maternal alleles at two different loci. With an RNA probe on the other hand, only a single allele could be detected indicating that only a single allele is expressed. Unlike in the immune system where antibody diversity and selection is achieved through alternative splicing, the ORs

are expressed from separate genes that are clustered in various regions of the genome. However, the molecular mechanisms responsible for the activation and regulation of a single OR allele are not clear. There is little evidence to indicate that a DNA recombination mechanism, like that seen in antibody expression in B lymphocytes, or a gene conversion mechanism, which is observed in the switching between two mating types in yeast, is involved. Serizawa *et al*. (2003) identified a 2 kb conserved region in mouse and humans located 75 kb upstream of the MOR28 gene (Serizawa *et al*., 2003). This homology (H) region was required for the expression of three OR genes on a yeast artificial chromosome. When the H region was deleted, they observed no expression of OR genes. The authors hypothesize the H region is a cis-acting element that stochastically activates expression of an OR gene and a subsequent negative feedback mechanism inhibits the expression of other OR genes (Serizawa et al. 2003; Lewcock et al. 2004). A similar model proposed by Magklara *et al*. (2011) states that all OR alleles are silenced prior to any transcription (Magklara *et al.*, 2011). An allele is then stochastically chosen to be activated and a feedback loop maintains the silencing of the other alleles.

Figure 4: Depicts the number of ORs genes (Intact, Truncated, or Pseudogenes) identified in whole genome sequences of the various species examined. The percentage of pseudogenes (P%) is calculated from the number of pseudogenes divided by the total number of ORs. Some truncated genes may in fact be pseudogenes, and the P% may be inaccurate for low-coverage genomes. Figure adapted from Niimura (2012).



1.5. Trace amines and their receptors

The biogenic amines are a group of molecules which have considerable biological effects by acting as chemical messengers. These include the classical examples of hormones and neurotransmitters such as adrenalin, dopamine, and serotonin. However, another group of amine compounds which constitute less than 1% of the total biogenic amines exists. They are found in the central nervous system of mammals in the 0.1-10 nM range and hence are termed 'trace amines.' These include p-tyramine, β-phenylethylamine and tryptamine, which are the decarboxylated products of the aromatic amino acids tyrosine, phenylalanine and tryptophan, respectively (Berry, 2004). Trace amines are thought to have a sympathomimetic effect in vertebrates through competitive inhibition with neurotransmitters, such as catecholamine and serotonin (reviewed in Zucchi *et al*. 2006). This point of view changed quite dramatically with the identification of a receptor class specific for these compounds (Borowsky *et al*. 2001; Bunzow *et al*. 2001). The two groups had independently identified a novel sequence from a multiplex PCR reaction based on primers for biogenic amine receptors. The protein encoded by this sequence was found to stimulate production of cAMP in the presence of trace amines and was hence named a trace amine-associated receptor (TAAR). The finding by Liberles and Buck (2006) that all TAAR subtypes, except for TAAR1, are expressed in the olfactory epithelium of mice indicated that both trace amines and their receptors have roles in olfaction. TAARs are typically encoded by a single exon approximately 1 kb in length and a fingerprint motif located in the seventh TM segment defined as NSXXNPXX[YH]XXX[YF]XWF is observed in every member of this receptor family (Lindemann *et al*. 2005). The TAAR repertoire in Atlantic salmon is

the primary focus of this thesis. The genomic organization and evolution of the TAARs will be discussed further in Chapter 4.

1.6. Gustatory System in Fish

There are many parallels that can be drawn between the olfactory and gustatory systems both from an anatomical perspective, but also in the manner of how and what types of information are processed. The two systems are very much complimentary to each other. Similar to how the olfactory system in fish can be broken down into medial and lateral olfactory tracts, the taste or gustatory system can also be divided into two subsystems: taste buds and solitary chemosensory cells (reviewed in Kasumyan and Døving, 2003).

The solitary chemosensory cells are very similar morphologically compared to their taste bud counterparts and, like many taste buds, are also innervated by the facial (VII) cranial nerve. Based on these observations, they were grouped as a part of the gustatory system; however, they have many distinguishing features compared to taste buds. Solitary chemosensory cells are single cells and are distributed relatively uniformly over the body of fish. They have a much more narrow range of stimuli that they respond to, such as bile salts, and do not respond to tactile stimuli. It is thought that like the medial olfactory tract, the solitary chemosensory cells are also involved with inter- and intra-species communications.

Since there are solitary chemosensory cell homologs, but no taste bud homologs, found in the living agnathans (jawless vertebrates), it is thought that taste

buds evolved from solitary chemosensory cells. Taste buds are oval or pear-shaped sensory organs found on the epidermis. They are not only found in the oral cavity, but also on the pharynx, gills, lips, fins, and all over the body, however at different densities. This broad distribution of taste buds is a unique feature amongst gnathostomes (jawed vertebrates) as this is only seen in fish and some amphibians. Fish are also reported to have the largest number of tastes buds of any animal. Taste buds are found on top of small projections in the epithelium called papilla. They are comprised of three cells types: gustatory receptor cells, supporting cells, and basal cells. Gustatory receptor cells and supporting cells are both long, cylindrical shaped cells that run parallel along the axis of the papilla and are exposed to the environment by a small pore in the epithelium. However, only the gustatory receptor cells express the taste receptor cells that are used to detect mainly food related stimuli such as amino acids, but also respond to tactile stimuli. This could be considered analogous to the function of the lateral olfactory tract. The supporting cells simply provide a buffer between gustatory receptor cells in the taste bud, whose number can vary depending on the species. The basal cells are found at the base of the taste buds and attach to the gustatory receptor cells and supporting cells via desmosomes. At the base of each taste bud, between the basal cells and the gustatory receptor cells, are nerve fibers, which innervate from the facial (VII), glossopharyngeal (IX), and the vagal (X) cranial nerves.

1.7. Signal Transduction

Signal transduction refers to the conversion of a stimulus into a biological signal that can be transmitted and ultimately responded to. In the olfactory system, stimuli are

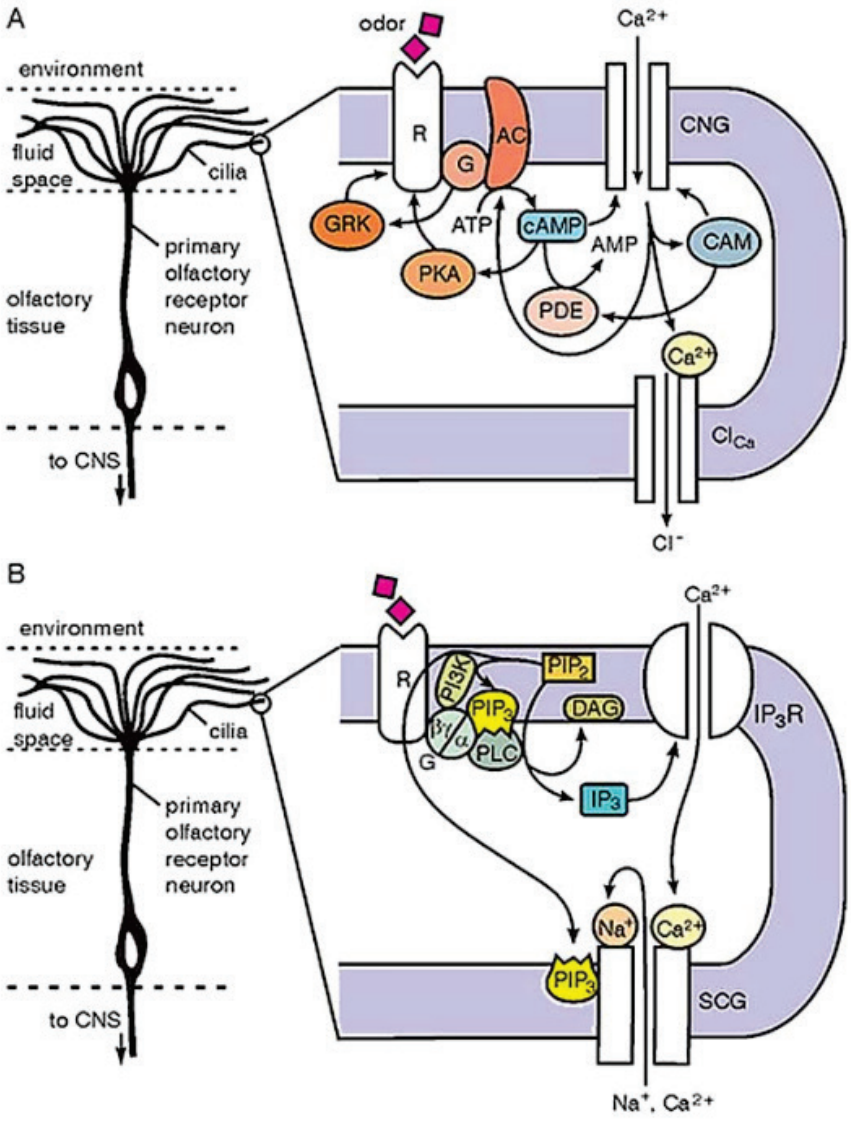
extracellular and therefore second messengers, such as cyclic 3'-5-adenosine monophosphate (cAMP) and inositol 1,4,5-triphosphate (IP3), are required to carry the response intracellularly (Ache and Zhainazarov, 1995). Second messengers also serve to amplify the signal in way that is proportional to the amount of stimuli. Lastly, since many signal transduction pathways use the same second messengers, it allows for cross talk and therefore a more tightly regulated response.

In both the cAMP and IP3 pathways, the second messengers target, amongst other proteins, ion channels (Figure 5). Both signalling pathways are found in eukaryotes, and there has been no evidence for one pathway being chosen over the other. Signal transduction is initiated with the binding of an odorant molecule to an appropriate olfactory receptor (reviewed in Ache and Young, 2005). In the case of the cAMP pathway, receptor binding recruits a G-protein that activates the type-III adenyl cyclase, which converts ATP to cAMP. cAMP causes cyclic-nucleotide gated ion channels (CNG) to open thereby increasing intracellular calcium (Ca^{2+}) concentrations. Ca^{2+} in turn opens chloride channels, which release Cl^- from the cell. The net effect is depolarization of the OR cell. The cAMP pathway is regulated by a number of enzymes. Phosphodiesterase (PDE) reduces cAMP concentrations by cleaving the phosphodiester bond in cAMP producing AMP. G-protein-coupled receptor kinase (GRK) belongs to the serine/threonine kinases and can modulate GPCRs by phosphorylating intracellular domains. Protein kinase A (PKA) performs a similar function as GRK, but is dependent on cAMP concentrations. Lastly, calmodulin (CAM) or calcium-modulated protein binds to calcium and alters its interactions with target proteins.

In the IP3 pathway, receptor binding also recruits a G-protein, but instead of activating type-III adenyl cyclase it activates two different enzymes: phospholipase C (PLC) and phosphoinositide 3-OH kinase (PI3K), which together convert phosphatidylinositol 4,5-bisphosphate (PIP2) into IP3, phosphatidylinositol 3,4,5-trisphosphate (PIP3) and diacylglycerol (DAG). Like cAMP, IP3 can also open CNG channels thereby increasing the intracellular Ca^{2+} concentration. Ca^{2+} and PIP3 are then free to activate sodium and calcium gated channels (SGC), which further increases intracellular Na^{+} and Ca^{2+}. Again, the net outcome is depolarization of the OSN (reviewed in Ache and Young, 2005).

The resting membrane potential of an OSN is between -80 and -60 mV (Lagostena and Menini, 2003). Depending on the affinity of an odorant for a receptor(s) and the concentration of odorant present, the amount of depolarization may reach a threshold. When this threshold is met it creates an action potential which causes the neuron to fire and transmit an electrical impulse. In OSNs, this threshold can be called the odorant threshold. When the odorant concentration increases, the neuron will fire more often. Therefore, the frequency of nerve impulses is proportional to the abundance of a particular odorant in the environmental, which allows information regarding both the presence and abundance of a particular stimulus to be detected (reviewed in Pifferi *et al*., 2010). Detecting the concentration of an odorant is critical in navigating its source.

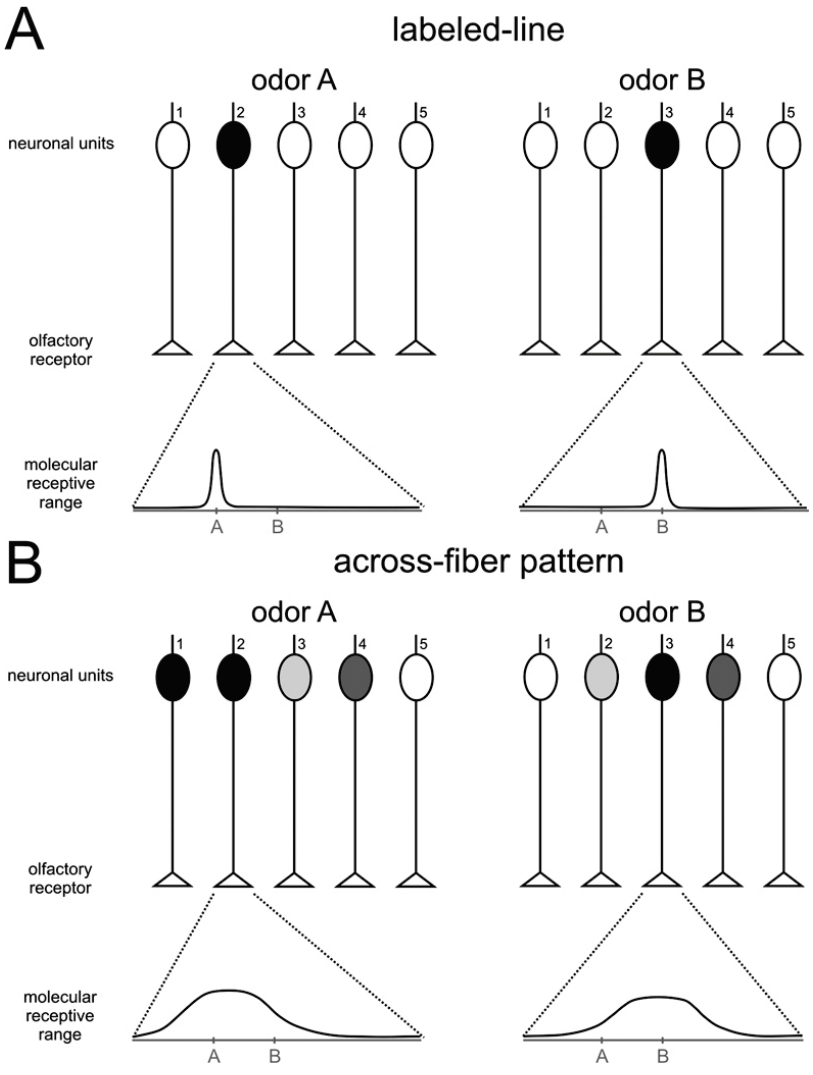
Figure 5: Schematic of the cAMP (A) and IP3 (B) secondary messenger signal transduction pathways utilized in the primary sensory neurons of the vertebrate olfactory system. R, olfactory receptor; G, G protein; cAMP, cyclic AMP; AC, adenylyl cyclase; CNG, cyclic nucleotide gated channel; GRK, G-protein-coupled receptor kinase; PKA, protein kinase A; PDE, phosphodiesterase; CAM, camodulin; PI3K, phosphoinositide 3-OH kinase; PLC, phospholipase-C; DAG, diacylglycerol; IP3, inositol 1,4,5-trisphosphate; PIP3, phosphatidylinositol 3,4,5- trisphosphate; PIP2, phosphatidylinositol 4,5-bisphosphate; IP3R, plasma membrane IP3



1.8. Olfactory and Gustatory Coding

There are two main hypotheses regarding the coding in olfactory and gustatory reception: labeled-line and the across-fiber pattern hypothesis. These hypotheses originate from early electrophysiological research in moths in which the authors state that some OSNs are specialized whereas others are more generalized (Boeckh *et al*., 1965). The labeled-line hypothesis suggests that each receptor is specific for a single or very limited number of stimuli, and upon recognition of that stimulus the information is sent in a direct 'line' to the brain indicating its presence. This mechanism would therefore provide very specific information about the stimulus and may be useful for detection and response to stimuli that have significant biological significance such as pheromones. However, this would imply that a receptor for every possible taste or odorant would need to be expressed. The across-fiber pattern hypothesis states that olfactory and taste receptors are less specific and respond to a broad range of stimuli. For example, the entire population of sensory receptors may respond to a single stimulus. Therefore, the combinatorial effect of using multiple receptors to recognize a single stimulus allows a much larger repertoire of stimuli that could be detected, but at the cost of specificity. These two mechanisms are not mutually exclusive and there is evidence for both depending on how the data are interpreted. It is also important to keep in mind that these connections are not a one-way street. Many of the higher brain centers are interconnected and able to modulate sensory input in a context dependant manner.

Figure 6: Schematic representation of the labelled-line (A) and across-fiber pattern (B) hypotheses in the coding of olfactory and gustatory information. In the labelled-line hypothesis, the receptors respond to a very narrow range of stimuli. Therefore, odour A and odour B activate different receptors with no overlap. This allows for more detailed information regarding a limited number of stimuli to be detected. In the across-fiber pattern hypothesis, the receptors are said to respond to a wide range of stimuli. Odour A and odour B both activate some of the same receptors, but to a different degree. Therefore, the combinatorial effects allow a much larger repertoire of stimuli that can be detected, but at the cost of sensitivity. Figure adapted from Sandoz (2007).



1.9. Whole Genome Duplication and the Atlantic Salmon Genome

Gene duplications and whole genome duplications are thought to be a major driver of evolution by providing redundant copies of genes that are no longer under the constraints of purifying selection (Ohno, 1970). Duplicated genes are free to accumulate mutations with little risk to the organism as a functional copy is still present. According to the duplication-degeneration-complementation (DDC) model, duplicated genes will fall into one of three evolutionary paths. The major fate of duplicated genes is loss of function, or pseudogenization. However, occasionally a duplicated gene may acquire a new function, neofunctionalization, or the function of the original ancestral protein can be divided between the two copies in a process known as subfunctionalization (Force *et al*., 1999). WGDs are known to have occurred in the ancestors of multiple vertebrate, plant, and fungal lineages, and are thought to have facilitated the vast species diversification observed in these phyla (reviewed in Van de Peer *et al*., 2009).

The Salmonidae family is comprised of 11 genera, which include Pacific and Atlantic salmon, trout, graylings, whitefishes, and char (Nelson, 2006). The closest ancestor to the Salmonidae are the Escodiae (pike) and Osmeridae (smelt) and together they make up the superorder Protcanthopterygii, which diverged from other teleosts, Ostariophysi (zebrafish) and Acanthopteryhgii (pufferfish), 217-290 million years ago (MYA) (Ishiguro *et al*., 2003; Yamanoue *et al*., 2006). There are substantial data indicating the common ancestor of the Salmonidae underwent a whole genome duplication. Evidence for a WGD occurring in the common ancestor of the Salmonidae

family includes: the abundance of duplicated gene families such as Hox and MHC genes, tetravalents observed in male meiosis and residual tetrasomic inheritance seen in some species, and homeologous chromosomal segments revealed from genetic maps (Ohno, 1970; Allendorf and Thorgaard, 1984; Danzmann *et al.*, 2005; Moghadam *et al.*, 2005; Lukacs *et al.*, 2007). Furthermore, Atlantic salmon have one of larger genome sizes (3-4.5 pg) amongst euteleosts with sister groups Escodiae and Osmeridae having genome sizes of 0.8-1.8 pg and 0.7 pg, respectively (Gregory *et al.*, 2007). Early estimates indicated this event took play between 25 – 100 MYA (Allendorf and Thorgaard, 1984), however more recent data indicate the WGD occurred 88 – 92 MYA (Macqueen and Johnston, 2014). Interestingly, Macqueen and Johnston (2014) also found that subfamilies, Thymallinae (whitefishes), Coregoniae (graylings), and Salmoninae (salmon, trout, and char), diverged only 40-50 MYA indicating that WGD may not be solely responsible for the species diversification within the Salmonidae (Macqueen and Johnston, 2014). This would be the fourth such event in the evolution of these species: two WGDs in common with the vertebrate lineage (1R and 2R), and a third WGD in common with the extant teleosts (3R) approximately 225 to 333 MYA (Amores *et al*. 1998; Meyer and Van de Peer, 2005). Therefore, Salmonidae are an important family for studying topics revolving around WGDs such as the fates of duplicated genes and determining how organisms cope with the tetraploid state. I hypothesize that the Salmonidae WGD may have major implications of the repertoire and genomic organization of the TAARs and other olfactory receptors in Atlantic salmon. The WGD would have doubled the number of genes so I expect the TAAR repertoire to be larger in Atlantic salmon than what is observed in teleosts such as zebrafish and medaka. Furthermore, the WGD would result in these duplicate genes being present on

homeologous chromosomes. However, there is substantial evidence indicating the salmon genome is undergoing rediploidization. The fusion and translocations that facilitate this process would likely play a large role in the genomic distribution of TAARs in Atlantic salmon.

Chapter 2.

Materials and Methods

2.1. Identification and mapping of the Atlantic salmon TAAR genes

Atlantic salmon TAAR genes and pseudogenes were identified in contig sequences from the first assembly of the Atlantic salmon genome (http://www.ncbi.nlm.nih.gov/assembly/313068/). I used the ``gene comparison`` function on ASalBase (www.asalbase.org) to generate a list of contigs that potentially contain TAAR genes. This function has BLASTx (Altschul *et al*. 1990) results of all the salmon contig sequences against a specific organism from an Ensemble database. Within that species, results for a specific protein of interest can be filtered out. In this case I used *Danio rerio* since it is arguably the closest related species with a well assembled and annotated genome. I used TAAR1a (ENSDARP00000012766) to filter the results. I then performed BLASTx for each contig identified against the entire NCBI protein database for all species to confirm that the top hit was a TAAR and not a closely related receptor. Distinguishing putatively functional genes and pseudogenes were straightforward as the TAARs are encoded by a single exon. I classified TAARs as putatively functional if (1) there was high sequence similarity to a known TAAR, (2) there was an open reading frame (ORF) of appropriate length (approximately 1 kb) in the correct reading frame and position as identified through BLASTx, and (3) the TAAR fingerprint NSXXNPXX[YH]XXX[YF]XWF was observed (Lindemann *et al*., 2005) (Table A3). I termed a TAAR a putative pseudogene if there was high sequence similarity to a known

TAAR, but there was no obvious ORF or if there was a disruption in the coding sequence that was predicted to produce a truncated or abnormal product (Table A4). Only in one case did I identify two TAARs within a single contig: a pseudogene and a functional TAAR in contig AGKD01084249, and these were distinguished by a/b suffixes. Although a simple BLASTn would likely have retrieved the same results, I decided to use BLASTx since the protein sequences should theoretically be more similar than nucleotide sequences given codon redundancy.

The TAAR genes and pseudogenes were either directly linked to the genetic map through markers within the contig, or indirectly through BLASTn of the contigs against BAC end sequences (www.asalbase.org) or scaffolds from the Centre for Integrative Genetics (CIGENE). A cutoff of 98% identify over a minimum of 300 bp was applied. Matches to a BAC end allowed the contig/TAAR to be assigned to a fingerprint scaffold (fps) on the physical map (Ng *et al*., 2005), which may or may not have been already linked with the genetic map (Danzmann *et al*., 2008) and chromosomes (Phillips *et al*. 2009) through a genetic marker (supporting information, Table A1). The CIGENE scaffolds represent a second assembly of the Atlantic salmon genome, which have been verified and genetically mapped through SNP data (Lien *et al*. in preparation). Lastly, if either of these methods failed I mapped microsatellite markers derived from the contig sequences containing TAAR genes or pseudogenes using the Br5 or Br6 Atlantic salmon mapping families (Danzmann *et al*., 2008) (Table 1, Table A1, Table A6, and Figure A1). PCR for microsatellites was carried out in a 6 µl reaction using 200 nM forward primer, 400 nM reverse primer, 200 uM dNTPs, 400 nM FAM-M13 primer, 1X Roche Taq Buffer, 0.289 Units Roche Taq, and 10 ng of template DNA. The PCR profile

for microsatellites consisted of 94°C for 2 minutes followed by 14 cycles of 94°C for 30 seconds, 62°C for 30 seconds, and 72°C for 30 seconds where the annealing temperature was decreased by 0.5°C (62°C to 55°C). Then another 18 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds was done, followed by 10 minutes at 72°C.

Table 1: Microsatellite markers used to map the Atlantic salmon contigs containing TAAR genes in the Br5 and Br6 mapping families. An M13 tag has been added to the forward primers.

Marker	**Source**	**Primer Sequence**
Ssa0008SCSFU	AGKD01004637	F: TGTAAAACGACGGCCAGTCATAGCACTCAACCAAGTATAGG R: TGCTGATGTAATATGCTCTTTGG
Ssa0014SCSFU	AGKD01073835	F: TGTAAAACGACGGCCAGTCGTTTGTTAATTTCTGTGCTCAC R: CCTTAAGTGGCAAACAGATACTC
Ssa0017SCSFU	AGKD01014617	F: TGTAAAACGACGGCCAGTCCTGTCACACATTATTTCATGC R: AATGGTGAGCAAACTGTAAATT
Ssa0019SCSFU	AGKD01073505	F: TGTAAAACGACGGCCAGTACTCAATGAACACACCTCCC R: CTGCCAAACACTTCCTTCTTAC

2.2. Identification of TAARs in other vertebrates

Using recursive search techniques of NCBI and ENSEMBL databases, I was able to compile the annotated TAAR repertoire from fourteen other vertebrate species including lamprey (*Petromyzon marinus*), elephant shark (*Callorhinchus milii*), coelacanth (*Latimeria chalumnae*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), zebrafish (*Danio rerio*), frog (*Xenopus tropicalis*), alligator (*Alligator mississippiensis*), lizard (*Anolis carolinensis*), chicken (*Gallus gallus*), mouse (*Mus musculus*) and human (*Homo sapiens*). The gene count may not represent the full TAAR repertoire for these species as I only included full-length TAAR sequences in my analyses. Gene names, genomic positions and accession numbers are available in the supplemental material (Table A2).

2.3. Alignment and Phylogenetic Analysis TAARs

All alignments and phylogenetic analyses were carried out using Geneious 6.0.6 (Biomatters) and plugins therein. Amino acid alignments were carried out using the MUSCLE plugin (Edgar, 2004). The Bayesian trees were constructed using the JTT + I + Γ model for amino acid substitution predicted by TOPALi v2. In Figure 7, human rhodopsin was used as an out-group (NM_001131055.1), and also included as additional out-groups are aminergic, serotonin, dopamine, and histamine H2 receptors from chicken, human and zebrafish (Table A2). Following the alignment of the vertebrate TAARs, I manually removed long overhangs introduced by the out-groups and used this trimmed alignment for the phylogenetic analysis. For phylogenetic analysis of Atlantic

salmon TAARs (Figure 8), no trimming of the amino acid alignment was done and zebrafish histamine receptor H2 (NM_001045338.1) was used as an out-group.

2.4. Secondary Structure and Ligand-Binding Prediction

The secondary structures of the Atlantic salmon TAARs were estimated using TMHMM2.0 (Krogh *et al*. 2001). The program uses a hidden Markov model to predict transmembrane domains based on amino acid properties (ie. hydrophobicity). I performed this analysis on each TAAR individually as well as the consensus sequence, but I only annotated the consensus sequence (Figure 9, 10). I used RaptorX-binding, a workflow of RaptorX structure prediction and DeepAlign (Källberg *et al*., 2012; Wang *et al*., 2013). This is a model-assisted program that uses current structure data in protein databases to predict amino acids in the ligand-binding domain. These predicted residues are annotated in both Figures 9 and 10.

2.5. RNA Isolation and cDNA synthesis

For information regarding sample collection, please refer to Johnstone *et al.* (2011). In cases where there was insufficient RNA remaining from Johnstone *et al.* (2011) to synthesize cDNA, RNA was first extracted from the remaining olfactory rosette using TRIzol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's procedures. Residual genomic DNA was then degraded by digesting the total RNA with DNase I (QIAGEN RNAse DNase I set) following Appendix C of the RNeasy® MinElute® Cleanup Kit Handbook (QIAGEN, Mississauga, ON). The DNase I treated total RNA was then cleaned using the RNeasy® MinElute® Cleanup Kit. The RNA quality was first

analyzed by using a ND-1000 Nanodrop spectrophotometer (Thermo Fisher Scientific, Ottawa, ON). The RNA was then recovered and electrophoresed on a 1% agarose gel stained with ethidium bromide. Only RNA for samples which had an $A_{260/280}$ > 2 and produced sharp ribosomal RNA bands on an agarose gel was used to make cDNA. First-strand cDNA synthesis was performed in a 20 ul reaction using 1 ug of purified total RNA with oligo$(dT)_{20}$ primers and Superscript III RT (Invitrogen) following the manufacturer's protocols. The samples were then diluted to a 5 ng/ul working stock and stored at -20°C.

2.6. Quantitative Real-Time PCR (qPCR) of TAAR genes in wild Atlantic salmon at different life stages from three distinct populations

In order to study the expression profile of the TAAR genes in Atlantic salmon, I performed qPCR using cDNA from the olfactory rosettes of wild Atlantic salmon collected at three different life stages (parr, smolts, and adults) from three different sites including both anadromous (Campbellton River and Exploits River) and non-anadromous (Seal Cove Brook) populations (Table 2). Gene specific primers were manually designed and melting temperatures and secondary structures analyzed using OligoAnalyzer 3.1 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) (Table 3). The specificity of primers was determined by melting curve and Sanger sequencing of the amplicons from PCR using genomic DNA template. The qPCR was carried out in 10 µl reactions using 2X PerfeCTa™SYBR® Green SuperMix, ROX (Quanta, Gaithersburg, MD), 200 nM forward and reverse primers, and 5 ng of salmon cDNA. Samples were run in triplicates on a 384-well plate with an Applied Biosystems (ABI) 7900HT qPCR machine. The qPCR profile consisted of 95°C for 3 minutes followed by

40 cycles of 95°C for 15 seconds, 61°C for 30 seconds, and 73°C for 30 seconds. Control gene primer pairs ($EF1A_{A3to4}$ and $EF1A_{A5to6}$) were also run for each individual in triplicate. A no template control for each primer pair, as well as an inter-run calibrator (IRC), was also tested in triplicate. Primer efficiencies were determined by performing qPCR with a five point serial dilution of pooled cDNA tested in triplicate.

Table 2: Summary of samples used for qPCR (Johnstone *et al*., 2011).

Population	**Life Stage**	**Number of Individuals**	**Life History Strategy**
Exploits River	Parr	9	Anadromous
	Smolts	10	
	Adults	10	
Campbellton River	Parr	14	Anadromous
	Smolts	17	
	Adults	6	
Seal Cove Brook	Juveniles (Class II)	10	Non-anadromous
	Adult (Class III)	8	

Table 3: Primer sequences for genes used in qPCR analysis

Gene	Primer Sequence
AGKD01103530	F: CACTGAATGTCATGTTGTGCAC R: CCCTCACAGGCAACATTTTTG
AGKD01044192	F: GAATTAAAATAACTGTTCAGGTTGTTC R: TAGGTAGATGCTAAGCATCCCG
$EF1A_{A3to4}$	F: CCTGTGGAAGTTTGAGACTGG R: GAGTCTGCCCGTTCTTTGAG
$EF1A_{A5to6}$	F: CCCCTCCAGGACGTTTACAAA R: CACACGGCCCACAGGTACA

2.7. Quantitative Real-Time PCR (qPCR) Data Collection

The TAAR gene expression levels were analyzed using the 7900HT Applied Biosystems Sequence Detection System (SDS) Version 2.3 and RQ Manager Version 1.2. If a technical replicate had a standard deviation (SD) greater than 0.7, which roughly corresponds to a cycle threshold (C_T) difference of 1, the outlier was removed. If the resulting SD was still greater than 0.7, that individual was not included in further analyses.

Relative quantification of gene expression was determined using the Comparative C_T method ($\Delta\Delta$CT) (Livak and Schmittgen, 2001; Pfaffl, 2001). The relative quantification (RQ) of a target gene is based on the difference in C_T (cycle threshold) between the target gene and a control gene, and assumes identical primer efficiencies between the target and control genes (see equation 1; Livak and Schmittgen, 2001) (RQ, relative quantification; CT, cycle threshold)

$$[1]\ RQ = 2^{-\Delta\Delta CT}$$

However, if the primer efficiencies between the target gene and the control gene are different, this must be taken into account. In this case, the expression of target gene is normalized to the geometric mean of the C_T from the two $EF1A_A$ primers. The IRC is used in order to make direct comparisons between individuals within a population that were run on different plates or on different days (see equation 2; Pfaffl, 2001) (RQ, relative quantification; CT, cycle threshold; E, primer efficiency).

$$[2]\ RQ = \frac{(E_{Ref})^{CT\ sample}}{(E_{Target})^{CT\ sample}} \div \frac{(E_{Ref})^{CT\ calibrator}}{(E_{Target})^{CT\ calibrator}}$$

Chapter 3.

Results

3.1. Introduction

Genes encoding the MOR, OlfC, and Ora olfactory receptors in Atlantic salmon have previously been identified by the Davidson lab (Johnstone *et al*., 2008, 2009, 2012); however, there has been no attempt to identify genes encoding the TAARs. The goals of this project were to: (1) identify the repertoire of TAAR genes in the Atlantic salmon genome, (2) determine their genomic organization, (3) provide phylogenetic analyses to decipher the evolution of this gene family in vertebrates, (4) predict key binding site residues, and (5) perform qPCR to study the expression of these genes in the olfactory tissue of wild Atlantic salmon.

3.2. The Atlantic Salmon TAAR Repertoire

Contigs from the first assembly of the Atlantic salmon genome (http://www.ncbi.nlm.nih.gov/assembly/313068/) were queried using the ``gene comparison`` function on ASalBase (www.asalbase.org) (see Chapter 2: Materials and Methods for details). I identified 27 putatively functional TAAR genes and 25 putative TAAR pseudogenes in the Atlantic salmon genome (Table A3 and A4). I did not name the TAARs according to the scheme proposed by Lindemann *et al*. (2005), which is based on the homology to human TAAR sequences; rather, the Atlantic salmon TAARs are labeled based on the sequence contig in which they were identified (Lindemann *et al*., 2005). Given the differential gene loss and extensive species-specific expansions in

the TAAR gene family, this naming scheme may not be appropriate. All putatively functional TAAR genes are encoded by a single exon varying in length from 948 bp to 1038 bp and contained the characteristic TAAR motif NSXXNPXX[YH]XXX[YF]XWF (Lindemann *et al*., 2005) (Table S3).

3.3. Genomic Organization of Atlantic Salmon TAARs

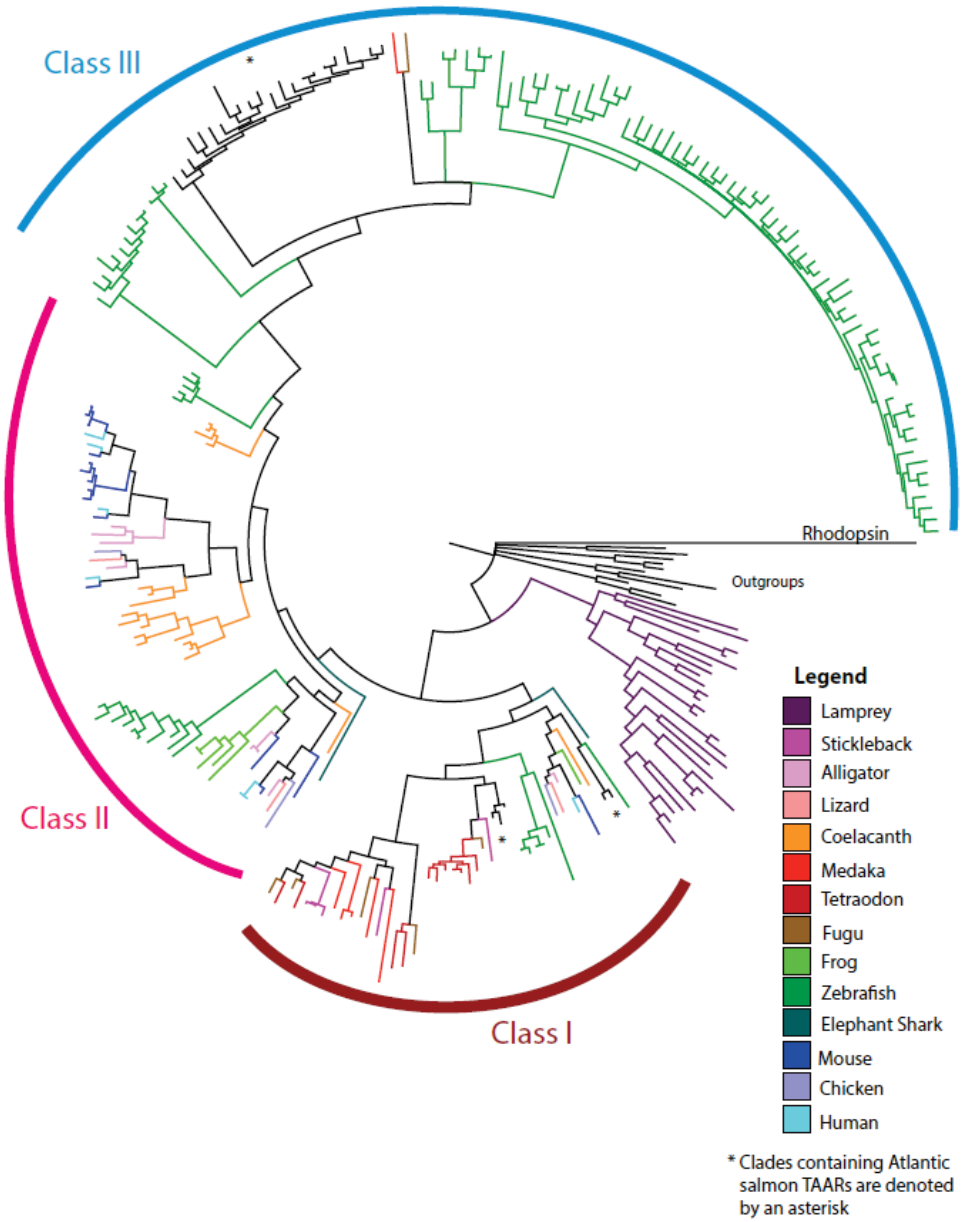
The TAAR genes and pseudogenes were positioned on either the Atlantic salmon physical map (Ng *et al*., 2005), the genetic map (Danzmann *et al*., 2008) or both through multiple independent techniques (see Chapter 2: Materials and Methods for details). It is important to note that the multiple mapping methods utilized yielded concordant results indicating that they are robust (Table A3 and Table A4).

The majority of mapped putatively functional TAAR genes and pseudogenes (20 of 52) are located on chromosome 21 (Ssa21) in adjacent fingerprint scaffolds: fps508 and fps943. A smaller subset (5 of 52) was found to cluster in fps798, which is located on Ssa15. A single TAAR gene or pseudogene was also found on Ssa01, Ssa02, Ssa04, Ssa13, Ssa14, and one of each found on Ssa06. Unfortunately, I was unable to map 20 of 52 TAAR genes or pseudogenes. The TAAR gene family, particularly those clustered on Ssa21, are located in a highly complex and repetitive region of the genome making the design of PCR primers in the flanking regions of di- and tri-nucleotide repeats impossible. In some instances where primer design was possible, I was still unable to map the contigs containing TAAR genes because the microsatellite alleles present in the sire and dam of the mapping families were not informative.

3.4. Evolution of Vertebrate TAARs

In order to examine the evolution of these olfactory receptors, I constructed a Bayesian phylogenetic tree (Biomatters; Huelsenbeck and Ronquist 2001) (Figure 7) from an alignment of 246 full-length TAAR amino acid sequences from fifteen vertebrate species including lamprey (*Petromyzon marinus*), elephant shark (*Callorhinchus milii*), coelacanth (*Latimeria chalumnae*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), zebrafish (*Danio rerio*), frog (*Xenopus tropicalis*), alligator (*Alligator mississippiensis*), lizard (*Anolis carolinensis*), chicken (*Gallus gallus*), human (*Homo sapiens*), mouse (*Mus musculus*) and Atlantic salmon (*Salmo salar*) (Table A2). Human rhodopsin was used as an out-group and the biogenic amine receptors, aminergic, serotonin, dopamine and histamine H2 receptors from chicken, human and zebrafish, were included as additional out-groups.

Figure 7: Bayesian phylogenetic analysis of 246 full-length putatively functional TAAR amino acid sequences from fifteen vertebrate species including lamprey (*Petromyzon marinus*), elephant shark (*Callorhinchus milii*), coelacanth (*Latimeria chalumnae*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), zebrafish (*Danio rerio*), frog (*Xenopus tropicalis*), human (*Homo sapiens*), mouse (*Mus musculus*), lizard (*Anolis carolinensis*), alligator (*Alligator mississippiensis*), chicken (*Gallus gallus*) and Atlantic salmon (*Salmo salar*) (see supporting Table A2 for summary). Rhodopsin from human was used as an out-group. The clades containing Atlantic salmon TAARs are indicated by an asterisk.



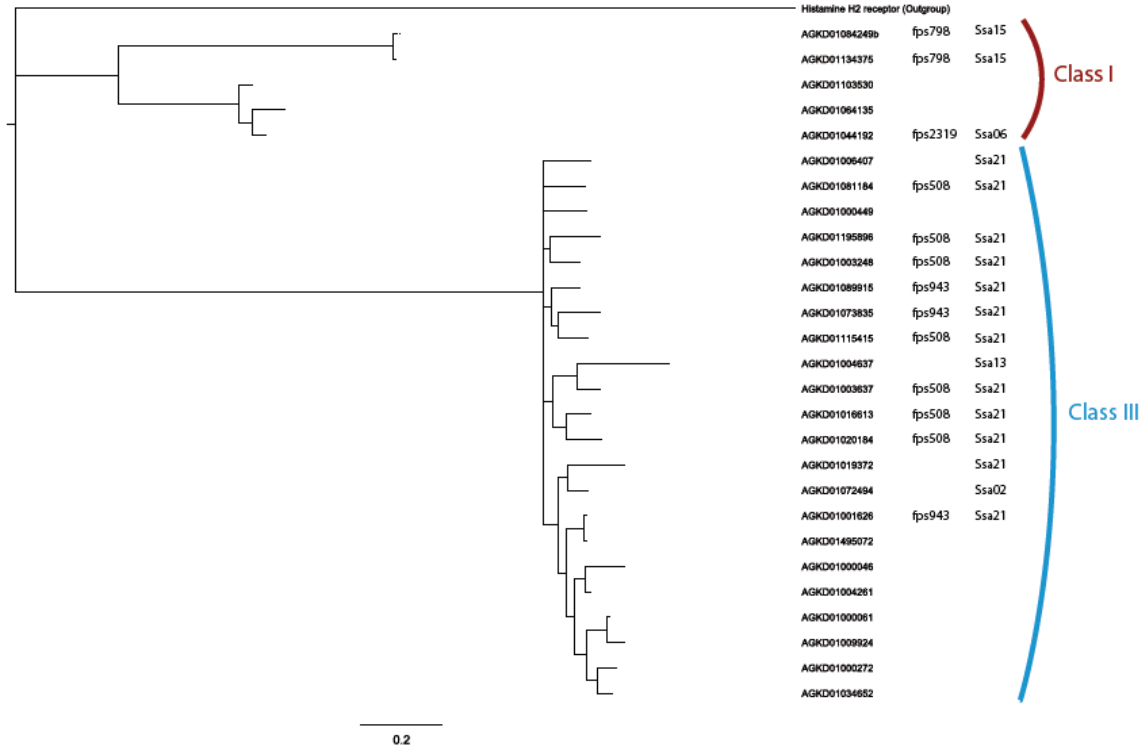
Interestingly, I found the putative lamprey TAARs do not cluster with the TAARs from other species. Instead, they form a species-specific clade that lies between the biogenic amine receptors and the TAARs from every other species examined.

The TAAR gene family can be separated into three main classes (class I, class II and class III) as identified by Hussain *et al*. (2009) (Hussain *et al.*, 2009). The TAAR initially emerged as two classes: class I and class II. Class I TAARs form a discrete clade that contains representatives from each species examined. Class III is comprised of TAARs from teleosts and shares a more recent common ancestor with class II TAARs. Within each class there are distinct species-specific expansions.

3.5. Phylogenetic Analysis of Atlantic Salmon TAARs

I performed Bayesian phylogenetic analysis (Biomatters; Huelsenbeck and Ronquist 2001) on the 27 putatively functional Atlantic salmon TAARs (Figure 8). Using amino acid sequences, I found the TAARs group into three clades. The first two smaller clades correspond to class I TAARs (5 of 27), whereas the third and largest belongs to the class III teleost-specifc TAARs (22 of 27). I also added the physical and genetic mapping data to the tip labels to emphasize that the genes comprising each clade are also physically clustered together in the genome (Figure 8).

Figure 8: Bayesian phylogenetic analysis of 27 full-length putatively functional Atlantic salmon (*Salmo salar*) TAAR amino acid sequences using the histamine receptor H2 from *Danio rerio* as an out-group. Information regarding physical and genetic location has been added beside the contig name when available. The clades correspond to those in Figure 1. Alignment provided in File S3.

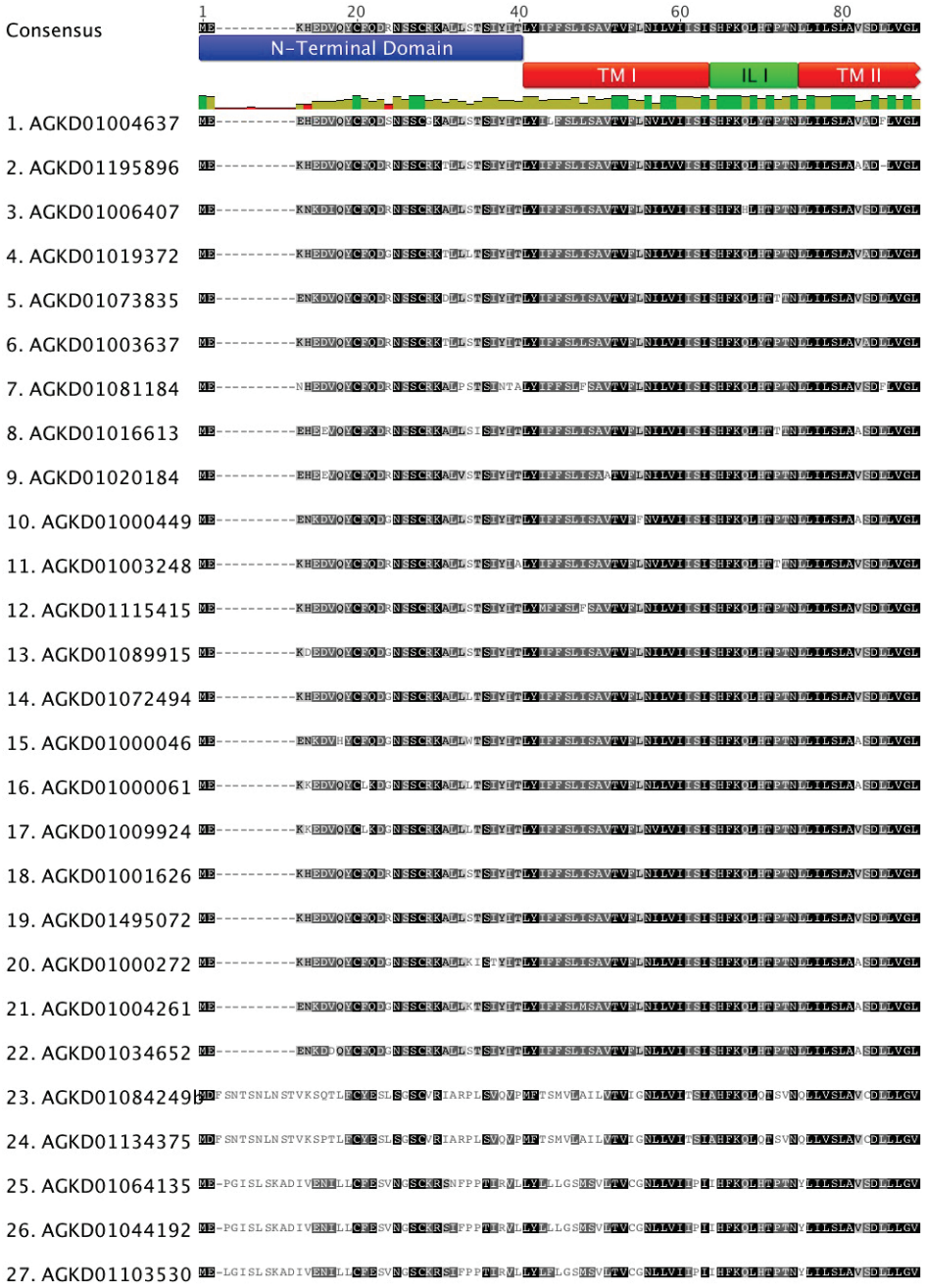


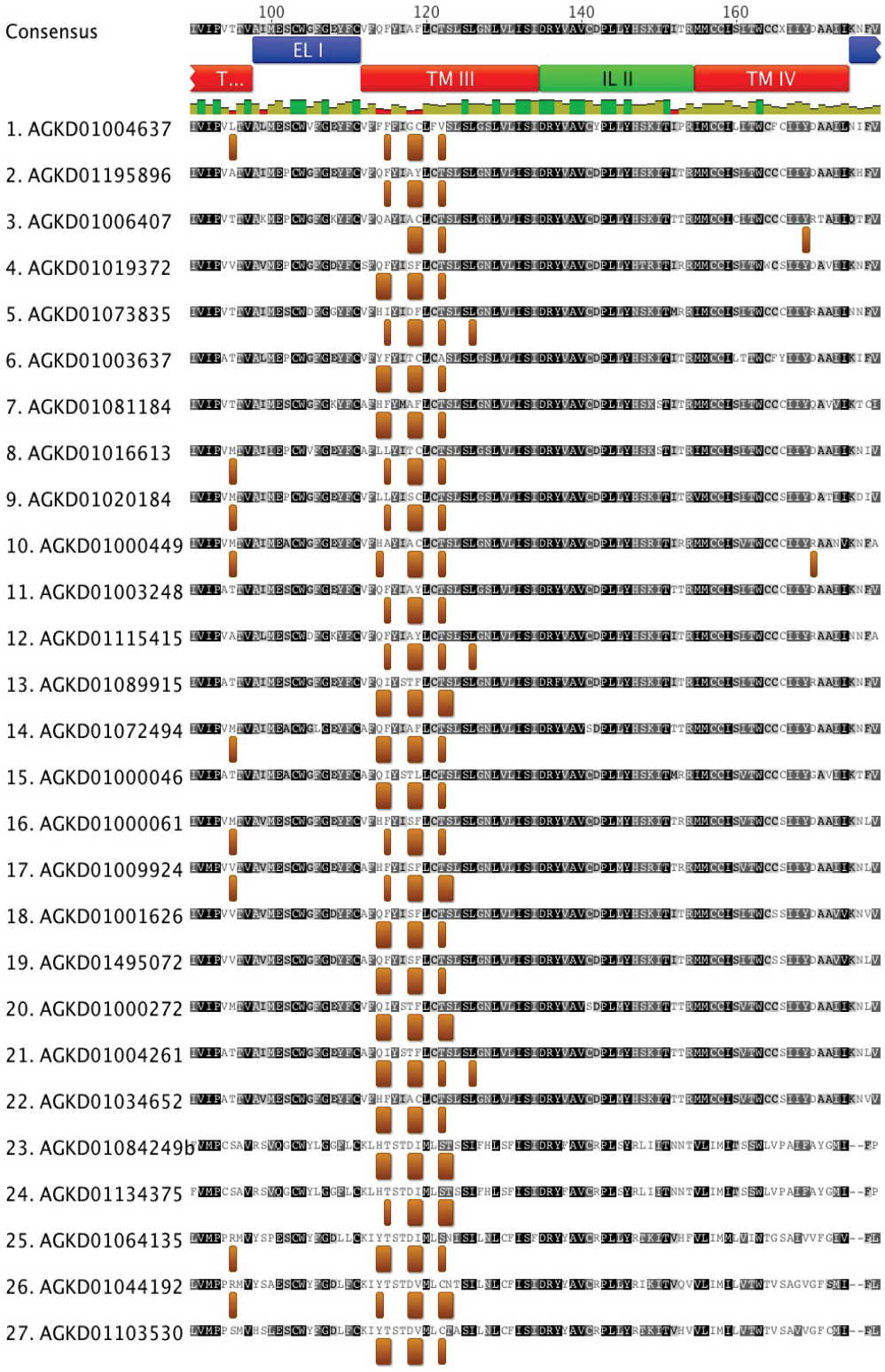
3.6. Predicted Structure and Binding Sites of Atlantic Salmon TAARs

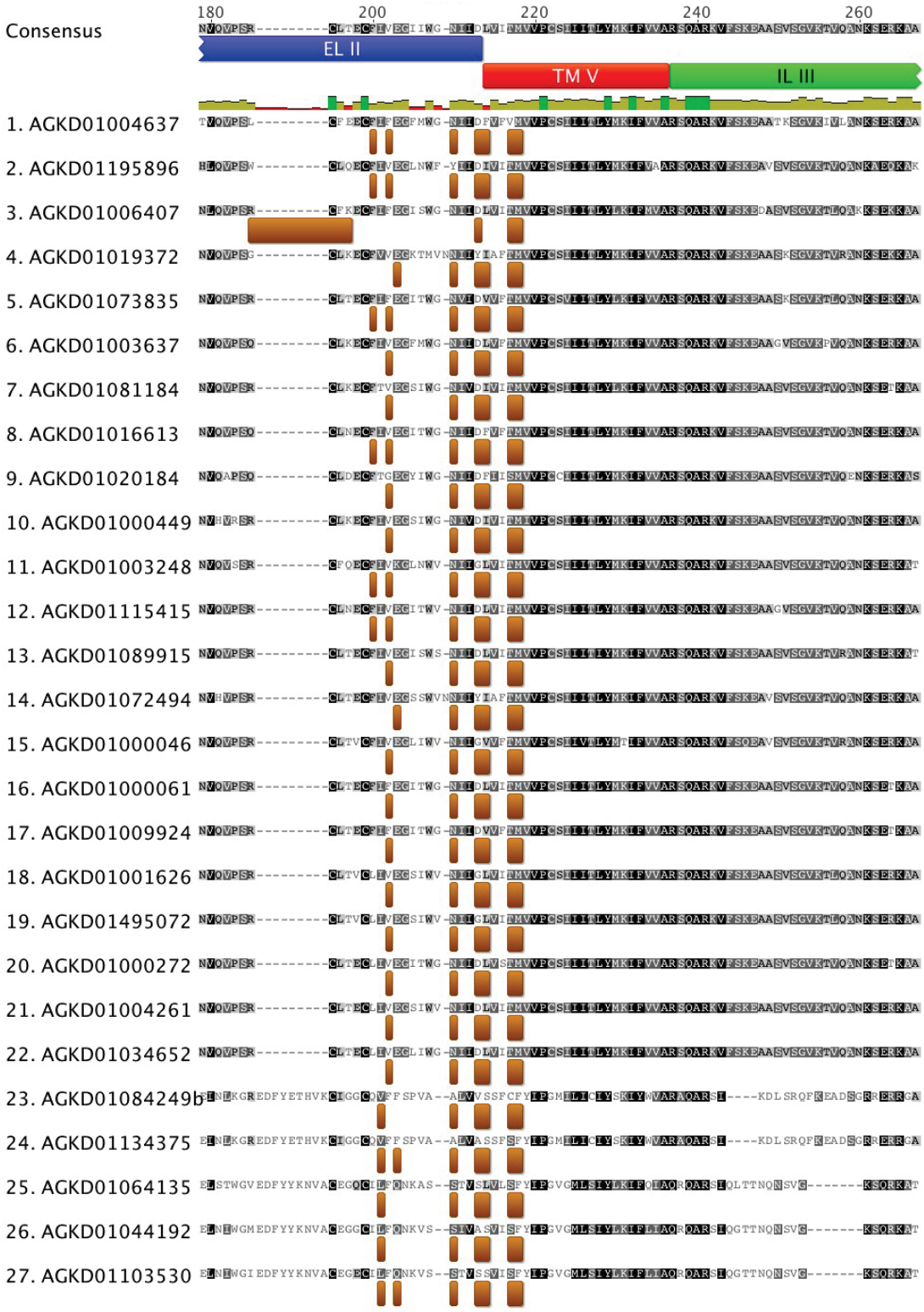
An alignment of the *in silico* translated Atlantic salmon TAARs indicated that on average they have 66.0% pair-wise identity, with 18.9% amino acid residues being invariant among all 27 putatively functional proteins (Figure 9, 10). All putatively functional gene products contain the TAAR fingerprint motif, NSXXNPXX[Y/H]XXX[Y/F]XWF, that was identified by (Lindemann *et al*., 2005) (Figure 9, 10). As the TAARs are members of the GPCR class of proteins, I used TMHMM2.0 to predict the secondary structures of the putatively functional salmon TAARs (Krogh *et al*., 2001). Each amino acid sequence was predicted to contain seven transmembrane (TM) segments with the C-terminal domain being located in the cytosol and the N-terminal domain located extra-cellularly (Figure 9, 10).

I used RaptorX-*binding* to predict which residues contribute to the ligand-binding domain (Källberg *et al*., 2012; Wang *et al*., 2013). The positions of predicted ligand-binding sites were fairly uniform across the TAARs in three of the clades. The predicted binding sites were located in TM III, EL II, TM V, TM VI, and TM VII (Figure 9, 10). The identity of these predicted ligand-binding sites are most conserved within each class; however, some sites are predicted at positions of low amino acid conservation (< 30%) across all three classes.

Figure 9: Amino acid alignment of putatively functional Atlantic salmon TAARs. Secondary structures were predicted using TMHMM2.0. The predicted extracellular segments have been highlighted blue, transmembrane segments red, and intracellular segments green. The TAAR fingerprint motif has been coloured yellow about the alignment. Predicted ligand-binding site residues from RaptorX-*binding* software have been highlighted brown below the alignment. TM, Transmembrane segment; IL, Intracellular Loop; EL, Extracellular Loop.




Consensus
100
120
140
160
EL I
T...
TM III
IL II
TM IV
1. AGKD01004637
2. AGKD01195896
3. AGKD01006407
4. AGKD01019372
5. AGKD01073835
6. AGKD01003637
7. AGKD01081184
8. AGKD01016613
9. AGKD01020184
10. AGKD01000449
11. AGKD01003248
12. AGKD01115415
13. AGKD01089915
14. AGKD01072494
15. AGKD01000046
16. AGKD01000061
17. AGKD01009924
18. AGKD01001626
19. AGKD01495072
20. AGKD01000272
21. AGKD01004261
22. AGKD01034652
23. AGKD01084249b
24. AGKD01134375
25. AGKD01064135
26. AGKD01044192
27. AGKD01103530


Consensus
180
200
220
240
260
EL II
TM V
IL III
1. AGKD01004637
2. AGKD01195896
3. AGKD01006407
4. AGKD01019372
5. AGKD01073835
6. AGKD01003637
7. AGKD01081184
8. AGKD01016613
9. AGKD01020184
10. AGKD01000449
11. AGKD01003248
12. AGKD01115415
13. AGKD01089915
14. AGKD01072494
15. AGKD01000046
16. AGKD01000061
17. AGKD01009924
18. AGKD01001626
19. AGKD01495072
20. AGKD01000272
21. AGKD01004261
22. AGKD01034652
23. AGKD01084249b
24. AGKD01134375
25. AGKD01064135
26. AGKD01044192
27. AGKD01103530

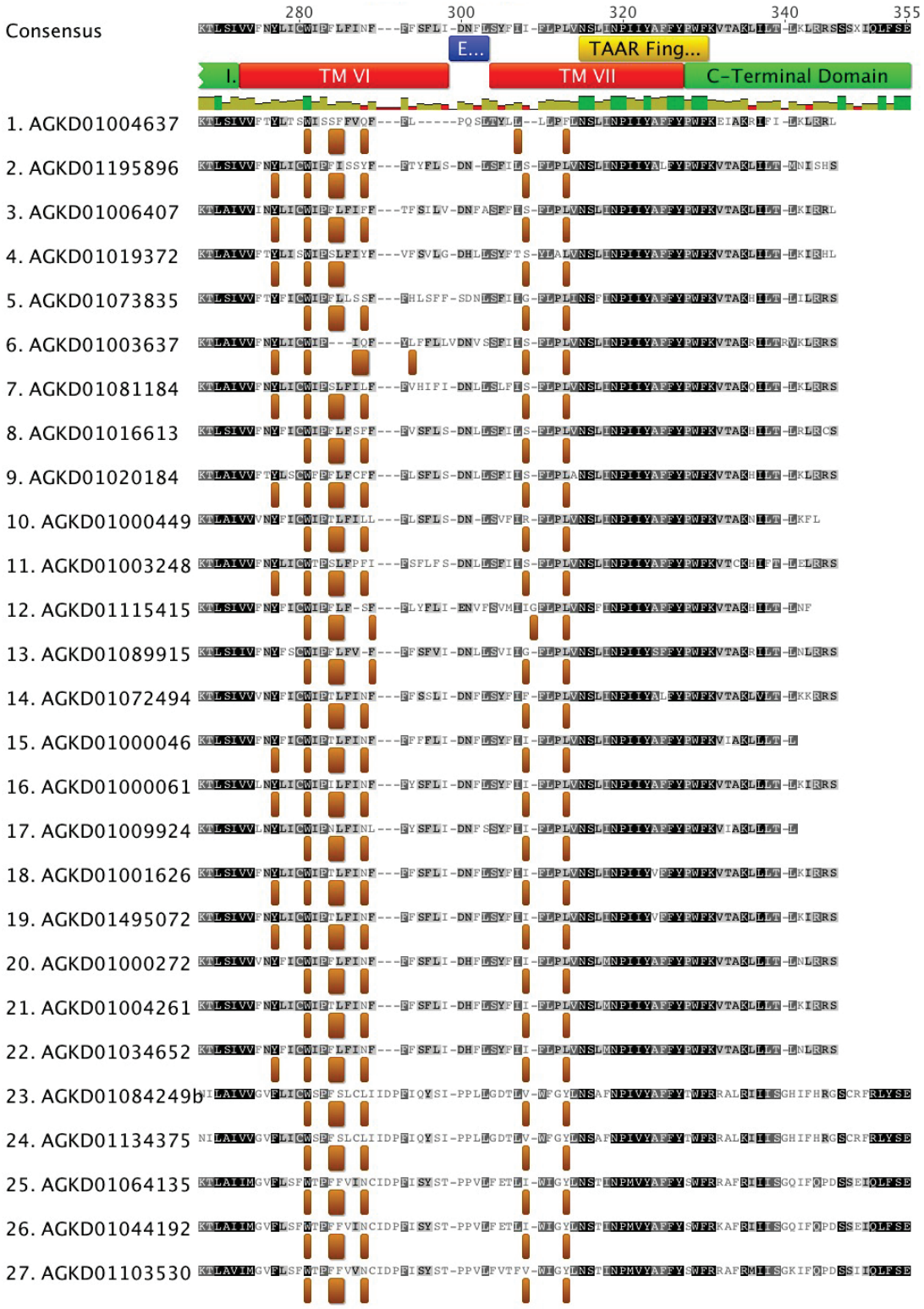
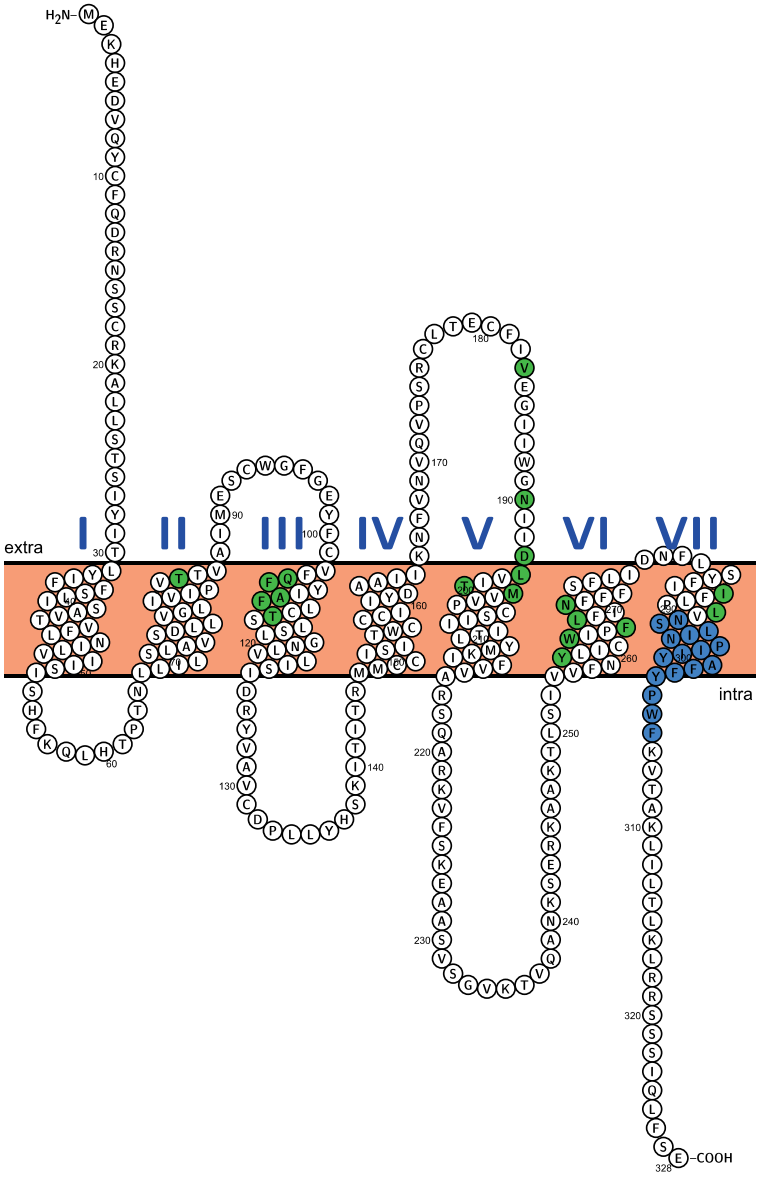

Consensus
280
300
320
340
355
E...
TAAR Fing...
I.
TM VI
TM VII
C-Terminal Domain
1. AGKD01004637
2. AGKD01195896
3. AGKD01006407
4. AGKD01019372
5. AGKD01073835
6. AGKD01003637
7. AGKD01081184
8. AGKD01016613
9. AGKD01020184
10. AGKD01000449
11. AGKD01003248
12. AGKD01115415
13. AGKD01089915
14. AGKD01072494
15. AGKD01000046
16. AGKD01000061
17. AGKD01009924
18. AGKD01001626
19. AGKD01495072
20. AGKD01000272
21. AGKD01004261
22. AGKD01034652
23. AGKD01084249b
24. AGKD01134375
25. AGKD01064135
26. AGKD01044192
27. AGKD01103530

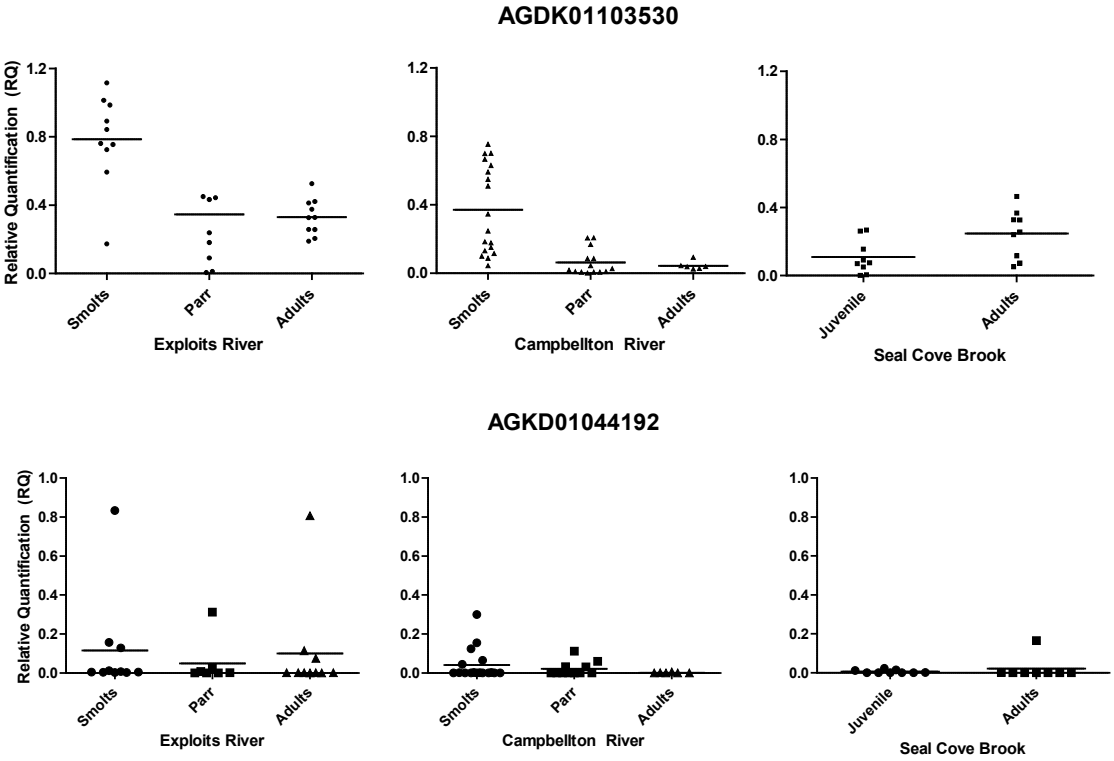
Figure 10: Schematic 2D structure of an Atlantic salmon TAAR based on the consensus sequence from the multiple alignment of putatively functional TAARs shown in Figure 9. The secondary structure was predicted using TMHMM2.0 and ligand-binding site residues predicted by RaptorX-*binding* (also shown in figure 9). The predicted binding sites have been coloured green. The TAAR fingerprint motif has been coloured blue.



3.7. Expression of TAAR genes at different life stages and in different populations of wild Atlantic salmon

I performed qPCR on olfactory tissue from three populations of wild Atlantic salmon populations with individuals at different life stages. Due to the high similarity within the TAAR gene family, only two pairs of gene specific primers were able to be designed. These include the TAAR genes identified in contigs AGKD01103530 and AGKD01044192. For TAAR AGKD01103530, both anadromous populations (Campbellton River and Exploits River) displayed greater overall expression in smolts compared to parr and adults, which showed no significant differences in expression. In the Seal Cove Brook, non-anadromous, population the adults had a slightly more expression than the juveniles, but overall the expression is low compared to anadromous smolts. In the second gene, AGKD0144192, I did not observe any differential expression in any of the populations (Figure 11 and Table A5).

Figure 11: Relative Quantification (RQ) of two TAAR genes in three populations of wild Atlantic salmon at different life stages.



Chapter 4.
Discussion

4.1. The Atlantic Salmon TAAR Repertoire

The TAAR gene count in Atlantic salmon is comparable to the numbers reported for other teleosts except for zebrafish, which possesses by far the largest reported TAAR repertoire (Table 4). The expected the number of TAAR genes in Atlantic salmon was expected to be greater than that of the other teleosts due to the 4R whole genome duplication (WGD), which occurred in the common ancestor of extant salmonids (Ohno 1970; Allendorf and Thorgaard, 1984). Moreover, if TAARs do indeed play a role in homing in salmon, it was anticipated that a large repertoire would be beneficial for discriminating among a wide variety of odorants. Instead, a large number of genes have become pseudogenized, which is the fate for most duplicated genes.

There has been little research into the extent of ligands the TAARs recognize and the downstream signalling mechanisms initiated upon binding. Therefore, it is unclear whether a larger olfactory repertoire directly translates to an increase in sensitivity to specific homing cues or is even necessary for the homing process.

It should be noted that five of the putative pseudogenes may in fact be functional, but have been temporarily classified as such due to the incomplete assembly of the salmon genome (Table A4).

Table 4: Number of putatively functional TAAR genes included in this report by species compared to the previously report number of putatively functional TAAR receptors

			Reported Functional TAARs	
Species	**Number of TAAR genes in analysis (Class I, Class II, Class III)**	**Database**	**2007[1]**	**2009[2]**
Chicken	3 (1, 2, 0)	NCBI	3	3
Human	7 (1, 6, 0)	NCBI	5	6
Mouse	15 (1, 14, 0)	NCBI	15	15
Frog	6 (1, 5, 0)	NCBI	6	3
Lamprey	27 (0)	Libants *et al*. 2009	21	0
Medaka	6 (5, 0, 1)	NCBI	25	25
Stickleback	5 (5, 0, 0)	Ensemble	49	48
Tetraodon	11 (11, 0 , 0)	Ensemble	N/A	18
Fugu	6 (5, 0, 1)	NCBI	13	18
Zebrafish	102 (7, 15, 80)	NCBI	109	112
Coelacanth	18 (1, 17, 0)	NCBI	N/A	N/A
Lizard	3 (1, 2, 0)	NCBI	N/A	N/A
Alligator	8 (1, 7, 0)	NCBI	N/A	N/A
Elephant Shark	2 (1, 1, 0)	NCBI	N/A	2
Rat	N/A	NCBI	N/A	17
Cow	N/A	NCBI	N/A	13
Opposum	N/A	NCBI	22	19
Atlantic salmon	27 (5, 0, 22)	This Thesis	---	---

4.2. Genomic organization of Atlantic Salmon TAARs

In tetrapods, TAARs localize to a single chromosome (chromosome 10 in mouse, chromosome 6 in human, chromosome 3 in chicken, chromosome 2 in opossum, chromosome 1 in rat) (Lindemann *et al*., 2005). The TAAR genes are mainly clustered on two chromosomes in zebrafish (chromosome 10 and chromosome 20) and also in medaka (chromosome 21 and chromosome 24) (Hussain *et al*. 2009). This is not surprising given the teleost-specific WGD (Amores *et al*., 1998; Meyer and Van de Peer 2005). As salmonids have undergone an additional WGD (Allendorf and Thorgaard 1984), one might have expected salmon TAARs to cluster to four distinct genomic locations, but this pattern of organization is not seen. Clusters of Atlantic salmon TAARs are found on Ssa21 and Ssa15, but an additional seven TAARs are distributed on six different chromosomes. The Atlantic salmon genome, like those of other salmonids, has experienced extensive changes since the salmonid-specific 4R WGD as it undergoes rediploidzation from the auto-tetraploid state (Phillips *et al.*, 2009; Phillips *et al*., 2013; Berthelot *et al*., 2014). The complex genetic organization of Atlantic salmon TAARs is likely a reflection of this ongoing genomic reorganization.

4.3. Evolution of Vertebrate TAARs

I performed phylogenetic analysis which incorporated TAARs from fifteen vertebrate species, including Atlantic salmon, in order to elucidate the evolution of this gene family (Figure 7). I selected species which represent the major classes within the chordate phylum: jawless fish (lamprey), cartilaginous fish (elephant shark), bony fish (medaka, fugu, stickleback, zebrafish, and Atlantic salmon), lobe-finned fish

(coelacanth), amphibians (frog), reptiles (lizard and alligator), birds (chicken) and mammals (mouse and human). This may not represent the complete TAAR repertoire in the species examined as I only included TAARs with full-length amino acid sequences in the analysis (see Chapter 2: Materials and Methods for details). Additional effort involving functional analyses is required to annotate these genes and confirm the presence or absence of introns in this gene family. In this report, all the TAAR genes included in the phylogenetic analysis are predicted to be encoded by a single exon, except for TAAR2 from human and mouse, which have been confirmed to contain a single intron by comparisons of the genomic sequences and their transcripts (Lindemann *et al*. 2005).

The lamprey genes predicted to be TAARs did not cluster with the TAARs from the other species examined. Instead, they formed a distinct clade that lies between the biogenic amine receptors and the TAARs from every other species examined (Figure 1). This finding has two major implications. First, as suggested by Hussain *et al*. (2009), these lamprey genes are in fact not TAARs, but likely some other aminergeric receptor (Hussain *et al.*, 2009). Secondly, the TAARs have likely arisen after the divergence of jawed and jawless vertebrates since every other species across vertebrate evolution was found to possess these genes to some extent. This makes the TAARs much younger from an evolutionary perspective compared to the other vertebrate olfactory receptors, such as the main olfactory receptors (MORs) and the vomeronasal type 1 and 2 receptors (ora and OlfC genes, respectively, in salmon).

My results are in agreement with the finding by Hussain *et al*. (2009) that the TAAR gene family can be represented by three classes: class I, class II, and class III

(Figure 1) (Hussain *et al.,* 2009). Class I forms a discrete clade and contains representatives from each species examined including Atlantic salmon. Class I also contains TAAR1, which is the only TAAR not expressed in olfactory epithelium (Lindemann *et al*., 2005). Therefore, class I TAARs may be the closest representation of the ancestral TAAR gene. Class II contains TAAR genes from all species examined, except for the teleosts: medaka, tetraodon, stickleback, fugu and Atlantic salmon. Interestingly, zebrafish has maintained class II TAARs whereas these have been lost in the other teleosts examined. Lastly, class III consists of TAARs only found in teleosts. They arose from class II TAARs, perhaps as a result of the teleost-specific (3R) WGD (Amores *et al*., 1998; Meyer and Van de Peer, 2005).

Within these classes, we observe species-specific expansions. This is contrary to what is seen with the other olfactory receptor classes in teleosts. For example, the genes encoding OlfCs do not form species-specific groups, but rather a species is likely to possess members from several sub-families in its repertoire with some expansions of a sub-family evident in a particular species (Johnstone *et al*., 2009). This suggests that the MORs, oras and OlfCs may perform general functions required by all teleosts, and therefore representatives of all sub-families within a class have been retained to some extent. In contrast, TAARs may perform specialized functions that may be differentially selected, and therefore have evolved in a species-specific manner. This is particularly clear in zebrafish and Atlantic salmon, both of which have a large expansion of class III TAARs, suggesting these may provide a particularly important role.

4.4. Phylogenetic Analysis of Atlantic salmon TAARs

Bayesian analysis of the Atlantic salmon TAARs revealed the genes comprising each class are also physically clustered in the genome (Figure 8). Although not all of the salmon TAAR genes could be positioned on the genetic map, the trend is still clear with those that have been successfully mapped. This information will be valuable to search for additional TAAR genes when a more complete assembly of the Atlantic salmon genome becomes available (Lien *et al*., in preparation). Furthermore, a large expansion has occurred in the class III (teleost-specific) TAARs relative to the class I TAARs. The same observation can be made for zebrafish. This extreme distribution between the two classes of Atlantic salmon TAARs is contrary to what I would expect from a gene family evolving solely on the basis of WGD. If this were the case, one would observe a more uniform distribution of genes between classes, and within each class genes would exist in pairs on the phylogenetic tree with each gene mapping to homeologous chromosomes. Interestingly, a large number of genes were found on Ssa21, but none on its well established homeologous chromosome, Ssa25 (Lien *et al*., 2011). These results suggest additional mechanisms for gene duplication apart from WGD, such as local tandem duplications, are acting on the TAARs and consequently impacting their genomic distribution and evolution.

4.5. Predicted Structure and Binding Sites of Atlantic Salmon TAARs

Although a few studies have identified possible ligands for various TAARs, no crystal structures are available for any TAAR and little is known about the ligand-binding

pocket (Borowsky *et al*., 2001; Bunzow *et al*., 2001; Lindemann *et al*., 2005; Zhang *et al*., 2013). Secondary structure was predicted using TMHMM2.0 to provide evidence, apart from sequence similarity, that the genes identified as TAARs are in fact GPCRs. As anticipated, each sequence was predicted to have 7 TM segments (Figure 9, 10) (Krogh *et al*., 2001).

A model-assisted program, RaptorX-*binding*, was used to predict putative binding sites in the Atlantic salmon TAARs (Källberg *et al*., 2012; Wang *et al*., 2013). Positions of predicted-binding sites are well conserved across all Atlantic salmon TAARs and were located in TM III, EL II, TM V, TM VI, and TM VII (Figure 9, 10). The identity of predicted ligand-binding sites are most conserved within each class. This suggests that TAARs within a class likely recognize similar molecules. Conversely, TAARs in different classes are likely to respond to different types of ligands. A small number of predicted sites are located in positions of low amino acid conservation (<30%) across all Atlantic salmon TAARs, and these may represent critical ligand binding residues that allow for the discrimination of odorants with similar chemical composition. Additional studies involving techniques such as site-directed mutagenesis will provide further insight into mapping the TAAR ligand-binding domain (Katada *et al*., 2005).

4.6. Expression of TAAR genes at different life stages and in different populations of wild Atlantic salmon

The mRNA expression of the MORs, OlfCs, and oras have previously been studied by Johnstone *et al.* (2011), but similar investigations into the expression of the TAARs have not been attempted (Johnstone *et al.,* 2011). The current Atlantic salmon

microarray (Koop *et al*., 2008) has few annotated ORs and therefore the expression of TAARs in the olfactory rosette of wild Atlantic salmon was measured using qPCR. The data are presented as a relative quantification based on the expression of $EF1A_A$, which has been identified as a control gene for qPCR in salmonids (Olsvik *et al.*, 2005) and previously been tested on these same samples with success (Johnstone *et al*., 2011). Primers encompassing $EF1A_A$ exons 3 to 4 and exons 5 to 6 were used in order to observe any cDNA degradation, but the two primer pairs gave consistent and concordant results indicating the integrity of the samples were good.

In order for homing to occur, juvenile salmon must be imprinted by odorant cues from their natal stream. Since PST has been identified as a critical time period for the imprinting process (Hasler and Scholz, 1983; Dittman *et al*., 1996), I hypothesized that I would observe differential expression of olfactory receptors in anadromous populations whereas in non-anadromous populations olfactory receptors would remain uniformly expressed. Due to the highly similar sequences within the TAAR gene family, gene specific primers could only be designed for two TAAR genes. Differential expression of TAAR AGKD01103530 was observed in anadromous populations with expression being greatest in smolts compared to the parr and adults (Figure 11). In the non-anadromous population, TAAR AGKD01103530 was slightly more expressed in the adults than juveniles but overall the expression level was low compared with smolts from the anadromous populations. The second TAAR, AGKD01044192, had low levels of expression across all life stages and populations (Figure 11). Therefore, if my hypothesis regarding the role of TAARs in homing is correct, these findings of differential expression in anadromous populations may indicate a particular importance of TAAR

AGKD01103530 compared to TAAR AGKD01044192. However, these results are very preliminary and further data regarding the rest of the TAAR repertoire are needed. After all, the odorant cue utilized by salmon during homing is likely a mixture of odorants rather than a single odorant, which would involve multiple olfactory receptors. It may also be useful to investigate changes in gene expression during smoltification rather than just the encompassing life stages. Such work has been done using a microarray, but it did not contain transcripts for ORs (Seear *et al*., 2009).

Chapter 5.

Conclusions and Future Directions

In this thesis I report a preliminary repertoire of the TAAR genes and pseudogenes in Atlantic salmon, which included 27 putatively functional genes (5 class I and 22 class III) and 25 putative pseudogenes. These genes were found to cluster primarily on Ssa21 and to a lesser extent on Ssa15. Additional TAARs were found on six other chromosomes, which is contrary to the single locus observed in tetrapods. This complicated genomic organization is likely a reflection of the Robertsonian translocations and fusions that have occurred in the Atlantic salmon genome as it undergoes rediploidization following the 4R WGD (Phillips *et al.*, 2009; Phillips *et al*., 2013; Berthelot *et al*., 2014). The genomic organization of TAARs is reflected in the phylogenetic analysis suggesting local duplications are also playing a role in the evolution of this gene family. There is conservation of predicted binding residues within each clade, but not between clades indicating each class of Atlantic salmon TAARs recognizes a different type of ligand. The predicted lamprey (jawless vertebrate) TAAR genes formed a monophyletic clade that was separate from those of the other vertebrates. From this phylogenetic analysis I can infer that this gene family arose after the divergence of jawed and jawless vertebrates since all jawed vertebrates contain TAARs. These genes form three classes with species-specific expansions occurring within each class. This pattern is contrary to what is observed with the other olfactory receptor types. These species-specific expansions are most pronounced in zebrafish and Atlantic salmon, whose expansions belong to class III TAARs. Whether this is

indicative of a particular importance of these receptors for the biology of these species or simply reflects that Atlantic salmon and zebrafish are the most closely related (although distantly) among the teleosts I examined, has yet to be determined. Preliminary data on the expression of two TAARs in the olfactory tissue of wild Atlantic salmon revealed that there is differential expression between the life stages of anadromous salmon populations whereas in non-anadromous salmon there is no significant changes in expression between juveniles and adults. However, the expression pattern of the remaining TAAR genes needs to be determined before any real inferences can be made. RNA-seq may be a useful alternative that overcomes the issues of designing gene specific primers and some preliminary results (not shown) appear promising. Another advantage of RNA-seq is that it may expand the current repertoire of identified olfactory receptors or even lead to the discovery of new olfactory receptors unique to Atlantic salmon.

There is increasing evidence that the odorant cues used by salmon during the final stages of homeward migration are derived from the amino acid composition of their natal stream water (Shoji *et al*. 2000, 2003; Yamamoto and Ueda 2009; Yamamoto *et al*. 2010, 2013). Therefore, receptors being capable of recognizing and distinguishing between these types of compounds would be crucial for this process to occur. The identification of the TAAR gene repertoire in Atlantic salmon will facilitate functional studies with these genes, such as the preliminary data presented on the expression at different life-stages in distinct populations (Johnstone *et al*., 2011), but also determining the range of allelic variation in olfactory receptors genes of different salmon populations.

References

Ache BW, and Zhainazarov A. 1995. Dual second-messenger pathways in olfactory transduction. Current Opinion in Neurobiology 5:461-466.

Ache BW, and Young JM. 2005. Olfaction: Diverse Species, Conserved Principles. Neuron 48:417-430.

Alioto TS, and Ngai J. 2006. The repertoire of olfactory C family G protein-coupled receptors in zebrafish: candidate chemosensory receptors for amino acids. BMC Genomics 7:309.

Allendorf FW, and Thorgaard GH. 1984. Tetraploidy and the evolution of Salmonid fishes. In Evolutionary Genetics of Fishes. Ed. Turner, B.J. New York, Plenum Press. pp. 55-93.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, and Postlethwait JH. 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711-4.

Bams RA. 1976. Survival and propensity for homing as affected by presence or absence of locally adapted paternal genes in two transplanted populations of pink salmon (*Oncorhynchus gorbuscha*). Journal of the Fisheries Research Board of Canada 33:2716-2725.

Berg OK. 1985. The formation of non-anadromous populations of Atlantic salmon, *Salmo salar* L., in Europe. Journal of Fish Biology 27:805-815.

Berry MD. 2004. Mammalian central nervous system trace amines. Pharmacologic amphetamines, physiologic neuromodulators. Journal of Neurochemistry 90:257-71.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, Aury JM, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff JN, Genêt C, Wincker P, Jaillon O, Roest Crollius H, and Guiguen Y. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature Communications 5:3657.

Biomatters. Geneious 6.0.6. Available from: http://www.geneious.com/

Boeckh J, Kaissling KE, and Schneider D. 1965. Insect olfactory receptors. Cold Spring Harbor Symposia on Quantitative Biology 30:263–280.

Borowsky B, Adham N, Jones KA, Raddatz R, Artymyshyn R, Ogozalek KL, Durkin MM, Lakhlani PP, Bonini JA, Pathirana S, Boyle N, Pu X, Kouranova E, Lichtblau H, Ochoa FY, Branchek TA, and Gerald C. 2001. Trace amines: identification of a family of mammalian G protein-coupled receptors. Proceedings of the National Academy of Sciences USA 98:8966–8971.

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

Bunzow JR, Sonders MS, Arttamangkul S, Harrison LM, Zhang G, Quigley DI, Darland T, Suchland KL, Pasumamula S, Kennedy JL, Olson SB, Magensis RE, Amara SG, and Grandy DK. 2001. Amphetamine, 3,4-methylenedioxymethamphetamine, lysergic acid diethylamide, and metabolites of the catecholamine neurotransmitters are agonists of a rat trace amine receptor. Molecular Pharmacology 60:1181–1188.

Chess A, Simon I, Cedar H, and Axel R. 1994. Allelic inactivation regulates olfactory receptor gene expression. Cell 78:823–834.

Couturier CY, Clarke L, and Sutterlin AM. 1986. Identification of spawning areas of two forms of Atlantic salmon (*Salmo salar* L.) inhabiting the same watershed. Fisheries Research 4:131-144.

Dadswell MJ, Spares AD, Reader JM, and Stokesbury MJ. 2010. The North Atlantic subpolar gyre and the marine migration of Atlantic salmon *Salmo salar*: the "Merry-Go-Round" hypothesis. Journal of Fish Biology 77:435-67.

Davidson WS, Koop BF, Jones SJM, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, and Omholt SW. 2010. Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome Biology 11:403.

Danzmann RG, Cairney M, Davidson WS, Ferguson MM, Gharbi K, Guyomard R, Leder E, Okamoto N, Ozaki A, Rexroad CE, Sakamoto T, Taggart JB, and Woram RA. 2005. A comparative analysis of the rainbow trout genome with two other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). Genome 48:1037-1051.

Danzmann RG, Davidson EA, Ferguson MM, Gharbi K, Koop BF, Hoyheim B, Lien S, Lubieniecki KP, Moghadam HK, Park J, Phillips RB, and Davidson WS. 2008. Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon). BMC Genomics 9:557

Dittman AH, Quinn TP, and Nevitt GA. 2006. Timing of imprinting to natural and artificial odors by coho salmon (Oncorhynchus kisutch) Can. Journal of Fisheries and Aquatic Sciences 53:434-442.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Fleming, IA. 1996. Reproductive strategies of Atlantic salmon: ecology and evolution. Reviews in Fish Biology and Fisheries 6: 379-416.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, and Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-45.

Gilad Y, Przeworski M, and Lancet D. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. PLoS Biology 2:E5.

Grant, SW (editor). 1997. Genetic effects of straying of non-native fish hatchery fish into natural populations: proceedings of the workshop. U.S. Dep. Commer., NOAA Tech Memo. NMFS-NWFSC-30, p130.

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, and Bennett MD. 2007. Eukaryotic genome size databases. Nucleic Acids Research 35:D332-8.

Grus WE, and Zhang J. 2009. Origin of the genetic components of the vomeronasal system in the common ancestor of all extant vertebrates. Molecular Biology and Evolution 26:407–419.

Hamdani EH, Alexander G, and Døving KB. 2001. Projection of sensory neurons with microvilli to the lateral olfactory tract indicates their participation in feeding behaviour in crucian carp. Chemical Senses 26:1139-1144.

Hamdani EH, and Døving KB. 2002. The alarm reaction in crucian carp is mediated by olfactory neurons with long dendrites. Chemical Senses 27:395-398.

Hamdani EH, and Døving KB. 2006. Specific projection of the sensory crypt cells in the olfactory system in crucian carp, *Carassius carassius*. Chemical Senses 31:63-67.

Hamdani EH, and Døving KB. 2007. The functional organization of the fish olfactory system. Progress in Neurobiology 82:80-86.

Hansen LP, and Quinn TP. 1998. The marine phase of the Atlantic salmon (*Salmo salar*) life cycle, with comparisons to Pacific salmon. Canadian Journal of Fisheries and Aquatic Sciences 55:104-118.

Hara TJ, Ueda K, and Gorbman A. 1965. Electroencephalographic studies of homing salmon. Science 149:884-85.

Hasler AD, and Wisby WJ. 1951. Discrimination of Stream Odors by Fishes and Its Relation to Parent Stream Behavior. American Society of Naturalists 85:223.

Hasler AD, and Scholz AT. 1983. Olfactory Imprinting and Homing in Salmon. Berlin, New York: Springer-Verlag.

Huelsenbeck JP, and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

Hussain A, Saraiva LR, and Korsching SI. 2009. Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts. Proceedings of the National Academy of Sciences USA 106:4313–4318.

Ishiguro, NB, Miya M, and Nishida M. 2003. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". Molecular Phylogenetics and Evolution 27:476-488.

Ishii T, Serizawa S, Kohda A, Nakatani H, Shiroishi T, Okumura K, Iwakura Y, Nagawa F, Tsuboi A, and Sakano H. 2001 Monoallelic expression of the odourant receptor gene and axonal projection of olfactory sensory neurons. Genes to Cells 6:71–78.

Johnson BA, and Leon M. 2007. Chemotopic Odorant Coding in a Mammalian Olfactory System. Journal of Comparative Neurology 503:1-34.

Johnstone KA, Ciborowski KL, Lubieniecki KP, Chow W, Phillips RB, Koop BF, Jordan WC, and Davidson WS. 2009. Genomic organization and evolution of the vomeronasal type 2 receptor-like (OlfC) gene clusters in Atlantic salmon, *Salmo salar*. Molecular Biology and Evolution 26:1117–1125.

Johnstone KA, Lubieniecki KP, Chow W, Phillips RB, Koop BF, and Davidson WS. 2008. Genomic organization and characterization of two vomeronasal 1 receptor-like genes (ora1 and ora2) in Atlantic salmon *Salmo salar*. Marine Genomics 1:23–31.

Johnstone KA, Lubieniecki KP, Koop BF, and Davidson WS. 2011. Expression of olfactory receptors in different life stages and life histories of wild Atlantic salmon (*Salmo salar*). Molecular Ecology 20:4059–4069.

Johnstone KA, Lubieniecki KP, Koop BF, and Davidson WS. 2012. Identification of olfactory receptor genes in Atlantic salmon *Salmo salar*. J. Fish Biol. 81:559–575.

Källberg M, Wang H, Wang S, Jian P, Wang Z, Lu H, and Xu J. 2012. Template-based protein structure modeling using the RaptorX web server. Nature Protocols. 7:1511–1522.

Kasumyan AO, and Døving KB. 2003. Taste preferences in fish. Fish and Fisheries 4:289-347.

Katada S, Hirokawa T, Oka Y, Suwa M, and Touhara K. 2005. Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site. Journal of Neuroscience 25: 1806–1815.

Klemetsen A, Amundsen PA, Dempson JB, Jonsson B, Jonsson N, O'Connell MF, and Mortensen E. 2003. Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12:1-59.

Koop BF, von Schalburg KR, Leong J, Walker N, Lieiph R, Cooper GA, Robb A, Beetz-Sargent M, Holt RA, Moore R, Brahmbhatt S, Rosner J, Rexroad III CE, McGowan CR, and Davidson WS. 2008. A salmonid EST genomic study: Genes, duplications, phylogeny and microarrays. BMC Genomics 9:545.

Krogh A, Larsson B, von Heijne G, and Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305:567–580.

Lagostena L, and Menini A. 2003. Whole-cell recordings and photolysis of caged compounds in olfactory sensory neurons isolated from the mouse. Chemical Senses 28:705-16.

Lewcock JW, and Reed RR. 2004. A feedback mechanism regulates monoallelic odorant receptor expression. Proceedings of the National Academy of Sciences USA 101:1069–1074.

Libants S, Carr K, Wu H, Teeter JH, Chung-Davidson Y-W, Zhang Z, Wilkerson C, and Li W. 2009. The sea lamprey *Petromyzon marinus* genome reveals the early origin of several chemosensory receptor families in the vertebrate lineage. BMC Evolutionary Biology 9:180.

Liberles SD, and Buck LB. 2006. A second class of chemosensory receptors in the olfactory epithelium. Nature 442:645-50.

Lindemann L, Ebeling M, Kratochwil NA, Bunzow JR, Grandy DK, and Hoener MC. 2005. Trace amine-associated receptors form structurally and functionally distinct subfamilies of novel G protein-coupled receptors. Genomics 85:372–385.

Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402-408.

Lukacs M, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Phillips RB, Miller KM, Davidson WS, and Koop BF. 2007. Genomic organization of duplicated Major Histocompatibility Complex class I regions in Atlantic salmon (*Salmo salar*). BMC Genomics 8:251.

Macqueen DJ, and Johnston IA. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proceedings of the Royal Society B. 281: 20132881.

Man O, Gilad Y, and Lancet D. 2004. Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. Protein Science 13:240-54.

Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, Kheradpour P, Mountoufaris G, Carey C, Barnea G, Kellis M, and Lomvardas S. 2011. An epigenetic signature for monoallelic olfactory receptor expression. Cell 145:555-70.

Matsui A, Go Y, and Niimura Y. 2010. Degeneration of olfactory receptor gene repertories in primates: no direct link to full trichromatic vision. Molecular Biology and Evolution 27:1192-200.

McCormick SD, Hansen LP, Quinn TP, and Saunders RL. 1998. Movement, migration, and smolting of Atlantic salmon (*Salmo salar*). Canadian Journal of Fisheries and Aquatic Sciences 55:77-92.

Meyer A, and Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27:937-945.

Moghadam, H.K., M.M. Ferguson and R.G. Danzmann. 2005. Evidence for Hox gene duplication in rainbow trout (*Oncorhynchus mykiss*): A tetraploid model species. Journal of Molecular Evolution 61:804-818.

Nei M, Niimura Y, and Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: Roles of chance and necessity. Nature Reviews Genetics 9:951-963.

Nelson J: Fishes of the world New York, Wiley; 2006.

Ng SH, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJ, de Jong PJ, Koop BF, Krzywinski MI, Lubieniecki K, Marra MA, Mitchell LA, Mathewson C, Osoegawa K, Parisotto SE, Phillips RB, Rise ML, von Schalburg KR, Schein JE, Shin H, Siddiqui A, Thorsen J, Wye N, Yang G, and Zhu B. 2005. A physical map of the genome of Atlantic salmon, *Salmo salar*. Genomics 86:396–404.

Niimura Y, and Nei M. 2007. Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS ONE 2:e708.

Niimura Y. 2012. Olfactory Receptor Multigene Family in Vertebrates: From the Viewpoint of Evolutionary Genomics. Current Genetics 13:103-114.

Nordeng H. 1971. Is the local orientation of anadromous fishes determined by pheromones? Nature 233:411–413.

Nordeng H. 1977. A Pheromone Hypothesis for Homeward Migration in Anadromous Salmonids. Oikos 28:155–159.

Ohno, S. 1970. Enormous diversity in genome sizes of fish as a reflection of nature's extensive experiments with gene duplication. Transactions of the American Fisheries Society 99:120-130.

Ohno S. 1970. Evolution by Gene Duplication. New York: Springer-Verlag.

Olsvik PA, Lie KK, Jordal AE, Nilsen TO, and Hordvik I. 2005. Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. BMC Molecular Biology 6:21.

Pennanen, JO. (n.d.). Atlantic salmon life cycle. Retrieved from: http://asf.ca/life-cycle.html

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29:e45.

Phillips RB, Keatley KA, Morasch MR, Ventura AB, Lubieniecki KP, Koop BF, Danzmann RG, and Davidson WS. 2009. Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). BMC Genetics. 10:46.

Phillips RB, Park LK, and Naish KA. 2013. Assignment of Chinook salmon (*Oncorhynchus tshawytscha*) linkage groups to specific chromosomes reveals a karyotype with multiple rearrangements of the chromosome arms of rainbow trout (*Oncorhynchus mykiss*). G3(Bethesda) 3:2289-95.

Pifferi S, Menini A, and Kurahashi T. 2010. Signal Transduction in Vertebrate Olfactory Cilia. In: Menini A, editor. The Neurobiology of Olfaction. Boca Raton (FL): CRC Press; Chapter 8. Available from: http://www.ncbi.nlm.nih.gov/books/NBK55986/

Power G. 1958. The evolution of the freshwater races of the Atlantic salmon (*Salmo salar* L.) in eastern North America. Arctic 11:86-92.

Putman NF, Lohmann KJ, Putman EM, Quinn TP, Klimley AP, and Noakes DLG. 2013. Evidence for geomagnetic imprinting as a homing mechanism in Pacific salmon. Current Biology 23:312–316.

Ressler KJ, Sullivan SL, and Buck LB. 1994. Information coding in the olfactory system: Evidence for a stereotyped and highly organized epitope map in the olfactory bulb. Cell 79:1245–55.

Rovati GE1, Capra V, and Neubig RR. 2007. The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state. Molecular Pharmacology 1:959-64.

Sandoz JC, Deisig N, Maria Gabriela de Brito Sanchez, and Giurfa M. 2007. Understanding the Logics of Pheromone Processing in the Honeybee Brain: From Labeled-Lines to Across-Fiber Patterns. Frontiers in Behavioral Neuroscience 1: 5.

Saraiva LR, and Korsching SI. 2007. A novel olfactory receptor gene family in teleost fish. Genome Res. 17:1448–1457.

Serizawa S, Miyamichi K, Nakatani H, Suzuki M, Saito M, Yoshihara Y, and Sakano H. 2003. Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. Science 302:2088-2094.

Scholz AT, Horrall RM, Cooper JC, and Hasler AD. 1976. Imprinting to chemical cues: The basis for home stream selection in salmon. Science 192:1247-1249.

Shoji T, Ueda H, Ohgami T, Sakamoto T, Katsuragi Y, Yamauchi K, and Kurihara K. 2000. Amino acids dissolved in stream water as possible home stream odorants for masu salmon. Chemical Senses 25:533–540.

Shoji T, Yamamoto Y, Nishikawa D, Kurihara K, and Ueda H. 2003. Amino acids in stream water are essential for salmon homing migration. Fish Physiology and Biochemistry 28:249–251.

Ueda K, Hara TJ, and Gorbman A. 1967. Electroencephalographic studies on olfactory discrimination in adult spawning salmon. Comparative Biochemistry and Physiology 21:133-43.

Van de Peer Y, Maere S, and Meyer A. 2009. The evolutionary significance of ancient genome duplication. Nature Reviews Genetics 10:725–732.

Vasemägi A, Gross R, Paaver T, Koljonen M-L, and Nilsson J. 2005. Extensive immigration from compensatory hatchery releases into wild Atlantic salmon population in the Baltic sea: spatio-temporal analysis over 18 years. Heredity 95:76-83.

Wang S, Ma J, Peng J, and Xu J. 2013. Protein structure alignment beyond spatial proximity. Scientific Reports 3:1448.

Wisby WJ, and Hasler AD. 1954. The effect of olfactory occlusion on migrating silver salmon (*O. kisutch*). Journal of the Fisheries Research Board of Canada 11:472-78.

Yamamoto Y, Hino H, and Ueda H. 2010. Olfactory imprinting of amino acids in lacustrine sockeye salmon. PLoS One 5:e8633.

Yamamoto Y, and Ueda H. 2009. Behavioral responses by migratory chum salmon to amino acids in natal stream water. Zoological Science 26:778–782.

Yamamoto Y, Shibata H, and Ueda H. 2013. Olfactory Homing of Chum Salmon to Stable Compositions of Amino Acids in Natal Stream Water. Zoological Science 30:607-612.

Yamanoue Y, Miya M, Inoue JG, Matsuura K, and Nishida M. 2006. The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. Genes and Genetic Systems 81:29-39.

Zhainazarov AB, and Ache BW. 1995. Odor-induced currents in Xenopus olfactory receptor cells measured with perforated-patch recording. Jouran of Neurophysiology 74:479-83.

Zhang J, Pacifico R, Cawley D, Feinstein P, and Bozza T. 2013. Ultrasensitive detection of amines by a trace amine-associated receptor. Journal of Neuroscience 33:3228–3239.

Zucchi R, Chiellini G, Scanlan TS, and Grandy DK. 2006. Trace amine-associated receptors and their ligands. British Journal of Pharmacology 149:967–978.

Supplementary Material

Appendix A

Table A1: Summary of markers used for mapping including their name, source and physical/genetic location. The details concerning primers and amplification conditions can be found at www.AsalBase.org.

Marker Name	Marker Source	Physical Location	Genetic Location
Ssa0907BSFU	S0152B01_SP6	fps798	Ssa15
Ssa1303BSFU	S0159E20_SP6	fps508	Ssa21
Ssa0975BSFU	S0085J07_SP6	fps943	Ssa21
Ssa0043BSFU	S0322O19_T7	fps2319	Ssa06
Ssa0008SCSFU	AGKD01004637	unmapped	Ssa13
Ssa0014SCSFU	AGKD01073835	fps943	Ssa21
Ssa0017SCSFU	AGKD01014617	unmapped	Ssa21
Ssa0019SCSFU	AGKD01073505	unmapped	Ssa14

Table A2: Accession numbers and genomic positions for vertebrate genes used in analyses.

Organism	Gene Name	Genomic Position	Accession Number	Source
Chicken	Gg_TAAR1	3:56100482..56108943	XM_004935740.1	NCBI
Gallus gallus	Gg_TAAR2	3:56116829..56117857	XM_003640968.1	NCBI
	Gg_TAAR5	3:56137725..56138753	XM_001231564.2	NCBI
Human	Hs_TAAR6	6:132891461..132892498	NM_175067.1	NCBI
Homo sapiens	Hs_TAAR1	6:132966123..132967142	NM_138327	NCBI
	Hs_TAAR2v1	6:132938289..132945414	NM_001033080.1	NCBI
	Hs_TAAR2v2	6:132938289..132945414	NM_014626.3	NCBI
	Hs_TAAR5	6:132909731..132910877	NM_003967.2	NCBI
	Hs_TAAR8	6:132873832..132874860	NM_053278.1	NCBI
	Hs_TAAR9	6:132859427..132860475	NM_175057.3	NCBI
Lamprey	STIG_44993.TAAR325			Libants *et al.* (2009)
Petromyzon marinus	STIG_38630.TAAR335			Libants *et al.* (2009)
	STIG_28657.TAAR352			Libants *et al.* (2009)

	STIG_22166.TAAR354			Libants *et al.* (2009)
	STIG_17331.TAAR351			Libants *et al.* (2009)
	STIG_16230.TAAR353			Libants *et al.* (2009)
	STIG_15712.TAAR345			Libants *et al.* (2009)
	STIG_15153.TAAR369			Libants *et al.* (2009)
	STIG_15153.TAAR353			Libants *et al.* (2009)
	STIG_14718.TAAR353			Libants *et al.* (2009)
	STIG_14718.TAAR345			Libants *et al.* (2009)
	STIG_14409.TAAR352			Libants *et al.* (2009)
	STIG_12707.TAAR348			Libants *et al.* (2009)
	STIG_9755.TAAR355B			Libants *et al.* (2009)
	STIG_9755.TAAR355A			Libants *et al.* (2009)
	STIG_9665.TAAR351			Libants *et al.* (2009)
	STIG_7446.TAAR346			Libants *et al.* (2009)
	STIG_6636.TAAR354			Libants *et al.* (2009)
	STIG_5267.TAAR362			Libants *et al.* (2009)
	STIG_3721.TAAR365			Libants *et al.* (2009)
	STIG_3721.TAAR351B			Libants *et al.* (2009)
	STIG_3721.TAAR351A			Libants *et al.* (2009)
	STIG_2594.TAAR358			Libants *et al.* (2009)
	STIG_2594.TAAR349			Libants *et al.* (2009)
	STIG_2594.TAAR340			Libants *et al.* (2009)
	STIG_1853.TAAR349B			Libants *et al.* (2009)
	STIG_1853.TAAR349A			Libants *et al.* (2009)
Mouse	Mm_TAAR1	10:23920406..23921404	NM_053205.1	NCBI
Mus musculus	Mm_TAAR2	10:23938572..23941583	NM_001007266.1	NCBI
	Mm_TAAR3	10:23949558..23950589	NM_001008429.1	NCBI
	Mm_TAAR4	10:23960494..23961537	NM_001008499.1	NCBI
	Mm_TAAR5	10:23970706..23971719	NM_001009574.1	NCBI
	Mm_TAAR6	10:23984609..23985646	NM_001010828.1	NCBI
	Mm_TAAR7a	10:23992405..23993481	NM_001010829.1	NCBI
	Mm_TAAR7b	10:23999939..24001015	NM_001010827.1	NCBI
	Mm_TAAR7d	10:24027222..24028298	NM_001010838.1	NCBI
	Mm_TAAR7e	10:24037614..24038690	NM_001010835.1	NCBI
	Mm_TAAR7f	10:24049510..24050586	NM_001010839.1	NCBI
	Mm_TAAR8a	10:24076500..24077534	NM_001010830.1	NCBI
	Mm_TAAR8b	10:24091260..24092294	NM_001010837.1	NCBI

	Mm_TAAR8c	10:24100843..24101951	NM_001010840.2	NCBI
	Mm_TAAR9	10:24108488..24109534	NM_001010831.1	NCBI
Fugu	Tr_TAAR1-like LOC101065853	scaffold_144:4873..5868	XM_003976993.1	NCBI
Takifugu rubripes	Tr_TAAR1-like LOC101068509	scaffold_2286:2698:3676	NW_004072942.1	NCBI
	Tr_TAAR1-like LOC101075198	scaffold_229:4033..5025	XM_003977307.1	NCBI
	Tr_TAAR1-like LOC101075670	8:5604511..5605530	XM_003966556.1	NCBI
	Tr_TAAR1-like LOC101076453	scaffold_682:7142..8143	XM003976898.1	NCBI
	Tr_TAAR1-like LOC101076685	scaffold_682:15010..15984	XM_003976899.1	NCBI
Medaka	Ol_TAAR 7e-like LOC101169023	2:26340511..26341482	XM_004066738.1	NCBI
Oryzias latipes	Ol_TAAR 1-like LOC101174969	24:10154632..10155600	XM_004083634.1	NCBI
	Ol_TAAR 1-like LOC101174242	24:10050386..10051363	XM_004083631.1	NCBI
	Ol_TAAR1-like LOC101163058	Un:2521..3516	XM_004085890.1	NCBI
	Ol_TAAR 1-like LOC101175700	24:10183943..10184941	XM_0040083637.1	NCBI
	Ol_TAAR 1-like LOC101175452	24:10173586..10174675	NC_019882.1	NCBI
Frog	Xt_TAAR1	scaffold_5:56592878..56593941	XM_002935811.2	NCBI
Xenopus tropicalis	Xt_TAAR4-like LOC100497949	scaffold_5:61512858..61513896	NW_004668237.1	NCBI
	Xt_TAAR4-like LOC100488258	scaffold_5:56562660..56563662	NW_004668237.1	NCBI
	Xt_TAAR4-like LOC100487949	scaffold_5:56584949..56585950	XM_002935812.2	NCBI
	Xt_TAAR4-like LOC100488107	scaffold_2751:3356..4363	XM_002935813.1	NCBI
	Xt_TAAR4-like LOC100485003	scaffold_5:56571031..56572038	XM004914870.1	NCBI
Zebrafish	Dr_TAAR1b	20:46197103..46198102	NM_001082904.1	NCBI
Danio rerio	Dr_TAAR10a	20:46166706..46167717	NM_001082898.1	NCBI
	Dr_TAAR10b	20:46187583..46188593	NM_001082903.1	NCBI
	Dr_TAAR10c	20:46170540..46171551	NM_001082900.1	NCBI
	Dr_TAAR10d	20:46178797..46179975	NM_001082902	NCBI
	Dr_TAAR11	20:46192811..46193795	NM_001083077.1	NCBI
	Dr_TAAR12a	20:46239853..46240885	NM_001083109.1	NCBI
	Dr_TAAR12b	20:46224045..46225059	NM_001083104.1	NCBI
	Dr_TAAR12c	20:46220574..46221591	NM_001083095.1	NCBI
	Dr_TAAR12e	20:46264043..46265058	NM_001083094.1	NCBI
	Dr_TAAR12f	20:46250999..46252013	NM_001082907.1	NCBI
	Dr_TAAR12g	20:46256084..46257095	NM_001082908.1	NCBI
	Dr_TAAR12h	20:46271973..46273008	NM_001082909.1	NCBI
	Dr_TAAR12i	20:46211225..46212248	NM_001083085.1	NCBI
	Dr_TAAR12j	20:46278935..46279958	NM_001082911.1	NCBI
	Dr_TAAR13b	20:46141163..46142188	NM_001083042.1	NCBI
	Dr_TAAR13c	20:46122478..46123505	NM_001083040.1	NCBI

Dr_TAAR13d	20:46133038..46134063	NM_001083041.1	NCBI
Dr_TAAR13e	20:46152019..46153044	NM_001083043.1	NCBI
Dr_TAAR14a	20:46087730..46088693	NM_001083036.1	NCBI
Dr_TAAR14b	20:46555069..46556055	NM_001082884.1	NCBI
Dr_TAAR14d	20:46542282..46543268	NM_001082912.1	NCBI
Dr_TAAR14e	20:46520053..46521039	NM_001082914.1	NCBI
Dr_TAAR14f	20:46537214..46538200	NM_001082913.1	NCBI
Dr_TAAR14h	20:46073613..46074577	NM_001083093.1	NCBI
Dr_TAAR14i	20:46069677..46070637	NM_001083037.1	NCBI
Dr_TAAR14j	20:46064040..46065004	NM_001083076.1	NCBI
Dr_TAAR14l	20:46079286..46080243	NM_001083103.1	NCBI
Dr_TAAR15	20:46117709..46118695	NM_001083039.1	NCBI
Dr_TAAR16a	10:41959548..41960522	NC_007121.5	NCBI
Dr_TAAR16f	10:41977904..41978899	NC_007121.5	NCBI
Dr_TAAR16g	10:41982677..41983672	NC_007121.5	NCBI
Dr_TAAR17a	10:41964902..41965909	XM_002663646.2	NCBI
Dr_TAAR18a	10:41889815..41890881	XM_001921119.3	NCBI
Dr_TAAR18c	10:41918143..41919165	XM_002663638.2	NCBI
Dr_TAAR18d	10:41897228..41898286	NC_007121.5	NCBI
Dr_TAAR18g	10:41894347..41895340	NC_007121.5	NCBI
Dr_TAAR18h	10:41927773..41929222	XM_002663640.3	NCBI
Dr_TAAR18i	10:41934817..41935885	NC_007121.5	NCBI
Dr_TAAR18j	10:41950041..41951072	XM_002663643.2	NCBI
Dr_TAAR19a	10:42103621..42104682	NC_007121.5	NCBI
Dr_TAAR19b	10:42140351..42141356	NC_007121.5	NCBI
Dr_TAAR19c	10:42147157..42148218	NC_007121.5	NCBI
Dr_TAAR19d	10:42134308..42135369	XM_001921481.1	NCBI
Dr_TAAR19e	10:42108791..42109853	NC_007121.5	NCBI
Dr_TAAR19h	10:42050839..42051900	NC_007121.5	NCBI
Dr_TAAR19i	10:42097939..42099000	NC_007121.5	NCBI
Dr_TAAR19j	10:42092480..42093541	NC_007121.5	NCBI
Dr_TAAR19m	10:42056442..42057541	XM_001342973.2	NCBI
Dr_TAAR19n	10:42066521..42067570	NC_007121.5	NCBI
Dr_TAAR19o	10:42043558..42044629	NC_007121.5	NCBI
Dr_TAAR19p	10:42036952..42038710	NM_001199914.1	NCBI
Dr_TAAR19q	10:42088337..42089398	NC_007121.5	NCBI
Dr_TAAR19s	10:42012869..42013930	NC_007121.5	NCBI
Dr_TAAR19t	10:42008622..42009707	NC_007121.5	NCBI

Dr_TAAR19u	10:42031263..42032312	NC_007121.5	NCBI
Dr_TAAR1-like LOC100535889	Un:6229..7240	NW_003336689.1	NCBI
Dr_TAAR20a	10:41698366..41699456	XM_001920656.2	NCBI
Dr_TAAR20a1	10:41839705..41840724	XM_002663628.2	NCBI
Dr_TAAR20b	10:41704278..41705383	XM_001920673.2	NCBI
Dr_TAAR20b1	10:41827971..41829023	NC_007121.5	NCBI
Dr_TAAR20c	Un:42..1040	BC107613.1	NCBI
Dr_TAAR20c1	10:41747212..41748268	NC_007121.5	NCBI
Dr_TAAR20d1	10:41845015..41846010	XM_002663629.2	NCBI
Dr_TAAR20f	10:41717303..41718355	NC_007121.5	NCBI
Dr_TAAR20g	10:41725305..41748271	NC_007121.5	NCBI
Dr_TAAR20h	10:41713185..41714183	XM_001341553.2	NCBI
Dr_TAAR20i	10:41731447..41732457	XM_001920721.2	NCBI
Dr_TAAR20n	10:41775937..41776989	NC_007121.5	NCBI
Dr_TAAR20o	10:41819593..41820646	NC_007121.5	NCBI
Dr_TAAR20r	10:41790993..41792046	NC_007121.5	NCBI
Dr_TAAR20u	10:41762407..41763433	XM_001920788.2	NCBI
Dr_TAAR20v	10:41808799..41809851	NC_007121.5	NCBI
Dr_TAAR20z	10:41849611..41850663	NC_007121.5	NCBI
Dr_TAAR32	10:41995697..41996758	NC_007121.5	NCBI
Dr_TAAR4-like LOC100150578	15:2040432..2041457	XM_001920809.1	NCBI
Dr_TAAR64	20:46148101..46149128	NM_001083102.1	NCBI
Dr_TAAR6-like LOC100002481	10:41834498..41835523	XM_001342220.1	NCBI
Dr_TAAR6-like LOC100002609	10:41902807..41903800	NC_007121.5	NCBI
Dr_TAAR6-like LOC100147966	10:42018541..42019952	NC_007121.5	NCBI
Dr_TAAR6-like LOC100147986	10:41790674..41791673	NC_007121.5	NCBI
Dr_TAAR9-like LOC101887086	10:41813764..41814763	NC_007121.5	NCBI
Dr_TAAR6-like LOC100150393	10:41856629..41857625	NC_007121.5	NCBI
Dr_TAAR6-like LOC100151046	10:42024091..42025099	NC_007121.5	NCBI
Dr_TAAR6-like LOC100330759	10:41910376..41911369	NC_007121.5	NCBI
Dr_TAAR6-like LOC100332495	13:305976..306950	NC_007124.5	NCBI
Dr_TAAR7a-like LOC100150419	20:46084062..46085048	NC_007131.5	NCBI
Dr_TAAR7c-like LOC100331259	10:41954907..41955899	NC_007121.5	NCBI
Dr_TAAR7d-like LOC100537518	10:42026878..42027874	NC_007121.5	NCBI
Dr_TAAR9-like LOC100151309	10:41879380..41880718	XM_002663633.2	NCBI
Dr_TAAR3-like LOC100331742	Un:10804..11790	NW_003336289.1	NCBI
Dr_TAAR2-like	20:Un:4860..5573	XM_001923755.4	NCBI

	Dr_TAAR20t	10:41767963..41769009	XM_002663626.3	NCBI
	Dr_TAAR20l	10:41771445..41772441	XM_005155417.1	NCBI
	Dr_TAAR1-like LOC100148217	10:41873631..41874632	NC_007121.5	NCBI
	TAAR 6-like LOC100148961	10:41868500..41869496	XM_001921070.3	NCBI
	Dr_TAAR20j	10:41781192..41782185	XM_002663627.2	NCBI
	TAAR 6-like LOC100149676	10:41864180..41865170	NC_007121.5	NCBI
	Dr_TAAR18b	10:41922296..41923295	XM_002663639.2	NCBI
	TAAR 9-like LOC100148476	10:41973844..41974831	NC_007121.5	NCBI
	Dr_TAAR6-like LOC100332366	13:295967..296969	NC_007124.5	NCBI
	Dr_TAAR9-like LOC100000110	13:300938..301940	NC_007124.5	NCBI
Tetraodon *Tetraodon nigroviridis*	Tn_TAAR1(6 of 15)	Un:85474851..85475831	ENSTNIT00000003225	Ensembl
	Tn_TAAR1(11 of 15)	Un:85467788..85468786	ENSTNIT00000002494	Ensembl
	Tn_TAAR1(14 of 15)	Un:35318087..35319082	ENSTNIT00000001436	Ensembl
	Tn_TAAR1(13 of 15)	Un:70258269..70259261	ENSTNIT00000007318	Ensembl
	Tn_TAAR1(15 of 15)	Un:70265697..70266692	ENSTNIT00000007319	Ensembl
	Tn_TAAR1(10 of 15)	Un: 85499769..85500764	ENSTNIT00000008083	Ensembl
	Tn_TAAR1(12 of 15)	14:1075051..1076047	ENSTNIG00000001506	Ensembl
	Tn_TAAR1(5 of 15)	14:1025315..1026317	ENSTNIG00000019179	Ensembl
	Tn_TAAR1(2 of 15)	14:1062370..1063270	ENSTNIG00000001657	Ensembl
	Tn_TAAR1(3 of 15)	14:1072819..1073800	ENSTNIG00000001360	Ensembl
	Tn_TAAR1(9 of 15)	Un:85495812..85496808	ENSTNIG00000005193	Ensembl
Stickleback *Gasterosteus aculeatus*	Ga_TAAR1(2 of 5)	groupXVIII:864898..865894	ENSGACG00000004284	Ensembl
	Ga_TAAR1(5 of 5)	groupXVIII:849379..850381	ENSGACG00000004279	Ensembl
	Ga_TAAR1(1 of 5)	groupXVIII:856337..857339	ENSGACG00000004280	Ensembl
	Ga_TAAR1(4 of 5)	groupXVIII:806425..807427	ENSGACG00000004274	Ensembl
	Ga_TAAR1(3 of 5)	groupI:27258528..2725921	ENSGACG00000015361	Ensembl
Anole Lizard *Anolius carolinensis*	Ac_TAAR1	Un:4463430..4464437	XM_003223314.1	NCBI
	Ac_TAAR2	Un:4474066..4475097	XM_003223315.1	NCBI
	Ac_TAAR5	Un:4484212..4485264	XM_003223316.1	NCBI
American Alligator *Alligator mississippiensis*	Am_TAAR1	Un:220473..221480	XM_006270537.1	NCBI
	Am_TAAR2	245497..246489	XM_006270538.1	NCBI
	Am_TAAR4-like LOC102559019	286493..287536	XM_006270529.1	NCBI
	Am_TAAR4-like LOC102558785	269855..270901	XM_006270528.1	NCBI
	Am_TAAR5-like LOC102561336	297408..298403	XM_006270539.1	NCBI
	Am_TAAR7a-like LOC102561788	335295..336326	XM_006270540.1	NCBI
	Am_TAAR9-like LOC102562260	358195..359220	XM_006270542.1	NCBI
	Am_TAAR9-like LOC102559257	305709..306737	XM_006270530.1	NCBI

African Coelacanth *Latimeria chalumnae*	Lc_TAAR4-like LOC102365654	Un:7900..8919	XM_006014231.1	NCBI
	Lc_TAAR4-like LOC102362127	Un:6731..7825	XM_006014219.1	NCBI
	Lc_TAAR9-like LOC102346275	Un:226..1260	XM_006014034.1	NCBI
	Lc_TAAR4-like LOC102346012	Un:4621..5655	XM_006014033.1	NCBI
	Lc_TAAR4-like LOC102362844	Un:16272..17312	XM_006014013.1	NCBI
	Lc_TAAR4-like LOC102358308	Un:34650..35687	XM_006013795.1	NCBI
	Lc_TAAR4-like LOC102346983	Un:43349..44395	XM_006013765.1	NCBI
	Lc_TAAR4-like LOC102362653	Un:81244..82287	XM_006013513.1	NCBI
	Lc_TAAR5-like LOC102362390	Un:53536..54543	XM_006013512.1	NCBI
	Lc_TAAR4-like LOC102361509	Un:24914..25954	XM_006013443.1	NCBI
	Lc_TAAR4-like LOC102358939	Un:98526..99563	XM_006013436.1	NCBI
	Lc_TAAR4-like LOC102367115	Un:100058..101794	XM_006013007.1	NCBI
	Lc_TAAR4-like LOC102366851	Un:54232..56373	XM_006013006.1	NCBI
	Lc_TAAR4-like LOC102366590	Un:17939..19252	XM_006013005.1	NCBI
	Lc_TAAR4-like LOC102364694	Un:107680..108717	XM_006012998.1	NCBI
	Lc_TAAR5-like LOC102348717	Un:134193..135206	XM_006011945.1	NCBI
	Lc_TAAR5-like LOC102348210	Un:47880..48881	XM_006011943.1	NCBI
	Lc_TAAR1-like LOC102360536	Un:202003..203031	XM_006010896.1	NCBI
Elephant Shark *Callorhinchus milii*	Cm_contig_19181 - ORF 1 (frame 3)	Un:6852..7886	AAVX02019181.1	NCBI
	Cm_contig_19181 - ORF 4 (frame 1)	Un:46282..47310	AAVX02019181.1	NCBI
Atlantic Salmon *Salmo salar*		21359..22306	AGKD01000046	ASalBase/NCBI
		16565..17527	AGKD01000061	ASalBase/NCBI
		35874..36836	AGKD01000272	ASalBase/NCBI
		6099..7052	AGKD01000449	ASalBase/NCBI
		2040..3002	AGKD01001626	ASalBase/NCBI
		16596..17558	AGKD01003248	ASalBase/NCBI
		13148..14107	AGKD01003637	ASalBase/NCBI
		9633..10595	AGKD01004261	ASalBase/NCBI
		12524..13471	AGKD01004637	ASalBase/NCBI
		7621..8583	AGKD01006407	ASalBase/NCBI
		8332..9279	AGKD01009924	ASalBase/NCBI
		9151..10113	AGKD01016613	ASalBase/NCBI
		3006..3971	AGKD01019372	ASalBase/NCBI
		15867..16829	AGKD01020184	ASalBase/NCBI
		10125..11087	AGKD01034652	ASalBase/NCBI
		859..1890	AGKD01044192	ASalBase/NCBI
		796..1821	AGKD01064135	ASalBase/NCBI
		1778..2743	AGKD01072494	ASalBase/NCBI

		2586..3548	AGKD01073835	ASalBase/NCBI
		3202..4164	AGKD01081184	ASalBase/NCBI
		1874..2911	AGKD01084249b	ASalBase/NCBI
		2433..3392	AGKD01089915	ASalBase/NCBI
		2628..3659	AGKD01103530	ASalBase/NCBI
		5286..6239	AGKD01115415	ASalBase/NCBI
		717..1823	AGKD01134375	ASalBase/NCBI
		9267..10224	AGKD01195896*	ASalBase/NCBI
		731..1693	AGKD01495072	ASalBase/NCBI
OUTGROUPS	Dr_Histamine Receptor H2	14:1618017..1619248	NM_001045338.1	NCBI
	Hs_Histamine Receptor H2	5:175084847..175136239	NM_001131055.1	NCBI
	Hs_Rhodopsin	3:129247482..129254187	NM_000539.3	NCBI
	Dr_Adrenergic Receptor Beta 3a	8:38812340..38850902	NM_001128335.1	NCBI
	Hs_Adrenergic Receptor Beta 3	8:37820513..37824184	NM_000025.2	NCBI
	Gg_Histamine Receptor H2	13:9783231..9794177	XM_004944905.1	NCBI
	Dr_Dopamine receptor D2a	15:21008238..21035946	NM_183068.1	NCBI
	Hs_Dopamine receptor D2	11:5001..70685	NM_000795.3	NCBI
	Gg_Dopamine receptor D2	24:5748287..5753356	NM_001113290.1	NCBI
	Dr_Serotonin receptor 1A b	21:19778036..19779277	NM_001145766.1	NCBI
	Hs_Serotonin receptor 1A	5:63255560..63257804	NM_000524.3	NCBI
	Gg_Serotonin receptor 1A	Z:19853067..19854401	NM_001170528.1	NCBI

*Contig with partial sequence. Complete TAAR was found in scaffold jcf2339333987_2792921-4584714_ssa21 (Lien *et al*., in preparation), which contains this contig

Table A3: Summary of putatively functional Atlantic salmon TAAR genes including predicted nucleotide length and physical/genetic location. Superscripts denote the various mapping methods utilized.

Contig	Length (bp)	Physical Location	Genetic Location
AGKD01000046	948	unmapped	unmapped
AGKD01000061	963	unmapped	unmapped
AGKD01000272	963	unmapped	unmapped
AGKD01000449	954	unmapped	unmapped
AGKD01001626	963	fps943[1]	Ssa21[1,2]
AGKD01003248	963	fps508[1]	Ssa21[1]
AGKD01003637	960	fps508[1]	Ssa21[1,2]
AGKD01004261	963	unmapped	unmapped

AGKD01004637	948	unmapped	Ssa13[2,3]
AGKD01006407	963	unmapped	Ssa21[2]
AGKD01009924	948	unmapped	unmapped
AGKD01016613	963	fps508[1]	Ssa21[1,2]
AGKD01019372	966	unmapped	Ssa21[2]
AGKD01020184	963	fps508[1]	Ssa21[1,2]
AGKD01034652	963	unmapped	unmapped
AGKD01044192	1026	fps2319[1]	Ssa06[1,3]
AGKD01064135	1026	unmapped	unmapped
AGKD01072494	966	unmapped	Ssa02[2]
AGKD01073835	963	fps943[1]	Ssa21[1,3]
AGKD01081184	963	fps508[1]	Ssa21[1,2]
AGKD01084249b	1038	fps798[1]	Ssa15[1,2]
AGKD01089915	960	fps943[1]	Ssa21[1,3]
AGKD01103530	1026	unmapped	unmapped
AGKD01115415	954	fps508[1]	Ssa21[1,2]
AGKD01134375	1038	fps798[1]	Ssa15[1,2]
AGKD01195896	957	fps508[1]	Ssa21[1]
AGKD01495072	963	unmapped	Unmapped

[1] BLASTn against BAC ends
[2] BLASTn against SNP mapped
[3] Microsatellite marker

Table A4: Summary of putative Atlantic salmon TAAR pseudogenes including physical/genetic location and predicted cause of pseudogenziation. Superscripts denote the various mapping methods utilized.

Name	Physical Location	Genetic Location	Predicted Cause
AGKD01001951	fps508[1]	Ssa21[1,2]	Indel / Frameshift Mutation
AGKD01036711	fps508[1]	Ssa21[1,2]	Indel / Frameshift Mutation
AGKD01087949	fps508[1]	Ssa21[1,2]	Partial match
AGKD01040598	fps798[1]	Ssa15[1,2]	Partial match
AGKD01084249a	fps798[1]	Ssa15[1,2]	Partial match
AGKD01222984	fps798[1]	Ssa15[1,2]	Nonsense mutation
AGKD01000376	unmapped	unmapped	Indel / Frameshift Mutation
AGKD01014617	unmapped	Ssa21[2,3]	Indel / Frameshift Mutation

AGKD01024914	unmapped	unmapped	Partial Match
AGKD01052335	unmapped	unmapped	Nonsense Mutation
AGKD01073505	unmapped	Ssa14[3]	Indel / Frameshift Mutation
AGKD01074920	unmapped	Ssa21[2]	Nonsense Mutation
AGKD01084072	unmapped	Ssa21[2]	Indel / Frameshift Mutation
AGKD01085575	unmapped	unmapped	Indel / Frameshift Mutation
AGKD01102951	unmapped	unmapped	more sequence data required
AGKD01137574	unmapped	Ssa06[2]	Partial Match
AGKD01143659	unmapped	unmapped	more sequence data required
AGKD01145155	unmapped	unmapped	more sequence data required
AGKD01162248	unmapped	Ssa04[2]	Partial Match
AGKD01165582	unmapped	unmapped	Frameshift Mutation / Nonsense Mutation
AGKD01284261	unmapped	Ssa01[2]	Partial Match
AGKD01437520	unmapped	Ssa21[2]	more sequence data required
AGKD01455862	unmapped	Ssa21[2]	Indel / Frameshift Mutation
AGKD01503373	unmapped	unmapped	more sequence data required
AGKD01507155	unmapped	unmapped	Partial Match

[1] BLASTn against BAC ends
[2] BLASTn against SNP mapped
[3] Microsatellite marker

Table A5: Relative Quantification (RQ) data for the expression of two Atlantic salmon TAAR genes

Gene	Location/Life Stage	Individual	RQ	Average RQ by Life Stage	SD by Life Stage	Fold Change Relative to Adult
AGKD01103530	CR_Smolt	S1	0.2151	0.05851	0.06814	11.29
		S2	0.0666			
		S3	0.0414			
		S5	0.0068			
		S6	0.0437			
		S8	0.0407			
		S9	0.0228			
		S11	0.0158			
		S12	0.2246			
		S13	0.0492			
		S14	0.0095			

		S15	0.0112			
		S16	0.0460			
		S17	0.0085			
		S18	0.0541			
		S19	0.0066			
		S20	0.1321			
	CR_Parr	P5	0.0167	0.01568	0.03783	3.03
		P6	0.0004			
		P8	0.0015			
		P9	0.0004			
		P10	0.0006			
		P12	0.0000			
		P13	0.0062			
		P14	0.1440			
		P17	0.0191			
		P18	0.0005			
		P19	0.0016			
		P21	0.0032			
		P22	0.0006			
		P24	0.0249			
	CR_Adult	A2	0.0025	0.005183	0.007079	1
		A3	0.0022			
		A4	0.0195			
		A5	0.0015			
		A6	0.0013			
		A7	0.0041			
AGKD01044192	CR_Smolt	S1	0.1543	0.0410	0.0817	37.52
		S2	0.0442			
		S3	0.0018			
		S5	0.0001			
		S6	0.0016			
		S8	0.0016			
		S9	0.0004			
		S11	0.0007			
		S12	0.1243			
		S13	0.0650			
		S14	0.0002			
		S15	0.0003			
		S16	0.0022			
		S17	0.0001			
		S18	0.0007			
		S19	0.0001			
		S20	0.3000			
	CR_Parr	P5	0.0310	0.0214	0.0360	19.53

		P8	0.0001			
		P9	0.0000			
		P10	0.0000			
		P13	0.0002			
		P14	0.1117			
		P17	0.0323			
		P19	0.0002			
		P21	0.0002			
		P22	0.0000			
		P24	0.0593			
	CR_Adult	A2	0.0009	0.0011	0.0022	1
		A3	0.0001			
		A4	0.0054			
		A5	0.0000			
		A6	0.0000			
		A7	0.0001			
AGKD01103530	EX_Smolt	1	1.1117	0.3617	0.2882	1.94
		2	0.3649			
		3	0.2268			
		6	0.1710			
		7	0.2118			
		9	0.2312			
		10	0.2455			
		11	0.2397			
		12	0.2359			
		13	0.5784			
	EX_Parr	16	0.0424	0.1166	0.1071	0.63
		18	0.0018			
		19	0.0540			
		20	0.0022			
		21	0.3241			
		23	0.2108			
		24	0.1595			
		25	0.1597			
		26	0.0946			
	EX_Adult	14	0.0866	0.1860	0.2752	1
		27	0.0699			
		28	0.1003			
		29	0.0586			
		30	0.1211			
		31	0.9539			
		32	0.0604			
		33	0.2405			
		34	0.1103			
		35	0.0587			

AGKD01044192	EX_Smolt	1	0.8338	0.1155	0.2589	1.1508
		2	0.1571			
		3	0.0116			
		6	0.0021			
		7	0.0044			
		9	0.0044			
		10	0.0031			
		11	0.0037			
		12	0.0067			
		13	0.1284			
	EX_Parr	16	0.0004	0.0496	0.1162	0.4945
		19	0.0003			
		20	0.0000			
		21	0.0079			
		23	0.3124			
		24	0.0255			
		26	0.0010			
	EX_Adult	14	0.0752	0.1004	0.2515	1
		27	0.0007			
		28	0.0011			
		29	0.0008			
		30	0.0015			
		31	0.8069			
		32	0.0009			
		33	0.1149			
		34	0.0010			
		35	0.0009			
AGKD01103530	SC_Juvenile	44	0.0130	0.0193	0.0208	1.58
		48	0.0422			
		53	0.0114			
		54	0.0035			
		56	0.0634			
		57	0.0247			
		58	0.0090			
		59	0.0027			
		60	0.0039			
	SC_Adult	49	0.0079	0.0123	0.0075	1
		50	0.0231			
		51	0.0026			
		52	0.0036			
		61	0.0152			
		62	0.0138			
		64	0.0107			
		65	0.0211			

AGKD01044192	SC_Juvenile	44	0.0002	0.0059	0.0087	0.28
		48	0.0227			
		53	0.0002			
		54	0.0000			
		56	0.0123			
		57	0.0155			
		58	0.0022			
		59	0.0000			
		60	0.0001			
	SC_Adult	49	0.0016	0.0211	0.0583	1
		50	0.0007			
		51	0.0000			
		52	0.0001			
		61	0.0003			
		62	0.0003			
		64	0.0002			
		65	0.1653			

Table A6: Microsatellite Allele calls for mapping of contigs containing TAARs. Source images are provided as supplementary figures.

Family	Marker	Sire	Dam	1	2	3	4	5	6	7	8	9	10	11
BR5	Ssa0008SCSFU	11	12	12	11	11	11	12	11	11	11	12	11	12
BR5	Ssa0014SCSFU	23	11	-	12	13	12	12	12	12	12	-	12	13
BR5	Ssa0019SCSFU	13	24	-		12	12	-	-	-	12	34	-	-
BR6	Ssa0017SCSFU	22	12	22	22	12	22	12	22	22	12	12	-	12
BR6	Ssa0019SCSFU	11	12	12	11	-	11	11	11	11	11	12	11	11

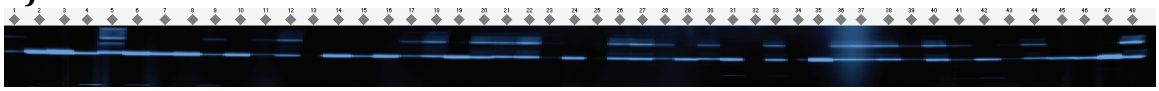
Family	Marker	12	13	14	15	16	17	18	19	20	21	22	23	24
BR5	Ssa0008SCSFU	12	-	11	11	11	12	12	11	12	12	12	12	11
BR5	Ssa0014SCSFU	12	-	12	12	12	13	13	13	12	13	13	-	12
BR5	Ssa0019SCSFU	-	-	-	-	-	12	34	12	34	12	12	12	-
BR6	Ssa0017SCSFU	22	22	22	22	22	-	12	-	12	12	22	22	12
BR6	Ssa0019SCSFU	12	11	12	-	12	-	12	-	12	12	12	11	11

Family	Marker	25	26	27	28	29	30	31	32	33	34	35	36	37
BR5	Ssa0008SCSFU	-	12	12	12	11	12	11	-	12	-	11	12	12
BR5	Ssa0014SCSFU	-	12	12	12	12	13	12	13	12	-	12	12	12
BR5	Ssa0019SCSFU	-	34	-	-	34	12	-	12	-	-	-	12	12
BR6	Ssa0017SCSFU	12	12	22	22	12	12	12	12	22	22	12	-	12
BR6	Ssa0019SCSFU	-	12	11	12	12	12	11	11	-	11	12	-	12

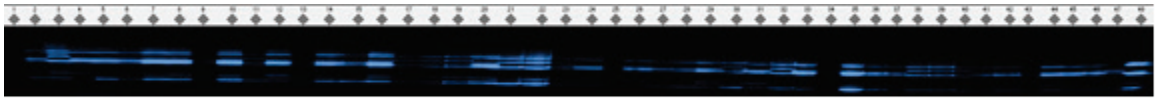
Family	Marker	38	39	40	41	42	43	44	45	46	Chromosome
BR5	Ssa0008SCSFU	12	12	12	12	11	11	12	11	11	Ssa13
BR5	Ssa0014SCSFU	13	13	13	12	12	-	12	12	13	Ssa21
BR5	Ssa0019SCSFU	34	-	12	34	-	-	-	12	-	Ssa14
BR6	Ssa0017SCSFU	12	22	12	-	-	12	-	22	12	Ssa21
BR6	Ssa0019SCSFU	12	12	-	-	-	12	-	11	12	Ssa14

Figure A1: Microsatellite gel images for genetic mapping of Atlantic salmon contigs containing TAARs. A – Ssa0008SCSFU-BR5; B – Ssa0014SCSFU-BR5; C – Ssa0019SCSFU-BR5; D – Ssa0017SCSFU-BR6; E – Ssa0019SCSFU-BR6. Allelic calls shown in Table S5.

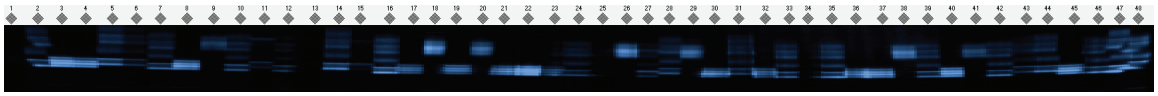
A)



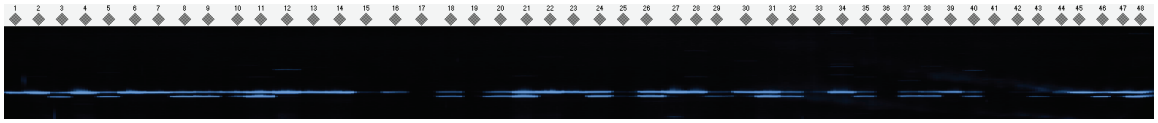
B)



C)



D)



E)